Archives of Biochemistry and Biophysics
Vol. 381, No. 2, September 15, pp. 173–180, 2000
doi:10.1006/abbi.2000.1962, available online at http://www.idealibrary.com on IDEAL®

Molecular Cloning, Expression, and Characterization of Amorpha-4,11-diene Synthase, a Key Enzyme of Artemisinin Biosynthesis in *Artemisia annua* L.[1]

Per Mercke,* Marie Bengtsson,† Harro J. Bouwmeester,‡ Maarten A. Posthumus,§ and Peter E. Brodelius[¶,2]

*Department of Plant Biochemistry, Lund University, P.O. Box 117, 22100 Lund, Sweden; †Department of Plant Protection Sciences, Swedish University of Agricultural Sciences, P.O. Box 44, 230 53 Alnarp, Sweden; ‡Plant Research International Business Unit Cell Cybernetics, P.O. Box 16, 6700 AA Wageningen, The Netherlands; §Department of Biomolecular Sciences, Laboratory of Organic Chemistry, P.O. Box 8026, 6700 EG Wageningen, The Netherlands; and ¶Department of Chemistry and Biomedical Sciences, University of Kalmar, P.O. Box 905, 39129 Kalmar, Sweden

Received March 3, 2000, and in revised form June 9, 2000

In plants, sesquiterpenes of different structural types are biosynthesized from the isoprenoid intermediate farnesyl diphosphate. The initial reaction of the biosynthesis is catalyzed by sesquiterpene cyclases (synthases). In *Artemisia annua* L. (annual wormwood), a number of such sesquiterpene cyclases are active. We have isolated a cDNA clone encoding one of these, amorpha-4,11-diene synthase, a putative key enzyme of artemisinin biosynthesis. This clone contains a 1641-bp open reading frame coding for 546 amino acids (63.9 kDa), a 12-bp 5′-untranslated end, and a 427-bp 3′-untranslated sequence. The deduced amino acid sequence is 32 to 51% identical with the sequence of other known sesquiterpene cyclases from angiosperms. When expressed in *Escherichia coli*, the recombinant enzyme catalyzed the formation of both olefinic (97.5%) and oxygenated (2.5%) sesquiterpenes from farnesyl diphosphate. GC–MS analysis identified the olefins as (*E*)-β-farnesene (0.8%), amorpha-4,11-diene (91.2%), amorpha-4,7(11)-diene (3.7%), γ-humulene (1.0%), β-sesquiphellandrene (0.5%), and an unknown olefin (0.2%) and the oxygenated sesquiterpenes as amorpha-4-en-11-ol (0.2%) (tentatively), amorpha-4-en-7-ol (2.1%), and α-bisabolol (0.3%) (tentatively). Using geranyl diphosphate as substrate, amorpha-4,11-diene synthase did not produce any monoterpenes. The recombinant enzyme has a broad pH optimum between 7.5 and 9.0 and the K_m values for farnesyl diphosphate, Mg^{2+}, and Mn^{2+} are 0.9, 70, and 13 μM, respectively, at pH 7.5. A putative reaction mechanism for amorpha-4,11-diene synthase is suggested. © 2000 Academic Press

Key Words: **amorpha-4,11-diene synthase; cDNA cloning; GC–MS; bacterial expression; sesquiterpenes monoterpenes; *Artemisia annua*.**

Human malaria remains an important cause of mortality in tropical regions and it is reported to affect 300–500 million people world wide annually. Infection with *Plasmodium falciparum* accounts for 200–250 million cases of which around 2 million are fatal (1). Multiresistance in *P. falciparum* to the commonly used antimalarial agents is becoming more and more widespread and poses serious threats to conventional or current therapeutic treatments. An alternative class of antimalarial compounds based on the sesquiterpenoid endoperoxide artemisinin (Fig. 1) from *Artemisia annua* L. has been developed from a traditional Chinese herbal remedy (Qinghaosu). The plasmodicidal effects of artemisinin and its derivatives, dihydroartemisinin and artemether, have been confirmed in clinical trials against various strains of *P. falciparum* (2).

It is not currently possible to chemically synthesize artemisinin on a commercial basis, and the yield of artemisinin from the plant itself is low (ca. 0.1% of dry weight). Attempts to obtain high-yielding strains by classical breeding and selection techniques have not improved this significantly (3, 4). The development of a

[1] The nucleotide sequence data reported in this article have been deposited with the GenBank/EMBL Database under Accession No. AF 138959.

[2] To whom correspondence should be addressed. Fax: +46-480-446262. E-mail: peter.brodelius@hik.se.

0003-9861/00 $35.00

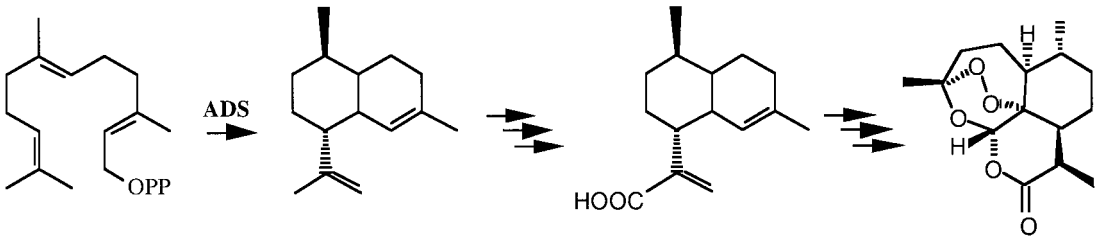

FIG. 1. Hypothetical biosynthetic pathway from farnesyl diphosphate to artemisinin in *Artemisia annua.* Redrawn from (8).

new strain of *A. annua,* genetically engineered for increased production of artemisinin, which could be grown as a cropping plant in the area where the drug is actually required, would thus be of great value.

Sesquiterpenoids are synthesized from farnesyl diphosphate (FDP),[3] a reaction that is catalyzed by a sesquiterpene cyclase (or synthase). In *A. annua,* a number of sesquiterpene cyclases are expressed since sesquiterpenes belonging to different structural classes have been isolated from the plant (5–8). Some examples of sesquiterpenoids that are present in most chemotypes of the plant are germacrene D, β-farnesene, β-caryophyllene, α-copaene, artesiminic acid, and artemisinin. From a number of plant species, a variety of sesquiterpene synthases have already been cloned and the primary structures of these enzymes display a relatively high homology. Some examples are *epi*-cedrol synthase from *A. annua* (5, 9), germacrene C synthase from tomato (10), (*E*)-β-farnesene synthase from peppermint (11), and δ-cadinene synthase (12), *epi*-aristolochene synthase (13), and vetispiradiene synthase (14). The latter three enzymes are involved in the biosynthesis of phytoalexins in cotton, tobacco, and *Hyoscyamus muticus,* respectively.

Even though a number of investigations on the sesquiterpenoid content of leaf tissue of *A. annua* have been carried out (5–7, 15), it was not until recently that the presence of amorpha-4,11-diene in extracts from *A. annua* was reported and its intermediacy in artemisinin biosynthesis postulated (8). A biosynthetic route from FDP to artemisinin as shown in Fig. 1 has been suggested. While the conversion of amorpha-4,11-diene to artemisinic acid is still hypothetical, the conversion of artemisinic acid to artemisinin via dihydroartemisinic acid has been shown (16). The low abundance of amorpha-4,11-diene suggests that the cyclization of FDP is a limiting step in artemisinin biosynthesis (8). Sesquiterpene cyclases have been shown to be regulatory enzymes in other plants (17–19). Thus, a possible strategy for increasing the formation of artemisinin in *A. annua* would be the upregulation of amorpha-4,11-diene synthase by introduction of extra copies of the gene. Here we report on the cloning, expression, and characterization of amorpha-4,11-diene synthase from *A. annua.*

[3] Abbreviations used: DTT dithiothreitol; FDP farnesyl diphosphate; GDP geranyl diphosphate.

MATERIALS AND METHODS

Reagents. [1-^{3}H]FDP (16 Ci/mmol) was purchased from Amersham–Pharmacia Biotech and [1-^{3}H]GDP (15 Ci/mmol) from American Radiochemical Co. Semisynthetic amorpha-4,11-diene was kindly supplied by Dr. Ben Jansen. All other biochemicals and reagents were purchased from Sigma or Merck unless otherwise noted.

Plant growth. A. annua was grown in a greenhouse at 20°C with a 16-h light period. Plants were vegetatively propagated from cutting of mature plants.

Isolation of a cDNA clone encoding amorpha-4,11-diene synthase. The cDNA library used for isolation of a clone encoding amorpha-4,11-diene synthase (NAC) has previously been described (5). This library was screened for sesquiterpene cyclase cDNA clones using as probe the full-length cDNA encoding *epi*-aristolochene synthase from tobacco (13) (a kind gift from J. Chappell, University of Kentucky). A sesquiterpene clone containing 1100 bp was selected [part of *epi*-cedrol synthase (5)]. Using this clone as a probe, a number of clones were isolated. In the next round of screening a 291-bp probe, prepared by PCR amplification of a region of the tobacco cDNA (bp 746–1036) corresponding to amino acids 223 to 319, was used. Primers for this PCR were 5-TTA CTT CGA TTT GCC AAA-3′ (forward) and 5′-CTT TGT ATG GCA TCT GTG-3′ (reverse). cDNA library screening identified a putative full-length synthase clone (NAC) that was further characterized.

DNA sequencing. DNA sequencing of the selected cDNA clone was performed using a DNA BigDye terminator cycle sequencing kit (Perkin–Elmer) for labeling of the sequencing reactions. Analyses were then carried out on an ABI PRISM 310 genetic analyzer. Oligonucleotides (17-mers) were synthesized according to sequence information and used directly as primers for further sequencing.

Cloning of amorpha-4,11-diene synthase into pBluescript SK and pET23c. The open reading frame of the NAC clone was cloned in frame with the β-galactosidase gene using the restriction sites *Eco*RI (forward) and *Xho*I (reversed). The complete open reading frame of the NAC cDNA was amplified by PCR using the following primers: forward primer, 5′-GAC GAA TTC CAT GTC ACT TAC AGA AGA-3′; and reverse primer, 5′-CGA CTC GAG TCA TAT ACT CAT AGG ATA-3′. PCR was carried out in a total volume of 100 μl with the following reagents: 1× cloned Pfu polymerase buffer (Stratagene), 0.2 mM dNTPs (Pharmacia), 40 pmol of each primer, and 2 units of Turbo Pfu polymerase (Stratagene). PCR cycling was 94°C (2 min); 25 cycles of 94°C (30 s), 56°C (1 min), 72°C (2 min); and 72°C (5 min).

Cloning of amorpha-4,11-diene synthase into the expression vector

pET23c was performed using *Nde*I and *Xho*I restriction sites. PCR amplification was performed as described earlier but using the forward primer CGT CTA CAT ATG TCA CTT ACA GAA GAA in combination with the above reverse primer.

The PCR fragments obtained were cloned into vectors pBSK and pET23c using the *Eco*RI and *Nde*I sites, respectively, in combination with the *Xho*I site. Expression in *Escherichia coli* of the pBSK construct results in a fusion protein while the pET23c construct gives a recombinant protein without any fusion peptide attached. The cloned fragments were sequenced as described above.

Bacterial expression and partial purification of sesquiterpene cyclase. Bacterial expression and partial purification of the amorpha-4,11-diene in fusion with the β-galactosidase were carried out as described (5).

E. coli BL21 (DE3) pLysS cells (Novagen) harboring the pET23c-NAC were grown at 37°C in LB medium (15 ml) with ampicillin (50 μg/ml) to an OD_{600} of 0.5. Induction was achieved by addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were maintained for 3 h at 25°C and then harvested by centrifugation (2000*g*, 10 min). Subsequently, the bacteria were suspended in 25 mM Hepes buffer, pH 7.5, (2.5 ml) containing 10% glycerol, 10 mM β-mercaptoethanol, and 2 mM $MgCl_2$. Sonication of the cells was carried out for 2 min in 5-s pulses with 5 s between pulses on ice (Vibracell power sonicate setting 12–15 W). Streptomycin sulfate (1% w/v) was added and the slurry was centrifuged at 17,400*g* for 10 min. The supernatant was run on a PD-10 column (Amersham–Pharmacia Biotech) preequilibrated with 3 mM Hepes buffer, pH 7.5, containing 10% glycerol and 1 mM DTT. Aliquots of the recombinant cyclase were stored at −80°C until used.

Preparations of plant enzyme extracts for sesquiterpene cyclase assays. Fresh young leaves of *A. annua* (5.0 g) were washed in cold tap water and ground in liquid nitrogen. Polyvinylpolypyrolidone (PVPP) (1.25 g) and 100 mM sodium phosphate buffer, pH 7.0 (25 ml), containing 0.25 M sucrose, 25 mM ascorbic acid, 25 mM sodium metabisulfite, and 10 mM β-mercaptoethanol were added. The homogenate was stirred at 0°C for 30 min and then centrifuged at 8000*g* for 10 min. The supernatant was subsequently centrifuged at 50,000*g* for 60 min. An aliquot of the supernatant (2.5 ml) was applied to a PD-10 gel-filtration column (Amersham–Pharmacia Biotech) preequilibrated with 100 mM Tris–HCl, pH 7.0, containing 10% (w/v) glycerol, 10 mM $MgCl_2$, and 10 mM β-mercaptoethanol. Aliquots of the plant extract were stored at −80°C until used.

Standard sesquiterpene cyclase assay. Determination of sesquiterpene cyclase activity in plant and bacterial extracts as well as for partially purified recombinant enzyme was carried out as described (5).

Preparation of enzyme products for structural analysis. An aliquot of bacterial (400 μl) or plant (100 μl) extract was mixed with Hepes buffer, pH 7.5, in a 10-ml Teflon-capped screw-top test tube to give the final concentrations of 30 mM Hepes, 10% glycerol, 3 mM β-mercaptoethanol, 0.27 mM DTT, 3.7 mM $MgCl_2$, and 15 μM FDP in total volumes of 1.5 and 1.0 ml, respectively. The reaction mixtures were overlaid with hexane (1 ml) and incubated overnight at 25°C. The mixtures were extracted with hexane (2 × 1 ml) and the pooled hexane phases were analyzed by GC–MS as described below.

GC–MS analysis. GC–MS analysis was carried out on two different systems. Routine GC–MS (70 eV) analysis was performed on a Hewlett–Packard Model 5970B GC–MS system equipped with a 5970B computer system upgraded for DOS and interfaced with a Hewlett–Packard Model 5890 GC. The column used was a 30 m × 0.25-mm-i.d. DB-Wax column (J & W Scientific, Folsom, CA), programmed from 50°C (hold 5 min) at 8°C/min to 230°C, with helium as the carrier gas. EI mass spectra of all products were recorded.

GC–MS analysis for product identification was performed by injecting a sample (splitless mode) onto a BP5 capillary column (30 m × 0.25 mm i.d., 0.25-μm film thickness). After an initial oven temperature of 40°C during the injection period, the oven temperature was programmed from 80 to 260°C at a rate of 3°C/min. The column was connected to the ion source of a Finnigan MAT 95 mass spectrometer, operating in the 70-eV EI ionization mode and scanning from *m*/*z* 24 to 300.

Identification of sesquiterpenoids. The compounds were identified by comparing the mass spectra and retention indices with those in the Wageningen database of natural products (ca. 5000 entries) and in other reference libraries (20, 21). The identification of amorph-4-en-7-ol is based on a recent paper on the volatile constituents of *Eupatorium coelestinum* (22).

Characterization of amorpha-4,11-diene synthase. The pH optimum of recombinant amorpha-4,11-diene synthase was determined in buffers of 50 mM Bis–Tris, 50 mM Hepes, and 50 mM Tris containing 5 mM $MgCl_2$, 2 mM β-mercaptoethanol, 250 mM KCl, and 10% glycerol, in the pH range from 6.0 to 9.3 using [1-^{3}H]FDP (10.7 μM) as substrate. Samples were incubated at 25°C for 10 min. K_m for FDP was determined at pH 7.5 in the same buffer as was used for pH optimum determinations. Determination of K_m for divalent cations was carried out in 50 mM Hepes buffer, pH 7.5, containing 10% glycerol, 5 mM β-mercaptoethanol, and 250 mM KCl.

RESULTS AND DISCUSSION

cDNA isolation and sequence analysis. In a first screen with a 1100-bp fragment of the *epi*-cedrol synthase, a number of putative cyclase clones were isolated from an *A. annua* leaf cDNA library. A few of these clones encoding *epi*-cedrol synthase have been characterized (5). A PCR-amplified 291-bp fragment of the tobacco *epi*-aristolochene synthase (bp 746–1036) (13) was used as probe in a second screen of the putative cyclase clones isolated. The selection of this fragment was based on the high similarity between the nucleotide sequences of *epi*-cedrol synthase and *epi*-aristolochene synthase (67%). The nucleotide identity of the two complete open reading frames of the two sesquiterpene cyclases was 44%. The cDNA clone (NAC) isolated showed a somewhat lower identity (64%) to the tobacco probe used for the screening.

The new full-length clone (NAC) contained an insert of 2080 bp with a putative translation starting sequence (AAATCATGT) at positions 9–17 followed by an open reading frame of 546 amino acids. The deduced mass of the encoded protein is 63.9 kDa and the p*I* is calculated to be 5.47. The estimated molecular mass is somewhat higher than the 56 kDa reported for the purified amorpha-4,11-diene synthase (8). The p*I* and the molecular weight are similar to those reported for other plant sesquiterpene cyclases (11–13, 23). It is generally assumed that sesquiterpene cyclases, in contrast to mono- and diterpene cyclases, are exclusively localized to the cytosolic compartment and in accordance with this theory the isolated cDNA clone does not contain a similar plastidial targeting sequence.

The deduced amino acid sequence of this *Artemisia* enzyme shows relatively high homology to other sesquiterpene synthases and especially to those from angiosperm species. An alignment of eight sesquiterpene synthases from angiosperms, including the new *Arte-*

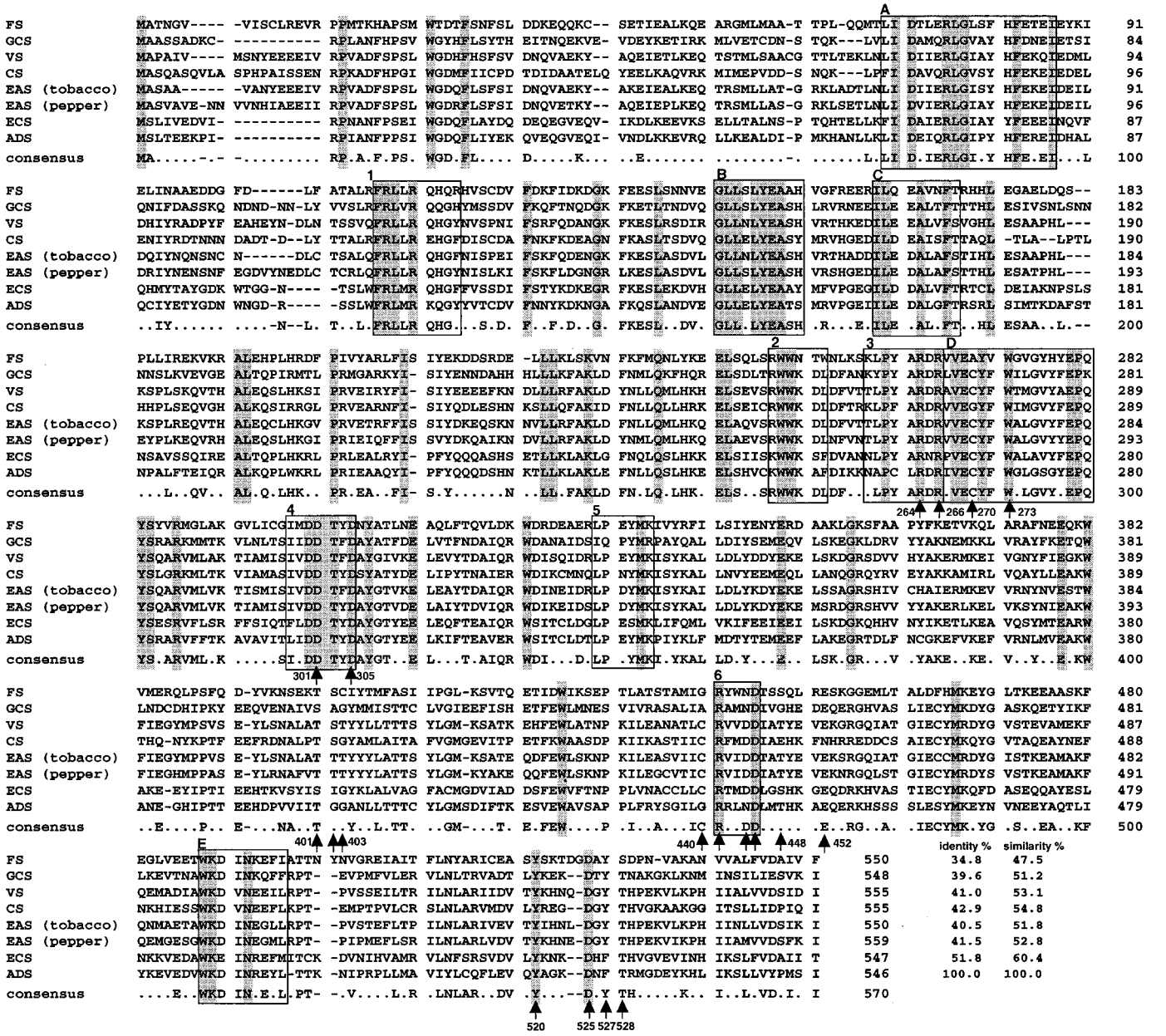

FIG. 2. Alignment of the deduced amino acid sequence of eight sesquiterpene synthases from angiosperms. FS, farnesene synthase from peppermint (9); GCS, germacrene C synthase from cherry tomato (8); VS, vetispiradiene synthase from *H. muticus* (12); CS, δ-cadinene synthase from cotton (10); EAS, *epi*-aristolochene synthase from tobacco (11) and pepper (18); ECS, *epi*-cedrol synthase from *A. annua* (5); ADS, amorpha-4,11-diene synthase from *A. annua*. The six regions of high similarity found among terpene synthases and identified by Chen *et al.* (10) are boxed and numbered. Five additional regions of high similarity in sesquiterpene synthases from angiosperms are boxed and labeled A to E. Arrows indicate amino acid residues in *epi*-aristolochene synthase from tobacco involved in substrate binding and catalysis (24). The consensus sequence shows amino acid residues conserved in at least five of the eight synthases.

misia cyclase, has been carried out (Fig. 2). The consensus sequence in Fig. 2 shows a number of conserved regions in these enzymes. This high homology between different sesquiterpene synthases makes modeling of the three-dimensional structure of the various synthases possible using the crystal structure of *epi*-aristolochene synthase, a protein that consists entirely of α-helices and short connecting loops and turns organized in two structural domains (24).

Chen and co-workers have defined six highly conserved regions in plant terpene synthases (boxed and numbered in Fig. 2) (12). The highly conserved aspartate-rich motif DDXXD (region 4), which is found in all terpenoid synthases (25), including those of microbial

origin, as well as in prenyltransferases which operate by a related mechanism on the common prenyl diphosphate substrates (26), is of particular interest. Based on data obtained by X-ray structural investigations and directed mutagenesis experiments, it has been concluded that this motif is involved in the coordination of the substrate-bound divalent metal ions (Mg^{2+} and Mn^{2+}) (24). The definition of the six conserved regions was based on a relatively limited number of deduced synthase amino acid sequences (12). From the alignment in Fig. 2 it is evident that sesquiterpene synthases from angiosperms show additional regions with high conservation of sequence (labeled A to E in Fig. 2). Using the crystal structure of *epi*-aristolochene synthase, it is concluded that regions 1 and A to C are grouped together in the N-terminal glycosyl hydrolase domain (residues 36–230 of the enzyme). The function of this region is not known but it may play a role in the interaction with other macromolecular structures. The presence of conserved structure in this domain, however, appears to be essential. The other conserved regions (2 to 6, D and E) are located in the C-terminal catalytic domain and contain many of the amino acids participating in substrate binding and catalysis in *epi*-aristolochene synthase (shown by arrows in Fig. 2).

The amino acid sequence homology between the new *Artemisia* cyclase and other sesquiterpene synthases varies (47.5–60.4% similarity and 34.8–51.8% identity) as detailed in Fig. 2. The highest homology is not surprisingly observed between the two *Artemisia* sesquiterpene synthases (5). In general, the amorphadiene synthase shows the same pattern of sequence homology to other terpene synthases as *epi*-cedrol synthase (5).

Bacterial expression. The pBSK system has been used successfully for expression of terpene cyclases (5, 10, 11). Protein extracts from *E. coli* expressing the β-galactosidase fusion enzyme, however, showed relatively low sesquiterpene synthase activity even when partially purified as previously described (5). Nevertheless, structural analysis of sesquiterpene products by GC–MS was possible (see below). Even though we did not test for levels of expressed recombinant fusion enzyme, the N-terminal fused β-galactosidase could be reducing the specific sesquiterpene synthase activity of NAC.

Extracts of *E. coli* harboring pET23-NAC displayed a high cyclase activity when [1-^{3}H]FDP was added. Analysis by GC–MS revealed one major (91.2% of total sesquiterpenes) and eight minor sesquiterpenoid products (Fig. 3A). Six sesquiterpene olefins, (*E*)-β-farnesene (0.8% of the olefins), amorpha-4,11-diene (93.6%), amorpha-4,7(11)-diene (3.8%), γ-humulene (1.1%), β-sesquiphellandrene (0.5%), and an unknown olefin (0.2%), together comprised 97.4% of the products generated in the assay from FDP. The oxygenated sesquiterpenes (2.6% of total sesquiterpene generated) consisted of amorpha-4-en-11-ol (tentative) (7.7%), amorpha-4-en-7-ol (80.8%), and α-bisabolol (11.5%). The major product exhibited a MS spectrum as shown in Fig. 3B. No significant difference in product distribution was observed between the two bacterially expressed constructs. Crude protein extracts from fresh leaves of *A. annua* showed relatively high sesquiterpene cyclase activity. GC–MS analysis revealed that the major product obtained with an enzyme extract of leaves exhibited the same GC retention time and mass spectrum as the major product from the recombinant enzyme (data not shown) indicating that amorpha-4,11-diene synthase is a major sesquiterpene synthase in *A. annua* leaves. The major sesquiterpenoid product was identified as amorpha-4,11-diene by comparison to the semisynthetic standard (Fig. 3C) (8). No monoterpene synthase activity could be detected in extracts of *E. coli* harboring pET23-NAC when [1-^{3}H]GDP was used as substrate under standard conditions.

Characterization of amorpha-4,11-diene synthase. The recombinant, partially purified amorpha-4,11-diene synthase exhibited maximum activity at a pH above 7.0. The enzyme exhibits high activity in a relatively broad pH range. Essentially the same enzyme activity is observed at pH 9.0 and 7.5. Also for the partially purified amorpha-4,11-diene synthase from leaves, a broad pH optimum was reported (8). For other sesquiterpene synthases, pH optima in the range of 6.5–8.5 have been reported (5, 11, 12).

The K_m value for FDP with the recombinant enzyme was calculated to be 0.9 μM at pH 7.5, which is very similar to 0.6 μM reported for the purified amorpha-4,11-diene synthase (8) and is typical also for other sesquiterpene synthases of plant origin (5, 11, 27). K_m values of 70 and 13 μM were determined for Mg^{2+} and Mn^{2+}, respectively. These values are similar to those reported for other terpene cyclases (5, 11).

Enzyme mechanism. Most terpene synthases generate multiple products from the prenyl diphosphate substrate (5, 11, 23, 27). This is most likely a consequence of the electrophilic reaction mechanism in which a series of highly unstable carbocationic intermediates are generated. These intermediates may be deprotonated or quenched by H_2O and released from the enzyme resulting in the formation of multiple products. Examples of enzymes generating a large number of products are γ-humulene synthase and δ-selinene synthase from grand fir which produce 52 and 34 discrete sesquiterpenes, respectively (27). This is achieved by variations upon several different cyclization routes in combination with deprotonation of many carbocationic intermediates to generate stable olefinic end products. The two enzymes in grand fir have ap-

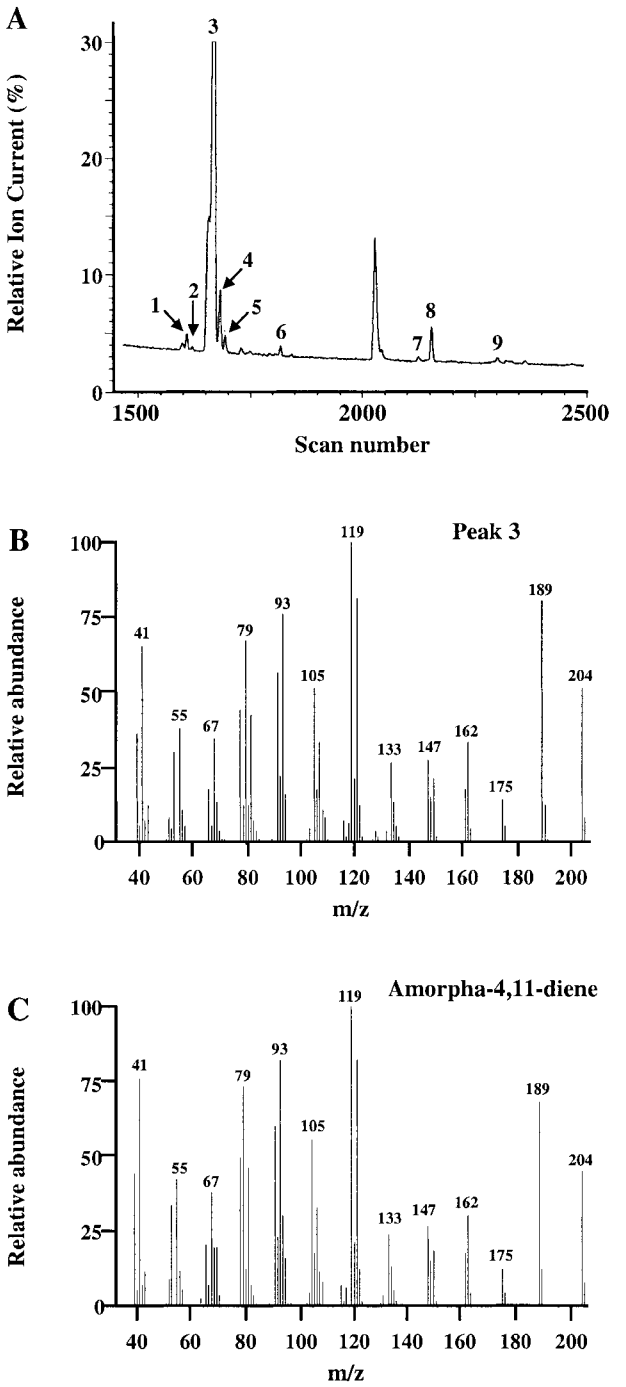

FIG. 3. GC–MS analysis of the sesquiterpene product of the recombinant enzyme encoded by NAC. (A) Total ion chromatogram of sesquiterpene products generated from FDP by the recombinant enzyme. The numbered peaks denote the sesquiterpenes (*E*)-β-farnesene (1), unknown (2), amorpha-4,11-diene (3), amorpha-4,7(11)-diene (4), γ-humulene (5), β-sesquiphellandrene (6), amorpha-4-en-11-ol (7) (tentative), amorpha-4-en-7-ol (8), and α-bisabolol (9) (tentative). The shoulder on peak 3 is due to an injection artifact caused by hexane. (B) Mass spectrum of the major biosynthetic product (peak 3). (C) Mass spectrum of authentic amorpha-4,11-diene.

parently evolved to produce the maximum number of terpene products using the minimum genetic and enzymatic machinery (27).

Most sesquiterpene synthases are, however, characterized by the generation of one dominant product (>80% of total terpenes produced) along with traces of a varying number of by-products. Examples are *epi*-aristolochene synthase from tobacco (93% *epi*-aristolochene) (13), (*E*)-β-farnesene synthase from peppermint [85% (*E*)-β-farnesene] (11), germacrene C synthase from tomato (>64% germacrene C) (23), δ-cadinene synthase from cotton (100% δ-cadinene) (12), *epi*-cedrol synthase from *A. annua* (93% *epi*-cedrol) (5), (*E*)-α-bisabolene synthase from grand fir [100% (*E*)-α-bisabolene] (28), and (−)-α-gurjunene synthase from *Solidago canadensis* [91% (−)-α-gurjunene] (29). Amorpha-4,11-diene synthase (91% amorpha-4,11-diene) also belongs to the latter category of sesquiterpene synthases. The major products of these enzymes are often intermediates in a biosynthetic route (often of phytoalexins) and consequently a more specific cyclization is required for optimum production of end product(s). Examples of such end products are capsidiol in tobacco (from *epi*-aristolochene), solavetivone and lubimin in *H. muticus* (from vetispiradiene), gossypol and lacinilene C in cotton (from δ-cadinene), artemisinin in *A. annua* (from amorpha-4,11-diene), todomatuic acid in grand fir (from α-bisabolene), and cyclocolorenone in *S. canadensis* (from α-gurjunene). By-products have not evolved, in these cases, possibly because they have no biological significance.

Identification of the minor by-products obtained in an enzymatic reaction may, however, give insight into the catalytic route from FDP to the final product as has been suggested in studies on recombinant sesquiterpene synthases (5, 10, 27). In the case of amorpha-4,11-diene, a central question is if the first cyclization step is a 1,6-closure resulting in a bisabolyl cation, which then undergoes a 1,10-closure to the final carbon structure or if it is a 1,10-closure resulting in a germacradienyl cation, which subsequently experiences a 1,6-closure. The identification of α-bisabolol (0.3%) and β-sesquiphellandrene (0.5%) as by-products in the recombinant enzyme reaction proves that an initial 1,6-cyclization can occur. These minor products may, however, be produced by multiple conformers of the enzyme that arise from bacterial expression or the *in vitro* assay conditions. It may also be pointed out that no minor products formed through a 1,10-closure (e.g., the germacrenes) could be found in the product mixture, suggesting that the recombinant enzyme preferentially carries out an initial 1,6-closure. In addition to this, the formation of amorph-4-en-7-ol and amorph-4-en-11-ol (tentative) is difficult to explain with an initial germacrene formation. Consequently, based on this reason-

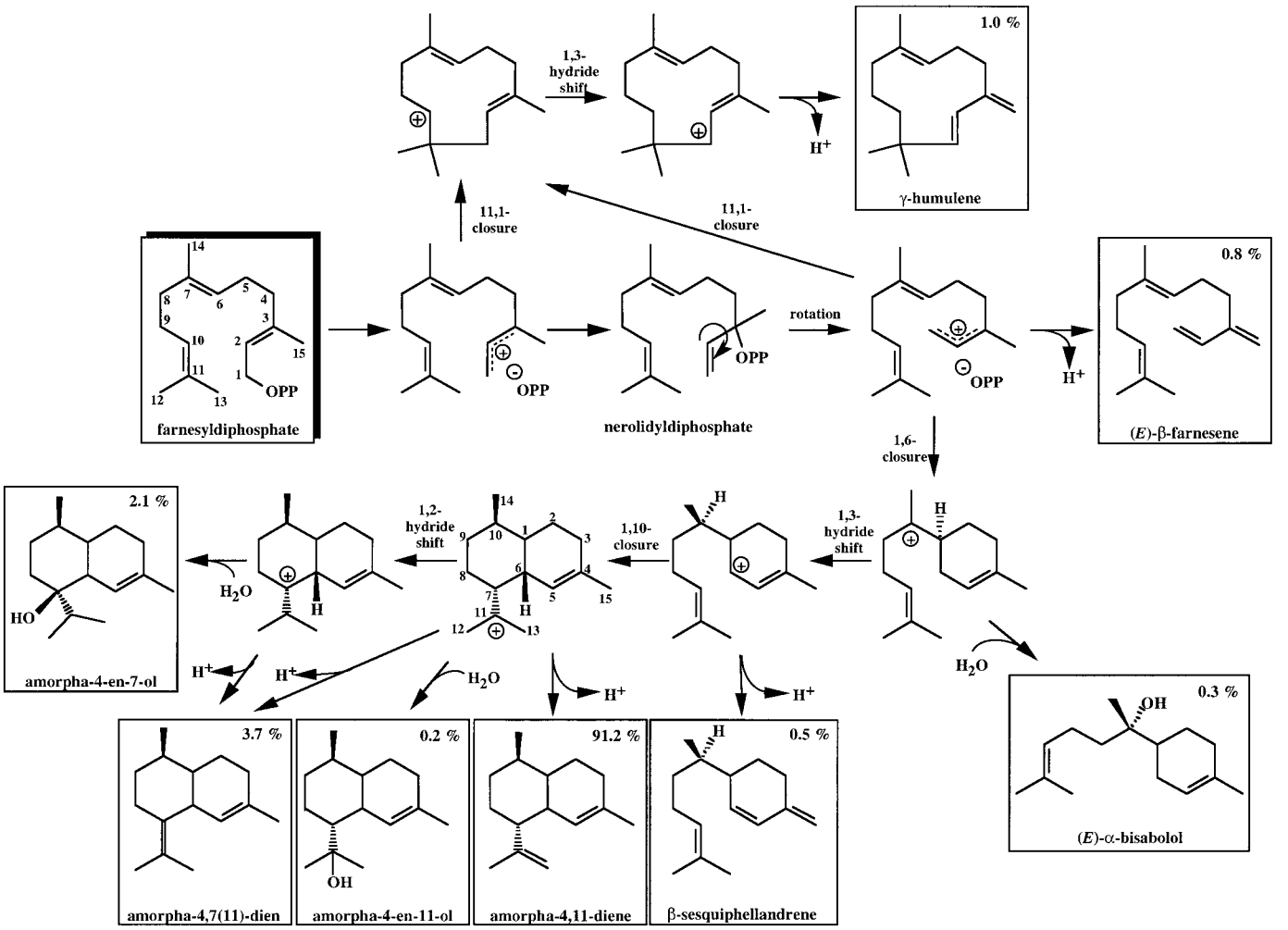

FIG. 4. Proposed mechanism for the formation of sesquiterpenes from FDP by amorpha-4,11-diene synthase. Structures enclosed in boxes have been identified by GC–MS except for amorpha-4-en-11-ol and α-bisabolol, which are only tentatively identified. The relative amount of each product is given in the boxes. (*E*)-β-Farnesene could also be derived by ionization of FDP and deprotonation of the resulting cation. OPP denotes diphosphate moiety.

ing, we suggest that amorpha-4,11-diene synthase first catalyzes a 1,6-closure with a subsequent 1,10-closure. The appearance of traces of γ-humulene in the product mixture (around 1% of total sesquiterpenoids) shows that the recombinant enzyme also carries out an initial 1,11-closure.

Using the information about products formed by the recombinant enzyme and the reported crystal structure of *epi*-aristolochene synthase from tobacco (24), a putative mechanism for amorpha-4,11-diene synthase may be suggested. Binding of FDP to the tobacco enzyme results in the closure of two loops over the active site that shield the developing carbocation from solvent water. This closure places the Tyr^{527} of the reaction pocket in proximity with the substrate and Trp^{273}; these residues are matched by Phe^{525} and Trp^{271} in amorpha-4,11-diene synthase. Two basic residues (Arg^{264} and Arg^{441}) in the tobacco enzyme, corresponding to Arg^{262} and Arg^{440} in amorpha-4,11-diene synthase, are brought close to each other by the loop movement. The positive charges assist in ionization of the substrate diphosphate ester and help to move the diphosphate anion away from the developing carbocation, with the assistance of the divalent metal cation(s) coordinated by the DDXXD motif present in all terpene synthases (25). After ionization of FDP by amorpha-4,11-diene synthase, the paired diphosphate anion is captured at C3 to form nerolidyl diphosphate, which allows rotation around C2–C3 to the *cisoid* conformer necessary for C1,C6-closure (Fig. 4). Ionization of nerolidyl diphosphate followed by ring closure generates the bisabolyl cation while deprotonation of farnesyl and/or nerolidyl cation results in the formation of (*E*)-β-farnesene. A 1,11-closure of either of these cations

followed by a 1,3-hydride shift and deprotonation yields γ-humulene (Fig. 4) in analogy with the mechanism suggested for γ-humulene synthase from *Abies grandis* (27).

The bisabolyl cation generated by amorpha-4,11-diene synthase may capture a water, releasing α-bisabolol as a minor product, or undergo a 1,3-hydride shift to yield a cationic intermediate which may deprotonate directly, releasing β-sesquiphellandrene as a by-product, or attack on C10 (FDP numbering) to generate the carbon skeleton of the final products. Deprotonation at either C12 or C13 of the amorphyl cation will afford the main product amorpha-4,11-diene and deprotonation at C7 will provide the by-product amorpha-4,7(11)-diene, whereas water capture gives rise to amorpha-4-en-11-ol. The presence of amorpha-4-en-7-ol indicates that an additional 1,2-hydride shift may occur before a water molecule is captured. Amorpha-4,7(11)-diene may also arise from deprotonation at C11 of this intermediate. In conclusion, it is possible that the formation of amorpha-4,11-diene is achieved by an initial 1,6-closure with a subsequent 1,10-closure.

The mechanism suggested for amorpha-4,11-diene synthase is similar to that suggested for *epi*-cedrol synthase from *A. annua* (5). In both cases, the bisabolyl cation is an intermediate but different hydride shifts on this intermediate result in a C1,C10-closure by the former enzyme and a C6,C10-closure by the latter enzyme. These two *Artemisia* enzymes of high sequence similarity offer an excellent experimental system to study structure–activity relationships in terpene synthases.

ACKNOWLEDGMENTS

The financial support to P.E.B. from the Swedish Research Council for Engineering Sciences and the Swedish Council for Forestry and Agricultural Research and to M.B. from Foundation for Strategic Environmental Research (MISTRA) is gratefully acknowledged. We thank Dr. Ben Jansen for the synthesis of amorpha-4,11-diene.

REFERENCES

1. White, N. J., Nosten, F., Looareesuwan, S., Watkins, W. M., Marsh, K., Snow, R. W., Kokwaro, G., Ouma, J., Hien, T. T., Molyneux, M. E., Taylor, T. E., Newbold, C. I., Ruebush, T. K., II, Danis, M., Greenwook, B. M., Anderson, R. M., and Olliaro, P. (1999) *Lancet* **353,** 1965–1967.
2. van Agtmael, M. A., Eggelte, T. A., and van Boxtel, C. J. (1999) *Trends Pharm. Sci.* **20,** 199–204.
3. Pras, N., Visser, J. F., Batterman, S., Woerdenbag, H. J., and Malingr, T. M. (1991) *Phytochem. Anal.* **2,** 80–83.
4. Jain, D. C., Mathur, A. K., Gupta, M. M., Singh, A. K., Verma, R. K., Gupta, A. P., and Kumar, S. (1996) *Phytochemistry* **43,** 993–1001.
5. Mercke, P., Crock, J., Croteau, R., and Brodelius, P. E. (1999) *Arch. Biochem. Biophys.* **369,** 213–222.
6. Woerdenbag, H. J., Bos, R., Salomons, M. C., Hendriks, H., Pras, N., and Malingré, T. M. (1993) *Flavour Fragr. J.* **8,** 131–137.
7. Ahmad, A., and Misra, L. N. (1994) *Phytochemistry* **37,** 183–186.
8. Bouwmeester, H. J., Wallaart, T. E., Janssen, M. H. A., van Loo, B., Jansen, B. J. M., Posthumus, M. A., Schmidt, C. O., De Kraker, J.-W., König, W. A., and Franssen, M. C. R. (1999) *Phytochemistry* **52,** 843–854.
9. Hua, L., and Matsuda, S. P. T. (1999) *Arch. Biochem. Biophys.* **369,** 208–212.
10. Colby, S. M., Crock, J., Dowdle-Rizzo, B., Lemaux, P. G., and Croteau, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2216–2221.
11. Crock, J., Wildung, M., and Croteau, R. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12833–12838.
12. Chen, X. Y., Chen, Y., Heinstein, P., and Davidsson, V. J. (1995) *Arch. Biochem. Biophys.* **324,** 255–266.
13. Facchini, P. J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11088–11092.
14. Back, K., and Chappell, J. (1995) *J. Biol. Chem.* **270,** 7375–7381.
15. Holm, Y., Laakso, I., Hiltunen, R., and Galambosi, B. (1997) *Flavour Fragr. J.* **12,** 241–246.
16. Wallaart, T. E., van Uden, W., Lubberink, H. G., Woerdenbag, H. J., Pras, N., and Quax, W. J. (1999) *J. Nat. Prod.* **62,** 430–433.
17. Vögeli, U., and Chappell, J. (1990) *Plant Physiol.* **94,** 1860–1866.
18. Zook, M. N., and Kuc, J. A. (1991) *Physiol. Mol. Plant Pathol.* **39,** 377–390.
19. Back, K., He, S., Kim, K. U., and Shin, D. H. (1998) *Plant Cell Physiol.* **39,** 899–904.
20. Adams, R. P. (1995) Identification of Essential Oil Components by Gas Chromatography/Mass Spectroscopy, Allured Publishing Corp., Carol Stream, IL.
21. Joulain, D., and Koenig, W. A. (1998) The Atlas of Spectral Data of Sesquiterpene Hydrocarbons, EB-Verlag, Hamburg, FRG.
22. Dung, N. X., Cu, L. D., Khien, P. V., Muselli, A., Casanova, J., Barthel, A., and Leclercq, P. A. (1998) *J. Essent. Oil Res.* **10,** 478–482.
23. Colby, S. M., Crock, J., Dowdle-Rizzo, B., Lemaux, P. G., and Croteau, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2216–2221.
24. Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997) *Science* **277,** 1815–1820.
25. McGarvey, D. J., and Croteau, R. (1995) *Plant Cell* **7,** 1015–1026.
26. Cane, D. E. (1990) *Chem. Rev.* **90,** 1089–1103.
27. Steele, C. L., Crock, J., Bohlmann, J., and Croteau, R. (1998) *J. Biol. Chem.* **273,** 2078–2089.
28. Bohlmann, J., Crock, J., Jetter, R., and Croteau, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6756–6761.
29. Schmidt, C. O., Bouwmeester, H. J., Bülow, N., and König, W. A. (1999) *Arch. Biochem. Biophys.* **364,** 167–177.