Bioorganic Chemistry **29,** 293–307 (2001)
doi:10.1006/bioo.2001.1216, available online at http://www.idealibrary.com on IDEAL®

Hexanoate Synthase, a Specialized Type I Fatty Acid Synthase in Aflatoxin B_1 Biosynthesis

Timothy S. Hitchman,* Eric W. Schmidt,* Frances Trail,† Matthew D. Rarick,‡ John E. Linz,‡ and Craig A. Townsend*,[1]

**Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland; and †Department of Botany and Plant Pathology, and ‡Department of Food Science and Human Nutrition, Michigan State University, East Lansing, Michigan*

Received February 27, 2001

In fungi, fatty acids are biosynthesized by large multifunctional enzyme complexes, the fatty acid synthases (FASs), which catalyze chain assembly in an iterative manner. Many fungal secondary metabolites contain fatty acid moieties, and it is often unclear whether they are recruited from primary metabolism or are biosynthesized *de novo* by secondary metabolic FASs. The most convincing evidence of such a dedicated FAS comes from the biosyntheses of aflatoxin (AF) and sterigmatocystin (ST) in certain species of the filamentous fungus *Aspergillus.* Incorporation studies in AF and genetic analyses of ST and AF biosynthesis strongly suggest that their biosyntheses begin with the production of a C_6 fatty acid by a specialized FAS. The genes encoding the alpha (*hexA*) and beta (*hexB*) subunits of this hexanoate synthase (HexS) from the AF pathway in *Aspergillus parsiticus* SU-1 were cloned and both their gDNAs and cDNAs were sequenced and their transcriptional ends analyzed. Translated amino acid sequences are predicted to result in proteins of 181.3 and 210.5 kDa, for HexA and HexB, respectively. Comparison of the HexA and HexB sequences with those of the ST FAS subunits and primary metabolic FASs indicated that the secondary metabolic enzymes are members of a well-defined subclass of the FAS family. Phylogenetic predictions and an analysis of GC-bias in AF and ST pathway genes compared with primary metabolic *Aspergillus* genes were used as a basis to propose a route for the evolution of the AF and ST clusters.

Key Words: filamentous fungi; polyketide; DNA sequence; natural product biosynthesis; yeast; polyketide synthase; aflatoxin B_1; sterigmatocystin.

INTRODUCTION

Significant advances in understanding the genetics and enzymology of fatty acid and polyketide biosynthesis have taken place in the past decade (*1*). The body of research on polyketide natural products has focused on prokaryotes, building on the

[1] To whom correspondence and reprint requests should be addressed. Department of Chemistry, The Johns Hopkins University, 3400 North Charles Street, Baltimore MD 21218. Fax: (410) 261-1233. E-mail: townsend@jhunix.hcf.jhu.edu.

0045-2068/01 $35.00

extensive knowledge of fatty acid biosynthesis in bacteria developed over several decades (*2*). In contrast, fatty acid and polyketide biosynthesis in eukaryotes such as the filamentous fungi remain poorly characterized at both the genetic and the enzymatic levels. The enzymes that catalyze the biosynthesis of these natural products in fungi are large multidomain, multifunctional proteins that operate iteratively (*3,4*). These fungal type I synthases are relatively difficult to handle and the iterative nature of their enzymatic programming at each round of chain elongation and modification is particularly hard to unravel.

The best-studied fungal fatty acid synthase (FAS)[2] is the palmitate synthase from *Saccharomyces cerevisiae* (*4,5*). This FAS comprises two subunits, alpha and beta, which form a megasynthase complex of $\alpha_6\beta_6$ stoichiometry having six sites of palmitate synthesis. The active site domains necessary for fatty acid assembly are shared across the two subunits. Thus, the α-protein contains the acyl carrier protein (ACP) domain together with the β-ketoreductase (KR), β-ketosynthase (KS), and phosphopantetheinyl transferase (PPT) active sites. The five remaining activities of acyl transferase (AT), enoylreductase (ER), dehydratase (DH), and malonyl/palmitoyl transferase (MT/PT) are located on the larger β-subunit. Evidence from electron microscopy of the purified *S. cerevisiae* FAS suggests that the whole complex adopts an overall barrel shape, sequestering substrates within (*4*).

Genes encoding fungal type I FASs have been identified in several organisms, including *S. cerevisiae*. All the deduced amino acid sequences share high levels of similarity and reveal the same domain organization in the translated proteins (*6,7*). Most of these fungal FAS genes code for primary metabolic enzymes essential for cell growth. There are several known secondary metabolites, however, that appear to contain fatty acid chains or derivatives thereof. Furthermore, there is evidence in a few cases that a dedicated FAS may be involved in provision of the fatty acyl units. For instance, piliformic acid, isolated from *Poronia piliformis*, is assembled from an octanoate moiety and a C_3 unit (*8*) (Fig. 1A). Recent chemical evidence indicates that the octanoate is not derived by β-oxidation of a larger fatty acid, suggesting the possible involvement of an octanoate synthase in the fungus (*9*). Similarly, HC toxin is a cyclic tetrapeptide produced by *Cochliobolus carbonum* in which the fourth amino acid is the fatty acid-derived 2-amino-9,10-epoxy-8-oxo-decanoic acid (Fig. 1A). In a recent report the identification of a genomic locus is described that is present only in HC-toxin-producing strains (*10*). Part of the novel locus *tox2*$^+$ encodes a protein ToxC with extensive amino acid sequence similarities to the known β-subunits of fungal FASs. Disruption of the *toxC* gene did not affect growth of the fungus, but

[2] Abbreviations used: ACP, acyl carrier protein; AF, aflatoxin; AT, acyl transferase; AUAP, abridged universal amplification primer from Gibco-BRL; bp, base pair(s); cDNA, DNA complementary to RNA; DH, dehydratase; DMSO, dimethyl sulfoxide; dNTP, deoxyribonucleotide triphosphate; ds, double strand(ed); ER, enoyl reductase; FAS, fatty acid synthase; FMN, flavin mononucleotide; GC, guanosine or cytosine; gDNA, genomic DNA; HexS, hexanoate synthase; kb, kilobase; KR, β-ketoreductase; KS, β-ketosynthase; LB, Luria–Bertani medium; MT, malonyl transferase; NA, norsolorinic acid; NADH, reduced form of nicotinamide-adenine dinucleotide; nt, nucleotide(s); PKS, polyketide synthase; PPT, phosphopantetheinyl transferase; PT, palmitoyl transferase; RACE, rapid amplification of cDNA ends; RNase, ribonuclease; RT, reverse transcriptase; ss, single strand(ed); ST, sterigmatocystin.

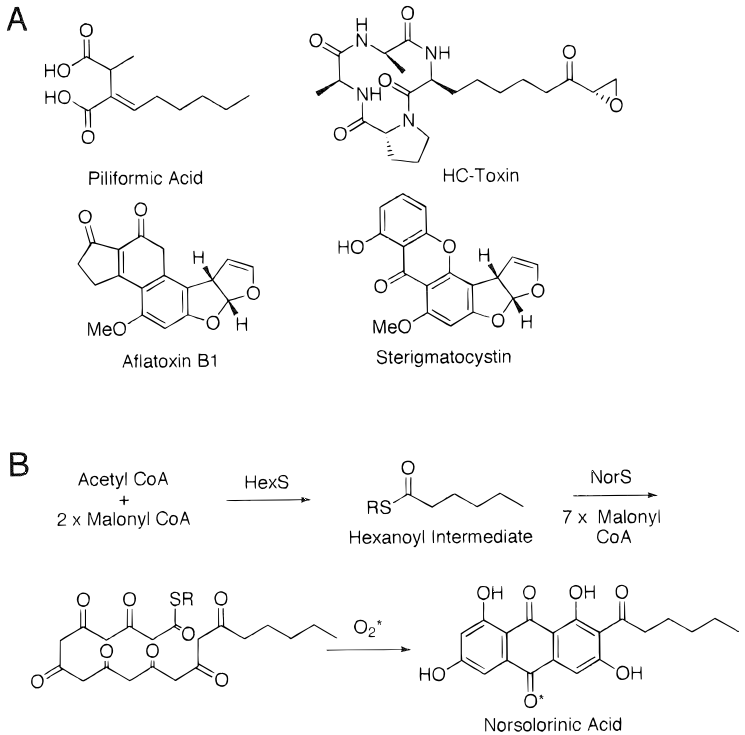

FIG. 1. Structures of fungal metabolites and biosynthesis of norsolorinic acid. (A) Natural products that have been proposed to contain fatty acid moieties assembled by dedicated secondary metabolic FAS enzymes. (B) Proposed biosynthesis of norsolorinic acid (NA), the first isolable intermediate from the aflatoxin (AF) and sterigmatocystin (ST) pathways. A dedicated FAS, HexS, assembles an hexanoyl intermediate from acetyl CoA and malonyl CoA, and the NorS complex of HexS and PksA constructs NA from the hexanoyl intermediate and malonyl CoA.

did result in loss of HC-toxin production, suggesting a purely secondary metabolic role for the FAS of which ToxC is a part.

The most convincing evidence for the existence of FAS enzymes with a specialized role in secondary metabolism comes from studies of the aflatoxin (AF) and sterigmatocystin (ST) biosynthetic pathways in the genus *Aspergillus* (Fig. 1A). AF and ST are potent mycotoxins produced by several species of these filamentous fungi, which infect food and feed supplies. ST is, in fact, a late biosynthetic precursor to AF, but in several species of fungus the AF pathway is truncated, producing ST as the final metabolite (reviews: *11*,*12*). The first isolable intermediate en route to ST and AF is the anthraquinone norsolorinic acid (NA), which undergoes several skeletal rearrangements to form first the ST (xanthone) and then the AF (coumarin) nuclei. Inspection of the hypothetical polyketide precursor to NA (Fig. 1B) indicates that it should comprise three fully reduced acetate units while seven C_2 units remain fully oxidized. These observations suggest a combined fatty acid–polyketide biogenesis for NA rather than a pure polyketide pathway. Incorporation of labeled hexanoate

into the AF biosynthetic intermediate averufin provided the first experimental evidence for use by a PKS of a starter unit other than acetate or propionate (*13,14*). Later studies using [1-^{13}C]-hexanoyl *N*-acetyl cysteamine (NAC) thiolester provided further support for this hypothesis, with up to 59% incorporation of label into the 1′-position of AF intermediates (*15*). Genetic evidence has also been gathered using a series of blocked mutants generated in *Aspergillus parasiticus*. Among these Dis-3, which lacked all of the PKS, one of two nearby FAS genes and part of another still grew and sporulated normally, but produced no aflatoxin (*16*).

Sequencing of the entire ST biosynthetic gene cluster in *Aspergillus nidulans* identified two genes (*stcJ* and *stcK*) predicted to encode yeast-like FAS proteins (*17*) as well as a gene predicted to code for the expected polyketide synthase (PKS) (*18*). When *stcJ* and *stcK* were individually disrupted, ST production ceased, but could be restored by adding exogenous hexanoic acid (*19*). Further evidence that StcJ and StcK are dedicated to ST biosynthesis came from the discovery of a separate gene pair encoding a primary metabolic FAS in *A. nidulans* (*19*). Therefore, it seems likely that the early stages of AF (and ST) biosynthesis begin with the assembly of a C_6 fatty acid from acetyl CoA and two units of malonyl CoA, catalyzed by a dedicated FAS: hexanoate synthase (HexS). Extension of the hexanoyl moiety by reaction with seven further malonyl CoA units and cyclization, both presumably catalyzed by the norsolorinic acid synthase PKS (PksA), would yield the first pathway intermediate NA (Fig. 1B).

It is our aim to use the aflatoxin pathway in *A. parasiticus* as a model system to study polyketide biosynthesis in fungi and to explore the role of HexS and its association with PksA to form the norsolorinic acid synthase complex, NorS. Feng and Leonard have previously reported accurate sequence of gDNA encoding PksA (*20*), providing a useful starting point. Sequence of the genomic region encoding the HexS β-subunit has also previously been reported (*16*), but it contains significant ($>$10%) errors. Here we report accurate and full-length sequences of the genes coding for both the α- and β-proteins of HexS, characterization of the transcriptional ends and analysis of the encoded amino acid sequences, which lay the groundwork for examining the first steps of AF biosynthesis. The sequence of the β-subunit will replace that of Mahanti *et al.* (*16*).

MATERIALS AND METHODS

Strains, media, and plasmids. *A. parasiticus* SU-1 was cultured on Adye-Mateles medium (*21*) at 28°C and 150 rpm. *Escherichia coli* strains were grown in LB medium supplemented with ampicillin (100 μg/ml) where required. The cosmid norA is as described previously (*22*). The vector pAL3 was a gift from Professor C. R. Hutchinson, University of Wisconsin, and was used to create the plasmid pAL3N by insertion of a *Not*I linker sequence (Stratagene) into the *Sma*I site. All other plasmids and reagents were obtained from the suppliers listed in the text and used according to manufacturer's specifications unless otherwise stated. Oligonucleotides were prepared either by Sigma-Genosys or at the Synthesis and Sequencing Facility, Johns Hopkins University School of Medicine, and DNA sequencing was also performed at the latter.

Cloning of hexA gDNA. The cosmid norA (20 μg) was digested with *Eco*RI and the unique 4.6-kb fragment containing all but the 5′-end of *hexA* was purified and

ligated into the *Eco*RI site of pBluescript IISK(−) (Stratagene), creating pBSFAS23. The oligonucleotides APFAS202 and APFAS203 (Table 1) were then used to prime the amplification by PCR of a 1975-bp segment of norA containing the 5′-end of *hexA*. The PCR mixture contained 20 pmol of each primer, 10 nmol of each dNTP, 10% (v/v) DMSO, 1.25 U cloned *Pfu* DNA polymerase (Stratagene) and 50 ng of cosmid norA DNA in a final volume of 50 μl. Temperature was cycled as follows: 4 min at 95°C and then 30 s at 95°C, 30 s at 55°C, and 2 min at 72°C, repeated 30 times. The purified PCR product was cloned into pT7Blue3 using the Perfectly Blunt cloning kit (Novagen), yielding pT7FAS253. Then, *hexA* was reconstituted by excision of the insert from pT7FAS253 with *Kpn*I and *Apa*I and ligation into the same sites of pBSFAS23. The resultant vector pBSFAS2A was used as the template for sequencing of *hexA*.

Cloning of hexB gDNA. The region of norA containing the entire *hexB* gene was amplified from the cosmid norA by PCR in sections, which were then subcloned and ligated together to reconstitute the gene. Conditions for PCR were as described for the cloning of *hexA*, except that DMSO was omitted and an extension time of 2.5 min at 72°C was used. Fragment FAS1S1 (2496 bp) was synthesized using primers FAS1S1F2 and FAS1S2R, purified, digested with *Eco*RI, and ligated into the *Sma*I/*Eco*RI sites of pBluescript IISK(−) to generate pBSFAS1S1. Fragment FAS1S2 (2408 bp) was amplified using primers FAS1S2F and FAS1S3R and blunt-cloned into pT7Blue3 to create pT7FAS1S2. Fragment FAS1S3 (2276 bp) was amplified with the primers FAS1S3F and FAS1S4R2 and similarly cloned into pT7Blue3, yielding pT7FAS1S3. The insert was excised from pT7FAS1S2 using *Eco*RI and ligated into

TABLE 1

Sequences of Oligonucleotide Primers Described in the Text

Primer	5′-Sequence-3′
APFAS202	GGTACCATGAGACCAGAAGCCATGG
APFAS203	GATGTCGTCGCCGGCTCCCTGCTC
FAS1S1F2	ACGCGTGCGGCCGCATATCGATACTCC
FAS1S2R	GGATAGGTTGACCGAATTCCG
FAS1S2F	CGGAATTCGGTCAACCTATCC
FAS1S3R	CCTCCGCTACGCGTTCTTCCG
FAS1S3F	CGGAAGAACGCGTAGCGGAGG
FAS1S4R2	GAATTCAATGATTCCCTGTAAACC
FAS1KPNI	GGTACCGGATGGGTTCCGTTAGTAGGA
FAS1U194R	TCTGCTCCACAGCGCTATA
FAS1S4R3	CGGTCACTCAAAGCGGCCGCAGTTGTTTCAATAGGGG
FAS15′RT2	CGGCCTGGATGGGGATTGACTC
APAJ3P	TTATGAAGCACCAGACATC
FAS1ENR	TCATACGGGGATATGATCAC
APFAS2T71R	CTAGGCATCTTGGACGAGA
APFAS2T72	TCTCGTCCAAGATGCCTAG
APFAS2P202R	TATTGGGCGTTTGGGGTAG
FAS25350F	GAGATTGAGATCGAGCATGG
FAS1U192R	ACCGCGAGATAGCAAATGC

the *Eco*RI site of pBSFAS1S1, forming pBSFAS1S1-S2. Then, the 4.9-kb insert was excised from pBSFAS1S1-S2 with *Mlu*I and ligated into the *Mlu*I site of pT7FAS1S3. A clone pT7FAS1 with the correct relative orientation of the three fragments was used as the template for initial sequencing of *hexB*. Based on preliminary sequence data it was apparent that a large excess of noncoding sequence was present at the 5′-end of the insert. Therefore, the insert was excised from pT7FAS1 *Not*I and ligated into the same site of the *Aspergillus* expression vector pAL3N, creating pAL3N-FAS1. Then, a further PCR product FAS1S5 was synthesized using the primers FAS1KPNI and FAS1S2R and cloned into pT7Blue3, creating pT7FAS1S5. This clone was digested using *Kpn*I and *Mfe*I and the purified insert was ligated into the same dephosphorylated sites of pAL3N-FAS1, yielding pP3NF1B. The initial sequence analysis also suggested that the above cloning strategy had led to truncation of FAS1S2 owing to the presence of a second *Mlu*I site 300 bp upstream of the first. Therefore a 3′-end gene fragment, FAS1S6, which encompasses both *Mlu*I sites, was amplified from the cosmid norA by PCR using primers FAS1U194R and FAS1S4R3. The PCR product FAS1S6 was digested with *Not*I and *Xma*I and ligated into the same sites of pP3NF1B, yielding the entire *hexB* gene in the construct pP3NF1C.

Isolation of mRNA and cloning of full-length hexA and hexB cDNAs. Mycelia of *A. parasiticus* were harvested over four layers of cheesecloth after 45 h growth and flash-frozen. Total RNA was extracted from the frozen mycelia using the RNeasy Mini Kit (Qiagen). mRNA was further purified from the total RNA using the Oligotex Kit (Qiagen). Full-length, single-stranded cDNAs for *hexA* and *hexB* were synthesized from the pure mRNA (200 ng) using the Thermoscript RT-PCR System (Gibco-BRL) at 55°C, priming the reverse-transcriptase reactions with the oligonucleotides APAJ3P and FAS1ENR, respectively. Double-stranded cDNAs were then amplified from ss-cDNA by PCR. The reaction mixtures were as described for the cloning of *hexA* except Platinum *Taq* DNA Polymerase High Fidelity (Gibco-BRL) was used and the template was 5 μl of the RT reaction. For amplification of *hexA* the temperature was cycled as follows: 4 min at 95°C and then 30 s at 95°C, 30 s at 55°C, and 6 min at 72°C, repeated 30 times. Reaction conditions were identical for amplification of *hexB*, except that an extension time of 6.5 min was used. The purified ds-cDNAs were cloned into the plasmid pCR2.1 using the TOPO TA Cloning Kit (Invitrogen), creating pCRCFAS2 and pCRCFAS1, which were used as templates for sequencing the cDNAs.

5′-RACE analysis. 5′-RACE analysis of the *hexA* transcript was performed using a modification of the adapter-independent method of Eyal *et al.* (*23*). First, the oligonucleotide APFAS2T71R was 5′-end phosphorylated using T4 polynucleotide kinase (New England Biolabs). Aliquots (2 μl) of the phosphorylation reactions were then used to prime synthesis of the 5′-end ss-cDNA of *hexA*, using Thermoscript RT with approximately 400 ng mRNA as template. RNase H treatment of the RT reaction mixture yielded single stranded cDNA, which was purified using the QIAquick Spin Kit (Qiagen), eluting in 30 μl. Half of the purified ss-cDNA was then circularized by incubation with T4 RNA ligase (NEB) and used as template in the RACE PCR reaction. The PCR conditions were as described for the cloning of *hexA*, priming the reaction with APFAS2T72 and APFAS2P202R and using 4 μl of the ligation reaction as template. The purified PCR product was cloned into pCR2.1 and sequenced. A similar strategy was attempted with the *hexB* transcript. However, PCR products

were truncated, probably due to secondary structure within the single-stranded DNA template. 5′-RACE analysis of hexB was accomplished using ~400 ng mRNA, Thermoscript RT, the primer FAS15′RT2, and [α-^{35}S]dATP to radiolabel the products. Sequencing reactions were performed on *hexB* gDNA from pAL3N-FAS1 using the same primer and the Sequenase Version 2.0 DNA Sequencing Kit (USB), following the manufacturers instructions. Products from the sequencing and reverse transcriptase reactions were loaded on a sequencing gel and analyzed by autoradiography.

3′-RACE analysis. DNA complementary to *A. parasiticus* SU-1 transcripts was prepared using Thermoscript RT with purified mRNA as template and priming the reaction with the Adapter Primer (Gibco-BRL). The resultant ss-cDNA was treated with RNase H (Gibco-BRL) and used for amplification by PCR of the 3′-ends of the *hexA* and *hexB* transcripts. Synthesis of the *hexA* transcript 3′-end was primed by the oligonucleotide primers FAS25350F and Abridged Universal Amplification Primer (AUAP; Gibco-BRL) using the conditions described in Section 2.4, except that DMSO was omitted from the reaction. Similarly, the 3′-end portion of the *hexB* transcript was amplified using the primers FAS1U192R and AUAP. The purified ds-cDNAs were cloned into pCR2.1 and sequenced.

Sequence analysis. Initial analysis of gene sequences was performed using BLAST 2.1 (*24*). Amino acid sequence alignments were generated using ClustalX (*25*). These alignments were used to identify nonhomologous regions within the different sequences and these regions were excluded from the phylogenetic analyses. A total of 1510 and 1670 residues were used in the analyses for the α- and β-proteins, respectively. Maximum parsimony analyses were performed using the Protpars software of the PHYLIP package (version 3.57c;*26*) to generate phylogenetic trees. The final trees were evaluated by bootstrap analysis based on 100 resamplings. Similar trees were also obtained using the Protdist and Neighbor software from the PHYLIP package. GC content and biases were calculated using FramePlot (*27*).

Accession number. The *hexS* sequence is available through GenBank, Accession No. AF391094.

RESULTS

Naming of genes. Sequencing of the entire ST biosynthetic cluster of *A. nidulans* revealed 25 genes arranged across 60 kb of the genome (*19*). These genes were then labeled *stcA-X* with the exception of the regulatory gene *aflR*. Nearly all the genes so far identified in the AF clusters of *A. parasiticus* and *A. flavus* code for proteins homologous to those encoded by the ST cluster. However, there is no systematic nomenclature for the AF clusters and the literature remains inconsistent. In order to link the genes to the functions of their encoded enzymes, we refer to the genes encoding the α and β subunits of HexS as *hexA* and *hexB*, respectively. We have retained the name *pksA* (also *pksL1*) for the PKS, but refer to the complex of HexS and PksA as norsolorinic acid synthase, NorS.

Cloning of hexA and hexB. The cosmid norA (22) contains approximately 35 kb of the AF cluster from *A. parasiticus*, including the entireties of *hexA* and *hexB*, as well as most of *pksA*. Partial sequence of norA had previously been obtained (F.T., M.D.R., and J.E.L., unpublished results) which allowed the relative orientation and positioning of *hexA* and *hexB* to be broadly defined. The two-step strategy devised