PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Diterpene cyclases and the nature of the isoprene fold

Rong Cao,[1] Yonghui Zhang,[2] Francis M. Mann,[3] Cancan Huang,[1] Dushyant Mukkamala,[1] Michael P. Hudock,[1] Matthew E. Mead,[3] Sladjana Prisic,[3] Ke Wang,[2] Fu-Yang Lin,[1] Ting-Kai Chang,[2] Reuben J. Peters,[3]* and Eric Oldfield[1,2]*

[1] Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[2] Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[3] Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011

ABSTRACT

The structures and mechanism of action of many terpene cyclases are known, but no structures of diterpene cyclases have yet been reported. Here, we propose structural models based on bioinformatics, site-directed mutagenesis, domain swapping, enzyme inhibition, and spectroscopy that help explain the nature of diterpene cyclase structure, function, and evolution. Bacterial diterpene cyclases contain ~20 α-helices and the same conserved "QW" and DxDD motifs as in triterpene cyclases, indicating the presence of a βγ barrel structure. Plant diterpene cyclases have a similar catalytic motif and βγ-domain structure together with a third, α-domain, forming an αβγ structure, and in H^+-initiated cyclases, there is an EDxxD-like Mg^{2+}/diphosphate binding motif located in the γ-domain. The results support a new view of terpene cyclase structure and function and suggest evolution from ancient (βγ) bacterial triterpene cyclases to (βγ) bacterial and thence to (αβγ) plant diterpene cyclases.

Proteins 2010; 78:2417–2432.

Key words: isoprenoid; diterpene; triterpene; cyclase; farnesyl diphosphate; homology model.

INTRODUCTION

Terpenoids, or more generally isoprenoids, represent the largest (>50,000), most ancient, as well as the most structurally diverse class of small molecule natural products on earth.[1–5] Isoprenoid biosynthesis begins with the condensation of isopentenyl diphosphate (IPP) and/or dimethylallyl diphosphate (DMAPP), typically forming geranyl diphosphate (GPP), farnesyl diphosphate (FPP), or geranylgeranyl diphosphate (GGPP) [Fig. 1(a)]. The resulting diphosphates can then condense in a head-to-head fashion to form linear tri- and tetra-terpenes that then form for example, hopanoids, sterols, and carotenoids.[6] They can also cyclize to form smaller terpenes such as limonene, kaurene precursors to plant gibberellin growth hormones such as gibberellins[7]; or they can prenylate proteins such as Ras [Fig. 1(a)].[8]

The structures of many terpene cyclases are known to contain the highly α-helical "isoprenoid" or farnesyl diphosphate synthase (FPPS or FPPase) fold [here called α, Fig. 1(b)],[2,4] with catalysis being mediated via conserved DDxxD motifs using a Mg^{2+}-dependent "ionization-initiated" (class I) mechanism. But in triterpene cyclases such as squalene-hopene cyclase (SHC),[6] catalysis is fundamentally different and involves a "protonation-initiated" mechanism utilizing a catalytic DxDD motif, located in one of two $(\alpha/\alpha)_6$ barrels [here called βγ, Fig. 1(c)].[6] In plant mono- and sesquiterpene cyclases,[9–11] for example, tobacco *epi*-aristolochene synthase (TEAS) and bornyl diphosphate synthase (BS), the α-isoprenoid fold is again present [green, Fig. 1(d)], but these also contain an N-terminal domain [red, Fig. 1(d)] that while largely relictual, resembles the β domain in SHC.[6] This β-domain is also structurally similar to that seen in the β-subunit in protein prenyl transferases[8] [Fig. 1(e)], where it acts as the binding site for FPP or GGPP substrates. This β-domain is present in both plant and animal prenyltransferases. For clarity in the discussion that follows, these three putative domain structures are shown schematically in Figure 1(f) (α, green; β, red; and γ, cyan).

The nature of the catalytic mechanism in terpene cyclases (and other synthases) containing solely α-, αβ-, or βγ-domains are now beginning to be understood since x-ray crystallographic structures (in the presence of inhibitors) have been reported.[6,9–13] However, there are no structures of the enzymes that catalyze the

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: United States Public Health Service; Grant numbers: GM65307, GM076324
Rong Cao, Yonghui Zhang, and Francis M. Mann contributed equally to this work.
*Correspondence to: Reuben Peters. E-mail: rjpeters@iastate.edu or Eric Oldfield. E-mail: eo@chad.scs.uiuc.edu.
Received 15 December 2009; Revised 25 March 2010; Accepted 29 March 2010
Published online 14 April 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22751

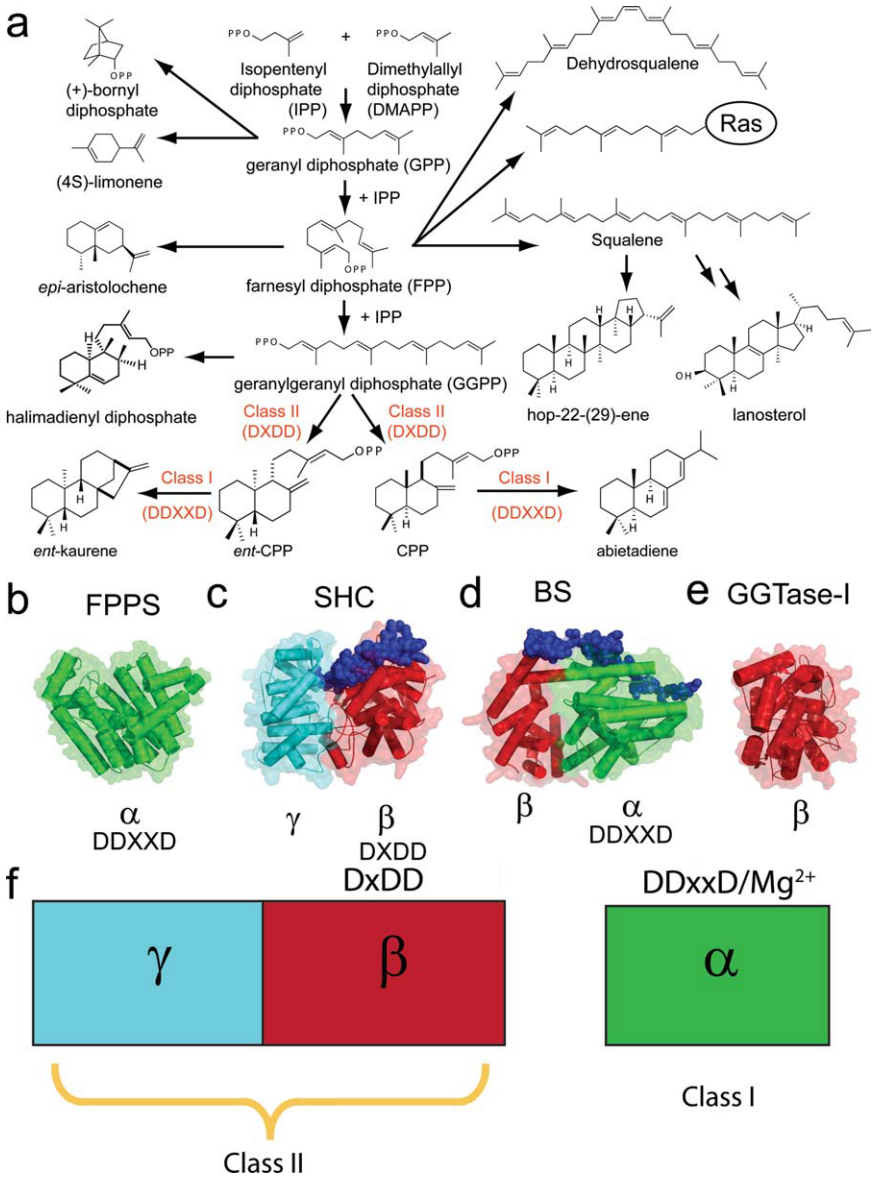


Figure 1

Overview of terpene biosynthesis and terpene synthase structures. (**a**) Structures of representative terpenes, all of which flow from condensation of IPP and DMAPP to form GPP, FPP and GGPP. (**b**)–(**e**) Modular building blocks for terpene/isoprenoid biosynthesis enzymes. (b) FPPase fold (α domain) in human FPPS, contains the catalytic DDxxD domains, carries out Mg^{2+}-dependent PPi-ionization initiated reactions. (c) βγ domains in squalene hopene cyclase, a triterpene cyclase. The β-domain (red) contains the catalytic DxDD. The two subunits are thought to have arisen via gene duplication of an ancestral $(\alpha/\alpha)_6$ barrel protein. The N-terminus is illustrated in blue and folds back from the γ domain and forms part of β domain. (d) αβ domains in bornyl diphosphate synthase (BS), a monoterpene cyclase. The catalytic α domain (green) contains DDxxD and resembles the FPPase fold, the β domain resembles the catalytic β domain in SHC. The N-terminus again folds back from the β domain and is important for catalysis. (e) The β-domain $(\alpha/\alpha)_6$ barrel in a protein geranylgeranyl transferase. This domain has structural homology to the β domain in SHC (and BS). (**f**) Cartoon diagram showing α, β, γ domains are their corresponding activities and key catalytic residues.

biosynthesis of cyclic diterpenes, such as *ent*-copalyl diphosphate, abietadiene and halimadienyl diphosphate [Fig. 1(a)]. What is known, however, is that in many diterpene cyclases, both DxDD as well as DDxxD motifs are present, and that these motifs are involved in protonation- (class II) and ionization- (class I) initiated reactions, respectively.[11–14] Plus, Mg^{2+} is involved in both reactions. But what are the key structural features of diterpene cyclase structure? How do they work? How is Mg^{2+} involved in the H^+-initiated cyclization reaction? How did diterpene cyclases originate? We explore these questions here by using a variety of experimental and computational methods, beginning with a consideration of structure.

MATERIALS AND METHODS

Site-directed mutagenesis

Site-directed mutagenesis was performed on the *Abies grandis* abietadiene synthase (*AgABS*) construct in pENTR-SD (Invitrogen Gateway System) as previously described,[15] with 18 bp primers for the single mutations and 45 bp primers for the quad-mutants. Accuprime Pfx (Invitrogen, Carlsbad, CA) was used according to the manufacturer's directions. PCR was performed under the following conditions: (1) $T = 95^{\circ}C$, 5 min; (2) $T = 95^{\circ}C$, 30s; (3) $T = 53^{\circ}C$, 1 min; (4) $T = 68^{\circ}C$, 5 min; (5) $T = 4^{\circ}C$, 10 min; 6) Hold at 15°C; repeating steps 2–4 for 20 times. PCR products were treated with Dpn1 overnight at 37°C to remove parent strands and transformed into DH5α *E. coli* cells. Plasmids were propagated and mutagenesis confirmed via complete sequencing. The mutated genes were transferred into protein expression vectors (pDEST14, Invitrogen) using directional recombination. Expression vectors containing the gene of interest were then transformed into C41 *E. coli* cells for protein expression. Protein expression was achieved by growth of cells to an OD_{600} of 0.6–0.8 followed by IPTG induction for 16 hours at 16°C. Cells were then pelleted and resuspended in 15 mL of lysis buffer (10 m*M* Tris-Cl, 10% glycerol, 10 m*M* $MgCl_2$ pH 6.8). Cells were lysed via sonic disruption, on ice. Lysates were clarified before enzymatic assay. Expression was checked via SDS-PAGE, and protein expression was determined to be constant for all constructs (data not shown).

Computational aspects

For the sequence alignments, the full sequence of abietadiene synthase (minus the plastid targeting segment) was searched against UniRef90[16] and an alignment was constructed with the PSI-BLAST module[17] embedded in the JPRED3 program[18] (three iterations, a first iteration e-value cut-off of 0.05 and 0.01 e-value cut-off for subsequent iterations), resulting in the alignment of 228 terpene synthase sequences. In addition, we used Clustal_W,[19] to investigate to what extent patterns of conserved catalytic residues could be determined between many different types of terpene cyclases, with particular attention being paid to the location of the catalytic (DDxxD, DxDD) domains, and the presence or absence of any likely metal binding sites, used in protonation-initiated (class II) or mixed function diterpene cyclase catalysis. We used JPRED3[18] and COUDES[20] programs to predict secondary structures (helix, sheet and turn contents). We also determined the pairwise rms deviations for Cα in 20 terpene synthases using the secondary structure matching (SSM)[21] program and incremental combinatorial extension (CE[22]) program. The Phylip[23] package was used to generate phylogenic trees using these pairwise rmsd values, normalized by the number of aligned positions. Alignments of the full sequence of *B. japonicum* CPS (Blr2149) with other proteins were also obtained by using the PSI-BLAST module[17] embedded in the JPRED3 program[18] (using three iterations, a first iteration e-value cut-off of 0.05 and 0.01 e-value cut-off for subsequent iterations), as well as with Clustal_W.[19] Normalized entropy scores [Shannon entropy is normalized so that conserved (low entropy) columns score 1 and diverse (high entropy) columns score 0] were calculated using the Scorecon server.[24]

Chimera construction

Chimeras were constructed from previously described pENTR-SD clones of truncated pseudo-mature *OsCPS2* (*Oryza sativa* copalyl diphosphate synthase-2)[25] and *OsCPS4* (*Oryza sativa* copalyl diphosphate synthase-4).[26] Overlapping primers carrying a half sequence from one gene and half from another at the following splicing sites (W544 for *OsCPS2*/W536 of *OsCPS4*) were used to amplify two fragments and combine them in a two-step PCR reaction, basically as described previously.[27] Briefly, in the first PCR reaction, pairs of primers were used (vector specific M13 forward primer and chimera reverse primer and *vice versa*, M13 reverse primer and chimera forward primer) to make two halves of a chimeric gene and in the second reaction, the halves were mixed and amplified with another (M13 forward and reverse primed) PCR reaction. We were not successful in transferring the final PCR product into pENTR/SD/D-TOPO vector via the TOPO cloning reaction, as we did with the original genes, so a recombination reaction was used instead, to transfer the chimeric genes to pDONR221, after PCR with att (recombination) sites added to gene-specific primers, according to the manufacturer's (Invitrogen) instructions. The resulting chimeric clones were verified by complete sequencing, then transferred via directional recombination to pDEST14 vectors, for expression.

Chimera assay conditions

Enzymes were assayed at 30°C for 1 hour. The assay buffer consisted of 50 m*M* HEPES, 10% glycerol, 10 m*M* $MgCl_2$, 1 m*M* KCl (pH 7.75). Directly before the assay, 5 μ*M* GGPP was added, followed by 100 μL of clarified lysate. After dephosphorylation with calf intestinal phosphatase, products were extracted into hexanes and analyzed via GC-MS as described previously[28] using authentic standards to confirm product identities. All assays were performed in duplicate. For the control assay, constructs were fed CPP of normal stereochemistry. This CPP was synthesized enzymatically from GGPP (in each tube) using 10 μ*M* rAgABS D621A purified protein.[29] Complete conversion to CPP was accomplished after

4 hours incubation (as described previously), before the introduction of the construct of interest. The assays were then performed as described.[29]

Abietadiene synthase inhibition assays

The activity of abietadiene synthase was monitored by a continuous spectrophotometric assay for diphosphate release.[30] Assays were carried out using 96 well plates with 200 μL of reaction mixture in each well, which contained 20 ng AgABS, 50 m*M* Hepes, 1 m*M* $MgCl_2$, 10 m*M* Chaps at pH 7.2 and various inhibitor concentrations. Either 25 μ*M* GGPP or CPP was used as substrate. The IC_{50} values were obtained by fitting the initial velocities in the presence of various concentrations of inhibitor to the dose-response curve in Origin 7.1 (OriginLab Corporation, Northampton, MA, www.OriginLab.com).

Protein modeling

Models for the bacterial diterpene cyclases *B. japonicum* CPS (Blr2149) and *B. japonicum* KS (Blr2150), were built in I-TASSER.[31] To build the (plant) diterpene cyclase, abietadiene synthase structure, we used a two-part approach. First, based on our earlier observations and those of Wendt and Schulz[32] (using the Dali program[33]), we took advantage of the fact that the β-domain in squalene hopene cyclase (SHC) and tobacco epi-aristolochene synthase (TEAS) superimposed within 3 Å, and are homologous to corresponding domains in abietadiene synthase. After aligning β-domains of SHC (PDB File 3SQC) and TEAS (PDB File 5EAU), we merged the SHC βγ-domain and TEAS α-domain to form an "αβγ" chimera, which after geometry optimization in MOE[34] was then used to construct a homology model of ABS using the "Homology Modeling" option in MOE.[34]

All the models were also externally validated using PROSA_WEB.[35]

Comparative molecular similarity indices analysis (CoMSIA) analysis

Inhibitor structures were generated and minimized using MOE,[34] then aligned by using the flexible alignment protocol. The resulting structures were then exported into Sybyl 7.3,[36] and CoMSIA[37] fields computed for the aligned structures, using default grid spacing, and probe atom types. Partial-least-square (PLS) regression was used to assign coefficients to grid points based on the experimentally determined abietadiene synthase pIC_{50} values (where $pIC_{50} = -\log_{10}(IC_{50}, [M])$. The optimum number of components in the model was determined by the SAMPLS[38] method implemented in Sybyl. The final model was selected based on cross-validated r^2 (q^2), r^2, error and number of components, such that a statistically robust and parsimonious model could be generated. Test-set calculations were performed using training sets containing 23 or 24 compounds.

M. tuberculosis tuberculosinol diphosphate synthase (Rv3377c) inhibition assays

The plasmids containing *M. tuberculosis Rv3377c* or *Rv3378c* were obtained from Professor Tom Alber (University of California at Berkeley). The expression and purification of the two proteins are based on protocols provided by Alber. Briefly, *E. coli* [BL21 (DE3)] transformed with the plasmid containing either *Rv3377c* or *Rv3378c* was inoculated and grown to an OD_{600} ~0.6 to 1. Expression was then induced with IPTG, followed by vigorous shaking at 18°C for 16 hours. The cells were harvested and disrupted by using the B-PER (bacterial protein extraction reagent, Thermo Scientific). The His-tagged proteins were purified using affinity column chromatography. The inhibition of various compounds against Rv3377c was carried out using a radioactive assay. 50 μL of assay buffer (10 m*M* Hepes, pH 7.0, 2 m*M* $MgCl_2$, 1 m*M* dithiothreitol, 1% Triton), 5 μg Rv3377c, and serial diluted inhibitors was heated to 37°C. The assay was initiated by the addition of 100μM GGPP and 100 n*M* [^{3}H]-GGPP. The reaction proceeded for 3 hours at 37 °C and was then terminated by heat deactivation of Rv3377c at 70°C. The reaction products were dephosphorylated by Rv3378c for another 3 hours at 37°C, then extracted with hexane. The resulting organic layer was transferred to a scintillation vial for counting. The specificity of Rv3378c was evaluated by performing the dephosphorylation assay using GGPP as substrate under the same conditions, where the reaction was confirmed to be much slower (<10%). The IC_{50} values were obtained by fitting the data to the dose-response curve in Origin 7.1 (OriginLab Corporation, Northampton, MA, www.OriginLab.com). Other details: Compound syntheses are reported in the SI Methods.

RESULTS

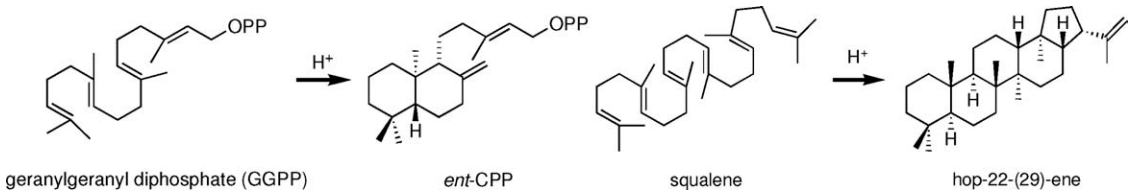
βγ Structure in bacterial diterpene cyclases

It has been known for some time[39] that many bacteria, in particular soil-dwelling bacteria, produce gibberellins, diterpenoids involved in plant growth regulation as phytohormones, although the relevant diterpene synthases have only recently begun to be elucidated. By way of example, consider the biosynthesis of kaurene in the bacterium *Bradyrhizobium japonicum*. Kaurene biosynthesis from GGPP is catalyzed by two enzymes: a (class II) copalyl diphosphate synthase (CPS; blr2149), which converts GGPP to CPP, and a (class I) kaurene synthase (KS; blr2150), which converts CPP to kaurene [Fig. 1(a)], with the two open reading frames overlapping by a single nucleotide. This and other bacterial class II diterpene

cyclases show homology to the previously identified eukaryotic enzymes, in particular conservation of the catalytic DxDD motif [the red box in Fig. 1(f)]. On the other hand, the bacterial class I diterpene synthases are composed of only a class I α-domain [the green box in Fig. 1(f)]. They contain two divalent metal binding (DDxxD like) motifs, and are homologous to the corresponding regions of plant class I terpene synthases. In plants, the same two products, CPP and kaurene, are also both made, but these biosynthetic reactions are carried out by much larger enzymes (containing ~900 vs. ~500 residues), leading to the idea that the plant enzymes might have arisen by fusion of the more ancient, bacterial class I and class II cyclases.[40] To investigate this possibility in more depth, we need to know more about the likely structures of the existing bacterial and plant enzymes.

In the case of the class I KS (blr2150), it was recently suggested, based on sequence alignments with a variety of other class I terpene cyclases, that there were two divalent cation binding motifs, required for catalysis. We thus sought to discover homologous structures by using the fold recognition feature embedded in the I-TASSER[31] program, to find proteins of known structure that also have good homology to the BjKS protein. There were three hits: *epi*-aristolochene synthase from the plant *Nicotiana tabacum* (TEAS; PBD File 5eau); 1,8-cineole synthase from the plant *Salvia fruticosa* (PDB File 2j5c), and pentalenene synthase fom the fungus *Streptomyces* sp. uc5319 (PDB File 1ps1). Partial sequence alignments are shown in Figure 2a, together with a homology model in Figure 2(b), and are consistent with the hypothesized conservation of the DDXXD motif, together with a second $NDX_6(D/E)$ sequence similar to that seen in other class I terpene cyclases.[41] Using the I-TASSER threading algorithm, we obtained a homology model [Fig. 2(b)] having a C-score of 0.26, a TM-score of 0.90 and an rmsd of 1.81A, all good scores,[31] plus, this structure was externally validated by PROSA_WEB[35] to be energetically favorable (a Z-score of −6.04). More remarkably, the observation that the top scoring structure for the bacterial class I diterpene cyclase was obtained with a plant sesquiterpene cyclase (TEAS) strongly supports the hypothesis that plant terpene cyclases may have originated from more ancestral, bacterial terpene cyclases, that is, by fusion of bacterial class I and class II diterpene cyclases. But what then would be the structures of the bacterial class II diterpene cyclases? And how did they arise?

What is particularly interesting about these class II bacterial enzymes (involved in diterpene biosynthesis) is that the DxDD (or very similar) catalytic motif is the same as that used by bacterial triterpene cyclases, such as SHC from *Acidobacillus acidocaldarius*, an enzyme whose three dimensional structure is known.[6] SHC produces hopane, the precursor to the hopanoid hydrocarbons found in very ancient sediments, potentially dated at ~2.7 Ga[42](gigaannum, SI, 10^9 years). This raises the intriguing possibility of a nexus between triterpene and diterpene cyclase structure, which might in fact be expected given that the reaction mechanisms for the bacterial diterpene and triterpene cyclases are quite similar:



with a catalytic DxDD protonating a double bond, in both GGPP and squalene. Supporting this structural connection, we found [Fig. 2(c) and Supporting Information Fig. S1] that this catalytic motif occupies the same position in bacterial class II diterpene and triterpene cyclase sequence alignments [Fig. 2(c) and Supporting Information Fig. S1]. These alignments also indicate the presence of two conserved "QW motifs" in the bacterial diterpene cyclases, and in the triterpene cyclases SHC and oxidosqualene cyclase (OSC). These "QW motifs" vary somewhat in composition but are typically QxxDGGWG or QxxDGSWG, and in the βγ triterpene cyclases, they lie between the outer helices of the $(\alpha/\alpha)_6$ barrels, Figure 2(d,e), and exhibit (in SHC) 6-fold symmetry in the β domain, [Fig. 2(d)]. The three dimensional structures of these sub-domains, or foldons, are very similar in the triterpene cyclases, with a ~0.7Å rmsd between their 24 backbone $NC_\alpha C'$ atoms, and ~0.8Å for all backbone and common side-chain atoms, and one representative comparison (between "QW motifs" in OSC and SHC) is shown in Figure 2(f) (OSC, green; SHC, cyan). This β-turn is also seen in the β-subunit of protein farnesyl transferases and geranylgeranyl transferases I and II, as shown in Figure 2(f). The observation that these "QW"-like motifs are present in all of the bacterial class II diterpene cyclases, in both triterpene cyclases, as well as in the β-domain of the protein prenyl transferases, and that they have the same spatial arrangement [Fig. 2(c,d)]

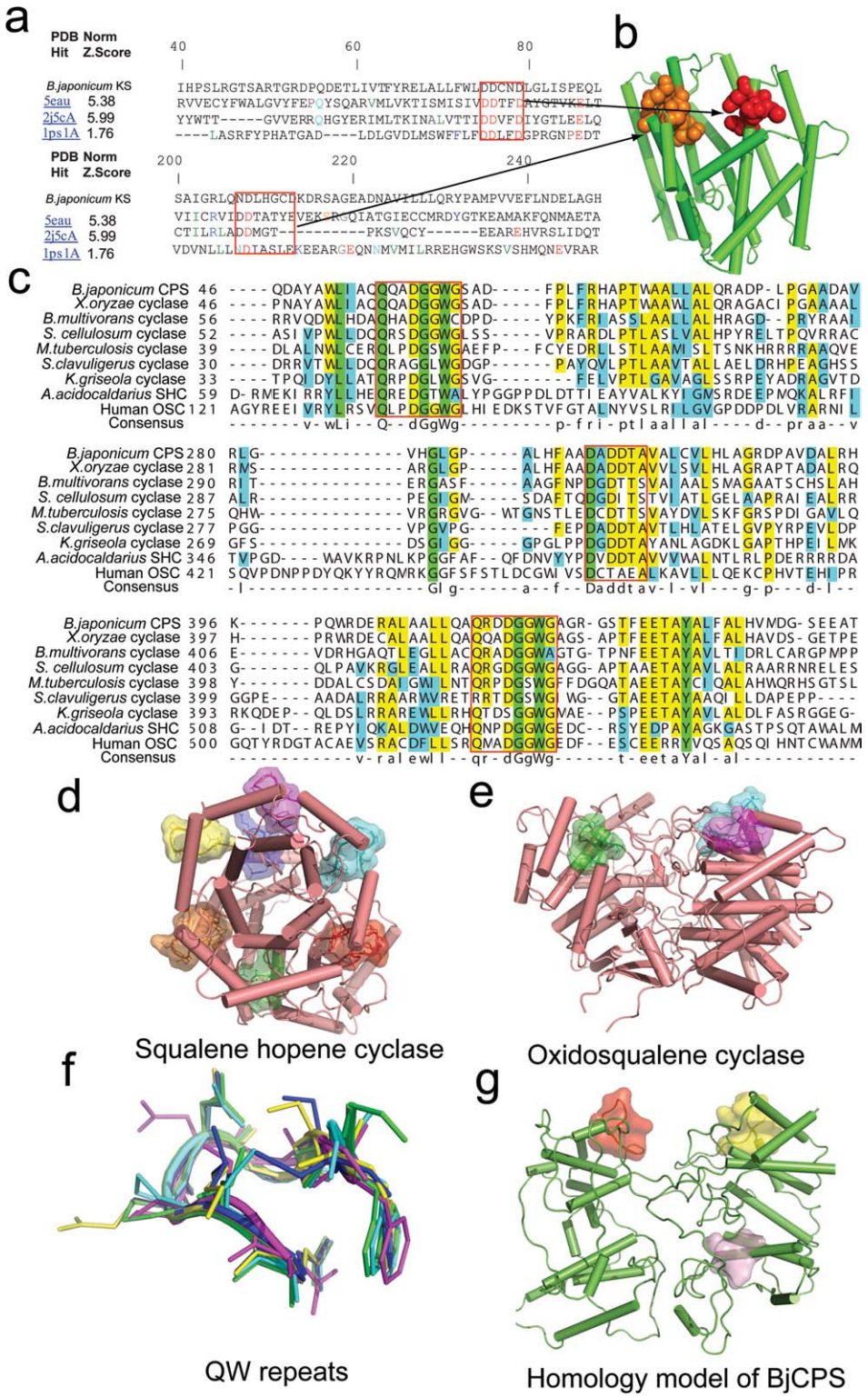


Figure 2

(**a**) Alignment of *Bradyrhizobium japonicum* Kaurene synthase (BjKS) with epi-aristolochene synthase (5EAU), 1,8-cineole synthase (2J5C), and pentalene synthase (1PS1A) by using the I-TASSER program. (**b**) Model of BjKS by using the I-TASSER program. The first DDXXD and the second $NDX_6(D/E)$ motifs are shown in space filling (orange and red, respectively). (**c**) ClustalW alignments of the di and triterpene cyclases showing "QW repeats" and the catalytic motifs. (**d**) Side-view of SHC showing six-fold symmetry of helices and "QW" repeats. (**e**) Front-view of OSC showing fewer "QW repeats." (**f**) Close-up view of QXXDGGWG repeats from β and γ domains showing very similar (<1 Å rmsd) 3D structures. (**g**) Homology model of *B. japonicum* CPS based on SHC (1sqc), showing "QW" repeats.

strongly suggests that the bacterial class II diterpene cyclases have a similar fold to that seen in the triterpene cyclases, and protein transferases.

Additional evidence for this comes from JPRED[18] or COUDES[20] predictions of helix and turn content. Theoretically, an $(\alpha/\alpha)_6$ β-barrel has 6 inner and 6 outer helices, 24 in all for the double $(\alpha/\alpha)_6$ β-barrel βγ structure. In SHC, 24 helices are predicted using the COUDES program, to be compared with 27 helices seen crystallographically (Supporting Information Table S1 and Fig. S2). In OSC, (which is larger than SHC), there are 26 helices predicted, as compared with 28 seen experimentally. In the *B. japonicum* CPS, 23 helices are predicted, in the *M. tuberculosis* halimadienyl diphosphate cyclase (Rv3377c gene product), 19. Similar results are seen when using the JPRED program.[18] The bacterial class II diterpene cyclases thus have a similar number of helices to those found in the two triterpene cyclases whose structures are known,[6,12] about twice as many as seen in the single β-domains of the protein prenyltransferases. We also find that the number of β-turns predicted in SHC, OSC and the three prenyl transferases are close to the number seen experimentally, and the predicted versus experimental result for the number of helices (N_h, blue) and the number of turns (N_t, red) for 10 α, β, or γ-domain containing proteins are shown in Supporting Information Figure S2 (slope = 0.98, R^2 = 0.97) with an average error of 1.7 (±0.8) helices or turns (Supporting Information Table S1). When taken together with the presence of the two aligned "QW repeats" and the positions of the catalytic motifs from the ClustalW sequence alignment between the bacterial class II diterpene cyclases and the triterpene cyclases [Fig. 2(c) and Supporting Information Fig. S1], we conclude that the diterpene cyclases also contain a βγ–domain structure, that is, two $(\alpha/\alpha)_6$ barrels with N_h ~19 to 23 and N_t ~ 19 to 26. To see this more graphically, we constructed a series of homology models for the bacterial class II diterpene cyclases, such as that for the *B. japonicum* CPS shown in Figure 2(g). The helix content (~20 helices) is close to that predicted from the neural network analysis (22 helices) and from the COUDES program (23 helices), the DxDD catalytic motif [pink, Fig. 2(g)] is in the position expected, as are the two (β,γ) "QW repeats" [orange, yellow, Fig. 2(g)], between the outer helices. But do plant diterpene cyclases have a similar organization? And how is Mg^{2+} involved in class II activity?

αβγ Structure in plant diterpene cyclases

Plant diterpene cyclases are much larger than bacterial diterpene cyclases, typically ~900 versus ~500 residues, but in many cases they contain the same DDxxD catalytic motif as found in terpene synthases such as FPPS, involved in ionization-initiated catalysis[4] [Fig. 3(a) and Supporting Information Fig. S3]. Plant monoterpene and sesquiterpene cyclases also use this motif, and the structures of several are known. What is surprising about these structures [such as bornyl diphosphate synthase (BS), Fig. 1(d), and TEAS] is that in addition to an FPPS-like catalytic α-domain, there is also a large β-domain, so these proteins have an αβ domain architecture. The β-domain appears to have the same $(\alpha/\alpha)_6$ fold as found in the triterpene cyclase SHC, with Wendt and Schulz[32] reporting (using the Dali program[33]) that C_α of 70 residues in SHC and TEAS superimposed within 3Å. Using this structure-based approach, it is apparent that the SHC βγ and TEAS αβ structures have a common β-domain [Fig. 3(b)], and in the region of the catalytic DxDD motif [DLCT in TEAS, Fig. 3(a)], we find that there is only a 1.1Å rmsd between the 18 backbone ($NC_\alpha C'$) atoms in the DxDD motif in SHC and the corresponding region in TEAS. Similar results are found between OSC and TEAS, and between OSC and BS or limonene synthase (LS; e.g., a 0.8Å rmsd between OSC and LS) and overall, the average deviation between all the aligned triterpene and mono- and sesqui-terpene synthase backbone structures in this region is only 1.1 ± 0.2 Å, indicative of good structural homology.

So, both bacterial diterpene and triterpene cyclases have βγ domain architecture, while some mono-terpene and sesqui-terpene cyclases have αβ architecture, and since the monoterpene and sesquiterpene cyclases have been shown to arise (in plants) via exon loss and recombination from ancestral diterpene cyclases,[14] we propose that the plant diterpene cyclases typically have an αβγ domain architecture. This is in good agreement with previous sequence alignment work[43] which indicated that the catalytic DxDD in the plant bifunctional (class I and II) diterpene cyclase abietadiene synthase (ABS) was located at the end of an α-helix in an $(\alpha/\alpha)_6$ barrel. This sequence alignment is now supported by a PSI-BLAST sequence alignments[17] of 228 terpene synthases [Fig. 3(a) and Supporting Information Fig. S3], plus, we find that there are 26 ± 1 helices predicted in the αβ domains of both TEAS and ABS, with the helices predicted in TEAS being the same as those seen crystallographically (PDB File # 5EAS). Also, circular dichroism results [Fig. 3(c,d)] show that both ABS and CPS from the plant *Arabidopsis thaliana* are highly helical (~65% helix content in ABS and 54% in *A. thaliana* CPS),[44,45] with <10% β-sheet content.

The Mg^{2+}-binding domain: site-directed mutagenesis

In the triterpene cyclases, there is no requirement for Mg^{2+}, but in the class II (H^+-initiated) diterpene cyclases, Mg^{2+} is required for effective catalysis.[28,29] There must, therefore, be a Mg^{2+}-binding motif present in the class II diterpene cyclases (which catalyze a H^+-initiated reaction), and the presence of this motif, which

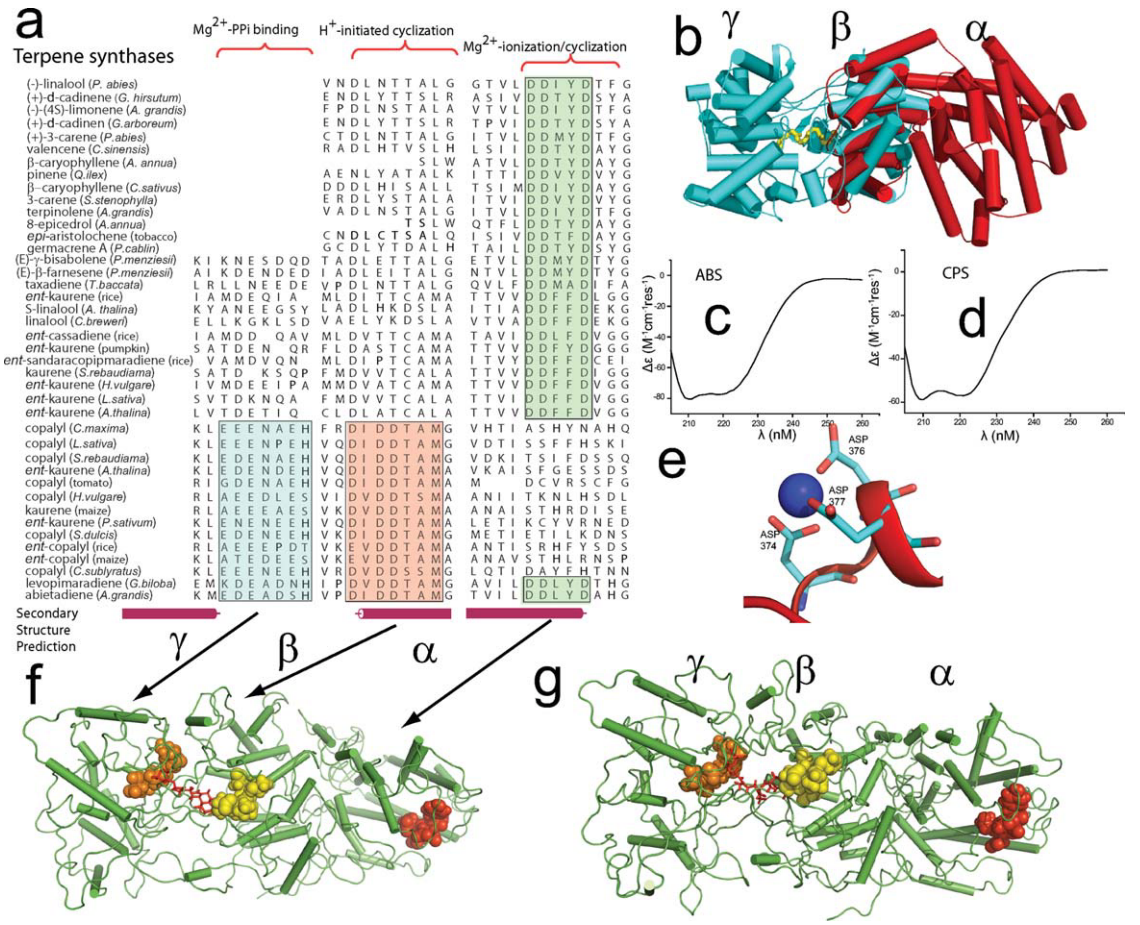


Figure 3

(**a**) Part of an alignment of 228 terpene synthases obtained from PSI-BLAST module in JPRED3[18] highlighting the catalytic domains in the α (green), β(red), and γ (cyan)-domains. The full alignment is shown in the Supporting Information Figure S3. The colored bars indicate residues thought to be important for activity: DDxxD in the α domain; DxDD (in general) in the β domain, and a D/E rich γ domain involved in Coulombic interactions with Mg^{2+} and the GGPP diphosphate. There is more plasticity in the organization of this domain since it just has to bind to diphosphate, but when there is a DxDD present there are on average 4 D/E residues here (cyan boxes), to be compared with only 2 when DVDD is absent (and catalysis is purely in the α domain). (**b**) Organization of terpene synthase structures showing α, β, γ domain structures. Superposition (EMBL SSM Program[21]) of squalene hopene cyclase (cyan) and (red), epiaristolochene synthase (PDB Files 1UMP, 5EAU) showing overlap of the β domains and relationship to location of SHC substrate-analog inhibitor, 2-azasqualene (in yellow). (**c**) circular dichroism spectra of *A. grandis* ABS. (**d**) circular dichroism spectra of *A. thaliana* CPS. (**e**) Schematic illustration showing binding of Mg^{2+} to the three active site Asp residues, obtained by using the MOE program.[33] Mg^{2+} is required for catalytic activity in H^{+}-initiated diterpene cyclases but at high levels can be inhibitory since it can bind to the catalytic Asp residues in *Ag*ABS. The diterpene cyclase model is based on the catalytic β domain in SHC, PDB File 1UMP. High levels of Mg^{2+} inhibit both the bacterial (*M. tuberculosis*) tuberculosinol diphosphate synthase as well as plant CPS and ABS diterpene cyclases. For convenience of reference, the numbering used is that reported for SHC. (**f**) Homology model of *A. grandis* ABS showing DDxxD (red), DxDD (yellow) and EDxxD (orange). Stick structure is CPP. (**g**) Homology model of *A. grandis* ABS with GGPP.

facilitates binding of GGPP, should correlate with the presence of the DxDD (or similar) motif, which initiates cyclization. As can be seen in Figure 3(a), a highly acidic "EDxxD-like" motif (cyan shading), capable in principle of binding Mg^{2+}, is in fact found in all plant class II diterpene cyclases, and it co-occurs with the DxDD motif (pink shading). This DxDD motif is found in the β-domain while, based on published sequence alignments[46,47] and our alignments from PSI-BLAST,[17] the EDxxD-like motif resides in the γ-domain (previously called the conifer diterpene internal sequence or "insertional" element), Figure 3(a). This highly anionic feature is absent in solely class I (ionization-initiated) diterpene cyclases that have a DDxxD containing α-domain, such as taxadiene synthase [TXS, Fig. 3(a)], as expected. It is also not readily apparent in the bacterial diterpene cyclases. This is not surprising, however, since when known, the cyclase products (copalyl diphosphate; halimadienyl diphosphate; and terpentedienyl diphosphate) have rather different structures, and the position of the diphosphate-binding site is likely to vary from one enzyme to another.

Table I
Effect of Single and Quad-Mutants on *A. grandis* ABS Catalytic Activity

Mutant	Substrates	Percentage of activity compared to wildtype[a]
D144A	GGPP	13%
	CPP	>90%
E140Q/D141N/E142Q/D144N	GGPP	2%
	CPP	>90%

[a]Data from duplicate assays.

To test the hypothesis that the acidic EDxxD-like domain is, in fact, important for catalysis in plant diterpene cyclases, we constructed two mutants in *Abies grandis* ABS. In the first (AgABS D144A), we mutated the last aspartic acid in the EDxxD motif to alanine, while in the second (AgABS E140Q/D141N/E142Q/D144N) we mutated the four carboxylic acid containing side chains to the corresponding amides. In the case of the single mutant, we found that enzyme activity decreased by a factor of 6 (using GGPP as substrate) (Table I), while in the case of the quad-mutant, there was a factor of 50 decrease in activity (Table I). When using CPP (copalyl diphosphate) as substrate, there was a <10% decrease in the activity of both mutants, because the second (class I) cyclization reaction (to abietadiene) occurs in the α-domain [the green box in Fig. 1(f)], and is thus not expected to be significantly affected by more distal mutations such as those in the γ–domain [the cyan box in Fig. 1(f)]. These results show that the EDxxD motif is involved in the class II (protonation-initiated) cyclization of GGPP to CPP in the plant diterpene cyclases, presumably facilitating binding of the diphosphate group of GGPP into the γ- domain active site region via electrostatic interactions with Mg^{2+}. It is also notable that this EDxxD-like motif is chemically similar to the ubiquitous DDxxD domains used for Mg^{2+}-dependent ionization-initiated catalysis in α-domain proteins such as FPPS. This raises the question as to how these similar motifs can be involved in diphosphate ester ionization in one case, but just Mg^{2+} binding in another. Apparently, location in a helix (e.g., DDxxD in FPPS) facilitates ionization-initiated catalysis, while location outside a helix, as in the case of the EDxxD domain, favors just Mg^{2+}-mediated diphosphate binding, due to a different spatial arrangement of the carboxyl groups. In addition, it is of course likely that two separate motifs (that bind multiple Mg^{2+}) are required for diphosphate ester ionization, as opposed to the single diphosphate binding interaction required in the class II (H^+-initiated) enzymes.

A similar Mg^{2+}-binding effect is also likely with the DxDD β-domain catalytic motif. In particular, although Mg^{2+} is important for catalysis in these class II (protonation-initiated) diterpene cyclases, its effects on activity can be complex, and in the case of ABS as well as CPS, Mg^{2+} at high concentrations actually has an inhibitory effect.[28] This is likely due to blocking of the H^+-initiation (DxDD) site by Mg^{2+} binding to two or three aspartate residues in the DxDD motif.[28,29] Based on the structures of the DVDD motif in SHC, Figure 3(e), we see that the three catalytic aspartates in the DxDD motif are well poised to interact with Mg^{2+}, forming a tridentate chelate, Figure 3(e), which would inhibit the H^+-initiated cyclization reaction (at high concentrations).

But how is the EDxxD motif in the γ-domain organized? Initial attempts at structure building based on homology with SHC or TEAS produced structures that had their EDxxD motifs at the end of an α-helix, as expected, but the distance between this motif and the catalytic DxDD was large and the EDxxD motif was solvent exposed since the actual sequence identity was very low (14.5%).[48] We thus next used a secondary structure based modeling approach, in MOE,[34] in which we first built, computationally, an SHC(βγ)-TEAS(α) chimera, then used this as a template to construct an ABS model, Figure 3(f,g). There is a 3.7 Å Cα rmsd for the ~600 aligned residues between the ABS model and SHC, and 3.8 Å Cα rmsd in the ~450 aligned residues between TEAS and the model. Plus, the z-score from PROSA_WEB structure analysis is −5.3.[35] In this model, we see that all three domains, α, β, and γ, contain their expected DDxxD, DxDD, or EDxxD motifs, and are well poised to interact with GGPP or CPP, Figure 3(f,g).

αβγ domain interactions in a plant diterpene cyclase

In previous reports[27] it was shown that individually expressed α- or βγ- domains of ABS had no class I (ionization-initiated) or class II (H^+-initiated) activity, unlike the situation found with the simpler bacterial βγ- domain containing enzymes, but when co-expressed (or simply mixed *in vitro*), the ABS α+βγ domains were active, suggesting the importance of an intact, correctly co-folded αβγ-domain architecture. This in turn suggested that it might be possible to construct "chimeras," such as $\alpha_1\beta_2\gamma_2$ and $\alpha_2\beta_1\gamma_1$, from $\alpha_1\beta_1\gamma_1$ and $\alpha_2\beta_2\gamma_2$ parents, and as a test of this hypothesis we investigated chimeras of the class II (H^+-initiated) *ent*-CPP and *syn*-CPP synthases from rice: OsCPS2 (*Oryza sativa* copalyl diphosphate synthase-2) and OsCPS4 (*Oryza sativa* copalyl diphosphate synthase-4), respectively) in which the α-domains were swapped. Based on the structural models described earlier, the predicted outcome of these experiments is that the chimeras would be active, but would only make CPP of the same stereochemistry as that produced by the parental CPS providing the βγ-domain. As predicted, $OsCPS2_{\beta\gamma}4_{\alpha}$ produced *ent*-CPP, while $OsCSP4_{\beta\gamma}2_{\alpha}$ produced *syn*-CPP (see Fig. 4), in agreement with the idea that while in plants the

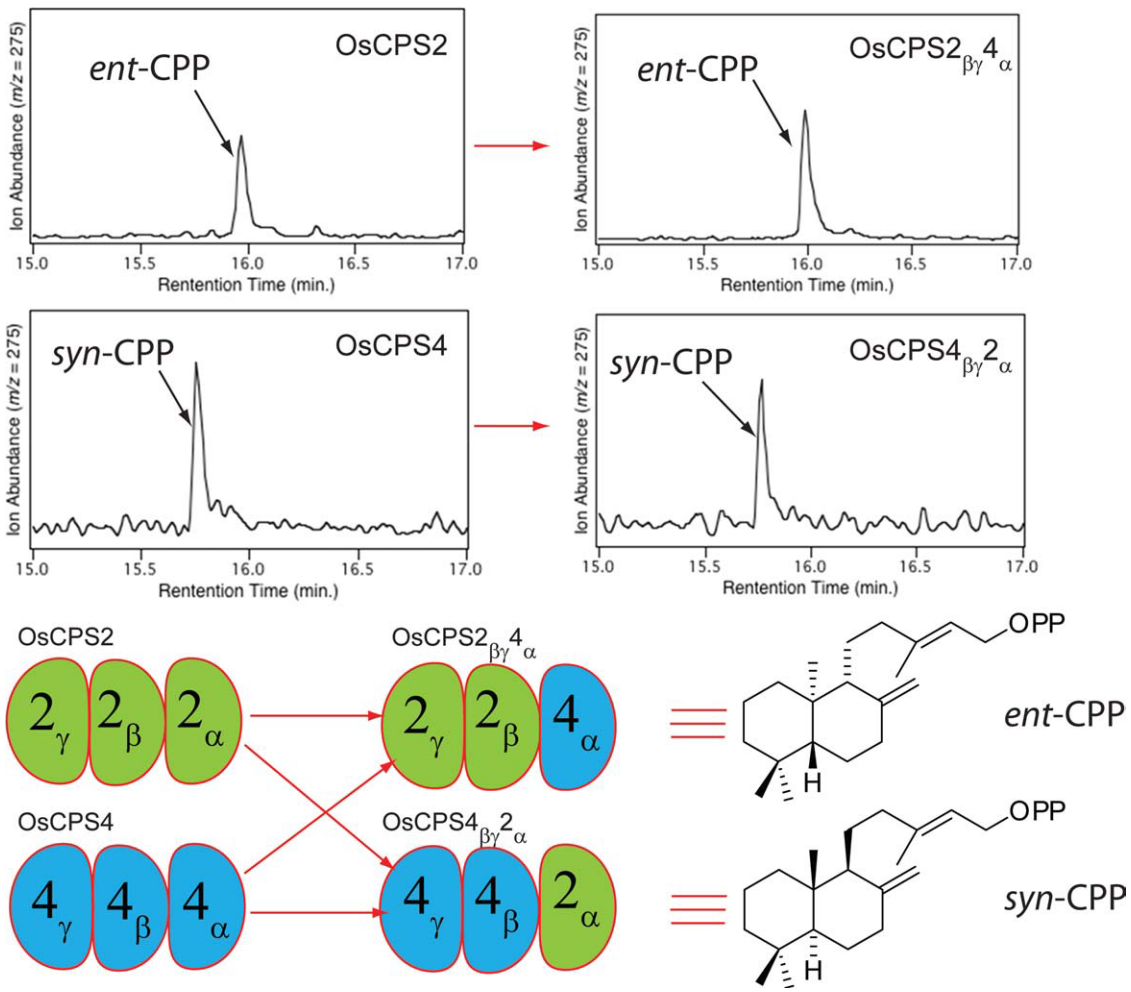


Figure 4

GC-MS ion-chromatograms showing (top) formation of *ent*-CPP and *syn*-CPP and (bottom), cartoon showing the αβγ domain in the chimeras coding for *ent*-CPP/*syn*-CPP biosynthesis. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

α-domain may be required for folding, it has no effect on the stereo-chemical outcome of the βγ-domain reaction.

Structure/function relationships for enzyme inhibition

We next considered how information on diterpene cyclase inhibition fits with the structure/function proposals described earlier. Such inhibitors are of interest as plant growth regulators, where for example, AMO-1618 inhibits gibberellin biosynthesis and 15-aza-GGPP is a tight binding transition state analog inhibitor of AgABS,[49] as well as in the development of anti-tuberculosis drugs inhibiting formation of the virulence factor edaxadiene.[50] Indeed, results obtained with the mycobacterial halimadienyl diphosphate synthase indicate strong binding interactions with 15-aza-GGPP and AMO-1618, providing additional support for the similar organization of eukaryotic and prokaryotic class II diterpene cyclases.[51,52] Of particular interest here is a comparison of the ability of known inhibitors of triterpene cyclases to inhibit the class II activity of diterpene cyclases, and to investigate these questions, we first determined the activity of 38 compounds (Supporting Information Fig. S4) which inhibit either Mg^{2+}-dependent ionization initiated (class I) or H^{+}-initiated (class II) activity, in ABS. We used a diphosphate-release coupled enzyme activity assay[30] to deduce the IC_{50} values for inhibition by all 38 compounds using GGPP as substrate (class II or I+II activity), together with a smaller subset of compounds using CPP as substrate, to assess class I activity. Selected results are shown in Figure 5. As can be seen in Figure 5 (Supporting Information Fig. S4 and Table S2), when using GGPP as substrate (Fig. 5, in blue), in essentially all cases the most potent inhibitors have cationic

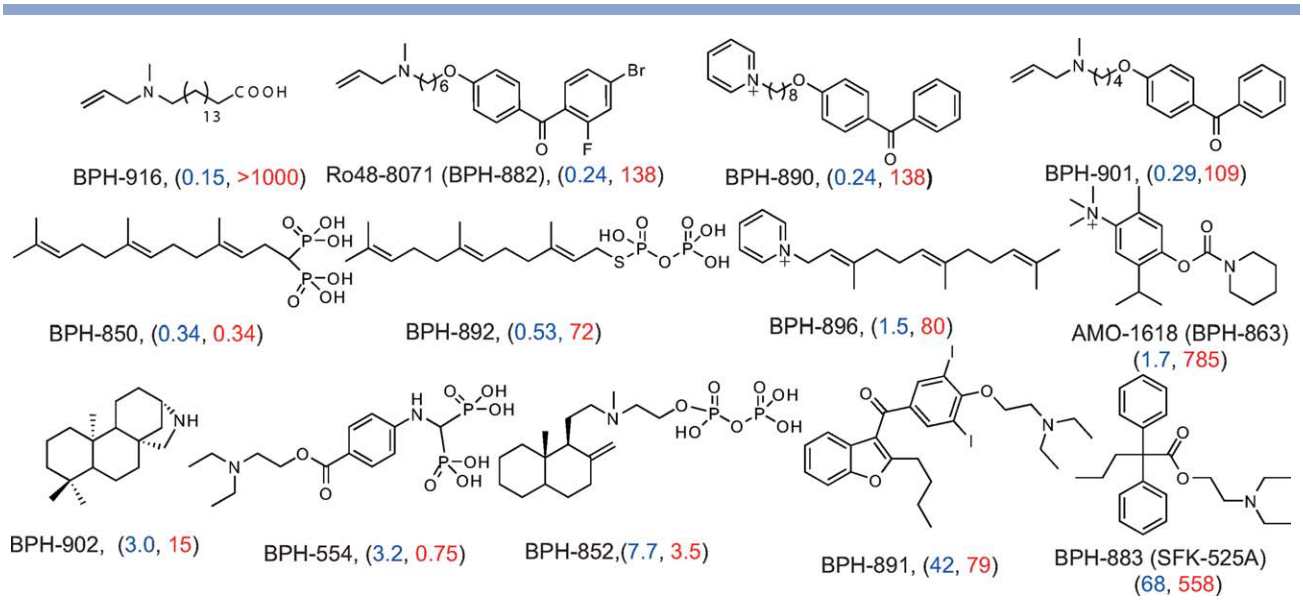


Figure 5

Inhibition results for *Arabidopsis thaliana* ABS: representative inhibitor structures together with IC_{50} values (in μM) for ABS inhibition: blue (GGPP substrate); red (CPP substrate).

(or basic) groups attached to long, hydrophobic side-chains. These inhibitors must bind to the catalytic DxDD motif in the βγ-domain (inhibiting class II activity), as they do when inhibiting squalene-hopene cyclase,[53] since they have essentially no inhibitory activity (see Fig. 5) when CPP is the substrate (Fig. 5, in red). The most potent inhibitor is the 15-aza analog of GGPP (Supporting Information Table S2), which can bind to both the DxDD motif (via the ammonium group) as well as the EDxxD-cluster (via Mg^{2+}). The second most potent inhibitor (when using GGPP as substrate) is the allylamine carboxylic acid BPH-916, but the activity of this compound (IC_{50} = 150 n*M*) is similar to that of its corresponding ester (BPH-922, IC_{50} = 250 n*M*). These results show, therefore, that the major requirement for inhibiting class II (H^+-initiated) cyclization is the presence of a hydrophobic, cationic species, consistent with the observation that mutation of the aspartate residues in the DxDD motif results in up to a 10^5 decrease in activity,[54] while mutations in the EDxxD-like diphosphate binding motif are much smaller (~50x). The cationic species also have little activity in inhibiting class I (ionization-initiated) cyclization (Fig. 5, in red). For example, the IC_{50} for AMO-1618, (BPH-863) is 1.7 μM in the H^+-initiated reaction, but ~1 m*M* for the ionization-initiated reaction. Interestingly, BPH-850 (farnesylmethylene bisphosphonate) has about the same IC_{50} using GGPP or CPP as substrate, but based on the results discussed earlier, this simply means that it inhibits primarily the class I site (in the α-domain). We further find that potent inhibitors of SHC and OSC such as BPH-882 (Ro48-8071) (SHC K_i = 6.6 n*M*; OSC K_i = 22 n*M*)[53] are also potent inhibitors of ABS (K_i = 4 n*M*), since the key requirements are the presence of hydrophobic and cationic features. The activities of the 28 class II site inhibitors (having measurable activity) can be quite well predicted (r^2 = 0.99, q^2 = 0.68) by using a comparative molecular similarity QSAR (quantitative structure activity relationship) method,[37] with activities being predicted within a factor of 4 (Supporting Information Table S2).

α, β, γ domain evolution

In the case of the *M. tuberculosis* diterpene cyclase (Rv3377c) that produces halimadienyl diphosphate, as expected, this enzyme is potently inhibited by aza-GGPP (5 n*M*), as well as by the most potent ABS inhibitor, BPH-916 (IC_{50} = 340 n*M*, structure shown in Supporting Information Fig. S4). On the other hand, hydrophilic bisphosphonates such as zoledronate and minodronate, potent inhibitors of α-domain isoprenoid synthases (such as FPPS), have no activity against the *M. tuberculosis* protein, consistent with the relatively weak binding of diphosphates in the γ-domain, and the lack of an α-domain. More surprising is the observation that these bisphosphonates also have no effect on ABS activity, so they do not inhibit either the diphosphate binding site in the γ-domain, or the catalytic site in the α-domain. The α-domain in these systems appears, therefore, to have a rather different organization to that found in more "conventional" isoprenoid synthases, such as FPPS. This divergence in structure is readily seen from the 3D-struc-

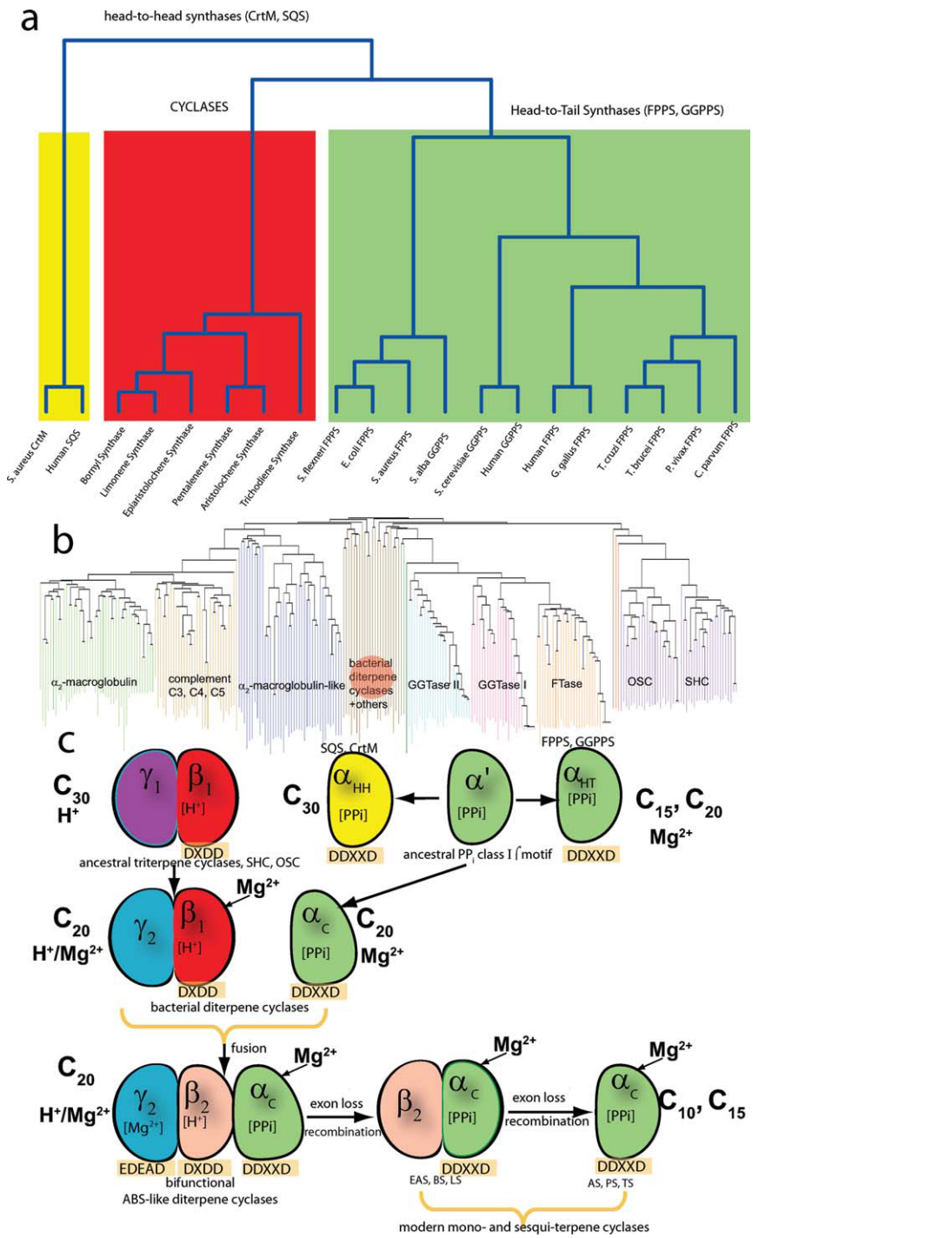


Figure 6

Proposed structural evolution of terpene cyclases and acyclic isoprenoid synthases. (**a**) Structure dendrogram based on data in Supporting Information Table S3 for α and αβ proteins showing strong clustering of GGPPS/FPPS, separate from that seen with the mono- and sesquiterpene cyclases. This suggests no direct evolutionary link from FPPS to terpene cyclases. (**b**) phylogenetic tree for *B. japonicum* CPS obtained by using the CDD conserved domain search program[55] showing homology of β-domain subunit in *B. japonicum* CPS and SHC, OSC, FTase, GGTase I, GGTase II, α_2-macroglobulins and complement C3-5 β-domain sequences. (**c**) Formation of an ancestral mixed-function diterpene cyclase by fusion of a βγ-domain protonation-initiated bacterial diterpene cyclase with an ancestral ionization-initiated α-domain synthase. The catalytic β-domain in the cyclase contains a DxDD catalytic motif, the α-domain has DDxxD. An EDxxD-like Mg^{2+} binding-motif is present in the plant diterpene cyclases and resides in the γ-domain and is responsible for the Mg^{2+}-dependence of the protonation-initiated cyclization. Loss of exons 4-6 in modern bifunctional diterpene cyclases and recombination leads[14] to αβ proteins such as TEAS. An evolution from βγ bacterial triterpene cyclases to βγ (soil) bacterial diterpene cyclases to αβγ plant diterpene cyclases is an attractive possibility, given the structural and functional similarities described in the text, and the observation that gibberellin producing soil bacteria live symbiotically with plants.

ture dendrograms [Fig. 6(a) and Supporting Information Fig. S5] that can be constructed from a matrix (Supporting Information Tables S3 and S4) of structure homology rmsd values[21,22] between 20 head-to-head, and head-to-tail terpene synthases, and mono- and sesquiterpene cyclases, which shows a clear separation between the three different structure types [Fig. 6(a) and Supporting Information Fig. S5].

As to the possible origins of the βγ–domain: the results of a PSI-BLAST[17] analysis (Supporting Information Fig. S6) show that the β-domain of for example, the *B. japonicum* CPS has homology to that seen in SHC/OSC, while the γ-domain shares homology with the plant diterpene cyclases, such as CPS. Moreover, the β-domain in *B. japonicum* CPS is homologous to the β–domain in all three protein prenyltransferases, as well as the α_2-macroglobulin and complement C3,4,5 proteins, Figure 6(b), pointing to a common $(\alpha/\alpha)_6$ barrel connection. The γ-domain does not share this similarity, indicating a separate origin. The γ-domain in *B. japonicum* CPS does have, however, strong sequence similarity to other diterpene cyclases, such as the halimadienyl diphosphate synthase (Rv3377c) (26.8% identity and 52.8% similarity), and maize and rice CPS (29.6% identity and 52.4 similarity with maize CPS, and 30.9% identity and 53.0% similarity with rice CPS), as well as other diterpene synthases (Supporting Information Fig. S6). The β-domain appears very ancient, because the SHC products, hopanoids, are found in very ancient sediments.[42] This domain catalyzes class II (H^+-initiated) cyclization in both diterpene and triterpene cyclases, while the γ-domain is more divergent, with the Mg^{2+} binding EDxxD motif seen only in diterpene cyclases, which based on the chemical fossil record are much more recent (≈200 Ma).[56]

DISCUSSION

The results presented above provide detailed new insights into diterpene cyclase structure and function, and are particularly illuminating in the context of diterpene cyclase and hence plant, terpene genesis. We find that the primary structures of bacterial class II diterpene cyclases have key sequence similarity (two aligned "QW repeats" and a catalytic DxDD motif) to that found in bacterial triterpene cyclases, where in one case the three dimensional structure is known.[6] This structure (of SHC) is similar to that found in the human triterpene cyclase, oxidosqualene cyclase, which catalyses the conversion of oxidosqualene to lanosterol, so these features are present in highly divergent triterpene cyclases, as well as in the bacterial class II diterpene cyclases. This points to the conserved nature (βγ) of the fold (arising from two ancestral $(\alpha/\alpha)_6$ barrels) for such H^+-initiated (class II) cyclizations. We thus propose that bacterial class II diterpene cyclases have a βγ-domain structure, similar to that found in bacterial triterpene cyclases such as SHC, based on: (1) the nature of the catalytic site (∼DxDD); (2) the number of helices (∼21 ± 2) and β-turns (∼23 ± 4); (3) the presence of sequence aligned "QW repeats" such as QxxDGGWG that are highly conserved in the bacterial class II diterpene and triterpene cylases, and in human OSC; and (4) enzyme inhibition results. These observations apply equally to the plant diterpene cyclases, where in addition, CD results support the expected overwhelmingly helical structure.

However, overall sequence similarity in the bacterial diterpene and triterpene cyclases is greater in the β-domains than in the γ domains (e.g., <10% identity and 25% similarity in the γ domain, and 16% identity and 32% similarity in the β domain between a bacterial diterpene cyclase (BjCPS) and a triterpene cyclase (*A. acidocaldarius* SHC)). This can also be seen for example, from the sequence alignments (Supporting Information Fig. S2) based on the bacterial class II diterpene cyclases, where we find that the β-domains show sequence similarity with the β-domain in the triterpene cyclases, as well as with other β-domains, while the γ-domain only shows significant sequence similarity with the plant class II diterpene cyclases (Supporting Information Fig. S7). Similar conclusions can be drawn from results obtained by using the CDD conserved domain search program,[55] as shown for the *B. japonicum* CPS in Figure 6(b), which again reflects the homology in β-domains from many varied class II terpene cyclases and other $(\alpha/\alpha)_6$-fold containing proteins. While these observations may at first seem surprising, they are actually not unexpected, based on chemical considerations. Specifically, the β-domains contain the catalytic DxDD motif, which carries out essentially the same reaction: protonation of a terminal isoprenoid unit, of squalene in SHC, or of GGPP in the class II diterpene cyclases. Since these reactions are quite similar, it seems reasonable that the basic barrel structure is conserved, although slightly fewer helices might be expected in the γ-domain in the diterpene cyclases, since the substrate is smaller (C_{20} vs. C_{30}), as are the protein molecular weights (∼50 k vs. ∼65 k Da). Overall sequence identity between the class II diterpene and triterpene cyclases is low, but in the two conserved "QW repeats" and in the catalytic domains, sequence conservation is readily apparent from normalized entropy scores.[24] For example, the average entropy score in the two QW motifs and the DxDD motif are 0.73, 0.75 and 0.71, respectively, compared to 0.37, the average entropy score in the other aligned regions.

The γ-domain in the bacterial class II diterpene cyclases shows, on the other hand, considerable overall homology with the γ-domain in plant class II diterpene cyclases (Supporting Information Figs. S3 and S7), but no significant homology (outside the conserved motifs) with the triterpene cyclases. This similarity again arises

because the plant and bacterial diterpene cyclases accommodate only a C_{20} species, not a C_{30} species. But why then is the β-domain homology between bacterial and plant diterpene cyclases not as high? One possibility is that fusion of ancestral α-domain (DDxxD-containing) synthases with ancestral βγ-domain cyclases resulted in αβγ hybrids in which the β-domain structure underwent reconstruction during fusion, although it is also possible that the plants acquired a different ancestral β (or βγ) domain, but in either case, the final fold appears to be similar, especially in the catalytic motif.

αβγ cyclases are present in even the most primitive plants. For example, in the bryophtye (moss) *Physcomitrella patens*, there is a bifunctional terpene cyclase[57] that has very strong sequence identity to the conifer *A. grandis* ABS, with common QxxDGxWGE and EDxxD-like motifs in the γ-domain, catalytic DxDD motifs in the β-domain, and a common DDxxD catalytic motif in the α-domain. This enzyme produces a mixture of 16α-hydroxy-*ent*-kaurene and *ent*-kaur-16-ene in a ~6 to 1 ratio, with the moss producing concentrations of the hydroxy species as high as 1 m*M*.[58] In higher plants, *ent*-kaurene is an intermediate in gibberellic acid (GA) biosynthesis, but in *P. patens*, key GA-biosynthesis enzymes are absent[59] and it appears that the diterpenes produced are used to regulate spore germination.[60]

But how did plant diterpene cyclases originate? A close association with ancient soil dwelling bacteria producing diterpenes is one possible route, and it is of interest here that in the soil bacterium *B. japonicum*, the CPS gene is part of a "symbiotic genome compartment" that contains a large cluster of genes involved in gibberellin biosynthesis.[61] And in the Mycobacteria, the Rv3377c gene involved in virulence factor formation is part of a genomic island in which there appears to have been horizontal gene transfer between *Rhizobium* or *Agrobacterium* spp. into environmental Mycobacteria.[62] The *B. japonicum* CPS is located in an operon adjacent to the KS (kaurene synthase) gene (indeed, the relevant open reading frames overlap by a single nucleotide), so ancestral bifunctional αβγ cyclases could have arisen from fusion of these or similar, ancient α and βγ terpene synthases and cyclases, Figure 6(c). A logical extension of these proposals is that the plant diterpene cyclases will have similar structures to the bacterial enzymes, since they use the same catalytic machinery to make the same molecules as do some bacterial diterpene cyclases: copalyl diphosphate and kaurene, with as proposed by Trapp and Croteau,[14] modern plant sesquiterpene and monoterpene cyclases then arising via exon loss and recombination [Fig. 6(c)]. Triterpene cyclase activity appears to be particularly ancient, because the SHC products, hopanoids, are found in very ancient sediments, potentially dating to $\sim 2.7 \times 10^9$ years,[42] while diterpene cyclase products appear much more recently (≈200 Ma) in the chemical fossil record.[55] Thus, triterpene cyclases, or closely related enzymes such as sporulene synthase[63] similarly operating on olefinic substrates (i.e., other than squalene), are the ancestral class II enzymes that then gave rise to class II diterpene cyclases as well. The bacterial class II diterpene cyclases are more modern, and have a modified γ-domain structure. These enzymes are mainly found in nitrogen-fixing and soil bacteria, such as *B. japonicum* and *Rhizobium* species and interestingly, these and other related proteobacteria such as *Rhodopseudomonas* spp., also make hopanoids, as well as gammacerane series triterpenoids,[61] emphasizing their close phylogenetic relationships.

The bacterial class II diterpene cyclases may thus represent a bridge between the ancient triterpene cyclases and modern diterpene (and thence, sesqui- and monoterpene) cyclases, with the observation that many of these bacteria are soil dwelling and/or are now associated with plants suggesting a route whereby their DNA might have become incorporated into plant DNA, in particular, into plant plastids, which are thought to have a bacterial origin. This bridge may also extend to the acquisition of more than just diterpene cyclase genes but also their targets, since in the case of modern plants, the GID1 gibberellin receptor has extensive sequence as well as three dimensional structural identity to bacterial esterases,[64] many of which are found in the proteobacteria.

ACKNOWLEDGMENTS

The authors thank Professor Tom Alber for providing plasmids for *M. tuberculosis* Rv3377c and Rv3378c. They also thank Professors Robert M. Coates, David Christianson, David Cane, Ann Pearson, Tanja Bosak, Carl Woese, and Jake MacMillan for their helpful comments, and R. M. Coates, Lise-Lotte Gunderson for providing several inhibitors. Figures were produced by using PyMol v0.99 2006, Delano Scientific LLC. (http://www.pymol.org).

REFERENCES

1. Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA 1998;95:4126–4133.
2. Christianson DW. Chemistry. roots of biosynthetic diversity. Science 2007;316:60–61.
3. Thulasiram HV, Erickson HK, Poulter CD. Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. Science 2007;316:73–76.
4. Tarshis LC, Proteau PJ, Kellogg BA, Sacchettini JC, Poulter CD. Regulation of product chain length by isoprenyl diphosphate synthases. Proc Natl Acad Sci USA 1996;93:15018–15023.
5. Greenhagen B, Chappell J. Molecular scaffolds for chemical wizardry: learning nature's rules for terpene cyclases. Proc Natl Acad Sci USA 2001;98:13479–13481.
6. Wendt KU, Poralla K, Schulz GE. Structure and function of a squalene cyclase. Science 1997;277:1811–1815.
7. MacMillan J, Beale MH.Diterpene biosynthesis. comprehensive natural products chemistry: isoprenoids including carotenoids and steroids, 1st ed., Vol.2. Oxford: Elsevier Science Ltd.; 1999. pp 217–244.

8. Maurer-Stroh S, Washietl S, Eisenhaber F. Protein prenyltransferases. Genome Biol 2003;4:212.211–212.219.
9. Starks CM, Back K, Chappell J, Noel JP. Structural basis for cyclic terpene biosynthesis by tobacco 5-*epi*-aristolochene synthase. Science 1997;277:1815–1820.
10. Hyatt DC, Youn B, Zhao Y, Santhamma B, Coates RM, Croteau RB, Kang C. Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. Proc Natl Acad Sci USA 2007;104:5360–5365.
11. Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA 2002;99:15375–15380.
12. Thoma R, Schulz-Gasch T, D'Arcy B, Benz J, Aebi J, Dehmlow H, Hennig M, Stihle M, Ruf A. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. Nature 2004; 432:118–122.
13. O'Maille PE, Malone A, Dellas N, Andes Hess B, Jr, Smentek L, Sheehan I, Greenhagen BT, Chappell J, Manning G, Noel JP. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. Nat Chem Biol 2008;4:617–623.
14. Trapp SC, Croteau RB. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 2001;158: 811–832.
15. Cyr A, Wilderman PR, Determan M, Peters RJ. A modular approach for facile biosynthesis of labdane-related diterpenes. J Am Chem Soc 2007;129:6684–6685.
16. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 2007;23:1282–1288.
17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
18. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics 1998;14:892–893.
19. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
20. Fuchs PF, Alix AJ. High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 2005;59: 828–839.
21. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004;60 (Part 12):2256–2268.
22. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
23. Felsenstein J. PHYLIP—Phylogeny inference package (version 3.2). Cladistics 1989;5:164–166.
24. Valdar WS. Scoring residue conservation. Proteins 2002;48:227–241.
25. Prisic S, Xu M, Wilderman PR, Peters RJ. Rice contains disparate *ent*-copalyl diphosphate synthases with distinct metabolic functions. Plant Physiol 2004;136:4228–4236.
26. Xu M, Hillwig ML, Prisic S, Coates RM, Peters RJ. Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. Plant J 2004;39:309–318.
27. Peters RJ, Carter OA, Zhang Y, Matthews BW, Croteau RB. Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. Biochemistry 2003;42:2700–2707.
28. Prisic S, Peters RJ. Synergistic substrate inhibition of *ent*-copalyl diphosphate synthase: a potential feed-forward inhibition mechanism limiting gibberellin metabolism. Plant Physiol 2007;144:445–454.
29. Prisic S, Xu J, Coates RM, Peters RJ. Probing the role of the DXDD motif in Class II diterpene cyclases. Chem Bio Chem 2007;8:869–874.
30. Webb MR. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. Proc Natl Acad Sci USA 1992;89:4884–4887.
31. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9:40.
32. Wendt KU, Schulz GE. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. Structure 1998;6:127–133.
33. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
34. Molecular Operating Environment (MOE). Montreal, Quebec: Chemical Computing Group, Inc.; 2006. http://www.chemcomp.com
35. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35 (Web Server issue):W407–W410.
36. Sybyl 7.3. St. Louis, MO: Tripos, Inc.
37. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 1994;37:4130–4146.
38. Bush BL, Nachbar RB. Sample-distance partial least-squares—PLS optimized for many variables, with application to COMFA. J Comput-Aided Mol Des 1993;7:587–619.
39. MacMillan J. Occurrence of gibberellins in vascular plants, fungi, and bacteria. J Plant Growth Regul 2001;20:387–442.
40. Morrone D, Chambers J, Lowry L, Kim G, Anterola A, Bender K, Peters RJ. Gibberellin biosynthesis in bacteria: separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. FEBS Lett 2009;583:475–480.
41. Christianson DW. Structural biology and chemistry of the terpenoid cyclases. Chem Rev 2006;106:3412–3442.
42. Brocks JJ, Buick R, Summons RE, Logan GA. A reconstruction of archean biological diversity based on molecular fossiles from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup. Hamersley Basin Western Australia Geochim Cosmochim Acta 2003;67:4321–4355.
43. Peters RJ, Croteau RB. Abietadiene synthase catalysis: conserved residues involved in protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate. Biochemistry 2002;41: 1836–1842.
44. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 2008;89:392–400.
45. Zhou K, Peters RJ. Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants. Phytochemistry 2009;70:366–369.
46. Kawaide H, Imai R, Sassa T, Kamiya Y. *Ent*-kaurene synthase from the fungus *Phaeosphaeria sp.*. L487 c DNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase in fungal gibberellin biosynthesis J Biol Chem 1997;272:21706–21712.
47. Martin DM, Faldt J, Bohlmann J. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. Plant Physiol 2004;135: 1908–1927.
48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
49. Peters RJ, Ravn MM, Coates RM, Croteau RB. Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. J Am Chem Soc 2001;123:8974–8978.
50. Mann FM, Xu M, Chen X, Fulton DB, Russell DG, Peters RJ. Edaxadiene: a new bioactive diterpene from *Mycobacterium tuberculosis*. J Am Chem Soc 2009;131:15726–15727.
51. Nakano C, Hoshino T. Characterization of the Rv3377c Gene Product, a type-B diterpene cyclase, from the *Mycobacterium tuberculosis* H37 genome. ChemBioChem 2009;10:2060–2071.

52. Mann FM, Prisic S, Hu H, Xu M, Coates RM, Peters RJ. Characterization and inhibition of a class II diterpene cyclase from *Mycobacterium tuberculosis*: implications for tuberculosis. J Biol Chem 2009; 284:23574–23579.
53. Abe I, Zheng YF, Prestwich GD. Photoaffinity labeling of oxidosqualene cyclase and squalene cyclase by a benzophenone-containing inhibitor. Biochemistry 1998;37:5779–5784.
54. Feil C, Sussmuth R, Jung G, Poralla K. Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase. Eur J Biochem 1996;242:51–55.
55. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 1988;85:2444–2448.
56. Marynowski L, Otto A, Zaton M, Philippe M, Simoneit BR. Biomolecules preserved in ca. 168 million year old fossil conifer wood. Naturwissenschaften 2007;94:228–236.
57. Hayashi K, Kawaide H, Notomi M, Sakigi Y, Matsuo A, Nozaki H. Identification and functional analysis of bifunctional *ent*-kaurene synthase from the moss *Physcomitrella patens*.. FEBS Lett 2006;580: 6175–6181.
58. Von Schwartzenberg K, Schultze W, Kassner H. The moss *Physcomitrella patens* releases a tetracyclic diterpene. Plant Cell Rep 2004;22: 780–786.
59. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin IT, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 2008;319:64–69.
60. Anterola A, Shanle E, Mansouri K, Schuette S, Renzaglia K. Gibberellin precursor is involved in spore germination in the moss *Physcomitrella patens*. Planta 2008;229:1003–1007.
61. Gonzalez V, Bustos P, Ramirez-Romero MA, Medrano-Soto A, Salgado H, Hernandez-Gonzalez I, Hernandez-Celis JC, Quintero V, Moreno-Hagelsieb G, Girard L, Rodriguez O, Flores M, Cevallos MA, Collado-Vides J, Romero D, Davila G. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. Genome Biol 2003;4:R36.
62. Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P. Contribution of horizontally acquired genomic islands to the evolution of the *Tubercle bacilli*. Mol Biol Evol 2007;24:1861–1871.
63. Kontnik R, Bosak T, Butcher RA, Brocks JJ, Losick R, Clardy J, Pearson A. Sporulenes, heptaprenyl metabolites from *Bacillus subtilis* spores. Org Lett 2008;10:3551–3554.
64. Shimada A, Ueguchi-Tanaka M, Nakatsu T, Nakajima M, Naoe Y, Ohmiya H, Kato H, Matsuoka M. Structural basis for gibberellin recognition by its receptor GID1. Nature 2008;456:520–523.