Bifunctional Abietadiene Synthase: Mutual Structural Dependence of the Active Sites for Protonation-Initiated and Ionization-Initiated Cyclizations†

Reuben J. Peters,‡,§,|| Ora A. Carter,‡,§ Yan Zhang,⊥,@ Brian W. Matthews,⊥ and Rodney B. Croteau*,‡

Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340, and Institute of Molecular Biology, Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon 97403-1229

Received July 25, 2002; Revised Manuscript Received November 18, 2002

ABSTRACT: Abietadiene synthase from grand fir catalyzes two sequential, mechanistically distinct cyclizations, of geranylgeranyl diphosphate and of copalyl diphosphate, in the formation of a mixture of abietadiene isomers as the committed step of diterpenoid resin acid biosynthesis. Each reaction is independently conducted at a separate active site residing in what were considered to be structurally distinct domains typical of terpene cyclases. Despite the presence of an unusual 250-residue N-terminal insertional element, a tandem pair of charged residues distal to the insertion was shown to form a functional part of the C-terminal active site. Because abietadiene synthase resembles the ancestral plant terpene cyclase, this observation suggests an early evolutionary origin of catalytically important positively charged residues at the N-terminus of enzymes of this general class. A series of N- and C-terminal truncations of this enzyme were constructed and characterized, both alone and as mixtures of adjacent polypeptide pairs, to assess the proposed domain architecture, the function of the insertional element, and the role of presumptive interdomain contacts. These studies indicated a requirement for the insertional element in functional folding and allowed definition of the minimum primary structure of N- and C-terminal active site peptides. Most importantly, the results showed that, although the two active sites of abietadiene synthase are catalytically independent, substantial contact between the two regions is essential for the functional competence of this enzyme. Thus, the two cyclization sites of abietadiene synthase cannot be dissected into catalytically distinct domains, and, therefore, abietadiene synthase is unlikely to have arisen by fusion of two previously independent genes.

A primary wound response in conifers is the secretion of oleoresin (pitch), a complex mixture of roughly equal amounts of volatile monoterpene olefins (turpentine) and diterpene resin acids (rosin), along with smaller amounts of sesquiterpenes (*1*, *2*). In grand fir (*Abies grandis*), rosin is composed of the tricyclic carboxylic acid abietic acid (**8**) and its positional isomers (**9** and **10**) (*3*), all of which are derived by oxidation of the corresponding olefin precursors produced by abietadiene synthase (*4*) (Scheme 1). This enzyme (AS)[1] is bifunctional in catalyzing two mechanistically different and independent cyclization reactions at separate active sites (Scheme 1) (*5*−*7*). At the first active site, protonation-initiated cyclization of the universal diterpene precursor geranylgeranyl diphosphate (GGPP, **1**) to the stable bicyclic intermediate (+)-copalyl diphosphate (CPP, **2**) occurs in a reaction analogous to that catalyzed by (−)-copalyl diphosphate synthase (CPPS) of the gibberellin biosynthetic pathway (*8*, *9*). In the second active site, (+)-CPP undergoes an ionization-initiated cyclization and rearrangement of the diphosphate ester to produce the mixture of tricyclic olefins (**5**−**7**) in a reaction analogous to that catalyzed by kaurene synthase (KS) of gibberellin biosynthesis with (−)-CPP as a substrate (*8*, *9*); this latter reaction type is common in terpene biosynthesis (*10*). The formation of kaurene in higher plants requires the action of two distinct enzymes, CPPS and the CPP cyclase KS (*8*); however, an unrelated bifunctional kaurene synthase (FCPPS/KS) of fungal origin has been shown to contain two separate active sites in two distinct domains (*11*, *12*), somewhat similar to the proposed organization of the bifunctional AS (*5*−*7*).

On the basis of the only relevant crystal structure now available, that of tobacco 5-epiaristolochene synthase (5-EAS) (*13*), a model has been generated for AS (*14*, *15*). Because 5-EAS does not contain an unusual ∼250-amino acid insertional element found near the N-terminus of AS

† This work was supported by National Institutes of Health Grant GM 31354 (R.B.C.) and a Postdoctoral Fellowship from the Jane Coffin Childs Memorial Fund for Medical Research (R.J.P.).

* To whom correspondence should be addressed. Phone: (509) 335-1790. Fax: (509) 335-7643. E-mail: croteau@mail.wsu.edu.

‡ Washington State University.

§ These authors contributed equally to this work.

|| Current address: Department of Biochemistry, Biophysics and Molecular Biology, Molecular Biology Building, Room 4108, Iowa State University, Ames, IA 50011.

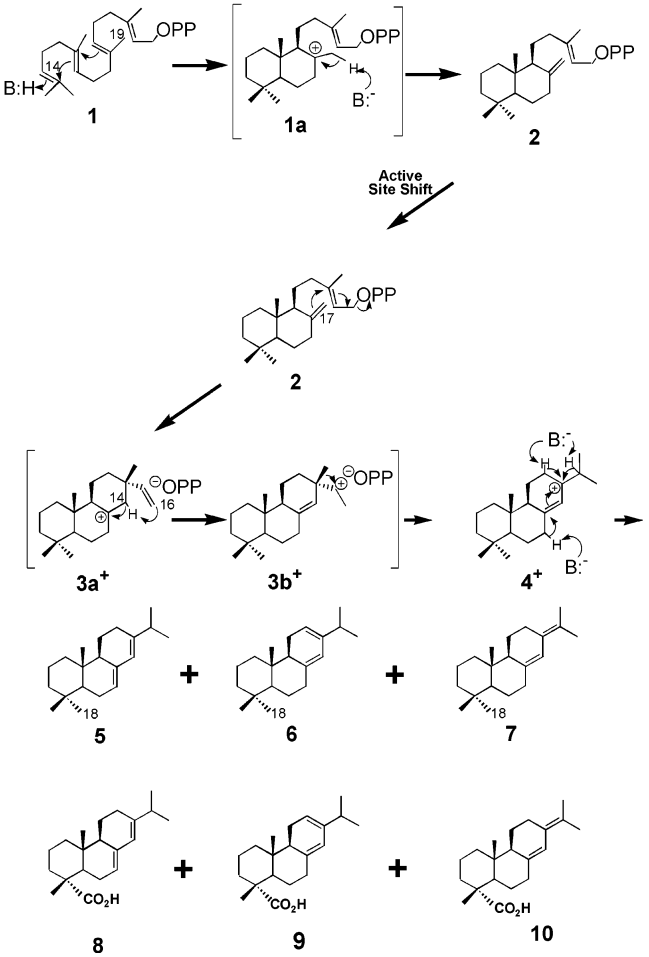
⊥ University of Oregon.

@ Current address: Department of Molecular Biology, BCC206, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037.

[1] Abbreviations: AS, abietadiene synthase (the prefix r denotes the recombinant "pseudomature" enzyme from which the transit peptide has been deleted); CPP, copalyl diphosphate; CPPC, copalyl diphosphate cyclase; CPPS, (−)-copalyl diphosphate synthase; FCPPS/KS, bifunctional fungal CPPS and KS; GC, gas chromatography; GGPP, (*E*,*E*,*E*)-geranylgeranyl diphosphate; KS, kaurene synthase; MS, mass spectrometry; PCR, polymerase chain reaction; 5-EAS, 5-epiaristolochene synthase.

10.1021/bi020492n CCC: $25.00
Published on Web 02/01/2003

Scheme 1: Proposed Reaction Mechanism for Abietadiene Synthase[a]



[a] GGPP (**1**) is bound at the central region active site and protonated at C14 to initiate bicyclization followed by deprotonation at C19 of **1a** (*28*) to yield (+)-CPP (**2**). (+)-CPP then diffuses to the C-terminally located active site (*7*) where ionization of the diphosphate ester initiates anti-S_{N}' cyclization (*29*) to the C8-sandaracopimerenyl cation (**3a**), which undergoes intramolecular proton transfer from C14 to C16 in **3b** (*28*), 1,2-methyl migration to **4**, followed by deprotonation at C7 to abietadiene (**5**), or at alternative positions to give rise to the other abietadiene isomers (**6** and **7**) (*4*). Subsequent oxidation at the A ring α-methyl affords abietic acid (**8**) and related resin acids (**9** and **10**) (*30*).

and a few other terpene cyclases (*16*, *17*), the modeled structure bears only the two helical bundle regions corresponding to the proposed N- and C-terminal domains of plant terpene synthases (*13*) (starting at His348 of AS, and suggesting an approximate boundary between the insertional element and the central region bearing the first cyclization site; see Figure 1 for a schematic diagram). Mutational analysis of the putative active sites for each reaction conducted by AS has verified the general features of the modeled structure and confirmed the catalytic independence of the two sites. The first active site, responsible for the protonation-initiated cyclization of GGPP to CPP (the CPP synthase site), was shown to reside principally in the central cavity of the middle helical bundle (*14*) corresponding to the N-terminal domain of 5-EAS (*13*). The central cavity of the C-terminal domain helical bundle of 5-EAS was identified as the active site for the ionization-initiated cyclization reaction catalyzed by this enzyme (*13*). Similarly, the second active site of AS, which conducts an analogous ionization-initiated cyclization (the CPP cyclase site), was located in the same relative position in the modeled structure (*15*).

The crystal structure of 5-EAS, obtained in the absence and presence of substrate analogues, demonstrated that the N-terminus folds back upon the C-terminal active site, and that this region, along with the A−C and J−K loops, assumes an ordered conformation over the active site only when substrate is bound (*13*). This transition upon substrate binding is thought to provide shielding from water of the highly reactive carbocationic intermediates generated in the course of the cyclization reaction. The presence of N-terminal elements at the C-terminal active site has also been demonstrated by truncation and mutation of other sesquiterpene synthases (*18*) and of the monoterpene cyclase limonene synthase (*19*), all of which catalyze similar ionization-initiated cyclizations. In the latter case, a tandem pair of arginine residues at the N-terminus, which are completely conserved in monoterpene synthases (*16*, *17*), was demonstrated to be required for the initial ionization (and isomerization) of the substrate diphosphate ester (*19*). However, the presence of the insertional element leaves unclear the location of, as well as any possible role for, the N-terminus of AS. In addition, little work has been done to elucidate the role of the insertional element found in AS, which is conserved in sequence and position among all other diterpene synthases that catalyze similar protonation-initiated cyclizations and in a few other terpene synthases that catalyze ionization-initiated reactions; possible roles for the insertional element in stabilization, targeting, and regulation have been suggested (*16*, *17*). Database searching reveals no homologues for the insertional element outside of the terpene synthases, thus offering no other clues about the possible function. Because the modeled structure of AS offers no obvious means of excluding water from the site of the protonation-initiated cyclization, we suggested that the cryptic insertional element may perform this function (*14*); however, the basis of the interaction of this region with the two active sites remains undefined. In an attempt to clarify the interactions of the regions containing the two active sites of AS, and the function of the insertional element, we have constructed and analyzed a series of truncations of the enzyme based upon both limited proteolysis and the previously defined exon structure (*17*). The structural and functional implications of these findings for the complex, two-step cyclization sequence catalyzed by AS are discussed. Given that AS is considered a close relative of the ancestral terpene synthase (*16*, *17*), these findings also have a bearing on the evolutionary origin of this enzyme class.

EXPERIMENTAL PROCEDURES

Materials and General Procedures. Wild-type recombinant AS (rAS) was the original clone (*6*) from which 5′-codons specifying the N-terminal plastidial transit peptide were deleted and a new starting ATG was appended to translate Met(1)Val85−Ala868 (*4*). This modification permits expression from pSBET (*20*) of a "pseudomature" version of the synthase that kinetically resembles the proteolytically processed native enzyme from grand fir xylem (*5*). Bioanalytical procedures, including liquid scintillation counting, SDS−PAGE, and product analysis by GC−MS, were carried out as previously described (*5*−*7*). The preparations of (*E*,*E*,*E*)-

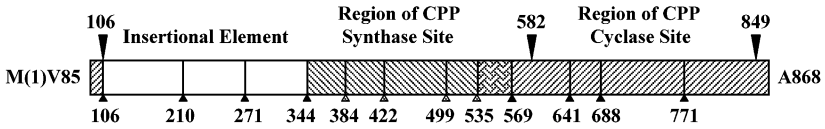


FIGURE 1: Schematic diagram of abietadiene synthase illustrating general structural features. The insertional element of AS and related terpenoid synthases has been described previously (*16*, *17*). The overall structure is modeled on 5-epiaristolochene synthase (*13*). On the basis of this modeled structure, the boundary between the insertional element and the central CPP synthase active site region lies near His348, and the A helix (residues 535−569) spans the central and C-terminal region (CPP cyclase site), preventing assignment of a precise boundary between these two apparent structural units. The numbering is based on the preprotein; however, the pseudomature recombinant enzyme has had the plastidial transit peptide deleted and thus starts with Met (1), and then Val85−Ala868. The large arrowheads and numbers above the diagram denote the limited proteolysis sites, and the small arrowheads and numbers beneath denote exon boundaries (*17*), at which the corresponding truncations were made.

[1-^{3}H]geranylgeranyl diphosphate (120 Ci/mol) and (+)-[1-^{3}H]CPP (120 Ci/mol) have been previously described (*4*, *5*). Proteolysis was carried out using 0.1 unit of chymotrypsin/mg of purified rAS (at ~10 mg/mL) in 50 mM Tris (pH 8.0) containing 50 mM NaCl and 5 mM $CaCl_2$, with reaction times from 4 to 16 h at 15 °C. The resulting peptides were separated on a 15% polyacrylamide gel, electroblotted to an NC membrane in buffer consisting of 20 mM Tris, 150 mM glycine, and 10% (v/v) methanol, and stained with 0.5% Ponceau S in 0.1% acetic acid. After destaining in distilled water had been carried out, the peptide bands were excised, and the sites of proteolysis were determined by N-terminal microsequencing of half of the sample and mass spectrometry (MALDI-TOF, PerSeptive Biosystems) of the remainder. The sample for MALDI-TOF was dissolved in acetone containing 1% trifluoroacetic acid before mixing with a matrix for MS analysis (*21*). Gel filtration on a calibrated Superdex 200 column (Pharmacia) was carried out with a BioLogic chromatography system (Bio-Rad) using a running buffer containing 50 mM Bis-Tris (pH 6.8), 150 mM KCl, and 10 mM $MgCl_2$, at a rate of 0.3 mL/min.

Mutant Construction and Evaluation. Truncations were produced by PCR amplification of the appropriate gene fragment with the introduction (or retention as appropriate) of a 5′-*Nde*I site (overlapping the initiating Met codon) and a 3′-*Bam*HI restriction site (beyond the newly introduced or original termination codon). The resulting PCR products were subcloned back into the pSBET expression vector using the *Nde*I and *Bam*HI restriction sites as described previously (*4*). Point mutants were produced as previously described (*7*), and all constructs were verified by complete sequencing.

For coexpression, selected truncations were subcloned into pET16b using the same *Nde*I and *Bam*HI restriction sites, and were then cotransformed with the pSBET construct containing the adjoining fragment using double-antibiotic (kanamycin/carbenicillin) selection by the standard protocol as previously described (*22*). All constructs were expressed in *Escherichia coli* BLR at 15 °C in 50 mL NZY cultures, and the recombinant enzyme was extracted as described previously (*4*) and was partially purified by binding to and elution from ceramic hydroxyapatite with 180 mM sodium phosphate (pH 6.8) containing 10% glycerol. The protein fraction that was obtained was spin filtered (0.2 μm), and expression of the AS fragments contained therein was quantified by immunoblotting (using polyclonal antibodies raised against wild-type rAS) and densitometric comparison to the wild-type rAS standard by a procedure previously described (*18*). Selected constructs were expressed on a larger scale (1 L cultures), and the derived recombinant proteins were purified (to >98%) and quantified by established methods (*4*).

Enzyme Assays. Kinetic assays with freshly prepared enzyme were performed as previously described (*4*), with the exception that reaction times and rAS concentrations were increased as necessary to accurately measure the activity of impaired mutant enzymes. To determine product distribution, preparative assays with the partially purified enzyme were employed at a saturating substrate concentration (1 mM). GC−MS analysis of pentane extracts of the reaction mixtures (*15*) was sufficient for identification and quantification of diterpene olefin and alcohol products by reference to the authentic standards (*23*). Following pentane extraction of the reaction mixtures, the residual aqueous phases were treated with acid (0.3 M HCl, 15 min), and the resulting solvolysis products of any allylic diphosphate esters present (i.e., hydroxyGGPP or hydroxyCPP derivatives from premature water trapping in the first or second cyclization, respectively) were extracted into pentane for GC−MS analysis as described previously (*7*). The ionization-initiated reaction (second step, conversion of CPP to olefinic products) was directly assessed using (+)-CPP as a substrate by established protocols (*4*). To independently assess the protonation-initiated reaction [first step, conversion of GGPP to (+)-CPP], a previously described (*7*, *15*) coupled assay was utilized, in which the AS D404A mutant (specifically deficient in the corresponding cyclization of GGPP to CPP) was employed to ensure complete conversion of intermediate CPP to the readily measured olefinic products.

RESULTS

Preliminary Studies. Modeling AS based on the crystal structure of 5-EAS (*13*) implicated an organization of two helical bundles typical of plant terpene synthases, to which the sites for the first and second cyclizations catalyzed by AS could be assigned by mutational analysis (*7*, *14*, *15*). Because 5-EAS does not contain the unusual ~250-residue insertional element found in AS (Figure 1) and several other terpene synthases (*16*, *17*), the modeled structure could not address the possible interaction of this region or of the N-terminus with the two catalytically independent active sites. A crystal structure of AS is not yet available. Thus, alternate approaches to defining the structural interactions of the regions containing the two sites, and the function of the insertional element, were required. As the starting point, a kinetically competent, pseudomature form of the recombinant enzyme, which contains the insertional element but from which the approximate plastidial targeting sequence (84 N-terminal residues) has been deleted (*4*), was employed.

In an attempt to dissect this bifunctional enzyme into catalytically active pseudosubunits for the two cyclization

Table 1: Kinetic Analysis of Abietadiene Synthase Trunctions and Mutants Relative to the Pseudomature Wild-Type Enzyme (rAS:85−868)

	CPP[a]	
construct	K_m (μM)	k_{cat} (s^{-1})
wild type (rAS:85−868)	0.4 ± 0.2	2.2 ± 0.3
ASΔ:107−868	0.7	0.04
ASΔ:85−849	ND[b]	$<10^{-5}$
rAS:K86A/R87A	0.5	0.05
rAS:D96A	0.9	2

[a] Kinetic constants are accurate to within ±50% of the reported value. In all cases where measurement was possible, the CPP synthase activity of the first site was unaffected. [b] Could not be determined.

reactions, and to provide core fragments that might be suitable for crystallization, limited proteolysis was carried out using chymotrypsin. Short reaction times resulted in degradation of the N-terminus; a slightly longer reaction additionally allowed degradation of the C-terminus, while overnight incubation yielded two large fragments indicating a specific, internal proteolytic event. The N-termini of the resulting fragments were identified by microsequencing, and the C-termini were defined by the exact mass of the proteolytic fragments. Thus, degradation of the N-terminus leaves Lys107 as the terminal residue, and C-terminal degradation leaves Leu849. The internal site of proteolysis was determined to occur between Phe582 and Thr583, and thus was close to the presumed overlap zone between the central and C-terminal regions bearing the two active sites (Figure 1). Gel filtration of the internally cleaved product demonstrated that the fragments associate (corresponding to a size of ~90 kDa like the intact protein), thus indicating that the interpeptide contacts, including those of the insertional element, are sufficient for maintaining structural integrity (as heterodimeric pseudosubunits) after loss of the covalent linkage. These findings held promise that further analysis could be employed to evaluate the interactions of the pseudosubunits and the importance of the contacts between them, and, potentially, to reveal the function of the insertional element.

Roles of the N- and C-Termini in the Ionization-Initiated Cyclization. Recombinant versions of AS corresponding to proteolytic removal of the N- and/or C-terminal residues were constructed (i.e., ASΔ:107−868, ASΔ:85−849, and ASΔ:107−849) and tested. Loss of the C-terminal residues seemingly compromised folding of the corresponding proteins, as both ASΔ:85−849 and ASΔ:107−849 were produced almost entirely as inclusion bodies. Nevertheless, kinetic characterization of ASΔ:85−849 was possible and demonstrated that, as might be expected, C-terminal truncation influenced only the second, ionization-initiated, cyclization (CPP cyclase site) occurring in the C-terminal region; this truncated version was completely unable to conduct the cyclization of CPP, yet catalyzed the first, protonation-initiated, cyclization of GGPP to CPP at a rate equivalent to that of the wild-type enzyme (rAS) (Table 1). Surprisingly, ASΔ:107−868 exhibited similar behavior in converting CPP to products at a rate of only 2% of that of the wild type, while converting GGPP to CPP (first cyclization, CPP synthase site) at a rate equivalent to that of the wild-type enzyme. This result demonstrates the involvement of N-terminal element(s) at the C-terminal active site for the second, ionization-initiated (CPP cyclase) reaction. Product analyses of these truncated enzymes demonstrated no alteration in profile from that observed with wild-type rAS; however, analysis of the product distribution of an overnight incubation of ASΔ:85−849 with GGPP (this reaction was expected to go to completion) revealed that only ~5% of the substrate had been turned over to CPP. This unexpected result suggests that the CPP synthase site for the first, protonation-initiated, cyclization is subject to product inhibition. Substrate inhibition at this site was previously described (*4*, *7*); this phenomenon of product inhibition can only be observed in the absence of the second cyclization step for consumption of CPP.

Functional Significance of N-Terminal Residues. There are two readily apparent sequence motifs at the N-terminus of AS that might account for the loss of second site activity upon truncation, $K_{86}R_{87}$ and $D_{96}DXXD_{100}$. The N-terminal aspartate-rich sequence resembles a highly conserved C-terminal motif of terpene cyclases that catalyze ionization-initiated reactions, and is required for metal ion chelation-dependent substrate binding and ionization (*10*, *13*). Furthermore, this motif superficially resembles, in both sequence and location, a motif in the bifunctional FCPPS/KS that has been demonstrated to play a role in the second reaction (ionization-initiated) but not the first reaction [protonation-initiated cyclization of GGPP to (−)-CPP] catalyzed by this fungal enzyme (*12*), as observed here with AS. The KR pair was previously noted to resemble in placement the functionally important N-terminal RR motif that is conserved in the monoterpene synthases (*19*).

Earlier indirect evidence had suggested that the D_{96}-$DXXD_{100}$ motif was of little significance in AS catalysis (*4*), and mutation of Asp96 to alanine [this alteration renders the DDXXD element inoperative in other terpene synthases (*10*)] had little effect (Table 1), indicating that this motif is not functionally important. In contrast, replacing both K_{86} and R_{87} with alanine yielded an AS mutant with kinetics that resembled those of the ASΔ:107−868 truncation (Table 1). This result suggests that the penultimate KR motif is largely responsible for the functional influence of the N-terminus on the second, ionization-dependent, reaction occurring in the C-terminal region. Neither mutation of the DDXXD or KR element affected the first reaction, protonation-initiated cyclization of GGPP to CPP, nor did either alter product distribution.

Truncation and Demonstration of Functional Dependence. Mutational analysis of rAS has indicated that the two active sites of this bifunctional enzyme are fully independent, in that point mutations influence one reaction or the other, but not both (*14*, *15*). Furthermore, truncation analysis of the bifunctional fungal kaurene synthase demonstrated that the two analogous reactions of this distantly related enzyme could be separated into functionally independent polypeptides which presumably formed structurally distinct catalytic domains (*12*). Although the N-terminus of AS performs some substantive role at the C-terminal active site for the ionization-initiated cyclization, it is not absolutely required for catalysis of this second reaction with CPP (Table 1). It therefore seemed possible that AS could be dissected into functional polypeptides which would permit probing the structural interactions of the regions containing the two active sites of this enzyme. To address this possibility, a series of

Table 2: Kinetic Analysis of Abietadiene Synthase Fragments Expressed Alone[a]

Construct	Expression Condition	GGPP[b] k_{cat} (s^{-1})	CPP[b] k_{cat} (s^{-1})
rAS:85-868	Soluble	2.2 ± 0.3	2.2 ± 0.3
ASΔ:107-868	Soluble	0.04	0.04
ASΔ:211-868	Insoluble	-	-
ASΔ:345-868	Insoluble	-	-
ASΔ:385-868	Insoluble	-	-
ASΔ:423-868	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:500-868	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:536-868	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:583-868	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:85-422	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:85-499	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:85-535	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:85-582	Soluble	$<10^{-5}$	$<10^{-5}$
ASΔ:85-641	Insoluble	-	-
ASΔ:85-688	Insoluble	-	-
ASΔ:85-771	Insoluble	-	-

[a] For the full schematic of the indicated regions, see Figure 1. [b] No entry in these columns indicates that k_{cat} could not be determined.

N- and C-terminal truncations were prepared corresponding to proteolysis sites and to exon boundaries (*17*), since several of the latter could approximate regional borders and those of the insertional element, and others might correlate with structural motifs (Figure 1). Progressive N-terminal truncations were made inward to the internal proteolysis site (i.e., Phe582/Thr583). Similarly, sequential C-terminal truncations of the pseudomature wild-type recombinant enzyme were made as far inward as the exon border at residue 422. In this way, it was possible to assess the functional capability of the individual polypeptide fragments as well as the combination of adjoining pairs (each of which necessarily bore an introduced initiating methionine).

Many of the truncated forms were expressed exclusively in the form of inclusion bodies, including the version devoid of only the insertional element (ASΔ:345−868), thus indicating misfolding (Table 2). Nevertheless, a sufficient number of polypeptide pairs (corresponding to cleavage at amino acids 422, 499, and 535 and the internal proteolysis site at amino acid 582) were expressed in soluble form to permit analysis. Although apparently stable in soluble form, none of the isolated polypeptides was able to conduct either the first or second cyclization reaction catalyzed by wild-type rAS ($k_{cat} \leq 10^{-5}$ s^{-1}). Because the N-terminus forms part of the C-terminal CPP cyclase active site, two doubly truncated forms (missing the region containing the CPP cyclase active site) were also prepared that contained only the insertional element and the central region, either with or without the "A" helix that connects to the C-terminal region of the second active site (i.e., ASΔ:107−535 and ASΔ:107−582); however, neither of the resulting soluble polypeptides exhibited the anticipated first site activity for the conversion of GGPP (data not shown), suggesting that contact with the C-terminal (CPP cyclase) region was needed for first site (CPP synthase) function.

To determine if the lack of activity of the separate polypeptide pairs was due to the loss of the normal interregional contacts and/or to the misfolding of the isolated polypeptides, adjoining pairs were separately expressed, isolated, and mixed before the assays, or they were coexpressed in the same cells (by transfer of the C-terminally truncated versions to an orthogonal vector) to allow the partners to fold in the presence of each other before isolation and assay. In all cases, coexpression of the adjoining polypeptide pairs resulted in the partial recovery of the ability to catalyze the second, ionization-initiated (CPP cyclase), reaction, but only the ASΔ:85−582 and ASΔ:583−868 pair recovered limited ability to conduct the first, protonation-initiated, cyclization of GGPP to CPP (Table 3). Mixing of the separately expressed polypeptides had little influence on the catalytic capability of the pairs separated at residues 422 and 423, and 499 and 500; however, the mixed pairs separated at residues 535 and 536 and at residues 582 and 583 were quite seriously impaired relative to the coexpressed forms. This further loss of activity by separate expression and subsequent mixing of partners, relative to coexpression, suggested that the isolated pairs, although expressed as soluble polypeptides, nevertheless were misfolded. The association behavior of the ASΔ:85−582 and ASΔ:583−868 pair (the most productive of the four pairs) was therefore examined by gel permeation chromatography (Figure 2), whereupon it was demonstrated that the coexpressed pair

Table 3: Kinetic Analysis of Abietadiene Synthase Fragments Expressed Separately and Mixed (in Vitro), or Coexpressed (in Vivo)

construct	mixing condition	k_{cat} (s^{-1}) GGPP[a]	k_{cat} (s^{-1}) CPP[a]
ASΔ:85−422 and ASΔ:423−868	in vivo	$<10^{-5}$	4×10^{-4}
ASΔ:85−422 and ASΔ:423−868	in vitro	$<10^{-5}$	4×10^{-4}
ASΔ:85−499 and ASΔ:500−868	in vivo	$<10^{-5}$	3×10^{-3}
ASΔ:85−499 and ASΔ:500−868	in vitro	$<10^{-5}$	2×10^{-3}
ASΔ:85−535 and ASΔ:536−868	in vivo	$<10^{-5}$	0.06
ASΔ:85−535 and ASΔ:536−868	in vitro	$<10^{-5}$	2×10^{-4}
ASΔ:85−582 and ASΔ:583−868	in vivo	0.02	0.02
ASΔ:85−582 and ASΔ:583−868	in vitro	$<10^{-5}$	$<10^{-5}$

[a] Turnover rates are accurate to within 3-fold of the reported value because of uncertainties in the linearity of the response to the polyclonal antibody preparation used to quantify the various protein fragments.

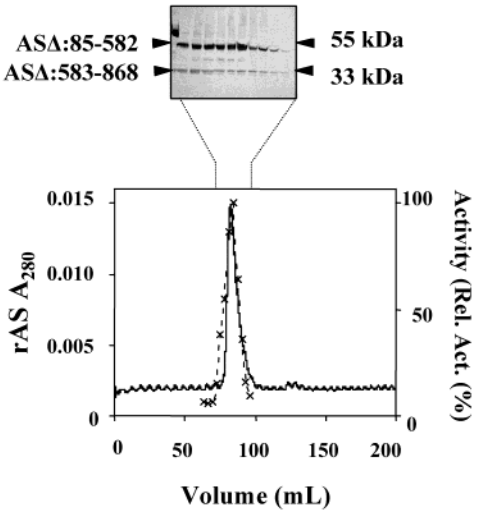


FIGURE 2: Elution behavior of pseudomature wild-type rAS (measured at A_{280}) compared to that of the coexpressed heterodimer composed of the polypeptides ASΔ:85−582 and ASΔ:583−868 (measured by an assay with CPP as the substrate). The inset shows immunoblot analyses of fractions corresponding to the indicated portion of the chromatogram, thereby confirming the presence of both the ASΔ:85−582 (55 kDa) and the ASΔ:583−868 (33 kDa) peptides.

eluted coincidentally with wild-type AS (i.e., at a volume corresponding to a size of ~90 kDa), and thus maintained sufficient integrity to function as a catalytically productive heterodimer of pseudosubunits. In contrast, when these two polypeptides were separately prepared and mixed, the complex eluted in the void volume and thus as an association of higher-order oligomers indicative of misfolding. These results showed that the two active site regions could be separated but that contacts between them were required for productive folding and catalytic function. In all cases in which the coexpressed "heterodimers" possessed sufficient activity to reliably measure, the product profiles using CPP as the substrate were essentially identical to that of wild-type rAS.

DISCUSSION

Modeling of AS against the structure of 5-EAS (*13*) allowed folding alignment of the central (CPP synthase site) and C-terminal (CPP cyclase site) helical bundles with the homologous N-terminal and C-terminal (active site) domains of 5-EAS, but was unable to resolve the location and structure

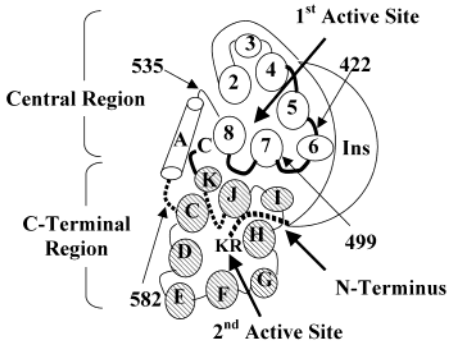


FIGURE 3: Topographical projection of the central (numbered) and C-terminal (lettered) helical bundle regions of AS with the placement of active sites indicated. The circles correspond to individual helices, and the truncation sites are noted. The thick dotted lines indicate the sites of limited proteolysis. Ins represents the insertional element. Although the bundle regions are shown schematically here as parallel, they are at approximately at right angles. The central and C-terminal regions of AS correspond to the respective N- and C-terminal helical bundle domains of 5-EAS (*13*) upon which AS was modeled.

of the unusual ~250-residue insertional element (and N-terminus) of AS that is absent from the modeled structure (Figures 1 and 3). Three regions of 5-EAS, the N-terminus, the A−C loop, and the J−K loop, undergo disordered to ordered transitions upon substrate binding at the C-terminal domain active site of this sesquiterpene synthase to provide solvent shielding for the highly reactive carbocation intermediates formed in the cyclization sequence (*13*). This type of lid mechanism would be expected to be conserved in all terpene synthases which catalyze such electrophilic reactions. Limited proteolysis of AS in the absence of substrate resulted in cleavage at three sites corresponding to the lid regions of 5-EAS (between residues 106 and 107, 582 and 583, and 849 and 850 of AS). Because proteolytic access is correlated with structural flexibility (*24*), this result suggests a similar transition of these elements to lid the corresponding C-terminal active site of AS. However, the product distribution of constructs corresponding to loss of the N-terminus (ASΔ:107−868) or the C-terminus (ASΔ:85−849, back into the J−K loop) or cleavage within the A−C loop (coexpressed polypeptide pair separated at residues 582 and 583) was similar to that of the wild type; no abortive reaction products (terpenols or allylic terpenyl diphosphates), resulting from premature water capture of carbocation intermediates, were observed. Thus, no direct evidence for a "lidding" function was obtained for these structural features. Since deletion of the insertional sequence yielded a misfolded enzyme (expressed to inclusion bodies), it was not possible to assess the potential lidding function for this element.

Truncation of AS at positions corresponding to the proteolytic sites at the N- and C-termini both selectively affected the second (ionization-initiated) reaction catalyzed by AS in the C-terminal CPP cyclase active site. Deletion of 19 C-terminal residues completely abolished this second activity, without affecting the first (protonation-initiated) cyclization catalyzed at the central CPP synthase active site. N-Terminal truncation of 22 residues unexpectedly had a similar effect in impairing the second site, with negligible influence on the first site (Table 1). This result demonstrated that the N-terminus, as with other terpene cyclases lacking

the insertional element (*13*, *18*, *19*), folds back to form part of the C-terminal second site, and thus defines an approximate N-terminal boundary for the insertional element. This insertional element boundary is close to that determined for (−)-CPPS of plant origin, with which it was found that truncation (to remove the transit peptide) beyond Ile100 (corresponding to Ile123 of AS) abolished activity (*25*). Notably, proteolytic cleavage was not observed within the insertional element (defined as residues 107−344), indicating sufficient structural coherence which limits proteolytic access to this region.

Mutational analysis of the N-terminus implicated the presence of the $K_{86}R_{87}$ pair at the C-terminal, second active site; however, the functional significance of these residues is uncertain because the structurally and mechanistically very similar levopimaradiene synthase from *Ginkgo biloba* (which also bears the insertion element) lacks this motif (*26*). Nevertheless, similar N-terminally charged pairs are known to participate at the C-terminal active sites of other terpene synthases (ionization-dependent types) that lack the insertional element (*18*, *19*). On the basis of its genomic, intron-rich, structure, AS is considered to resemble the ancestral terpene cyclase that, through sequential intron loss and further divergence, gave rise to the various classes of terpene synthases (*17*). Therefore, the above results suggest an early origin for the participation of the N-terminus in the C-terminal active site of most types of descendent terpene synthases (*10*), a feature that has apparently persisted after the evolutionary loss of the insertional element. In "modern" (ionization-dependent) terpene synthases of plant origin, the N-terminus is conveyed through the intermediacy of the prototypical N-terminal domain (*13*, *27*) without the intervening insertional element, thus providing little guidance as to the functional role of the insertional element in terpene synthases that contain it.

To evaluate in greater detail AS architecture and the interactions of the regions containing the two sites, more extensive truncation studies were undertaken, along with recombination studies using pairs of adjacent polypeptide partners as "pseudosubunits". From evaluation of the N-terminal truncation series, it was clear that even partial elimination of the insertional element resulted in misfolding of the remaining polypeptides, most of which were produced exclusively in inclusion bodies. Furthermore, constructs that were devoid of the insertional element but that contained the intact C-terminal region (i.e., from residue 423 to the C-terminus), although expressed in soluble form, were inactive in the expected second site, ionization-dependent, reaction *except* in the presence of the adjacent N-terminal polypeptide containing the intact insertional element. Notably, the loss of the N-terminal KR pair only reduces second site activity by 50-fold, rather than the $<10^{-5}$ rate reduction observed upon more extensive N-terminal truncation, suggesting that loss of this motif cannot, in itself, account for the dramatic catalytic impairment observed upon deletion into the insertion sequence.

From evaluation of the C-terminal truncation series (Table 2), it was clear that partial elimination of the C-terminus (i.e., deletion of 97 or 227 residues) led to misfolding of the remaining polypeptide, as these versions yielded only inclusion bodies. Even the deletion of just 19 residues, comprising the "K" helix and part of the J−K loop (i.e., truncated after Leu849; see Figure 3), substantially reduced the level of expression of a soluble polypeptide. Furthermore, polypeptides that were completely or largely devoid of the C-terminal helical bundle (i.e., truncated after residue 535, or after residue 582 which contains the "A" helix spanning the two helical bundle regions), although expressed in soluble form, were unexpectedly inactive in the first site for protonation-initiated cyclization. Only the polypeptide truncated after residue 582 (containing the insertional element, central region, and connecting A helix) recovered this CPP synthase activity when coexpressed with its C-terminal pseudosubunit partner (i.e., the pair dissected between residues 582 and 583) (Table 3).

Deletion of just the K helix (in ASΔ:85−849) influences only the CPP cyclase reaction occurring in the C-terminal helical bundle region, so this contact appears to play a lesser role in interactions between the regions containing the two active sites (Figure 3). The remaining contacts, alteration of which influences both reactions (Table 3), appear to be robust if allowed to form, based on the observation that proteolysis between residues 582 and 583 (after loss of 22 residues from the N-terminus and 19 from the C-terminus, including the K helix) leaves the polypeptides associated. Furthermore, coexpression of the corresponding pair of polypeptides (with intact termini) promotes sufficient productive association to form a functional heterodimer (Figure 2). Partial retention of the modeled contacts between the helical bundles (resulting from dissection between residues 422 and 423 and between residues 499 and 500; see Figure 3) allowed some productive association of the resulting polypeptide pairs, both in vivo (coexpression) and in vitro (separate expression and mixing). The recovery of activity upon mixing in vitro indicates that the pairs can productively refold only in the presence of their heterodimeric pseudosubunit "partner". In contrast, separation of these contacts in the dissected polypeptides (by cleavage between residues 535 and 536 and between residues 582 and 583) negates productive association in vitro, and therefore, these solubly expressed polypeptides are irreversibly misfolded. Nevertheless, these recombination studies also implicated a role in catalysis for the insertional element and defined minimal contiguous sequence requirements for the CPP synthase and CPP cyclase active sites. Thus, any disruption in the central (helical bundle) region, including the A helix (spanning the two helical bundles), abolished CPP synthase activity, which also required the insertional element (i.e., only the pair dissected between residues 582 and 583 could recombine to recover this activity). These results indicate the importance of these two regions (insertional element and central helical bundle region) in forming this first active site. With regard to the CPP cyclase reaction, essentially all of the carboxyl-terminal helical bundle was required for this activity (i.e., ASΔ:85−849 cannot catalyze this reaction), whereas all of the polypeptide pairs representing dissections up to residue 583 recovered some activity for this second, ionization-dependent, reaction (Table 3). This result suggests that the minimal functional unit for this active site consists of contiguous residues 583−868, and is consistent with studies of monoterpene synthase chimeras with which it was found that exchange of the corresponding C-terminal helical bundle regions was sufficient to direct choice of product formation in the ionization-dependent cyclization reactions catalyzed

by these enzymes (R. J. Peters and R. B. Croteau, manuscript in preparation).

In conclusion, a charged pair of residues at the N-terminus of AS was found to play a functional role at the second active site in the C-terminal region. Given its resemblance to the ancestral terpene synthase (*16*, *17*), this finding with AS implies an early origin for the participation of the N-terminus at the C-terminal helical bundle, ionization-dependent, active site found in most terpene cyclases (*13*, *18*, *19*). Truncation analysis of AS demonstrated an essential role for the insertional element in folding, and indicated important functional role(s) for interregional contacts, which are required for catalysis. Recombination of adjacent pairs of polypeptides further defined the minimal contiguous sequence requirements for each active site; the CPP synthase reaction required contiguous residues 107−582, while the CPP cyclase reaction required contiguous resides 583−868. Finally, it is clear from the dissection and recombination experiments that the catalytically independent CPP synthase and CPP cyclase active sites do not reside in readily separable, classic structural domains for each half-reaction, but rather within structurally interdependent regions. This latter finding implies that AS separately evolved its two active sites and did not arise by gene fusion of distinct CPP synthase and CPP cyclase genes and, therefore, is consistent with the previous suggestion that abietadiene synthase may be the closest extant relative to the ancestral terpene cyclase (*17*). Presumably, this ancestral gene encoded both activities and gave rise to the mechanistically distinct, protonation-initiated (i.e., CPP synthase) and ionization-initiated (i.e., CPP cyclase), classes of terpene cyclases through the separation of these activities in the course of sequential intron loss and further divergence.

REFERENCES

1. Phillips, M. A., and Croteau, R. (1999) *Trends Plant Sci. 4*, 184−190.
2. Johnson, M. A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W. D., Eds.) pp 76−91, American Chemical Society, Washington, DC.
3. Lewinsohn, E., Savage, T. J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal. 4*, 220−225.
4. Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. W., Lee, H.-J., Coates, R. M., and Croteau, R. B. (2000) *Biochemistry 39*, 15592−15602.
5. LaFever, R. E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys. 313*, 139−149.
6. Stofer Vogel, B., Wildung, M., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem. 271*, 23262−23268.
7. Peters, R. J., Ravn, M. M., Coates, R. M., and Croteau, R. (2001) *J. Am. Chem. Soc. 123*, 8974−8978.
8. Duncan, J. D., and West, C. A. (1981) *Plant Physiol. 68*, 1128−1134.
9. Saito, T., Abe, H., Yamane, H., Sakurai, A., Murofushi, N., Takio, K., Takahashi, N., and Kamiya, Y. (1995) *Plant Physiol. 109*, 1239−1245.
10. Davis, E. M., and Croteau, R. (2000) in *Topics in Current Chemistry*: *Biosynthesis-Aromatic Polyketides, Isoprenoids, Alkaloids* (Leeper, F., and Vederas, J. C., Eds.) pp 53−95, Springer-Verlag, Heidelberg, Germany.
11. Kawaide, H., Imai, R., Sassa, T., and Kamiya, Y. (1997) *J. Biol. Chem. 272*, 21706−21712.
12. Kawaide, H., Sassa, T., and Kamiya, Y. (2000) *J. Biol. Chem. 275*, 2276−2280.
13. Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997) *Science 277*, 1815−1820.
14. Peters, R. J., and Croteau, R. B. (2002) *Biochemistry 41*, 1836−1842.
15. Peters, R. J., and Croteau, R. B. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 580−584.
16. Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 4126−4133.
17. Trapp, S. C., and Croteau, R. B. (2001) *Genetics 158*, 811−832.
18. Little, D. B., and Croteau, R. B. (2002) *Arch. Biochem. Biophys. 402*, 120−135.
19. Williams, D. C., McGarvey, D. J., Katahira, E. J., and Croteau, R. (1998) *Biochemistry 37*, 12213−12220.
20. Schenk, P. M., Baumann, S., Mattes, R., and Steinbiss, H. H. (1995) *BioTechniques 19*, 196−200.
21. Liang, X., Bai, J., Liu, Y. H., and Lubman, D. M. (1996) *Anal. Chem. 68*, 1012−1018.
22. Burke, C. C., Wildung, M. R., and Croteau, R. (2000) *Proc. Natl. Acad. Sci. U.S.A. 96*, 13062−13067.
23. Lee, H. J., Ravn, M. M., and Coates, R. M. (2001) *Tetrahedron 57*, 6155−6167.
24. Hubbard, S. J. (1998) *Biochim. Biophys. Acta 1382*, 191−206.
25. Smith, M. W., Yamaguchi, S., Ait-Ali, T., and Kamiya, Y. (1998) *Plant Physiol. 118*, 1411−1419.
26. Schepmann, H. G., Pang, J., and Matsuda, S. P. T. (2000) *Arch. Biochem. Biophys. 392*, 263−269.
27. Lesburg, C. A., Caruthers, J. M., Paschall, C. M., and Christianson, D. W. (1998) *Curr. Opin. Struct. Biol. 8*, 695−702.
28. Ravn, M. M., Coates, R. M., Jetter, R., and Croteau, R. (1998) *J. Chem. Soc., Chem. Commun.*, 21−22.
29. Ravn, M. M., Coates, R. M., Flory, J. E., Peters, R. J., and Croteau, R. (2000) *Org. Lett. 2*, 573−576.
30. Funk, C., and Croteau, R. (1994) *Arch. Biochem. Biophys. 308*, 258−266.

BI020492N