Aristolochene Synthase

ISOLATION, CHARACTERIZATION, AND BACTERIAL EXPRESSION OF A SESQUITERPENOID BIOSYNTHETIC GENE (Aril) FROM PINKILLIUM ROQUEFORTI

(Received for publication, August 21, 1992)

Robert H. Proctor and Thomas M. Hohn‡
From the Mycotoxin Research Unit, United States Department of Agriculture/Agricultural Research Service, National Center for Agricultural Utilization Research, Peoria, Illinois 61604

Aristolochene is the likely precursor of the sesquiterpenoid toxins produced by a number of filamentous fungi. One of these, PR-toxin, is produced by Penicillium roqueforti. We report here the isolation of a gene (Aril) coding for the sesquiterpene cyclase, aristolochene synthase (AS), from P. roqueforti. Nucleotide sequence analysis of genomic and cDNA clones revealed that the Aril gene contains two introns. A Protein A/AS fusion enzyme was expressed in Escherichia coli and shown to have sesquiterpene cyclase activity. Analysis of the Protein A/AS fusion enzyme reaction mixtures by TLC and gas chromatography/mass spectrometry identified aristolochene as a major product. The Aril gene encodes a polypeptide of molecular weight 39,200. Expression of Aril occurs in stationary phase cultures of P. roqueforti and appears to be transcriptionally regulated.

Sesquiterpenoids comprise a structurally diverse group of cyclic terpenoids, some of which are known to function as insect pheromones (1) and plant defensive compounds (2). The functions of most fungal sesquiterpenoids are unknown. Many plant pathogenic fungi produce phytotoxic sesquiterpenoids which, in a few cases, have been shown to contribute to their virulence (3, 4). Aristolochene is an eremophilane-type sesquiterpenoid that is produced by a variety of organisms. The (+)-enantiomer has been identified from plant (5) and insect (6) sources, whereas the (+)-enantiomer has been identified in extracts from the fungus Aspergillus terreus (7). Aristolochene is also the likely parent compound for a number of sesquiterpenoid toxins (Fig. 1) produced by filamentous fungi (7-12). A family of sesquiterpenoids, which appear to be derived from aristolochene, is produced by the blue cheese mold, Penicillium roqueforti (13). The best characterized of these sesquiterpenoids, PR-toxin, has been implicated in incidences of mycotoxicoses resulting from the consumption of contaminated grain (14).

Sesquiterpene cyclases catalyze the first unique step in the biosynthesis of most sesquiterpenoids and, as a result, may represent an important regulatory step in sesquiterpenoid biosynthesis. They cyclize the isoprenoid pathway intermediate, trans,trans-farnesyl pyrophosphate (FPP),1 to yield as many as 200 different sesquiterpenoids (15). In addition, their products serve as the carbon skeletons of pathway end products. Previously, we isolated aristolochene synthase (AS) from P. roqueforti and showed that its properties are characteristic of sesquiterpene cyclases (16). The presence of a (+)-AS has recently been demonstrated in A. terreus (17). Investigations using a partially purified preparation of this enzyme have provided details concerning the stereochemical course of the aristolochene cyclization reaction (17, 18).

1The abbreviations used are: FPP, farnesyl pyrophosphate; AS, aristolochene synthase; GC/MS, gas chromatography/mass spectrometry; MOPS, 3-(N-morpholino)propanesulfonic acid; nt, nucleotide(s); PCR, polymerase chain reaction.
Studies of sesquiterpene and other terpene cyclase reactions indicate that the divalent metal-assisted ionization of FPP is a common mechanistic feature of this group of enzymes (19, 20). However, despite this and other mechanistic similarities, the relationships between these enzymes remain unclear. The physical and chemical properties of sesquiterpene cyclases suggest that they may not be closely related. Sesquiterpene cyclases from microbial (16, 21), plant (22–24), and streptomy- cete (25) sources differ in molecular weight over a range of 48,000 to 80,000 and appear to have different subunit compositions. The gene coding for the sesquiterpene cyclase, trichodiene synthase, has recently been isolated from two different fungi (26, 27). It is presently the only terpene cyclase for which the complete primary structure has been reported.

We have isolated a gene coding for AS (Aril) from P. roqueforti. Here we report the sequence of Aril and its expression in Escherichia coli. The regulation of Aril expression in P. roqueforti is also discussed.

Procedure

**EXPERIMENTAL PROCEDURES**

**Materials**—AS was purified from P. roqueforti as described previously (16). A sample of purified enzyme was submitted to the Protein and Carbohydrate Structure Facility at the University of Michigan where following a 5% C digestion, peptides were isolated by high performance liquid chromatography and sequenced. The complete sequencing of both DNA strands was performed with the Sequenase Version 2.0 kit purchased from U. S. Biochemical Corp. (Cleveland, OH), using double-stranded plasmid templates. Both [3H] trans,trans-farnesyl pyrophosphate (682 mCi/mmol) and (±)-aristolochene were kindly provided by Professor D. E. Cane of Brown University. All DNA probes were labeled with [α-32P]dCTP obtained from Du Pont-New England Nuclear using the Prime-a-Gene System Kit (Promega Biotech). The plasmid pGEM7zf+ was purchased from Promega Biotech.

**Strains, Media, and Culture Conditions**—P. roqueforti NRRL 849 was obtained from the USDA/ARS Culture Collection at the National Center for Agricultural Utilization Research, Peoria, IL. Cultures were grown in YPG medium (0.3% yeast extract, 1% peptone, 2% glucose) for DNA isolation and in GYEP medium (5% glucose, 0.1% yeast extract, 0.1% peptone) for RNA and enzyme analyses. All cultures were inoculated with 106 conidia/ml and incubated at 28 °C on a gyratory shaker (180 rpm). Dry weight analysis was performed as described in Ref. 28. The E. coli strains SURE and XL1-Blue (Stratagene Inc.) were used for all cloning procedures, except for some of those involving pHIT2T (Promega Biotech Inc.), which employed strain X4380-1 as the host.

**Polymerase Chain Reaction**—In PCR experiments designed to amplify a portion of the Aril gene, partially degenerate oligonucleotides were synthesized based on the amino acid sequences of peptides isolated from purified AS. The primers for PCR were 5'-CCGAAATTCAATGCTACCTCACAAGAAAC-3' and 5'-GCTGC ACTACATTCTAGTGAGCGACG-3'. The conditions for PCR were as described above. The resulting PCR product was extracted from an agarose gel band with GeneClean (BIO 101) and used as a template for a second PCR step using the two primers and conditions described above. The second PCR product was extracted as above and cloned into a plasmid derived from pGEM7zf+ (pTA3) that had been modified to permit the generation of 3'-T overhangs following digestion with XcmI.

**Enzyme Assays**—To prepare cell homogenates of P. roqueforti, the fungus was grown and harvested as described in Ref. 16. Mycelial mats were ground in liquid N2 and then suspended in Buffer T (16). This suspension was first centrifuged at 8,500 g at 4 °C for 10 min, and the resulting supernatant was centrifuged again at 15,600 g at 4 °C for 25 min. For the expression of AS in E. coli, cells were grown and harvested as described previously (32), except that expression was induced by changing the incubation temperature from 30 to 42 °C. Cells were suspended in Buffer T (4 °C) and broken with the microtip probe of a sonic disrupter. Cell homogenates were centrifuged at 15,600 g and 4 °C for 25 min, and were immediately used in enzyme assays or stored at −65 °C. Enzyme activity was assayed as previously described (16), except that reactions contained 25 mM MOPS (pH 7.0), 2 mM MgCl2, 1 mM dithiothreitol, 10 mM [3H] FPP, and 10 μl of cell homogenate in a total volume of 150 μl. Reactions were incubated at 30 °C for 30 min. Partial purification of AS was as described in Ref. 16 through the Mono Q (Pharmacia LKB Biotechnology Inc.) step. Protein concentrations were determined by the method of Bradford (33) using the Bio-Rad protein assay with bovine γ-globulin as the standard.

**Analysis of Reaction Products**—Protein A/AS fusion enzyme reactions were stopped by the addition of an equal volume of ethanol (1.0 ml), and products were extracted with 4.7 ml of hexane. Material that could be eluted from a silica gel mini-column (1.0 g of silica gel) with hexane was concentrated under a gentle stream of nitrogen. For TLC analysis, products were separated on reverse-phase TLC plates (Chromatogram Whatman International Ltd.) developed in methanol. The plates were sprayed with En' Hance (Amersham Corp.) and exposed to Kodak X-AR film for 48 h. For GC/MS analysis, samples were injected into a 30-m fused silica capillary column (DB-5, 0.25 W Scientific) directly connected to the source of a quadrupole tandem mass spectrometer (Finnegan TSQ 700). Aristolochene was identified by comparison with a semisynthetic preparation of (±)-aristolochene.

**Computer Analyses**—Comparative analyses of specific sequences were performed with the protein matrix homology program within the IBI Fastall sequence analysis package. Comparative analyses involving protein data bases were performed with the FASTA program (34) for the GENPEPT 69.0 and SWISS-PROT 19.0 data bases and the Bionet IFIND program (IntelliGenetics) for the PIR 31 data base.

**RESULTS**

**Isolation and Characterization of Aril**—The sequences of the peptides isolated from a preparation of purified AS are shown in Fig. 2a. The presence of sequences of peptides AS2 and AS3, the partially degenerate oligonucleotides OAS2-1 and OAS3-1RC (Fig. 2b) were designed and used to amplify a 948-base pair fragment from P. roqueforti genomic DNA. The sequence of this fragment, designated PR1, encoded two additional AS peptides, AS1 and AS4 (Fig. 2a), that eluted from the high

T. M. Hohn, unpublished work.
Aristolochene Synthase Gene from P. roqueforti

**Fig. 2.** Sequences of peptides isolated from purified AS (a) and partially degenerate oligonucleotides used to amplify PR1 (b). The underlined portions of AS2 and AS3 represent the sequences used to design OAS2-1 and OAS3-1RC, respectively.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS1</td>
<td>Gln-Leu-Ser-Val-Val-Asn-Asp-Ile-Tyr-Ser-Tyr-Asp-Lys</td>
</tr>
<tr>
<td>AS2</td>
<td>Glu-Val-Gln-Asp-Glu-Val-Asp-Gly-Tyr-Phe-Leu-Glu-Asn-Trp-Lys</td>
</tr>
<tr>
<td>AS4</td>
<td>Ala-Val-Arg-Phe-Phe-Leu-Asp-Ala-Lys</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAS2-1</td>
<td>5'-CGGATCCGTYCARGAYGARGTYGAYGGYTA-3'</td>
</tr>
<tr>
<td>OAS3-1RC</td>
<td>5'-CTTSCWCCAYTGYTCRTT-3'</td>
</tr>
</tbody>
</table>

**Fig. 3.** Restriction map of the 4.5-kilobase insert in pPR2-6. The expanded portion of the figure represents the 1126-base pair Aril-coding region. Introns are indicated by the black boxes. The arrow indicates the direction of transcription.

**Table:** Peptide and Nucleotide Sequences

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS1</td>
<td>Gln-Leu-Ser-Val-Val-Asn-Asp-Ile-Tyr-Ser-Tyr-Asp-Lys</td>
</tr>
<tr>
<td>AS2</td>
<td>Glu-Val-Gln-Asp-Glu-Val-Asp-Gly-Tyr-Phe-Leu-Glu-Asn-Trp-Lys</td>
</tr>
<tr>
<td>AS4</td>
<td>Ala-Val-Arg-Phe-Phe-Leu-Asp-Ala-Lys</td>
</tr>
</tbody>
</table>

**Table:** Primer and Nucleotide Sequences

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAS2-1</td>
<td>5'-CGGATCCGTYCARGAYGARGTYGAYGGYTA-3'</td>
</tr>
<tr>
<td>OAS3-1RC</td>
<td>5'-CTTSCWCCAYTGYTCRTT-3'</td>
</tr>
</tbody>
</table>

Performance liquid chromatography column as a single peak.

After screening a λZAP genomic library with PR1, two positive phage were isolated. The plasmid, pPR2-6, was excised from one of these phage and shown to carry a 4.5-kilobase insert. Restriction mapping of the pPR2-6 insert (Fig. 3) localized the PR1 sequence to a 1562-base pair region. Sequence analysis of this 1562-base pair region indicated that it contained the entire coding region of the Aril gene (Fig. 4). To further characterize the Aril gene, the coding region was amplified by PCR from a cDNA template. Comparison of the genomic and cDNA sequences revealed the presence of two introns. Both of the Aril introns contained the 5', 3', and internal consensus sequences identified for introns from filamentous fungi (35). The start codon indicated in Fig. 4 is the first ATG in the open reading frame that encodes all four of the isolated AS peptides. This open reading frame specifies a 342-amino acid protein with a calculated molecular weight of 39,200 that is in agreement with the molecular weight of 37,000 determined for AS by SDS-polyacrylamide gel electrophoresis (16).

Southern analysis of genomic DNA digested with six different restriction enzymes that cut outside the Aril coding region revealed a single band (Fig. 5). Digestion with SacI, which cuts within the Aril coding region, produced two bands. These results suggest that Aril is present as a single copy in the P. roqueforti genome.

Expression of Aril in E. coli—To confirm the identity of the isolated gene, efforts were made to express Aril in E. coli. An open reading frame was constructed through the specific deletion of the intron sequences and inserted into the expression vector pRIT2T. Proteins expressed in pRIT2T contain an N-terminal sequence of 260 amino acids that codes for the IgG-binding domain of Protein A. Previously, we had observed that the expression of a Protein A/trichodiene synthase fusion protein resulted in a product that retained sesquiterpene cyclase activity.\(^3\)

Induction of fusion protein expression lead to the accumulation of low levels of sesquiterpene cyclase activity that were not observed in the control strains carrying pRIT2T. The level of sesquiterpene cyclase activity in cell homogenates was 0.6 nmol/h/mg protein. Analysis of reaction mixtures by TLC (Fig. 6) revealed a spot on autoradiographs that co-migrated with the product of purified AS from P. roqueforti. This spot was not observed in the reaction mixtures from control cultures. Fractionation of cell homogenates containing the fusion enzyme on a Mono Q column resulted in the elution of a single peak of sesquiterpene cyclase activity. Reaction mixtures from the pooled Mono Q fractions appeared identical with the column starting material when analyzed by TLC. Further analysis of reaction mixtures by GC/MS identified aristolochene as a product only in samples containing the fusion enzyme. The amount of aristolochene detected was within 30% of that predicted based on the radioactivity of the [1-\(^3\)H]FPP substrate and the radioactivity present in the sample analyzed by GC/MS.

**Regulation of Aril Expression**—AS activity was previously observed to occur in cultures of P. roqueforti grown in GYEP medium during the stationary growth phase (16). To investigate the regulation of AS expression, cultures grown in GYEP medium were harvested at various times between 0 and 70 h following inoculation and analyzed for Aril mRNA, AS activity, and dry weight. Northern blots of RNA isolated at 40, 50, 60, and 70 h postinoculation contained a single band of approximately 1350 nt when probed with the Aril coding region. The size of this mRNA is close to the predicted size for the Aril mRNA if the coding region of the Aril gene is 1126 nt. Changes in AS activity closely paralleled changes in the levels of Aril mRNA (Fig. 7). AS activity and mRNA levels were low at 30 h, peaked between 40 and 60 h, and were again low at 70 h. Both the increases in mRNA and AS activity occurred in the stationary growth phase as indicated by the leveling off of culture dry weight accumulation between 30 and 70 h.

**Homology with Other Sesquiterpene Cyclases**—The Aril sequence was compared with that of the trichodiene synthases from Fusarium sporotrichioides (26) and Gibberella pulicaris (27), the 5-epi-aristolochene synthase from tobacco (36), and a partial sequence of the pentalenene synthase from Streptogy

---

\(^3\) T. M. Hohn, unpublished result.
myces UC-5319.4 No significant homology was observed between any of these sequences. Direct comparison of Aril with the sequences of several prenyltransferases (37–40) and computer searches of protein sequence databases also failed to detect homologous sequences. The sequence (IL, V)XDDXXD occurs in two of the three homologous domains identified in prenyltransferases (41) and is thought to be involved in substrate binding. A search for this sequence in AS and available sesquiterpene cyclase sequences revealed that a single copy is present in all sesquiterpene cyclases except AS. Further analysis of AS identified the related sequence LXDDXXE starting at amino acid 113.

4 D. E. Cane, personal communication.

**Fig. 4. Nucleotide sequence of the Aril gene and the predicted amino acid sequence of AS.** Introns are indicated as untranslated regions (nucleotides 347–395 and 670–720). The nucleotide sequences used to amplify the Aril cDNA by PCR are underlined. The amino acid sequences of peptides isolated from purified AS are double underlined.

**DISCUSSION**

We have isolated and characterized the Aril gene that encodes the sesquiterpenoid biosynthetic enzyme aristolochene synthase. Two lines of evidence indicate that the gene encoded codes for AS: 1) the presence of nucleic acid sequences in Aril corresponding to the amino acid sequences of four peptides isolated from purified AS, and 2) the expression of an enzyme with AS activity in E. coli cells transformed with a chimeric gene containing the Aril coding sequence. In addition, the deduced amino acid sequence of Aril predicts a polypeptide that is within 10% of the molecular weight reported for AS (16). The Aril gene contains two introns and is present as a single copy in the genome of the *P. roqueforti* as revealed by cDNA sequence and Southern blot analysis.

The production of PR-toxin has been reported to occur in...
5, PstI; DNA digested with purified AS producing Protein A/AS fusion enzyme; products by PR1.

induction of PR-toxin in stationary phase cultures. Expression vector (pRIT2T) without the Aril-coding region; PR. farnesol prepared by apyrase treatment of Mono Q column run with sample from P. roqueforti; lune 3, cell homogenate from E. coli strain carrying the expression vector (pRIT2T) without the Aril-coding region; lane 7, farnesol prepared by apyrase treatment of [3H]FPP.

P. roqueforti cultures during the stationary growth phase (42, 43). We analyzed cultures for Aril expression and found that both mRNA and AS activity are present only in stationary phase cultures. It was also observed that changes in the levels of mRNA paralleled changes in enzyme activity. These results indicate that transcriptional controls may regulate the production of PR-toxin in stationary phase cultures.

The polypeptide encoded by Aril has a calculated pI of 5.05 and does not contain either membrane translocation sequences or membrane-spanning regions. This is consistent with the observation that AS and other terpene cyclases are cytosolic enzymes (15). Comparison of AS with the sequences of other sesquiterpene cyclases failed to identify regions of significant homology. The lack of homology with 5-epi-aristolochene synthase is particularly interesting since the products of these enzymes are structurally similar and their proposed mechanisms include several common intermediates (19). Prenyltransferases represent another group of enzymes that employ allylic pyrophosphates as substrates. The 1'-4 condensation reactions catalyzed by prenyltransferases involve an electrophilic mechanism analogous to that of terpene cyclases. Like sesquiterpene cyclases, they require metal cofactors such as Mg$^{2+}$ and Mn$^{2+}$ for activity. Several domains composed of conserved amino acids have been observed in prenyltransferases (41). In FPP synthetases from yeast and mammalian sources, the sequence (I,V,L)XDXXD occurs in two different domains. It has been proposed that these aspartate-rich sequences function in catalysis by forming salt bridges together with Mg$^{2+}$ and the pyrophosphate moiety of the prenyl substrates. AS is the only sesquiterpene cyclase that does not contain this sequence, although it does contain the related sequence LXDDXXE. The sesquiterpene cyclase sequences surrounding the putative Mg$^{2+}$ binding motif have little or no homology to each other or to the sequences surrounding this motif in prenyltransferases.

The apparent unrelatedness of sesquiterpene cyclases suggests these enzymes may be the result of convergent evolution and therefore do not have a common ancestor. Other examples of convergent evolution include serine proteases from mammals and bacteria (44) and phytoene desaturases from plants (45) and bacteria. These enzymes have different primary structures but are functionally similar with respect to substrate binding and catalysis. Alternatively, the lack of primary structure conservation in sesquiterpene cyclases may mean that the common mechanistic features of these enzymes, such as the FPP binding and ionization, are largely independent of primary structure.

Studies of AS and other sesquiterpene cyclases have been limited by enzyme availability. The isolation of the Aril gene will permit the development of an efficient expression system for AS. Preliminary results indicate that native AS can be expressed to high levels in E. coli using a vector based on the

---

phage T7 RNA polymerase/promoter system.

Acknowledgments—We thank R. Plattner for GC/MS analyses and Professors J. Chappell and D. E. Cane for providing us with previously unpublished data. We also thank Drs. T. Hayman and J. Miernyk for critically reviewing the manuscript.

REFERENCES


Aristolochene Synthase Gene from P. roqueforti