Isolation and functional expression of an animal geranyl diphosphate synthase and its role in bark beetle pheromone biosynthesis

Anna B. Gilg, Jeremy C. Bearfield, Claus Tittiger, William H. Welch, and Gary J. Blomquist*

Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557-0014

Edited by May R. Berenbaum, University of Illinois at Urbana–Champaign, Urbana, IL, and approved May 23, 2005 (received for review April 22, 2005)

Geranyl diphosphate synthase (GPPS) catalyzes the condensation of dimethylallyl diphosphate and isopentenyl diphosphate to form geranyl diphosphate. Geranyl diphosphate is the precursor of monoterpene molecules, a large family of natural occurring C10 compounds predominately found in plants. Similar to plants but unique to animals, some bark beetle genera (Coleoptera: Scolytidae) produce monoterpenes that function in intraspecific chemical communication as aggregation and dispersal pheromones. The release of monoterpane aggregation pheromone mediates host colonization and mating. It has been debated whether these monoterpenes are produced de novo through the mevalonate pathway or result from simple modifications of dietary precursors. The data reported here provide conclusive evidence for de novo biosynthesis of monoterpenic pheromone components from bark beetles. We describe GPPS in the midgut tissue of pheromone-producing male Ips pini. GPPS expression levels are regulated by juvenile hormone III, similar to other mevalonate pathway genes involved in pheromone biosynthesis. In addition, GPPS transcript is almost exclusively expressed in the anterior midgut of male I. pini, the site of aggregation pheromone biosynthesis. The recombinant enzyme was functionally expressed and produced geranyl diphosphate as its major product. The three-dimensional model structure of GPPS shows that the insect enzyme has the sequence structural motifs common to E-isoprenyl diphosphate synthases.

S
hort-chain E-isoprenyl diphosphate synthases (E-IPPSs) are a class of prenyltransferases that are central to isoprenoid metabolism. This group includes geranyl diphosphate synthase (GPPS; EC 2.5.1.1), farnesyl diphosphate synthase (FPPS; EC 2.5.1.10) and geranylgeranyl diphosphate synthase (GGPPS; EC 2.5.1.30) which synthesize geranyl diphosphate (GPP), farnesyl diphosphate, and geranylgeranyl diphosphate, respectively. These phosphorylated aliphatic products are the backbone components of >30,000 biologically essential isoprenoid-derived metabolites (e.g., terpenes, steroids, ether-linked lipids, dolichols, ubiquinones, retinoids, chlorophylls, and prenylated proteins) (1, 2). All E-IPPSs catalyze an ordered sequential prenyl transfer involving a 1’-4 condensation reaction between an allylic diphosphate and isopentenyl diphosphate (IPP) (3). The sequential addition of IPP to the growing hydrocarbon chain of the allylic substrate generates polyisoprenyl homologues of increasing chain length. The reaction proceeds and terminates precisely at a specific carbon chain length according to the enzyme’s product chain length specificity (4, 5).

The deduced amino acid sequences of all E-IPPSs have high similarity, and sequence alignments reveal five conserved sequence domains (I, II, III, IV, and V) (6). Domain II and IV contain the substrate-binding motifs, the two aspartate-rich regions (DDX2 ... DX represents any amino acid), which are flanked by other conserved residues. These two regions are defined as the first aspartate-rich motif (FARM) and the second aspartate-rich motif. Domain II, which includes the FARM, is designated as the chain length determination region because of the presence of amino acids that are recognized as determinants of product chain length specificity (7, 8).

FPPS and GGPPS are essentially ubiquitous in nature, whereas GPPS has more limited distribution in plants. GPPS catalyzes a single condensation of dimethylallyl diphosphate (DMAPP) with IPP to produce GPP, the C10 backbone component of monoterpene molecules (9). Monoterpane biosynthesis is primarily localized to plastids of plant cells (10–12), found in the essential oil and resin of plants they function in plant defenses, in allelopathic responses, and as attractants of pollinators (1). Recently, the isolation of cDNAs encoding GPPS from Mentha piperita (13), Arabidopsis thaliana (14), Abies grandis (15), Antirrhinum majus, and Clarkia breweri (11) has allowed for its functional characterization.

Monoterpane synthesis is usually not associated with animals. Subsequently in animals, GPP serves as an intermediate substrate for FPPS, which synthesizes FPP (C15), an essential precursor of cholesterol in vertebrates (1). In contrast, several bark beetle species produce monoterpane-derived aggregation pheromone to coordinate colonization and mating. Until the last decade, it was widely accepted that bark beetles, in contrast to most other insects (16), derived their monoterpane aggregation pheromones by simple modification of host tree dietary precursors (17, 18). However, it is becoming increasingly clear that most monoterpenes are synthesized de novo in the anterior midgut tissue (19–21) of pheromone-producing beetles through the mevalonate pathway (22, 23). The allylic C10 alcohol, ipsdienol, is the major monoterpane-derived aggregation pheromone of male Ips pini (Fig. 1) (24). Many mevalonate pathway genes involved in pheromone biosynthesis of I. pini have been identified, and the regulation of transcript expression has been characterized (25, 26).

In this study, we report the isolation of a GPPS of an animal and present evidence that it functions in bark beetle pheromone biosynthesis. A full-length GPPS cDNA from I. pini was cloned and functionally expressed. GPPS is a single-copy gene, and transcript levels are induced specifically in the anterior midgut of pheromone-producing male beetles. The three-dimensional model structure of GPPS, built on the crystal structure of the avian FPPS (27, 28), folded as a single α-helical domain with variable loop regions and showed the conserved structural features that are essential for E-IPPS activity.

This paper was submitted directly (Track II) to the PNAS office. Abbreviations: DMAPP, dimethylallyl diphosphate; E-IPPS, E-isoprenyl diphosphate synthases; FARM, first aspartate-rich motif; FPPS, farnesyl diphosphate synthase; GGPPS, geranylgeranyl diphosphate synthase; GPPS, geranyl diphosphate synthase; IPP, isopentenyl diphosphate.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY953508 (Ips pini GPPS), AY953507 (Ips pini FPPS), AY966009 (Dendroctonus jeffreyi FPPS), and AY966008 (Anthonomus grandis FPPS)).

*To whom correspondence should be addressed. E-mail: garyb@cabnr.unr.edu.
is stored at 8,000 g for 5 min by using a 10-ml conical centrifuge initial treatment with 10 μg of JH III or acetone. This time point was chosen because GPPS transcript levels are already induced at 8 h (data not shown). Each sample contained tissues pooled from five insects that were separated into head, anterior midgut, posterior midgut, hindgut, fat body, appendages (legs and antennae), and carcass. The tissues were kept at −80°C until total RNA was prepared by using the MasterPure RNA Purification Kit (Epicentre, Madison, WI). First-strand cDNA was synthesized by using Superscript III reverse transcriptase (Invitrogen) and an oligo(dt) primer (Promega) according to the manufacturer’s instructions. To normalize the template amounts and determine the range of linear amplification, PCR was performed with I. pini cytosolic actin primers (5′-CATGTTGTGAC-GAAAGATGAC-3′; 5′-GGCATACGCTTCTAGATGGG-3′) at several different cycle numbers. Once the template amounts were equalized, the same amounts were used with the GPPS primer set (5′-GTACCTCAGTACGGAATT-3′; 5′-CTCTTTAGT-GAAATCTGTTGA-3′) at the previously established number of cycles. All PCR products were separated on 1.5% agarose gels.

**Functional Expression of Recombinant GPPS.** A truncated version of IpiGPPS was generated for functional expression analysis. A cDNA coding region corresponding to amino acids P51-E416 was amplified by PCR with a 5′Flu Turbo Taq polymerase (Stratagene) and subcloned into the pET Trx Fusion System 32 (Novagen) in the pET-32c vector that had been previously digested with EcoRI and HindIII. The N-terminal His6 tagged truncated clone was generated by using an adapor oligonucleotide (5′- GCGAATTTCAAATCTGCAGGG-3′) as the upstream primer designed to create an EcoRI restriction site built into the start site and a downstream adaptor primer (5′-GCGAAGCTTCTACAGGG-3′) generating a HindIII restriction site at the terminal stop codon. The pET-32c construct was transformed into Escherichia coli BL21-Strain(ED3) pLysS cells (Invitrogen). The cell culture was grown to an OD600 of 0.5 and then induced in 1 mM isopropyl β-D-thiogalactoside for 3 h. Levels of protein expression were verified by SDS/PAGE (31).

**Prenyltransferase Activity and Product Distribution Analysis.** Prenyltransferase activity was measured by a standard acid liability assay that determined 1-[1-14C]IPP incorporation into polysoprenols (32). Samples were run in triplicate. Purified enzyme suspension (1–10 μg), clarified crude E. coli extract, or soluble beetle tissue extract (100–500 μg total protein) was incubated with 50 μM DMAPP and 10 μM 1-[1-14C]IPP (55 Ci/mmol, 1 Ci = 37 GBq) for 1 h at 32°C in a 200 μl reaction volume containing 20 mM Hepes (pH 7.2), 2.5 mM MgCl2, 5 mM KF, 1 mM DTT, 10% glycerol, and protease inhibitor mixture. Protein concentration was determined by the Bradford method (33) or estimated by SDS/PAGE. The identification of isoprenyl diphosphate product formation was performed by reverse phase-HPLC. To terminate the reaction before loading the sample onto the column, samples were boiled for 5 min and spun at (16,000 × g).
for 5 min). Clear homogenates (injection volume 200 μl: 100 μl of reaction sample mixed with 75 μl of buffer and 25 μl of authentic mixed standards [(DMAPP, GPP, FPP, GGPP; 1:1:0.5:0.5)] were directly loaded onto a Discovery 250 × 4.6 C18, 5 μm, silica column (Supelco) by using a HP 1050 series instrument. Phosphorylated short-chain isoprenyls were cleanly separated by using a solvent system of 25 mM NH4HCO3 pH 7.0: acetonitrile described by Zhang and Poulter (34). Samples were eluted at a flow rate 1 ml/min by using a linear gradient program: 100% 25 mM NH4HCO3 for 5 min and 0–100% acetonitrile for 30 min. Eluent was monitored by UV (214 nm). One milliliter fractions were collected by using a Amersham Pharmacia LKB-FRAC-100 Fraction Collector (Amersham Pharmacia Biotech), and fractions corresponding to IPP, DMAPP, GPP, FPP, and GGPP peaks (retention times 4, 16, 20, and 25 min, respectively) were assayed for radioactivity by using a TRI-CARB 2900 TR liquid scintillation analyzer (Packard).

Three-Dimensional Structural Modeling of GPPS. A three-dimensional model structure of GPPS from *I. pini* was constructed through knowledge-based methods by using SYBYL 6.91–7.0 (Tripos Associates, St. Louis) on Silicon Graphics (Mountain View, CA) Indigo2 workstations. The solution conformation was based on energy and statistical criteria by using a biased-threading algorithm [GENEFOLD (35), as implemented in SYBYL]. The crystallographic structure of avian mutant FPPS (Protein Data Bank ID code 1UBY) was used as a template to build the *I. pini* GPPS model structure. Gaps in the structurally conserved regions were closed by a loop search utility. Selection of loops from solved structures (Protein Data Bank ID codes 1UBY and 1FPS) was based on sequence homology, end-to-end distance and absence of severe steric clashes with the previously built sections of the insect target structure. The conformation of loops was refined by molecular dynamics and minimizations by using an AMBER7 force field. Pseudoenergy functions (MATCHMAKER) were used to evaluate energetically favorable conformers. The average positional energies were calculated, and the model was refined through energy minimizations (Powell, steepest descent) or molecular dynamics to obtain a local and global minimum, respectively. Pairwise alignments were made between the *I. pini* GPPS model structure and avian crystal structures (1FPS and 1UBY) to establish positions and nearby contacts of analogous catalytic residues located at or nearby the substrate binding site.

Results

Molecular Cloning and Sequence Analyses. A full-length cDNA clone encoding GPPS was isolated from a male *I. pini* cDNA library. The ORF is 1,248 bp, encoding a predicted 416-aa protein with a calculated molecular mass of 48.5 kDa and a pI of 8.33 (SwissProt/TrEMBL). The predicted translated product has the conserved sequence domains present in all *E*-IPPSs (6), including the FARM (DDIMD, D157–D161) and second aspartate-rich motif (NDFKD, N299–D303) (Fig. 2). Within the first 20 amino acids, residues from the start codon is a consensus N-terminal peroxisomal targeting sequence (S11–L19).

Localization of GPPS to the Midgut of Male *I. pini*. Northern analysis showed that GPPS is up-regulated by JH III in a dose- and time-dependent manner in male thoraces, similar to the transcriptionsal pattern described for other mevalonate pathway genes involved in pheromone biosynthesis (data not shown) (19, 36, 37). Semiquantitative RT-PCR demonstrated that the expression of GPPS was predominately localized to the anterior midgut tissue of pheromone-producing *I. pini* (Fig. 3). Notably, in the midgut of JH III-treated male beetles, GPP was the only short-chain isoprenyl diposphosphate product detected (Fig. 4A). Southern blot analysis showed a banding pattern consistent with a single-copy gene for GPPS of *I. pini* (data not shown).

Functional Expression and Product Formation. Induction of the transformed expression host with isopropyl-β-D-thiogalactoside consistent with a single-copy gene for GPPS of *I. pini* (data not shown).

![Fig. 2.](https://www.pnas.org/cgi/doi/10.1073/pnas.0503277102) The DNA coding region of GPPS is 416 amino acids. The conserved sequence domains (I–V) are underlined. The aspartate-rich motifs, common to all-ε-isoprenyl diphosphate synthases, are bold and underlined. A peroxisomal targeting sequence (italicized and underlined) is present at the N terminus.

![Fig. 3.](https://www.pnas.org/cgi/doi/10.1073/pnas.0503277102) A semiquantitative RT-PCR shows the tissue distribution of GPPS transcript from *I. pini* head (H), anterior midgut, AM; posterior midgut, PM; hindgut, HG; fat body, FB; appendages, AP; carcass, C.)

![Fig. 4.](https://www.pnas.org/cgi/doi/10.1073/pnas.0503277102) Southern blot analysis showed a banding pattern consistent with a single-copy gene for GPPS of *I. pini* (data not shown).
The major structural feature is a large central core surrounded by the insect enzyme is very similar to the avian crystal structure. The rms deviation of 2.2 Å, indicating the structural conformation of the avian FPPS (1FPS or 1UBY) has a respective atoms of the modeled GPPS superpositioned onto the crystal structure folds as a single all-\( /H11002\)-helices connected by loop regions (Fig. 5). The backbone atoms of the modeled GPPS superpositioned onto the crystal structure of the avian FPPS (1FPS or 1UBY) has a respective rms deviation of 2.2 Å, indicating the structural conformation of the insect enzyme is very similar to the avian crystal structure. The major structural feature is a large central core surrounded by 10 \( /H9251\)-helices. The all-\( /H9251\) and homo-\( /H9251\) substrate binding sites, the FARM and second aspartate-rich motif, respectively, are located in the largest of several solvent accessible channels.

To establish conserved function and catalytic competence of the proposed GPPS model, we identified and compared the location and positional orientation of reactive residues with the solved avian FPPS structure (27, 28). The conserved catalytic residues (38, 39) at the FARM and second aspartate-rich motif binding motifs in the avian template (1FPS and 1UBY) were aligned with analogous residues in the \( /H11021\) GPPS model structure. Importantly, the conserved aspartate residues (avian FPPS, D\(_{117}\), D\(_{121}\); \( /H11021\) GPPS, D\(_{157}\), D\(_{161}\)), which are involved in substrate binding at the FARM region were in isosteric positions within the active site (Fig. 5B).

Phylogenetic Analyses. To date, FPPS cDNAs have been cloned from three coleopterans. A multiple sequence alignment of \( /H20862\) FPPS (AY953507), \( /H20841\) Dendroctonus jeffreyi FPPS (AY966009) and \( /H20841\) Anthonomus grandis (boll weevil) FPPS (AY966008) has a relatively high amino acid sequence identity of 60%. When \( /H11021\) GPPS (AY953508) is included in the alignment, the sequence identity lowers to 21%. Phylogenetic analysis based on the degree of sequence similarity between short-chain \( /H20862\)-IPPSs shows that the \( /H11021\) GPPS shares a more recent ancestral origin with GGPPS (plant/animal) and GPPS (plant) rather than with FPPS (plant/animal) (Fig. 6A). Moreover, when comparing short-chain \( /H20862\)-IPPSs from several insects, the \( /H11021\) GPPS clusters with GGPPS from Drosophila melanogaster (Fig. 6B).

Discussion

Until now, the distribution of GPPS has been restricted to plants, which produce an abundant array of monoterpenes. Here, we report cDNA clone-encoding GPPS from an animal. GPPS of \( /H11021\) functions at a late stage in the mevalonate pathway. It is required to generate GPP, the \( /H11021\) precursor to ipsdienol, the major monoterpane-derived aggregation pheromone of male \( /H11021\).

JH III regulates pheromone biosynthesis by coordinate regulating gene expression of mevalonate pathway genes in pheromone-producing midgut tissue (40). Likewise, GPPS mRNA is up-regulated in midgut tissue of JH III-treated male \( /H11021\). Notably, there is an \( /H11021\)-fold higher basal level of GPPS in the midgut of males compared with female beetles, suggesting that GPPS may be modulated by sex-specific, developmental factors rather than directly by JH III (26). In turn, GPP was the only detectable short-chain isoprenyl diphosphate product produced in the midgut tissue of male beetles. Martin et al. (41) demonstrated the conversion of GPP to the acyclic monoterpene myrcene in male \( /H11021\). Moreover, monoterpane synthase activity was shown to be up-regulated by JH III. These data suggest that GPPS plays a regulatory role in controlling the flux of carbon into monoterpenoids.

We expressed the IpiGPPS clone in \( /H11021\) coli to characterize for functional activity. The recombinant enzyme expressed as a single subunit is shown to be active and is not suggested to be a heterodimer. The recombinant enzyme yielded GPP as its major product. Although FPP was also detected in trace amounts, GGPP and longer-chain products were not present under the assay conditions. Short-chain \( /H20841\)-IPPSs effectively use several allylic substrates to synthesize products with carbon chains longer than their native product specificities (8, 15, 42). Whether FPP production is due to the intrinsic activity of the recombinant GPPS or the in vitro assay conditions is unclear.

Southern blot analysis performed on genomic DNA of \( /H11021\) revealed GPPS as a single-copy gene. Nevertheless, it cannot be ruled out that several isoforms of GPPS translated from a single gene may exist for \( /H11021\). Bouvier et al. (14) proposed two isoforms of GPPS translate from the same gene in \( /H20862\) thaliana. Depending on the transcriptional start site, a plastid- or a cytosolic-targeted enzyme may be generated.

The deduced amino acid sequence of GPPS from \( /H11021\) has the conserved sequence domains required by all \( /H20841\)-IPPSs for substrate binding and catalytic activity (6), included is a peroxisomal substrate binding motif (SVX5QL; S\(_{94}\)-L\(_{102}\)) has been identified in 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-S) in \( /H11021\) jeffreyi (43).
Olivier et al. (44) showed this version of PTS-2 (SVX5QL) targeted cytosolic HMG-S and mevalonate diphosphate decarboxylase to the peroxisomes in Chinese hamster ovary cells. The absence of a peroxisomal signaling peptide for \textit{I. pini} FPPS indicates that the two short-chain E-IPPSs are possibly targeted to different subcellular locations for functional activity. Subcellular compartmentalization may function in the regulation of GPPS activity and pheromone production in bark beetles.

The \textit{I. pini} GPPS model structure has consensus with the avian FPPS crystal structure (27, 28). Reasonable assumptions were drawn from the deduced hypothetical structure based on the known mechanism of reaction and product chain length specificity of E-IPPSs (3, 8, 45). Importantly, analogous substrate binding and catalytic residues are similarly located and oriented in the \textit{I. pini} GPPS model and avian FPPS crystal structure.

A three-dimensional model structure of FPPS from \textit{I. pini} was also constructed through knowledge-based methods. Currently, ligand–protein interactions are being simulated at the binding site of both the \textit{I. pini} GPPS and \textit{I. pini} FPPS model structures. The substrate binding pocket of \textit{I. pini} GPPS model structure is smaller in size than that of both the avian FPPS crystal and \textit{I. pini} FPPS model structures, which reflects their known product chain-length specificities (2, 28). Preliminary analysis of steric interactions within the substrate binding pocket of the GPPS model structure shows that GPP is well fitted within the binding pocket, the distal end of the isoprene tail extends and is located at the pocket floor, whereas FPP is not well accommodated and tightly kinks within the binding pocket. Importantly, these comparative structural studies will help define determinants of product chain-length specificity between C_{10}– and C_{15}-producing enzymes within the same species.

Phylogenetic analyses support that IPPSs have evolved from a common ancestor into three distinct clades. The earliest divergence was a separation of the long-chain and short-chain E-IPPSs (6). It has been suggested that an archaea bacterial GGPPS-type enzyme is the origin of short-chain biosynthesis. GPPS of
plants are more evolutionary related to GGPPS than to FPPS. A phylogenetic tree based on the degree of sequence similarity between selected eukaryotic short-chain E-IFPSs show that FPPS (insect), GGPPS (plant/animal) and GPPS (plant/I. pini) all share a common origin. The next branching event within this cluster separates insect FPPS from GGPPS and GPPS. This event signifies that GPPS of I. pini evolved along the same lineage as GGPPS and GPPS of other eukaryotes. It should be noted that the GGPPS of the insect FPPS enzymes that is similar to what is found for GPPS of all share a common origin. The next branching event within this cluster separates insect FPPS from GGPPS and GPPS. This next event within this cluster separates insect FPPS from GGPPS and GPPS.

We thank the Whittell Forest management board for allowing us to collect beetle-infested trees. This work is supported in part by National Science Foundation Grants IBN 9906530 and IBN 0316370 and the Nevada Agricultural Experiment Station publication no. 05055518.