Chemistry & Biology *13*, 91–98, January 2006 DOI 10.1016/j.chembiol.2005.10.016

Engineering Cotton (+)-δ-Cadinene Synthase to an Altered Function: Germacrene D-4-ol Synthase

Yasuo Yoshikuni,[1,4] Vincent J.J. Martin,[2,4,6] Thomas E. Ferrin,[1,5] and Jay D. Keasling[1,2,3,4,*]
[1]UCSF/UCB Joint Graduate Group in Bioengineering
[2]Department of Chemical Engineering
[3]California Institute for Quantitative Biomedical Research
University of California, Berkeley
Berkeley, California 94720
[4]Synthetic Biology Department
Physical Bioscience Division
Lawrence Berkeley National Laboratory
Berkeley, California 94710
[5]Department of Pharmaceutical Chemistry and Biopharmaceutical Sciences
University of California, San Francisco
San Francisco, California 94143

Summary

The combined approaches of rational design and random mutagenesis were applied to generate a sesquiterpene synthase with an altered activity. Due to the lack of a convenient screen for sesquiterpene synthase activity, a high-throughput dual-activity screen was used by fusing (+)-δ-cadinene synthase to chloramphenicol acetyltransferase (CAT). The gene encoding (+)-δ-cadinene synthase was mutagenized using error-prone PCR. The resulting mutant fusion proteins were screened for CAT activity and altered sesquiterpene selectivity. Twenty-one clones producing (+)-δ-cadinene and germacrene D-4-ol in different ratios were isolated from the library. Analysis using a homology model of (+)-δ-cadinene synthase suggested that the G helix plays a very important role in (+)-δ-cadinene formation. Reconstruction of the G helix using site-directed, saturation mutagenesis yielded a mutant, N403P/L405H, that maintained its specific activity and showed higher selectivity to germacrene D-4-ol in vivo (up to 93%).

Introduction

Sesquiterpenes, 15-carbon isoprenoids, are a structurally diverse family of natural compounds [1]. There are large numbers of known sesquiterpene structures with different regio- and stereochemistry. Due to the wide variety of biochemical functions in organisms, such as antimicrobial, antifungal, herbicidal, and hormonal activities, many of these compounds have found use as medicines, pesticides, fragrances, and flavors. However, it is often difficult to isolate sesquiterpenes in large quantities and high purity, since plants usually produce extremely low quantities in complex mixtures with various structural isomers. To achieve mass production of terpenes in high purity, we described an alternative method involving overexpression of a sesquiterpene synthase and the mevalonate pathway genes in *Escherichia coli* [2, 3]. The method is solely dependent on having a synthase gene available for the synthesis of the targeted terpene. Although the number of cloned sesquiterpene synthase genes has increased dramatically over the last few years, these isolated genes fall short of covering all possible structures of interest. Therefore, we investigated an alternative method to create sesquiterpene synthases with altered activities using a combination of conventional directed evolution and rational design.

The formation of sesquiterpenes in the enzyme active site is initiated by the cleavage of pyrophosphate group from the substrate farnesyl pyrophosphate (FPP). The reaction is driven by surrounding amino acid residues and divalent metals located in the upper part of the active site cleft [4–7]. The resulting substrate cation is then shaped based on the inner structure of the active site cleft. This reaction is quenched by either nucleophilic attack by a water molecule or proton abstraction from the substrate. Due to higher reactivity and complexity of carbocation rearrangement, sesquiterpene synthases often show promiscuous catalytic function. Among those, γ-humulene synthase and δ-selinene synthase, constitutively expressed in *Abies grandis*, catalyze the formation of more than 50 and 30 different sesquiterpenes, respectively [8, 9]. On the other hand, (+)-δ-cadinene synthase from *Gossypium arboreum* is one of the most specific sesquiterpene synthases (>98% to (+)-δ-cadinene). In cotton, (+)-δ-cadinene is the primary precursor for phytoalexin production in response to invasion by a pathogen [10, 11].

(+)-δ-cadinene synthase was targeted for engineering because it showed the highest in vivo productivity of three enzymes previously tested in our laboratory [2]. Error-prone PCR (EP-PCR) was first applied to improve the desired function of the enzyme [12–14]. EP-PCR involves the iterative introduction of random mutations to a target gene through rounds of PCR, followed by screening of clones for the desired function. The method has been limited to enzymes for which an effective screen or selection exists, because the number of clones that needs to be screened to find a desired mutation is extremely large (10,000–1,000,000 or more). Unfortunately, the best method available to measure sesquiterpenes is gas chromatography-mass spectrometry (GC-MS), a low-throughput method that can analyze approximately 100 clones per day. Therefore, a complementary approach was necessary to reduce the number of clones screened to a reasonable number. Several groups have reported that, in fusing soluble reporter proteins to the target protein, the solubility of the fusion protein is likely proportional to the solubility of the target protein; the solubility of the target enzyme could be detected by measuring activity of the reporter protein [15–17]. Fusion reporter systems not only exclude clones that produced insoluble proteins, but also allow the elimination of nonsense mutations from

*Correspondence: keasling@berkeley.edu
[6]Present address: Department of Biology, Concordia University, 7141 Sherbrooke West, Montreal, Quebec H4B 1R6, Canada.

the pool. This system proved to be very useful for our target enzyme, since no simple screen or selection was possible.

Site-directed mutagenesis has been used to identify amino acids thought to be involved in the terpene synthase reaction mechanism [9, 18–25]. For terpene synthases, the consensus aspartate-rich motif is important in metal binding and catalytic activity. Mutations to these residues can alter the catalytic mechanisms of enzymes. For example, mutation of the aspartate (D100E) in trichodiene synthase attenuated the substrate-inducible closure of the active site cleft, which allowed the substrate to cyclize by alternative pathways in the increased catalytic space and generated a greater sesquiterpene structural diversity [18]. On the other hand, similar mutations to δ-selinene and γ-humulene sesquiterpene synthases resulted in reduction of product complexity (tricyclic products to macro- or monocyclic products) [9]. Although several other residues were known to be catalytically important, their mutation also resulted in significant loss of specific activity [9, 18–25]. In most cases, it is very difficult to rationalize the catalytically important residue-function relationship.

Here, we report the application of the combination of directed evolution and rational design to construct sesquiterpene synthases with altered catalytic properties, while maintaining specific activity. Since the structure of (+)-δ-cadinene synthase has not been elucidated, a homology structural model was built [26, 27] using the structure of 5-*epi*-aristolochene synthase [5] as a guide to rationalize the results obtained from random mutagenesis. The two sesquiterpene synthases share 45% identity at the primary sequence level. Homology modeling and mutagenesis identified the structural importance of the G helix in maintaining (+)-δ-cadinene synthase catalytic activity and successfully converted the enzyme to a germacrene D-4-ol synthase. In the future, the successful engineering of terpene synthases and their resulting products may have a significant impact on the design of novel terpenes for the flavors, fragrances, neutraceuticals, and pharmaceutical industries.

Results and Discussion

Selection Using (+)-δ-Cadinene Synthase-Chloramphenicol Acetyltransferase Fusions

Directed evolution is a powerful tool to alter enzymatic function. The method is, however, highly dependent on the availability of a high-throughput screening method. Although we previously developed an in vivo method to analyze the functions of terpene synthases using GC-MS, the method is extremely low throughput, processing ~100 samples/day at most [2]. To compensate for the drawbacks of screening by GC-MS, the 3′ end of mutagenized terpene synthase genes were fused to the 5′ end of the chloramphenicol acetyltransferase (CAT) gene, and the clones were screened for CAT activity. This method is based on the fact that the solubility of the C-terminal fused reporter protein is dependent on the solubility of the N-terminal fused protein [15–17]; hence, mutations producing variants with lower solubility could be excluded by this step.

The CAT gene was cloned into the vector pTrc99A to yield pTrcCAT. To test the solubility screening system, a green fluorescent protein (GFP)-CAT fusion was made by cloning the gene encoding a variant of GFP, GFPuv, 5′ to the gene encoding CAT in pTrcGFPCAT. A short, flexible linker consisting of the amino acids G and S was used to fuse the proteins. Although it is still unclear what kind of linkers should be used, short, flexible linkers with little or no secondary structure, or linkers that form a single α helix, are usually preferred [28, 29]. The strain harboring the GFP-CAT fusion fluoresced and was able to grow in the presence of 100 μg/ml chloramphenicol. The in vivo production of (+)-δ-cadinene using the wild-type (+)-δ-cadinene synthase fused with CAT reporter (pTrcCADCAT) was 4-fold lower than when using the native (+)-δ-cadinene synthase. Clones harboring the gene encoding the fusion protein grew on plates containing 10 μg/ml chloramphenicol. The chloramphenicol concentration that inhibited growth of 90% of the clones (IC_{90}) was 40 μg/ml, and this concentration was used to exclude mutations producing clones with poor CAT activity (less soluble mutants) and potentially find clones with improved CAT activity (more soluble mutants).

The optimal EP-PCR condition was established by determining the ratio of nonfunctional to functional enzyme as a function of the $MnCl_2$ concentration. A concentration of 0.2 mM $MnCl_2$, introducing 2–5 mutations per 1 kb of DNA polymerized, was used for the EP-PCR condition. Although this mutation rate is relatively high (up to 90% of the resulting clones significantly or completely lost in vivo sesquiterpene productivity), the high error rate should increase the likelihood of finding a critical amino acid involved in determining synthase product specificity. When plating cells transformed with EP-PCR libraries of CAD-CAT fusions on solid medium, several colony sizes were observed. When those with comparatively large size were screened, the ratio of nonfunctional to functional synthase was 1.2, whereas a ratio of 5 was obtained when those colonies with small sizes were screened. Thus, it is implied that approximately 87% of the nonfunctional enzymes also had reduced CAT activity (smaller colonies) and can be excluded by this step. This is a significant enrichment of functional enzymes. Therefore, the selection of clones using CAT activity and large colonies is reasonable for significantly enriching functional clones to be screened by GC-MS.

Directed Evolution of (+)-δ-Cadinene Synthase to Germacrene D-4-ol Synthase

After the initial screening for retained or improved CAT activity, the in vivo sesquiterpene production for each clone was measured using GC-MS. Screening of only 100 library clones identified a first mutation, L405S, producing a new sesquiterpenoid. The product was identified as germacrene D-4-ol by comparison of its MS spectrum to that of the pure compound found in the NIST MS spectrum library (version 1.6d) (Figures 1B and 1C). Using the L405S variant as a template, a second round of EP-PCR was carried out. The double-mutant, N403K/L405S, producing an increased amount of germacrene D-4-ol, was isolated by screening an additional 100 clones (Figures 2A and 2B: clone no. 20). Finally, the N403K/L405S variant was used as a template for a third round of EP-PCR. Screening of 400 library clones from

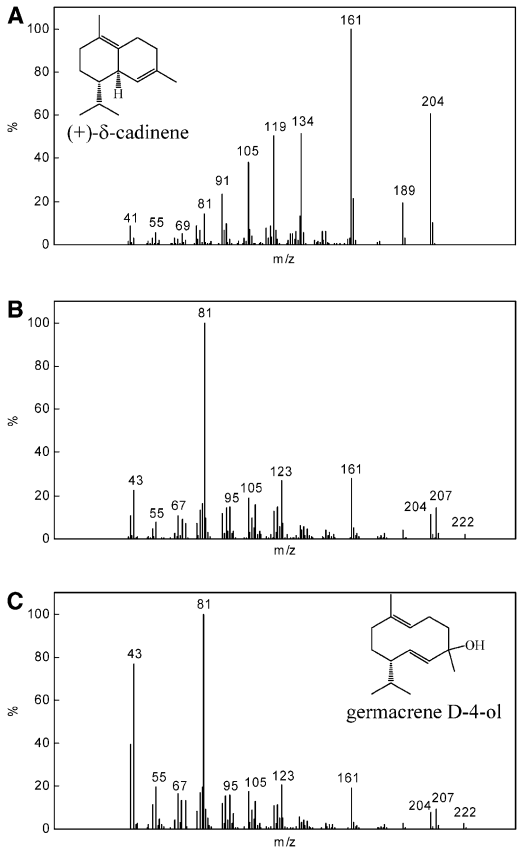

Figure 1. Mass Spectra of (+)-δ-Cadinene and Germacrene D-4-ol

Mass spectra are shown for (+)-δ-cadinene standard (A), the product obtained from the mutated (+)-δ-cadinene synthase (B), and from the NIST library for germacrene D-4-ol (C). Differences in the abundances of smaller molecular ions in (B) and (C) may be a result of a lower sensitivity in the library spectrum. Alternatively, the molecule represented by the spectra may be an epimer at the 4-hydroxyl group.

the third round of EP-PCR did not yield any further mutations resulting in increased productivity of germacrene D-4-ol. Since we sought to identify other mutations involved either in the formation of germacrene D-4-ol or other products, a second EP-PCR library using the wild-type (+)-δ-cadinene synthase gene as a template was screened. From 1000 EP-PCR library clones selected using the CAT assay, 20 more clones that showed selectivity toward germacrene D-4-ol were obtained (Figures 2A and 2B). DNA sequencing of the mutations determined that every clone involved in selectivity of (+)-δ-cadinene synthase was unique (Figures 2A and 2B).

Analysis of Mutations Using a Homology Structural Model for (+)-δ-Cadinene Synthase

A three-dimensional model for (+)-δ-cadinene synthase was built [26, 27] using the structure of 5-*epi*-aristolochene synthase (PDB code: 5EAT) as a guide [5]. Of the terpene synthases identified to date, 5-*epi*-aristolochene synthase has the highest degree of sequence identity (45%) to (+)-δ-cadinene synthase. Farnesyl hydroxy phosphate (FHP), used in the structural study of 5-*epi*-aristolochene synthase, was incorporated into the catalytic site of the modeled (+)-δ-cadinene synthase. Interestingly, most residues identified in clones 12–21, which had the highest fraction of germacrene D-4-ol production, were mapped in or near the G helix of the structural model (Figure 3).

We considered the possible reaction mechanisms for (+)-δ-cadinene synthase and the most likely route to germacrene D-4-ol formation (Figure 4). The cyclization reaction is initiated by cleavage of the pyrophosphate group from the hydrocarbon chain of FPP to yield the *transoid*-farnesyl cation. The C2-C3 bond is readily isomerized to form the *cisoid*-farnesyl cation. It then undergoes an initial cyclization and yields the 3*Z,* 7*E*-germacryl cation, the positive charge of which is delocalized over C2, C3, and C4. If the cation is accessible to a water molecule, the cyclization reaction is most likely quenched by hydroxylation at C4. On the other hand, if it is not accessible to a water molecule, the cation undergoes a second cyclization reaction, and the reaction is quenched enzymatically. When the location of FHP was compared to the modeled structure of (+)-δ-cadinene synthase (Figures 5A and 5C), the C4 of FHP was located under the G helix. These data suggest that the G helix plays an important role in protecting the substrate carbocation from exposure to solvent.

Reconstruction of the G Helix by Saturation Mutagenesis at F400, N403, and L405

The structural model and results of random mutagenesis for (+)-δ-cadinene synthase indicate that reconstruction of the G helix may allow alteration of product selectivity. Although D452N significantly affected product selectivity, this residue was not of interest because it is part of the conserved aspartate-rich motif and because mutations to this residue are often accompanied by significant reduction in specific activity [9, 22, 23, 25]. Therefore, F400, N403, and L405 were chosen as targets for site-directed saturation mutagenesis to randomly insert any of 20 amino acids at these positions. The single N403P mutation (G1 in Figures 2A and 2B) significantly altered the product selectivity to germacrene D-4-ol (52%). As the introduction of proline into α helices usually interferes with their formation, it is possible that the mutation disrupted the local secondary structure of the G helix, which allowed water access into the active site cleft and to the substrate carbocation. Introduction of mutation L405H (G2 in Figure 2) altered product selectivity from (+)-δ-cadinene to germacrene D-4-ol (53%). Interestingly, the introduction of L405H to N403P decreased k_{cat}/K_m but improved the in vivo productivity (Figures 2A–2D and Table 1). This result suggests that L405H improved the solubility of the enzyme.

The differences in the enrichment of germacrene D-4-ol relative to (+)-δ-cadinene between experiments performed in the presence (Figures 2A and 2B) and absence (Figures 2C and 2D) of a dodecane overlay can be explained by faster evaporation of (+)-δ-cadinene compared with germacrene D-4-ol. When dodecane was present in the enzyme assays, it trapped the (+)-δ-cadinene that would have evaporated from the assay mixture had the dodecane not been present. Hence, the germacrene D-4-ol produced in vivo using G2 was

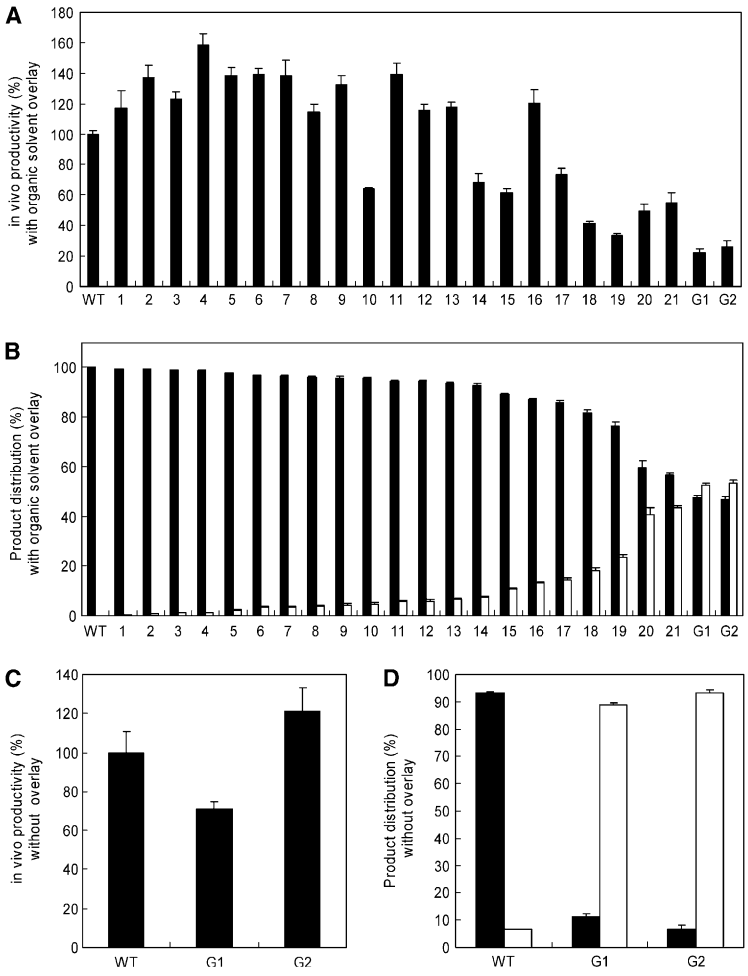

Figure 2. Product Selectivity of (+)-δ-Cadinene Synthase and Its Variants

The total in vivo productivity (A) and selectivity (B) for (+)-δ-cadinene and germacrene D-4-ol production for each of 21 mutations generated from random mutagenesis is shown. For (B) and (D), black and white bars represent (+)-δ-cadinene and germacrene D-4-ol product distributions, respectively. These cultures were overlaid with organic solvent to accumulate all products. 1, E457D; 2, V368A/E463G/K538Q; 3, T506A/V536A; 4, Y484C; 5, Q357L/R364G/I424V/D436G; 6, Q357L/R364G/I424V/D436G; 7, D436V/D464G; 8, E361V/F400S; 9, F400S; 10, D465G; 11, E336D/L405S; 12, W388R/L390H/T407A; 13, D452V; 14, A360D/F400L/K431E/H460L; 15, F400L/Y410C/F418Y; 16, F397L/A414V/K431V; 17, S417T; 18, C447G/S467T; 19, T407N; 20, N403K/L405; and 21, D452N. Two mutations generated from structural analysis and site-directed saturation mutagenesis are also shown. G1: N403P; and G2: N403P/L405H. The total in vivo productivity (C) and selectivity (D) for (+)-δ-cadinene and germacrene D-4-ol production for WT, G1, and G2 are shown. In the absence of an organic solvent overlay, G1 and G2 produced 93% pure germacrene D-4-ol. Error bars represent standard deviation from triplicate measurements.

of significantly higher purity (93%) than that produced in vitro using the same enzyme. To our knowledge, this is the first sesquiterpene synthase that produces germacrene D-4-ol as its major product.

Slightly above N403, there is a hydrophobic core formed by several hydrophobic residues (Figure 6). Site-directed saturation mutagenesis of one of these residues (F400) revealed that Phe was the best amino acid for this position, since all substitutions resulted in a significant decrease of the in vivo sesquiterpene productivity (data not shown). From these observations, it appears that the hydrophobic core, consisting of F397, F400, A468, C471, and Y472 (Figure 6A), plays an important role in tightening the two adjacent α helices together to avoid exposing the catalytic active site cleft to the outer environment. It is not clear whether the water molecule entered the active site before or after substrate binding. Our computational model and energy optimization calculations predicted only the water molecule-accessible hole at the G helix region (Figures 5B and 5D) proximal to C4 of the germacryl cation. In addition, it is interesting to note that nerolidol was not produced from the mutant N403P/L405H. Since nerolidol would be produced upon hydroxylation at the C3 of the farnesyl cation (corresponding to C4 of germacryl cation), it is inferred that the farnesyl cation was not exposed to the solvent upon activation of the pyrophosphate, and that the location of the water molecule within the enzyme cavity was specific, as only germacrene D-4-ol was produced by this mutant. However, from our data, we cannot establish if the water molecule was present before or enters after FPP binding to the active site, or if it enters the active site through the opening created at the G helix region.

Kinetic Analysis of (+)-δ-Cadinene Synthase and Its Variants

The wild-type (+)-δ-cadinene synthase and the G1 and G2 variants were expressed in *E. coli* BL21(DE3). The proteins were purified to ~95% purity, and specific activities were measured. Wild-type (+)-δ-cadinene synthase had higher k_{cat} and K_m than the G1 and G2 variants (Table 1). Interestingly, the specific activity (k_{cat}/K_m) of the wild-type enzyme was maintained in G1 and G2. We hypothesized that the screening of variants first by CAT activity followed by in vivo sesquiterpene productivity resulted in the selection of mutations that maintained a high specific activity, since the in vivo substrate concentration could be predicted to be significantly lower than the K_m for these enzymes.

Significance

Our primary objective was to develop an effective method to identify sesquiterpene synthase residues

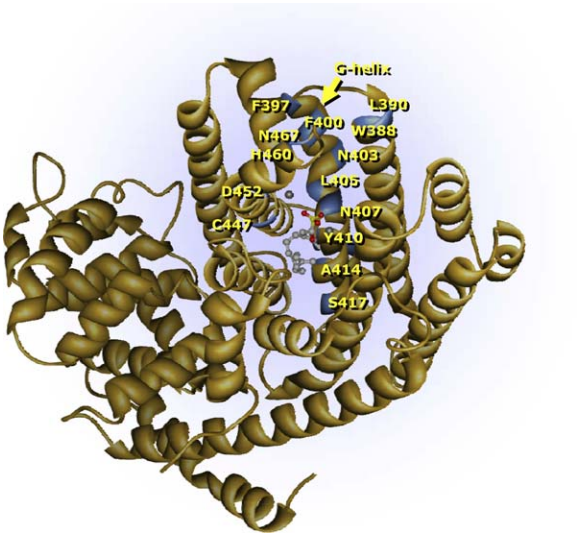

Figure 3. Homology Structural Model of (+)-δ-Cadinene and Structural Annotation of the Mutations Obtained from Random Mutagenesis

The homology structural model for (+)-δ-cadinene synthase was built using the crystal structure of 5-*epi*-aristolochene synthase as a guide. All mutations found in clones 12–21 are represented in blue. Interestingly, all mutations occurred in amino acid residues located in or near the G helix, indicating that the G helix plays an important role in protecting the substrate carbocation from exposure to the outer solvent.

involved in product selectivity and construct a sesquiterpene synthase with altered function. Using a combination of directed evolution and rational design, we obtained mutated enzymes, the final products of which were different from those of their parents but maintained their in vivo productivity. The combination of high-rate EP-PCR and the CAT fusion system significantly improved the library quality: up to 87% of nonfunctional clones generated from EP-PCR were excluded, and the population of functional clones in the library was enriched up to 7.5-fold. These methods allowed us to isolate 21 clones showing a similar phenotype (production of germacrene D-4-ol). Mapping the mutations on a homology structural model identified the G helix as playing an important role in protecting the substrate carbocation from exposure to the outer environment. When mutations were focused on the G helix, we obtained the N403P/L405H mutant, which produced germacrene D-4-ol with 53% selectivity while maintaining specific activity. In the absence of an organic solvent overlay to trap (+)-δ-cadinene, N403P/L405H produced 93% germacrene D-4-ol. We believe that the methods and results described here will aid in the production of novel and pure isoprenoids to be used as precursors for pharmaceuticals, neutraceuticals, pesticides, flavors, and fragrances.

Experimental Procedures

Plasmids and Microbial Strains

The analysis of sesquiterpene activity was performed in *E. coli* DH5α (F$^-$ φ80d *lac*ZΔM15 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17(r^-_k, m^+_k) *sup*E44 relA1 *deo*RΔ(*lax*ZYA-*arg*F)U196). All restriction enzymes and T4 DNA ligase were purchased from New England Biolabs.

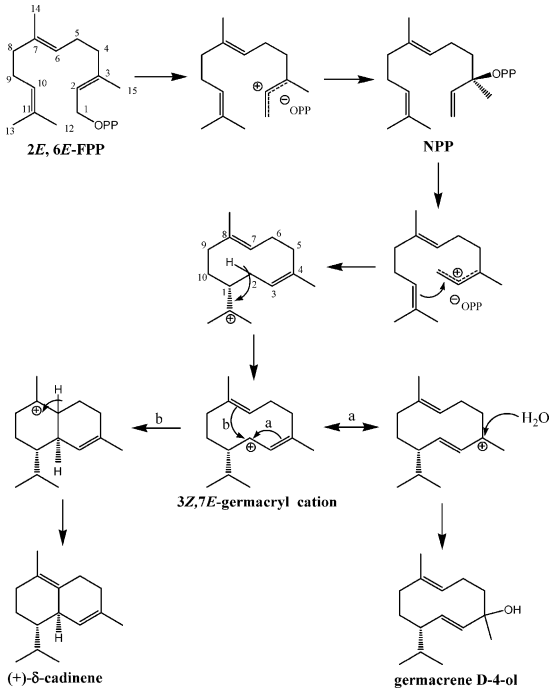

Figure 4. Predicted Cyclization Pathway for (+)-δ-Cadinene and Germacrene D-4-ol

The cyclization reaction is initiated by cleavage of the pyrophosphate group to yield the *transoid*-farnesyl cation. The C2-C3 bond is then readily isomerized to form the *cisoid*-farnesyl cation, which is then followed by the 10,1 cyclization reaction. The hydride on C2 shifts to C11 and forms the 3*Z*, 7*E*-gemacryl cation. In this carbocation intermediate, an electron is delocalized over C2, C3, and C4. The 3*Z*, 7*E*-gemacryl cation is predicted to be the branch point to (+)-δ-cadinene or germacrene D-4-ol. If a water molecule attacks C4, it forms germacrene D-4-ol. On the other hand, if another cyclization reaction occurs between C7 and C2, it yields the cadinyl cation. Then, proton abstraction from C7 results in the formation of (+)-δ-cadinene.

(+)-δ-cadinene synthase from *G. arboreum* was amplified by PCR from cad1-C1 as a template, and ligated into the expression vector pTrc99A. The PCR was carried out using Expand High Fidelity (HF) *Taq* polymerase (Roche), with 30 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 2 min. PCR primers were as follows (restriction sites are underlined and start/stop codons are in bold): 5′-G GAATTC**CATG**GCTTCACAAGTTTCTCAAAT-3′ and 5′-GC**TCA**AAGTGCAATTGGTTCAAT-3′. The amplified products were cut with EcoRI and XbaI and ligated into the high-copy-number plasmid pTrc99A under the control of the isopropyl 1-thio-β-D-galactoside (IPTG)-inducible trc promoter to yield pTrcCAD and transformed into *E. coli*.

GC-MS Analysis of Sesquiterpenes

For analysis of sesquiterpenes, a single colony of *E. coli* harboring pTrcCAD was inoculated into Luria Bertani (LB) medium containing 100 μg/ml ampicillin and 1 mM IPTG, and incubated overnight at 37°C. An aliquot (0.2 ml) of this seed culture was inoculated into fresh LB containing 100 μg/ml ampicillin (5 ml) overlaid with dodecane (10% v/v), and the culture was grown for 24 hr at 37°C on a rotary shaker at 200 rpm. For GC-MS analysis, 50 μl of dodecane was diluted with 200 μl of ethyl acetate. In addition, culture samples with no overlay were prepared and analyzed. An aliquot (0.2 ml) of the seed culture was inoculated into fresh LB containing 100 μg/ml ampicillin (5 ml), and the culture was grown for 7 hr at 37°C on a rotary shaker at 200 rpm. For GC-MS analysis, 700 μl of culture was extracted with equal volume of ethyl acetate. These prepared samples were subjected to gas chromatography electron impact-mass

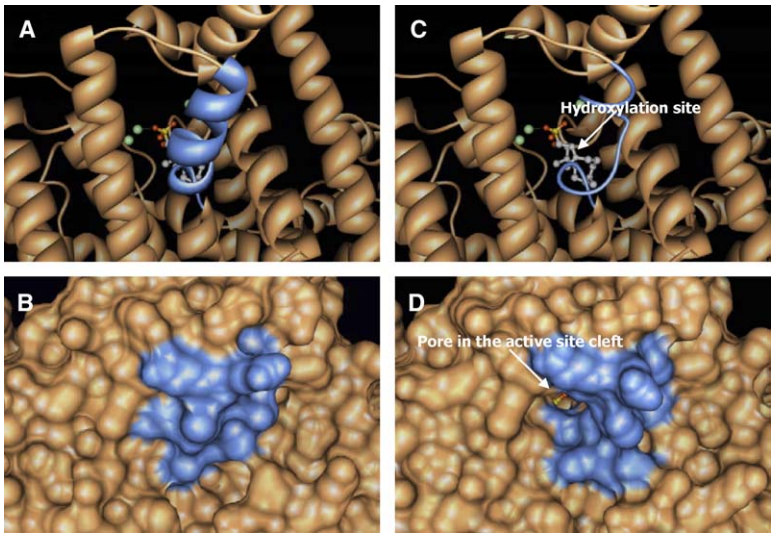

Figure 5. Predicted Structural Change in (+)-δ-Cadinene Synthase

The homology structural models around the G helix (represented in blue) for wild-type (+)-δ-cadinene synthase (A, ribbon diagram; B, filled model) and the G2 (+)-δ-cadinene synthase N403P/L405H variant (germacrene D-4-ol synthase) (C, ribbon diagram; D, filled model) are described. Modeling of the mutations predicted the opening of the active site cleft to make a pore that would allow a water molecule to approach the 3*Z*, 7*E*-germacryl cation. Based on this model, C3 of FHP (where hydroxylation occurs) is located directly under the predicted pore.

spectrometry (GC-EI-MS) using a Hewlett Packard HP6890 gas chromatograph equipped with a Hewlett Packard 5973 mass selective detector, a CyclosilB capillary column (30 min × 250 μm i.d. × 0.25 μm thickness; Agilent Technologies) and a Combi PAL autosampler-injector (LEAP Technologies). Sesquiterpenes from a splitless 1 μl injection were separated using a GC oven temperature program of 80ºC for 2 min, then ramping 130ºC/min to 150ºC, 5ºC/min to 190ºC, and 130ºC/min to 250ºC. The (+)-δ-cadinene was identified from its mass spectrum and GC retention time by comparison to a (+)-δ-cadinene standard (Fluka). Relative quantification of germacrene D-4-ol was determined using the ion abundances of (+)-δ-cadinene as a standard.

Design of CAT Fusion Selection Vector

To exclude nonsense mutants (possible stop codon, additions and deletions) and collapsed or insoluble proteins, a screening method based on chloramphenicol resistance was used. The CAT gene on plasmid pBBRMCS1 was mutated using the Quick Change site-directed mutagenesis kit (Stratagene) to remove the EcoRI restriction site from the CAT gene while maintaining the same amino acid sequence. Primers used were (mutation in bold) 5′-CTTTCATTGC CATACG**A**AATTCCGGATGAGCAT-3′ as a forward primer, and 5′-ATGCTCATCCGGAATTT**C**GTATGGCAATGAAAG-3′ as a reverse primer. Following removal of the EcoRI restriction site, the CAT gene was amplified using PCR with Expand HF *Taq* polymerase and 30 cycles at 95ºC for 30 s, 55ºC for 30 s, and 72ºC for 1 min. The PCR primers used were as follows: 5′-GCTCTAGAGGCGGCGG CAGCGGCGGCGGCAGCGGCGGCGGC**ATG**GAGAAAAAAATCACT G-3′ as a forward primer and 5′-CGTCTAGA**TCA**GAATCGGCCA ACGC-3′ as a reverse primer. The forward primer contained a flexible linker sequence (LRGGGSGGGSGGG) designed based on several studies of linkers for bifunctional protein fusion. The PCR products were cut with XbaI and ligated into pTrc99A to form pTrcCAT. The XbaI site on the reverse primer is methylated in DH5α, but the XbaI of the forward primer (N terminus of CAT) was reusable for cloning the synthase gene.

To confirm the suitability of this plasmid for screening, the gene encoding the GFP was fused to the 5′ end of the CAT gene, and GFP expression of chloramphenicol-resistant colonies was assayed. The PCR product encoding the GFP gene from pGFPuv (CLONTECH) was ligated into pTrcCAT to yield pTrcGFPCAT. PCR of GFP was performed using Expand HF *Taq* polymerase and 30 cycles at 95ºC for 30 s, 55ºC for 30 s, and 72ºC for 1 min. The PCR primers used were 5′-GGAATTC**ATG**GAAGAACAGTAAAGGA-3′ as a forward primer and 5′-GCTCTAGATCCTTTGTAGAGCTCATCC-3′ as a reverse primer. The cells harboring pTrcGFPCAT were spread onto LB agar plates containing 100 μg/ml ampicillin, 100 μg/ml chloramphenicol, and 1 mM IPTG.

After confirming that the CAT selection method would work, the (+)-δ-cadinene synthase gene (1.7 kb) was cloned 5′ of the CAT gene to yield pTrcCADCAT. The primers were 5′-GGAATTC C**ATG**GCTTCACAAGTTTCTCAAAT-3′ as a forward primer and 5′-GCTCTAGAAAGTGCAATTGGTTCA-3′ as a reverse primer. The PCR was carried out using Expand HF *Taq* polymerase with 30 cycles at 95ºC for 30 s, 55ºC for 30 s, and 72ºC for 2 min. The EP-PCR products were cut with EcoRI and XbaI and cloned into pTrcCAT. Transformants were selected for chloramphenicol resistance on LB plates containing 100 μg/ml ampicillin, 40 μg/ml chloramphenicol, and 1 mM IPTG.

EP-PCR and Screening of Library Clones

Variants of (+)-δ-cadinene synthase were generated using EP-PCR. EP-PCR was limited to the 1 kb region at the 3′ end of the gene, referred to as the terpene synthase fold, where the bulk of the active-site resides are found. EP-PCR was performed using 5′-ATGCAA CATATGAAGAGCTC-3′ as a forward primer and 5′-GCTCTAGA AAGTGCAATTGGTTCA-3′ as a reverse primer. The EP-PCR mixture contained 5 U of *Taq* polymerase (Invitrogen), 10 mM each of dCTP and dTTP, 2 mM each of dATP and dGTP, 4.0 mM $MgCl_2$, 0.2 mM $MnCl_2$, 3.5 μM each of the forward and reverse primers, and 10–100 ng of template (pTrcCAD) in total volume of 100 μl. A higher rate of mutation was chosen for EP-PCR (e.g., 2–5 mutations per kb polymerized) to increase the likelihood of identifying catalytically important residues. The PCR products were cut with ClaI and XbaI and ligated into pTrcCADCAT cut with the same enzymes. Single library clones were used for screening of the first three libraries, which yielded the L405S and N403K/L405S variants. For higher throughput, clones from the subsequent EP-PCR library were pooled in groups of 10 and screened. If the pool of 10 colonies was found to produce a new compound, an aliquot of the culture was spread over LB plates containing 100 μg/ml ampicillin and 50 μg/ml chloramphenicol, and individual colonies were screened for production of that compound; approximately 1000 mutants were screened in total. Twenty mutants producing the various levels of germacrene D-4-ol were discovered. All generated mutant genes were then PCR-amplified using the method in construction of pTrcCAD previously described here. These clones were analyzed

Table 1. Kinetics Study for Wild-Type (+)-δ-Cadinene Synthase and Its Mutants

Clones	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})
WT	$>4 \times 10^{-3}$	> 29	1.45×10^3
G1	2.60×10^{-3}	0.91	2.86×10^3
G2	1.86×10^{-3}	1.80	1.03×10^3

Standard deviations for all values were less than 10% of their values.

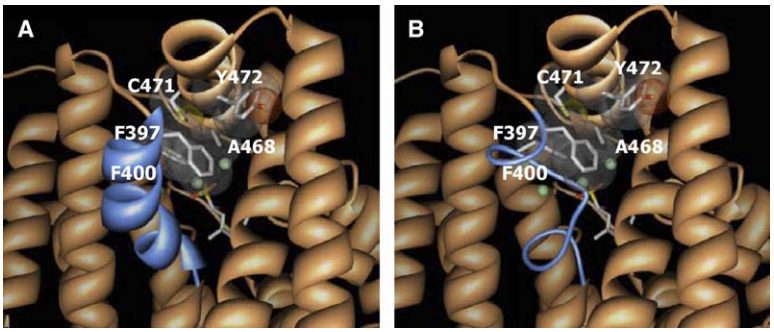

Figure 6. Hydrophobic Interactions Thought to Be Responsible for Production of Germacrene D-4-ol

The G helix is represented in blue. The amino acid positions F397, F400, A468, C471, and Y472 were found to be involved in hydrophobic interactions, which appear to tighten two adjacent α helices together (A). This interaction is thought to be maintained even after two mutations, N403P and L405H, were introduced (B). Additional mutations to F400 caused partial or significant loss of activity.

for terpene production, as previously described here. The results are shown in Figure 2.

Comparative Structural Modeling of (+)-δ-Cadinene Synthase

The structure of (+)-δ-cadinene synthase was developed based on the structure of 5-*epi*-aristolochene synthase using MODELLER, a fully automated comparative structure modeling software package developed by Sali and coworkers (http://www.salilab.org/modeller/modeller.html). The mutations resulting from EP-PCR were also modeled based on the model structure of (+)-δ-cadinene synthase. All protein structures were visualized using Chimera (http://www.cgl.ucsf.edu/chimera).

Saturation Mutagenesis

Based on the predicted (+)-δ-cadinene synthase structure, three amino acids (F400, N403, and L405) were selected for saturation mutagenesis using the Quick Change kit (Stratagene). First, a site-directed saturation mutagenesis was carried out at the codon corresponding to N403 of CAD using pTrcCAD as template. Primers used were 5′-CCATCATTCGAGGAGTTTAAGGCTNNNGCATTGC CAACTTGTGGTTATGCC-3′ as a forward primer and 5′-GGCATAAC CACAAGTTGGCAATGCNNNAGCCTTAAACTCCTCGAATGATGG-3′ as a reverse primer (N denotes any base). Then, 200 colonies were screened using the method described previously here. The clone with the best productivity and selectivity was selected for the next round of saturation mutagenesis (CAD-G1: N403P). The second and the third rounds of saturation mutagenesis were carried out at the codons corresponding to L405 and F400 of CAD-G1 (N403P) and CAD-G2 (N403P/L405H) gene as a template, respectively. The primers used were 5′-GAGTTTAAGGCTCCCGCANNNCCAACTTG TGGTTATGCC-3′ as a forward primer and 5′-GGCATAACCACA AGTTGGNNNTGCGGGAGCCTTAAACTC-3′ as a reverse primer for L405, and 5′-AAACCATCATTCGAGGAGNNNAAGGCTCCCGCACAT CCA-3′ as a forward primer and 5′- TGGATGTGCGGGAGCCTTNNN CTCCTCGAATGATGGTTT -3′ as a reverse primer for F400. The CAD variants containing mutation N403P and N403P/L405H were used as a template for site-directed saturation mutagenesis at L405 and F400, respectively. Sesquiterpene production from the mutants N403P and N403P/L405H were analyzed by methods previously described here.

Protein Expression and Purification

Wild-type (+)-δ-cadinene synthase and its variants were amplified by PCR from pTrcCAD and its variants as templates and ligated into the expression vector pET29. The PCR was carried out using *Pfu* DNA polymerase (Stratagene) with 30 cycles at 95ºC for 30 s, 55ºC for 30 s, and 72ºC for 3 min. PCR primers were as follows (restriction sites are underlined and start/stop codons are in bold): 5′-CATG CCATGGCTTCACAAGTTTC-3′ and 5′-AAATCTTCTCTCATCCG-3′ (complement to pTrc99A). The amplified products were cut with NcoI and HindIII and ligated into pET29 and transformed into *E. coli* BL21(DE3). Each transformant was inoculated into LB medium (5 ml) containing 50 μg/ml kanamycin and was grown overnight at 30ºC. An aliquot (2 ml) of this seed culture was inoculated into fresh LB medium containing 50 μg/ml kanamycin (500 ml), and the culture was grown at 30ºC. When the culture reached OD_{600nm} of 0.6–0.8, 0.1 mM IPTG was added and the culture was grown for an additional 5 hr. Cells were harvested by centrifugation at 6000 × g for 15 min. The pellet was suspended in 10–15 ml of BugBuster (Novagen) containing 20 U DNaseI and bacterial protease inhibitor (Sigma-Aldrich) and incubated for 2 hr at 4ºC. The solution was then centrifuged at 20,000 × g for 30 min and filtered through a 0.45 μm filter. The S-tag Thrombin purification kit (Novagen) was used for the purification following the protocol recommended by Novagen. All purifications were done at half scale. The eluted protein solution was dialyzed twice (3 kDa cut-off; Pierce Biotechnology) against 1 liter of buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 5% glycerol overnight. The protein concentration was measured by Bradford assay. We obtained approximately 3 ml of 25–500 μg/ml of protein solution with about 95% purity (confirmed by SDS-PAGE gel).

Enzyme Kinetics

Kinetics for each enzyme was measured in a 40 μl reaction containing 58.6 μM FPP (15 μCi/ml, [^{3}H]FPP), and 0.05–0.5 μM enzyme in buffer described previously here. The substrate concentration was varied, and every half concentration was used down to 0.23 μM using the same proportion of [^{3}H]FPP. The reaction was overlaid with 500 μl dodecane and incubated for 20 min at 31ºC. To stop the reaction, 40 μl of a solution containing 4 M NaOH and 1 M EDTA was added and mixed. To extract the sesquiterpene products, the reaction mixture was vortexed for 1 min, and 400 μl dodecane was taken from the solution and mixed with 10 ml of scintillation fluid. Radioactivity was measured by scintillation counting. k_{cat}, K_m, and k_{cat}/K_m were calculated using Enzyme Kinetics!Pro (ChemSW).

Acknowledgments

The authors would like to express their gratitude to Xiao-Ya Chen (Shanghai Institute of Plant Physiology) for the (+)-δ-cadinene synthase. This research was funded by the National Science Foundation (BES-9911463), the Office of Naval Research, Maxygen, and the University of California Discovery Grant Program.

Received: February 2, 2004
Revised: October 11, 2005
Accepted: October 12, 2005
Published: January 20, 2006

References

1. Glasby, J.S. (1982). Encyclopaedia of the Terpenoids (New York: Wiley & Sons).
2. Martin, V.J., Yoshikuni, Y., and Keasling, J.D. (2001). The in vivo synthesis of plant sesquiterpenes by *Escherichia coli*. Biotechnol. Bioeng. *75*, 497–503.
3. Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechnol. *21*, 796–802.
4. Lesburg, C.A., Zhai, G., Cane, D.E., and Christianson, D.W. (1997). Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science *277*, 1820–1824.

5. Starks, C.M., Back, K., Chappell, J., and Noel, J.P. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science *277*, 1815–1820.
6. Caruthers, J.M., Kang, I., Rynkiewicz, M.J., Cane, D.E., and Christianson, D.W. (2000). Crystal structure determination of aristolochene synthase from the blue cheese mold, *Penicillium roqueforti*. J. Biol. Chem. *275*, 25533–25539.
7. Rynkiewicz, M.J., Cane, D.E., and Christianson, D.W. (2001). Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. Proc. Natl. Acad. Sci. USA *98*, 13543–13548.
8. Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R. (1998). Sesquiterpene synthases from grand fir (*Abies grandis*): comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. J. Biol. Chem. *273*, 2078–2089.
9. Little, D.B., and Croteau, R.B. (2002). Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases delta-selinene synthase and gamma-humulene synthase. Arch. Biochem. Biophys. *402*, 120–135.
10. Essenberg, M., Grover, P.B., and Cover, E.C. (1990). Accumulation of antibacterial sesquiterpenoids in bacterially inoculated *Gossypium* leaves and cotyledons. Phytochemistry *29*, 3107–3113.
11. Chen, X.Y., Chen, Y., Heinstein, P., and Davisson, V.J. (1995). Cloning, expression, and characterization of (+)-delta-cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. Arch. Biochem. Biophys. *324*, 255–266.
12. Stemmer, W.P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–391.
13. Bornscheuer, U.T., and Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. Curr. Opin. Chem. Biol. *5*, 137–143.
14. May, O., Nguyen, P.T., and Arnold, F.H. (2000). Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. Nat. Biotechnol. *18*, 317–320.
15. Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. *17*, 691–695.
16. Maxwell, K.L., Mittermaier, A.K., Forman-Kay, J.D., and Davidson, A.R. (1999). A simple in vivo assay for increased protein solubility. Protein Sci. *8*, 1908–1911.
17. Wigley, W.C., Stidham, R.D., Smith, N.M., Hunt, J.F., and Thomas, P.J. (2001). Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. Nat. Biotechnol. *19*, 131–136.
18. Rynkiewicz, M.J., Cane, D.E., and Christianson, D.W. (2002). X-ray crystal structures of D100E trichodiene synthase and its pyrophosphate complex reveal the basis for terpene product diversity. Biochemistry *41*, 1732–1741.
19. Calvert, M.J., Ashton, P.R., and Allemann, R.K. (2002). Germacrene A is a product of the aristolochene synthase-mediated conversion of farnesylpyrophosphate to aristolochene. J. Am. Chem. Soc. *124*, 11636–11641.
20. Deligeorgopoulou, A., Taylor, S.E., Forcat, S., and Allemann, R.K. (2003). Stabilisation of eudesmane cation by tryptophan 334 during aristolochene synthase catalysis. Chem. Commun. (Camb.) 2162–2163.
21. Deligeorgopoulou, A., and Allemann, R.K. (2003). Evidence for differential folding of farnesyl pyrophosphate in the active site of aristolochene synthase: a single-point mutation converts aristolochene synthase into an (E)-beta-farnesene synthase. Biochemistry *42*, 7741–7747.
22. Felicetti, B., and Cane, D.E. (2004). Aristolochene synthase: mechanistic analysis of active site residues by site-directed mutagenesis. J. Am. Chem. Soc. *126*, 7212–7221.
23. Seemann, M., Zhai, G., de Kraker, J.W., Paschall, C.M., Christianson, D.W., and Cane, D.E. (2002). Pentalenene synthase: analysis of active site residues by site-directed mutagenesis. J. Am. Chem. Soc. *124*, 7681–7689.
24. Rising, K.A., Starks, C.M., Noel, J.P., and Chappell, J. (2000). Demonstration of germacrene A as an intermediate in 5-epi-aristolochene synthase catalysis. J. Am. Chem. Soc. *122*, 1861–1866.
25. Cane, D.E., Xue, Q., and Fitzsimons, B.C. (1996). Trichodiene synthase: probing the role of the highly conserved aspartate-rich region by site-directed mutagenesis. Biochemistry *35*, 12369–12376.
26. Sanchez, R., and Sali, A. (2000). Comparative protein structure modeling: introduction and practical examples with modeller. Methods Mol. Biol. *143*, 97–129.
27. Baker, D., and Sali, A. (2001). Protein structure prediction and structural genomics. Science *294*, 93–96.
28. Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng. *14*, 529–532.
29. Robinson, C.R., and Sauer, R.T. (1998). Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proc. Natl. Acad. Sci. USA *95*, 5929–5934.