Mol Gen Genet (1996) 253: 1–10
© Springer-Verlag 1996

ORIGINAL PAPER

I. Fujii · Y. Ono · H. Tada · K. Gomi
Y. Ebizuka · U. Sankawa

Cloning of the polyketide synthase gene *atX* from *Aspergillus terreus* and its identification as the 6-Methylsalicylic acid synthase gene by heterologous expression

Received: 12 March 1996/Accepted: 8 July 1996

Abstract Southern blot analysis of genomic DNAs of several fungi that produce polyketide compounds with the 6-methylsalicylic acid synthase (MSAS) gene of *Penicillium patulum* as a probe indicated the presence of an MSAS-homologous gene in the (+)-geodin-producing strain IMI 16043 of *Aspergillus terreus*. The gene, designated *atX* was cloned from an *A. terreus* genomic DNA library and 7588 bp of the gene together with its flanking regions were sequenced to reveal the presence of a 5.5 kb open reading frame coding for a protein of 1800 amino acids with 190 kDa molecular mass. The presence of a short (70 bp) intron near the N-terminus of the *atX* gene was predicted that contains the canonical GT and AG dinucleotides at its 5′- and 3′-splicing junctions. The predicted ATX polypeptide showed high homology with *P. patulum* MSAS along the whole sequence. On the other hand, slight homology was detected only around the β-ketoacyl synthase regions of *Aspergillus nidulans wA*, PKS^{ST} and *Colletotrichum lagenarium PKS1*. No transcription of *atX* was observed throughout the culture period by Northern blotting analysis. To identify the function of the polypeptide encoded by the *atX* gene, its coding region was introduced into the fungal expression vector pTAex3 under the control of the *amyB* promoter. The constructed expression plasmid was introduced into *A. nidulans*. The transformant produced significant amounts of 6-methylsalicylic acid, the structure of which was identified by physicochemical analysis. This result unambiguously demonstrated that the *atX* gene codes for MSAS of *A. terreus*.

Key words *Aspergillus terreus* · Polyketide synthase gene · 6-Methylsalicylic acid synthase · 6-Methylsalicylic acid · Expression

Communicated by C. van der Hondel

I. Fujii · Y. Ono · H. Tada · Y. Ebizuka (✉) · U. Sankawa[1]
Faculty of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

K. Gomi
National Research Institute of Brewing, 3-7-1, Kagamiyama, Higashihiroshima-shi, Hiroshima 739, Japan

Present address:
[1]Faculty of Pharmaceutical Sciences, Toyama Medical and Pharmaceutical University, Sugitani 2630, Toyama 930-01, Japan

Introduction

Polyketides are one of the most important groups of natural products because of their huge structural variety and versatile biological activities (O'Hagen 1991). In spite of their remarkable structural diversity, the initial reactions of polyketide biosynthesis follow a common scheme resembling fatty acid biosynthetic reactions catalyzed by polyketide synthases (PKSs), which catalyze β-polyketomethylene carbon chain formation from acyl primer and malonate C2 units and subsequent reduction, dehydration and/or cyclization reactions (O'Hagen 1992, 1993).

Molecular genetic analysis of polyketide biosynthesis genes, especially those of bacterial origin, has been extensively carried out. Thus, the presence of several types of PKSs has been recognized, such as multifunctional polypeptide PKSs (fungal PKSs), modular PKSs (bacterial macrocyclic PKSs), and multienzyme PKSs (bacterial aromatic PKSs) (Hopwood and Sherman 1990; Hopwood and Khosla 1992; Hutchinson and Fujii 1995).

So far, 6-methylsalicylic acid synthase (MSAS) of *Penicillium patulum* is the sole example of a fungal PKS that has been characterized at both the enzyme and gene levels (Dimroth et al. 1970; Beck et al. 1990; Spencer and Jordan 1992a, b). Bacterial PKS genes have been extensively studied, but PKS enzyme activities were detected only in a few cases: tetracenomycin

PKS by Shen and Hutchinson (1993) and 6-deoxyerythronolide B PKS by Khosla and colleagues (Pieper et al. 1995).

Hopwood and colleagues have indicated the presence of sequence similarity among the PKS genes of streptomycetes and the possible usefulness of PKS genes as probes for screening of new PKS genes (Malpartida et al. 1987). This has been proven in some cases. For example, we have succeeded in cloning the aklavinone biosynthesis gene cluster from *Streptomyces galilaeus* using *act* genes as probes (Tsukamoto et al. 1992, 1994). The similarity, especially, of the β-ketoacyl synthase gene, is ascribed to the inherent characteristics of PKSs whose key reaction is the iterative decarboxylative aldol condensation. Therefore, cloned PKS genes might be useful probes, in principle, for the cloning of PKS genes of all organisms.

We have been studying the biosynthesis of seco-anthraquinone, (+)-geodin in *Aspergillus terreus* at the enzyme and molecular genetic levels and successfully identified and characterized most of the enzymes involved (Fujii et al. 1982, 1983, 1987, 1988, 1991; Chen et al. 1992, 1995; Huang et al. 1995a). However, no polyketide synthase activity has been detected so far in spite of our extensive efforts. The only detectable fungal PKS enzyme activity that we have found is orsellinic acid synthase of *Penicillium cyclopium*, the stereochemical course of the reaction of which was approached by chiral malonate methodology (Woo et al. 1989).

To clarify further the nature of fungal PKSs, screening of the fungal PKS genes was carried out using the MSAS gene of *P. patulum* as a probe. We found a possible PKS gene, designated *atX*, in the (+)-geodin-producing strain IMI 16043 of *A. terreus*. In this paper, we describe the cloning of the *atX* gene and its identification as the *A. terreus* MSAS gene by heterologous expression in *Aspergillus nidulans*.

Materials and methods

Fungal strains

Polyketide-producing fungal strains, *A. terreus* IMI 16043 [(+)-geodin producer], *Myrothecium verrucaria* AH-14 [(−)-rugulosin producer], *Oospora sulphrea-ochracea* (asterric acid producer), *P. cyclopium* (penicillic acid producer), *Penicillium frequentans* CMI 96659 (asterric acid producer), *Penicillium islandicum* IN 487 (skyrin producer), and *Penicillium oxalicum* (secalonic acid producer) were stock cultures of our laboratory.

Strain FGSC 89 (*argB*, *biA*) of *A. nidulans* was maintained on arginine and biotin-supplemented *Aspergillus* minimal medium as described by Pontecorvo et al. (1953).

DNA and vectors

The MSAS gene of *P. patulum* was a generous gift from Prof. Schweizer (Beck et al. 1990). λEMBL3, λZAP II, and pBluescripts were from Stratagene. pGEM7-Zf(+) was purchased from Promega. pT7 Blue from Novagen was used for direct cloning of polymerase chain reaction (PCR) products. pTAex3 is a shuttle vector carrying the ampicillin resistance gene and the *argB* gene of *A. nidulans* for selection in *Escherichia coli* and arginine prototrophy selection in fungi, respectively (Fujii et al. 1995). The plasmid also carries the promoter and terminator region of the Taka-amylase A gene (*amyB*) of *Aspergillus oryzae* (Tada et al. 1991) flanking unique *Eco*RI and *Sma*I cloning sites.

Nucleic acid isolation and Southern blot analysis

Genomic DNAs of fungi were isolated by the modified procedure of Biel and Parrish (1986) as previously described (Huang et al. 1995b).

Genomic DNAs were subjected to restriction enzyme digestion and capillary-transferred from agarose gel onto Nytran membranes (Schleicher and Schuell). DNA probes were labeled with digoxigenin-dUTP using the DIG DNA labeling kit (Boehringer Mannheim, Germany) according to the manufacturer's instructions. Hybridization was carried out in a solution containing 5 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS, 2% blocking reagent (Boehringer Mannheim), 0.02% sodium N-lauroyl sarcosine at 65°C overnight. Filters were washed twice with 2 × SSC, 0.1% SDS at 65°C for 30 min. Enzyme-linked immunodetection was carried out using the DIG nucleic acid detection kit (Boehringer Mannheim) as recommended by the manufacturer.

Total RNAs were isolated by the phenol-SDS method (Chargwin et al. 1979) with additional LiCl precipitation. Poly(A)$^+$ RNA was separated by oligotex-dT30 (Daiichi Chemicals, Japan) according to the manufacturer's instructions.

Construction and screening of *A. terreus* partial genomic DNA library

Genomic DNA of *A. terreus* was digested with *Bam*HI, and the digestion products were separated by agarose gel electrophoresis. DNA fragments of approximately 8 kb recovered by electroelution were ligated with *Eco*RI and *Bam*HI-digested λEMBL3 phage DNA arms. Packaging of the ligation mixture gave a library comprising 1.4×10^6 plaque forming units.

The second genomic DNA library of 3.6×10^4 plaque forming units was constructed with approximately 8–10 kb *Eco*RI digested fragments and *Eco*RI-digested λZAP II phage DNA arms.

Screening was carried out by plaque hybridization with a digoxigenin-dUTP-labeled probe of the β-ketoacyl synthase region of MSAS (MKS). Detection was carried out using DIG nucleic acid detection kit (Boehringer Mannheim) and nitroblue tetrazolium salt and 5-bromo-4-chloro-3-indolyl phosphate.

DNA sequence analysis

DNA sequence analysis was carried out by the dideoxy chain termination method using Sequenase version 2.0 (Amersham) and fluorescein isothiocyanate-labeled primer or terminator (Yuki Gosei Kogyo Co., Japan) with an Hitachi SQ-3000 auto DNA sequencer. Fragments for sequencing were generated either by restriction digestion or deletion with exonuclease III and mung bean nuclease. DNA sequence data were organized and analyzed by using the DNASIS program (Hitachi Software Engineering Co., Japan) and DNA analysis program package of the Human Genome Center (Institute of Medical Science, The University of Tokyo). The final sequence of the cloned *atX* DNA was determined from both strands.

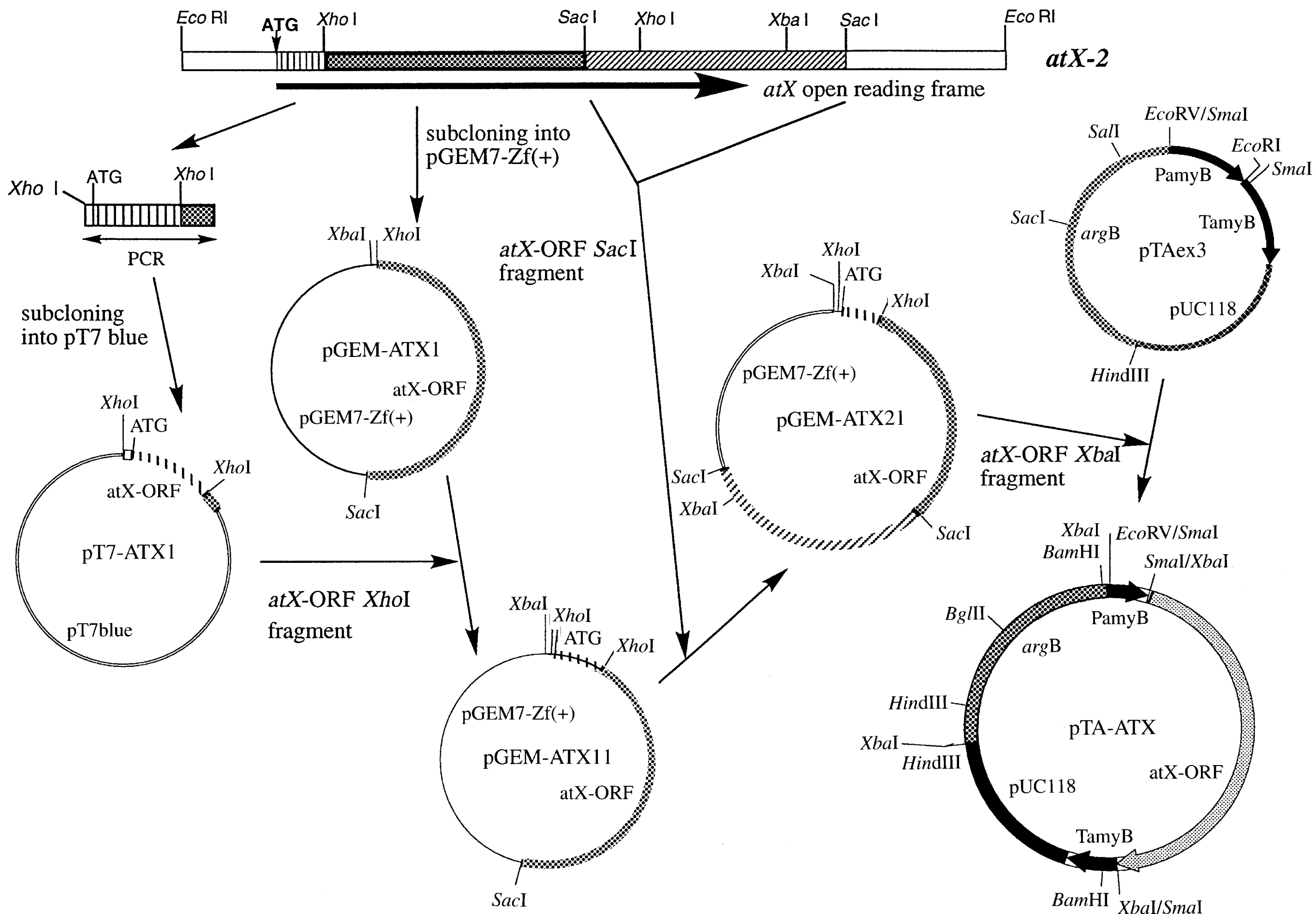

Fig. 1 Construction of *atX* expression plasmid pTA-ATX. (*PCR* polymerase chain reaction, *ORF* open reading frame, *Pamy B*, *Tamy B* promoter and terminator of *amyB*)

Nucleotide sequence accession number

The nucleotide sequence data reported in this paper have been deposited in the DDBJ, EMBL and GenBank nucleotide sequence databases with the accession number D85860.

Construction of the *atX* expression plasmid pTA-ATX

Construction of the *atX* expression plasmid pTA-ATX was carried out as summarized in Fig. 1. A 3.2 kb *Xho*I- *Sac*I fragment from the *atX* gene was subcloned into pGEM7-Zf(+) to form pGEM-ATX1. The N-terminus region from the ATG translation initiation codon to the *Xho*I site (480 bp) was amplified by PCR using 5′-sense primer: 5′-CTC GAG CAA GCC ATG GAG GTA CAT-3′ and 3′-antisense primer: 5′- ACA AAC TCT GCG GTG GGA TT-3′. In the 5′-sense primer, the designed *Xho*I site is underlined. Amplification with *Ex Taq* polymerase (Takara Shuzo, Japan) was carried out in the following conditions: template *atX-2* DNA, 100 pmol of each primer, 8 μl of 2.5 mM dNTP solution, 10 μl of $10 \times$ *Ex Taq* polymerase buffer (Takara Shuzo) and 2.5 units of *Ex Taq* polymerase were combined in a total volume of 100 μl. The cycle of 94° C for 30 s, 58° C for 1 min, and 72° C for 1 min was repeated 25 times. The amplified fragment was directly subcloned into pT7 Blue vector plasmid to construct pT7-ATX1. The absence of *Taq* polymerase fidelity error was confirmed by sequencing analysis. The *Xho*I fragment cut from pT7-ATX1 was inserted into the *Xho*I site of pGEM-ATX1. A plasmid with the correct orientation was selected and named pGEM-ATX11. The C-terminus of the *atX* gene (3.0 kb *Sac*I fragment) was inserted into the *Sac*I site of pGEM-ATX11 to construct pGEM-ATX21, which has a correctly assembled *atX* gene. After blunting by Klenow enzyme, the 6.2 kb *Xba*I fragment containing the whole *atX* coding region was inserted into the *Sma*I site of expression plasmid pTAex3 (Fujii et al. 1995). The plasmids thus constructed, pTA-ATX (right orientation) and pTA-ATXR (opposite orientation), were used to transform *A. nidulans* FGSC 89.

Transformation of *A. nidulans*

Protoplasts of *A. nidulans* were prepared by the method of Tilburn et al. (1983) and Huang et al. (1995b). Transformation was carried out according to the procedures described by Gomi et al. (1987). Stable transformants were isolated by repeated transfer of sporulating colonies on selection plates.

Isolation and identification of the ATX product

The *A. nidulans* pTA-ATX transformants were precultured for 3 days at 30° C on a rotary shaker in *Aspergillus* minimal medium containing glucose as a carbon source. Mycelia were collected by centrifugation and transferred into fresh medium containing soluble

starch as a carbon source. Cultures were continued for a further 3 days at 30° C. Culture medium was acidified and extracted with ethylacetate. After evaporation, residues were analyzed by thin layer chromatography (TLC plate Merck 16485; benzene : acetone = 3 : 1). The presence of a compound that was absent in *A. nidulans* FGSC 89 extract was detected as a main product. The compound was isolated by oxalic acid-impregnated silica gel column chromatography (benzene-acetone solvent system) and recrystallized from chloroform. From a 200 ml induction culture was obtained 65.5 mg ATX product as colorless needles. The following physicochemical data identified the ATX product as 6-methylsalicylic acid: mp, 174–5°C ($CHCl_3$); EI-MS, m/z = 152 (M^+), 134 ($-H_2O$); ^{1}H-NMR (acetone-d_6/TMS, 500 MHz) δ(ppm) 2.58 (3H, s), 6.78 (1H, d, J = 8 Hz), 6.79 (1H, d, J = 8 Hz), 7.33 (1H, dd, J = 8 Hz, 8Hz); ^{13}C-NMR (acetone-d_6, 125 MHz) δ(ppm) 24, 113, 116, 123, 135, 142, 164, 174.

Results

Genomic Southern blot analysis of polyketide-producing fungi

Genomic DNAs of several polyketide-producing fungi were analyzed by Southern blot hybridization using the *P. patulum* MSAS gene (Beck et al. 1990) as a probe. As shown in Fig. 2, a single hybridizing band was observed in the (+)-geodin-producing strain of *A. terreus* when the MKS probe was used. No homologous band was detected in genomic DNA of *P. cyclopium* of which the main metabolite is penicillic acid biosynthesized via orsellinic acid, a C-4 hydroxylated derivative of 6-methylsalicylic acid. Under the same stringency conditions, no homologous bands were detected even in *A. terreus* when probed with the acyl carrier protein region of the MSAS gene (data not shown). This result indicated the presence of an MSAS-like PKS gene in *A. terreus* that is not quite identical to that of *P. patulum*. This *A. terreus* gene was tentatively named *atX*.

Cloning of the *atX* gene from *A. terreus*

In the Southern analysis of genomic DNA performed using the MKS probe, single positive bands of about 8 and 10 kb were detected in *Bam*HI and *Eco*RI-digested *A. terreus* DNA, respectively. From the λEMBL3 and λZAP II partial genomic DNA libraries an 8 kb *Bam*HI fragment (*atX-1*) and a 10 kb *Eco*RI fragment (*atX-2*) that hybridized with the MKS gene were cloned. Restriction enzyme mapping of these fragments revealed an overlap of more than 5 kb in a total region of about 12 kb.

DNA and deduced amino acid sequence

The genomic DNA fragments *atX-1* and *atX-2* of *A. terreus* were subcloned as a series of overlapping

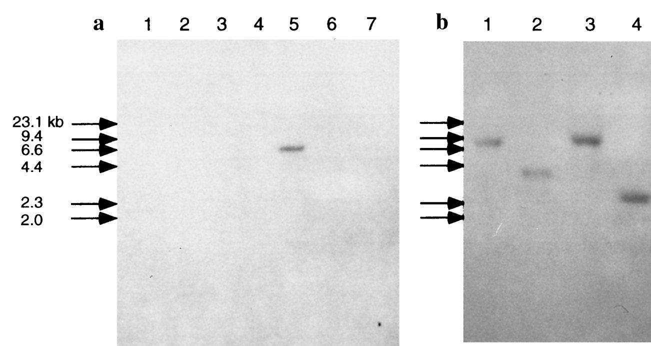

Fig. 2a, b Southern blot analysis of genomic DNAs of polyketide-producing fungi. **a** Approximately 1 μg of genomic DNA of each fungus was digested with *Bam*HI, separated on a 0.7% agarose gel, and then transferred to a nylon membrane filter. The blot was hybridized with the digoxigenin-labeled β-ketoacyl synthase region of the 6-methylsalicylic acid (MSAS) gene (MKS) in 5 × SSC at 65°C. Washing was carried out in 2 × SSC at 65°C. Lane 1 *Penicillium cyclopium*, 2 *P. frequentans*, 3 *P. islandicum*, 4 *P. oxalicum*, 5 *Aspergillus terreus*, 6 *Oospora sulphrea-ochracea*, 7 *Myrothecium verrucaria*. **b** Genomic DNA of *A. terreus* digested with restriction enzymes (lane 1 *Bam*HI, 2 *Bgl*II, 3 *Eco*RI, 4 *Pst*I) was hybridized with the MKS probe

DNA fragments. The nucleotide sequence of the MSAS-hybridizing region was determined over 7588 bp as shown in Fig. 4. The strategy applied for bidirectional sequencing of this DNA segment is shown in Fig. 3.

Open reading frame analysis indicated the presence of a small intron in the N-terminus region of the coding sequence. The presumed intron possesses typical 5′- and 3′-splicing signals of fungi (Gurr et al. 1987). The presumptive translation initiation codon **ATG** GAG occurs at nucleotide 954 and the methionine starts a 5.5 kb large open reading frame coding for a 1800 amino acid polypeptide of 190 kDa.

Sequence analysis of the 5′-flanking region revealed the presence of structural features considered to be important for gene transcription. A TATA box-like sequence (TATAA) was found 177 bp upstream from the translation start, and a CAAT element (CAAAT) was located at 238 bp upstream from the ATG.

Comparison of the ATX polypeptide with other polyketide synthases

The deduced amino acid sequence of the ATX polypeptide was compared with other polyketide synthase proteins to analyze the structural relationship among them.

As was expected, the deduced ATX polypeptide sequence displayed throughout a high homology with the *P. patulum* MSAS sequence (Beck et al. 1990) (Fig. 5). The degree of amino acid sequence identity between ATX and *P. patulum* MSAS is 62%. In particular, the

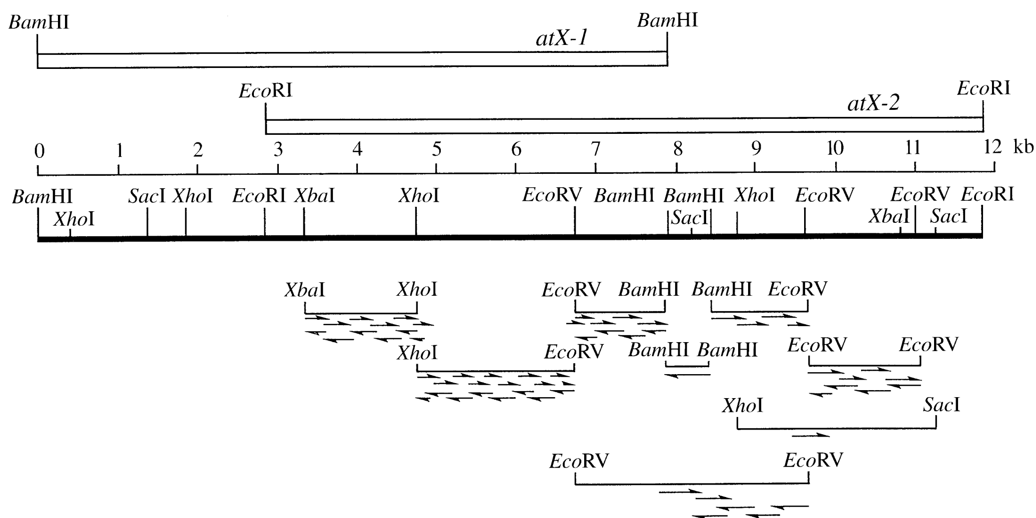

Fig. 3 Restriction map and sequencing strategy of the *A. terreus atX* gene. Genomic clones (*atX-1* and *atX-2*) were sucloned into eight fragments. Unidirectional deletions of each subclones were constructed with exonuclease and mung bean nuclease. The *arrows* indicated the extent and direction of sequencing

proposed active centers of MSAS were well conserved in ATX as shown in Fig. 6. The size of the encoded protein and the conserved active centers strongly indicated that *atX* is a gene coding for MSAS of *A. terreus*. On the other hand, the sequence homology of ATX with other fungal PKSs, *A. nidulans wA* (Mayorga and Timberlake 1990, 1992), PKS^{ST} (Yu and Leonard 1995) and *Colletotrichum lagenarium PKS1* (Takano et al. 1995), was restricted to the β-ketoacyl synthase region (Fig. 5), although the active site amino acid residues of these PKSs are preserved as shown in Fig. 6. Interestingly, fairly high homology of 35.4% was observed with *Saccharopolyspora erythrea eryAI* gene products (Cortes et al. 1990; Donadio et al. 1991; Bevitt et al. 1992; Donadio and Katz 1992), which are bacterial type-I modular polyketide synthases.

Expression of ATX in *A. nidulans*

No transcript of *atX* was observed by Northern blot analysis of all growth phases of *A. terreus* (data not shown). Thus the function of the *atX* gene cannot be assessed by experiments such as gene disruption. In order to identify its actual function, expression of the *atX* gene in a heterologous host was carried out by using fungal expression vector pTAex3 (Fujii et al. 1995), which has been successfully applied for the expression of dihydrogeodin oxidase of *A. terreus* in *A. nidulans* (Huang et al. 1995a). In this plasmid, expression of the cloned gene is controlled by the *amyB* promoter of *A. oryzae*. Expression under the *amyB* promoter is strictly regulated by the specific carbon source in the culture medium, that is, induction by starch or maltose, and repression by glucose (Tada et al. 1991).

The ATX expression plasmid pTA-ATX was constructed by insertion of the *atX* gene into the *Eco*RI site of pTAex3 without removing the intervening sequence (Fig. 1). *A. nidulans* was transformed with pTA-ATX, pTA-ATXR (*atX* gene in opposite direction) or the vector plasmid pTAex3 without an insert as a control. Integration of plasmid DNA into the genome of the host was confirmed by Southern blot analysis (data not shown). Transformants were cultured first in glucose-containing *Aspergillus* minimal medium (Pontecorvo et al. 1953) for 3 days and then in starch-containing medium for a further 3 days to induce gene expression.

From the culture medium of pTA-ATX transformants was isolated an acidic compound that was absent in the control cultures. The compound was identified as 6-methylsalicylic acid by physicochemical analysis. After recrystallization from $CHCl_3$, 65.5 mg of 6-methylsalicylic acid was obtained from 200 ml induction culture medium. This result unambiguously demonstrated that the *atX* gene codes for MSAS of *A. terreus*. This is to our knowledge the first time that a fungal polyketide synthase gene has been expressed in a heterologous fungal host.

Discussion

Hopwood and colleagues (Malpartida et al. 1987) first showed homology among the streptomycete PKS genes and demonstrated the possibility of cloning these genes with homologous PKS probes. In addition to streptomycetes, fungi are another rich source of polyketide compounds of diverse structural variety from single aromatic ring compounds, such as 6-methylsalicylic acid and orsellinic acid to highly modified aflatoxins (Turner and Aldridge 1983). To determine whether it would be possible to clone fungal PKS genes by homology, Southern blot analysis of genomic DNAs of several polyketide-producing fungi was carried out.

```
   1 TATCTCTTCCCTTAATGATCGGATGTCTAAAAGGAGCAGTGGATTTCGAGATCTAATCTA    60

  61 CACCCATCGGACTACTCAAAGGAGTCATACTATGGAGCACACGTAGCGCGATAGAGACGT   120

 121 AGATAAGATGGAGAGGGCGGGGATAAATTCTAGACCTGCAGATTGTGAGCCCCAATTACC   180

 181 CAAAAAGGCAAAAGACCCTGCACATTGACCTAAGCGGATGATAGTTATGCTCCGTGAATA   240

 241 TACGGAATAGTAGAATCTGACAAGCAGATTGGCACCCCCAGCTCTACCAATAACGATGGA   300

 301 ACGGGCATACCACGACACACGCAAGAGTCTGTCGCTAAGGGGAGCTACTGCACGGTGTCC   360

 361 CTTGGTCTCAAAGCGCTAGGGGAAACGGCGCGAAACAGATCCACAGTTCCCACAGTGACA   420

 421 GACTGTCTTCGGCTATCATTCCCTAACCGCGAATGAGGACCCATAATGAGACCACCATTC   480

 481 CAAGAATACCAAGAATACCAAGGATGCCAATTCTCACCATCGCAGACGCCGAGACTTCCA   540

 541 GACGTGCTTGATGCGACTTCGCGACGTTTTAGCGCCAGAAGCCATCACAGATTAATCATT   600

 601 CCTTTACCTAAAGGGGTTCTGCGGTGCAGGGCTCCGAGAGACAGGCGCCAATTCCTGGCG   660

 661 CTAGTCTCCTATCTGCACGTCCCGGGAAGATTCCGTAGGACTGGATGCCCGAATTGTTCG   720

 721 CGGGATGCAAGTGTTGACGAAGATCCACCAAGACTTGTTTGCACATGGCAGGTGCACTGG   780

 781 AGATTCTGATTCGAAACACTATCCACCATACCTCGCGTGGTACTGCTGAAGATGGTGATC   840

 841 AAGCGAATCAAGACAGAAGGAGTCAAATCCCTCGCGATATTCAGAAAACTGTAGTAATAT   900

 901 CAGTGAATCATATACTGTACCGAGTATAAGATGTAGAGCGACGCCGAAGATTTTGCGGAC   960

 961 TTTCAATCAAAGAAATTCCCTACCATACCATCTCTCCTCTTCAGGCGTTATAGGATTCTC  1020

1021 TCCTCTATCAGTTGATGCTCCATTGACAAAATCTTTTACCGGCTTCTCTCGCATAGACAG  1080

1081 CAAATATCAGCAAGAGCAGCCATGGAGGTACATGGAGATGAAGTGTTGTCAGTCGACTCT  1140
   1                       M  E  V  H  G  D  E  V  L  S  V  D  S     13

1141 GGCATCTCAACTCCGCCCTCGACAGGAAGTGGATTTCGGAGGCCACTAGAGACTCCCGGA· 1200
  14  G  I  S  T  P  P  S  T  G  S  G  F  R  R  P  L  E  T  P  G     33

1201 ACAGAAATCGGGAATCTCAATCTTgtgggtcttgctagctcccccaccgttctatgcact  1260
  34  T  E  I  G  N  L  N  L                                         41

1261 gtatctgacctctgggaaccattcagaaccctcagAATGAGGTTGCCGTTGTTGGAATGG  1320
  42                                     N  E  V  A  V  V  G  M      49

1321 CCTGCCGGCTTGCCGGGGGCAATCATTCTCCGGAAGAACTGTGGCAGTCCATTCTAAACA  1380
  50 A  C  R  L  A  G  G  N  H  S  P  E  E  L  W  Q  S  I  L  N      69

1381 GGAAGGATGCCTCTGGCGAGATCCCAAGCATGCGCTGGGAGCCGTACTACCGTCGGGACA  1440
  70 R  K  D  A  S  G  E  I  P  S  M  R  W  E  P  Y  Y  R  R  D      89

1441 TTCGCAACCCCAAGATCCTAGATCAAACGACAAAGCGCGGCTACTTCTTGGACCACATCG  1500
  90 I  R  N  P  K  I  L  D  Q  T  T  K  R  G  Y  F  L  D  H  I     109

1501 AGAATTTTGATGCCGCGTTCTTTGGCGTTTCCCCCAAAGAGGCCGAGCAGATGGACCCCC  1560
 110 E  N  F  D  A  A  F  F  G  V  S  P  K  E  A  E  Q  M  D  P     129

1561 AGCAGCGGTTGTCACTCGAGGTGACTTGGGAGGCCCTGGAAGACGCAGGAATCCCACCGC  1620
 130 Q  Q  R  L  S  L  E  V  T  W  E  A  L  E  D  A  G  I  P  P     149

1621 AGAGTTTGTCCGGCTCAGAAACAGCCGTGTTTATGGGGGTCAATTCGGATGATTATTCCA  1680
 150 Q  S  L  S  G  S  E  T  A  V  F  M  G  V  N  S  D  D  Y  S     169

1681 AGCTCTTACTGGAAGATATTCCGAACGTGGAGGCCTGGATGGGCATCGGCACTGCGTACT  1740
 170 K  L  L  L  E  D  I  P  N  V  E  A  W  M  G  I  G  T  A  Y     189

1741 GCGGAGTCCCGAACCGCATCTCCTACCACCTGAACCTCATGGGGCCCAGCACTGCCGTTG  1800
 190 C  G  V  P  N  R  I  S  Y  H  L  N  L  M  G  P  S  T  A  V     209

1801 ATGCCGCCTGTGCCTCCTCTCTCGTTGCCATCCATCACGGACGACAGGCCATCCTACAAG  1860
 210 D  A  A  C  A  S  S  L  V  A  I  H  H  G  R  Q  A  I  L  Q     229

1861 GCGAGAGCGAAGTCGCTATTGTCGGAGGAGTCAACGCTCTCTGCGGGCCAGGACTGACTC  1920
 230 G  E  S  E  V  A  I  V  G  G  V  N  A  L  C  G  P  G  L  T     249

1921 GCGTACTCGATAAGGCAGGAGCGACCTCCACGGAAGGTCGCTGTCTCTCTTTCGACGAAG  1980
 250 R  V  L  D  K  A  G  A  T  S  T  E  G  R  C  L  S  F  D  E     269

1981 ATGCGAAGGGCTACGGCCGTGGTGAAGGAGCTGCGGTGGTGATCTTGAAACGGCTGTCCA  2040
 270 D  A  K  G  Y  G  R  G  E  G  A  A  V  V  I  L  K  R  L  S     289

2041 CCGCCATCCGGGACGAGACCACATTCGCGGCCATCATCAAGGGTAGTGCCGTAGCACAGG  2100
 290 T  A  I  R  D  E  T  T  F  A  A  I  I  K  G  S  A  V  A  Q     309

2101 ATGGCAAAACCAACGGCATCATGGCTCCCAACGCCAAGGCACAAGAGCTTGTGGCATGGA  2160
 310 D  G  K  T  N  G  I  M  A  P  N  A  K  A  Q  E  L  V  A  W     329

2161 ATGCTCTTCGGACAGCCGGAGTCGACCCTCTGACGGTTGGGTATGTGGAAGCTCACGCAA  2220
 330 N  A  L  R  T  A  G  V  D  P  L  T  V  G  Y  V  E  A  H  A     349

2221 CGTCAACCCCTCTTGGCGATCCTACCGAGGTCAGCGCCGTCTCAGCAGTCTACGGCAAAG  2280
 350 T  S  T  P  L  G  D  P  T  E  V  S  A  V  S  A  V  Y  G  K     369

2281 GCAGACCGGAAGGGAATCCTTGCTTCATTGGCTCTGTCAAACCCAACGTGGGCCATTTGG  2340
 370 G  R  P  E  G  N  P  C  F  I  G  S  V  K  P  N  V  G  H  L     389

2341 AAGCGGGCGCTGGCGGCGTCGGTTTCATCAAAGCAGTCATGGCAGTTGAAAAGGCCACTT  2400
 390 E  A  G  A  G  G  V  G  F  I  K  A  V  M  A  V  E  K  A  T     409

2401 TCCCCCCACAAACCAACCTGAAGAGACTCAATTCTCGCATTGACTGGGACCAAGCCGGAG  2460
 410 F  P  P  Q  T  N  L  K  R  L  N  S  R  I  D  W  D  Q  A  G     429

2461 TGAAGGTCGTCCAGGAGACACTGGAATGGCCTGGCAATGAGGATGACGTCCGCCGAGCCG  2520
 430 V  K  V  V  Q  E  T  L  E  W  P  G  N  E  D  D  V  R  R  A     449

2521 GTGTTTGCTCTTACGGATATGGCGGTACGGTCTCCCATGCAATCATCGAGGAGTTTGCGC  2580
 450 G  V  C  S  Y  G  Y  G  G  T  V  S  H  A  I  I  E  E  F  A     469

2581 AACAGCTCCAACGGCCGACTACCAACACAACCGATGAAGAGCCTCTGCCTCGGATTCTTC  2640
 470 Q  Q  L  Q  R  P  T  T  N  T  T  D  E  E  P  L  P  R  I  L     489

2641 TGCTGTCGGCACCTCAAGAGAGACGCCTTGCTTTGCAGGCACGGACACAGGCCTCCTGGA  2700
 490 L  L  S  A  P  Q  E  R  R  L  A  L  Q  A  R  T  Q  A  S  W     509

2701 TTGCCGCGGAGGGCAGAAATAGAACCCTGGAGTCGATCGCAACCACCTTGAGCACTCGTC  2760
 510 I  A  A  E  G  R  N  R  T  L  E  S  I  A  T  T  L  S  T  R     529

2761 GTGGGCACCATGACTACCGGGCTGCCATCATCGCAGAGAACCATGATGACGCTGTGCAGA  2820
 530 R  G  H  H  D  Y  R  A  A  I  I  A  E  N  H  D  D  A  V  Q     549

2821 AACTGTCTGACATTGTCAATGGTAAAGCAGCCGAATGGACGACGTCGAGTCGTGTTCTCG  2880
 550 K  L  S  D  I  V  N  G  K  A  A  E  W  T  T  S  S  R  V  L     569

2881 ATGCCAGTTGCTCCAAGGACGTGGTGTGGGTTTTCTCCGGTCATGGCGCACAATGGACTG  2940
 570 D  A  S  C  S  K  D  V  V  W  V  F  S  G  H  G  A  Q  W  T     589

2941 CAATGGCTACGGATCTCCTCAAAGACATTGTGTTCTATCAAACAATCAGCCGTCTGGACC  3000
 590 A  M  A  T  D  L  L  K  D  I  V  F  Y  Q  T  I  S  R  L  D     609

3001 CGATTGTGGAGCGCGAAATGGGCTTCTCGGCATTGCATTCCCTTGCAAGTGGCGATTTCG  3060
 610 P  I  V  E  R  E  M  G  F  S  A  L  H  S  L  A  S  G  D  F     629

3061 AATCGTCCATCAAGGTGCAAGTGCTCACCTATCTCGTACAGGTGGGACTGGCTGCCATCT  3120
 630 E  S  S  I  K  V  Q  V  L  T  Y  L  V  Q  V  G  L  A  A  I     649

3121 TGCGCTCGAAGGGACTGGAGCCCCAGGCTGTCATCGGTCATTCAGTTGGCGAAATTGCCG  3180
 650 L  R  S  K  G  L  E  P  Q  A  V  I  G  H  S  V  G  E  I  A     669

3181 CCTCAGTCGCGGCCGGCTGTCTGACTGCAGAAGAAGGCGCCCTGATTGTCACCCGCAGAG  3240
 670 A  S  V  A  A  G  C  L  T  A  E  E  G  A  L  I  V  T  R  R     689

3241 CAAACCTCTATCGGCGTGTGATGGGCGCGGGCGCAATGGTTCTCGTCAACATTCCATTCG  3300
 690 A  N  L  Y  R  R  V  M  G  A  G  A  M  V  L  V  N  I  P  F     709

3301 CCGACATGGAGAAAGAGCTTCAAGGCCGGACGGACCTGGTGGCCGCCATTGACTCCTCGC  3360
 710 A  D  M  E  K  E  L  Q  G  R  T  D  L  V  A  A  I  D  S  S     729

3361 CGTCTTCATGTGTTGTTTCCGGTGCCACTGAAGCTGTCCTGGCGCTCGTGGAAGACCTCA  3420
 730 P  S  S  C  V  V  S  G  A  T  E  A  V  L  A  L  V  E  D  L     749

3421 AGTCTCGTGGTGTCAACGCTTTCCGGGTCAAGACGGATATTCCCTTCCACCACCCGATGC  3480
 750 K  S  R  G  V  N  A  F  R  V  K  T  D  I  P  F  H  H  P  M     769

3481 TGGATCAGCTGTCCGAGCCATTGCGAGAGGCCATGGAAGGGTCCCTGTCGCCACGCAAGC  3540
 770 L  D  Q  L  S  E  P  L  R  E  A  M  E  G  S  L  S  P  R  K     789

3541 CCAGAGTCCGTCTTTACTCGACGTCGGCAGAAGACCCACGCAGTATGGTTGCTCGGGATA  3600
 790 P  R  V  R  L  Y  S  T  S  A  E  D  P  R  S  M  V  A  R  D     809

3601 TCCATTACTGGACCAGCAACATGGTCAACCCGGTCCGGTTGACGGCCGCAGTGCAGGCAG  3660
 810 I  H  Y  W  T  S  N  M  V  N  P  V  R  L  T  A  A  V  Q  A     829

3661 CAGTGGACGATGGCCTGCGACTGTTCCTTGAAGTCTCTTCTCATCCCATTGTGTCTCACT  3720
 830 A  V  D  D  G  L  R  L  F  L  E  V  S  S  H  P  I  V  S  H     849

3721 CCGTCCGAGAGACCATGTTGGACCTGGGTGTGGAGGACTTCACCGTGACCAACACCATGG  3780
 850 S  V  R  E  T  M  L  D  L  G  V  E  D  F  T  V  T  N  T  M     869

3781 CTCGCAATAAGCCTGCCGACAAGACTATTCTGTCCAGCATTGCCCAGCTTCACTGTCGGG  3840
 870 A  R  N  K  P  A  D  K  T  I  L  S  S  I  A  Q  L  H  C  R     889

3841 GCGCCGTCGTCAATTGGAAGAAGCAGCTGCCGGGCCCTTGGGCGCTGGATGTGCCCTTGA  3900
 890 G  A  V  V  N  W  K  K  Q  L  P  G  P  W  A  L  D  V  P  L     909

3901 CGACGTGGGACCACAAGCCCTTCTGGCGGCATATTCACACTGGCCCTATCAGTGCCTCGA  3960
 910 T  T  W  D  H  K  P  F  W  R  H  I  H  T  G  P  I  S  A  S     929

3961 CTTTGCATGATGTGGACAAACACACGCTGTTGGGTCAGCGCGTTCCCGTTGCGGGAGAAA  4020
 930 T  L  H  D  V  D  K  H  T  L  L  G  Q  R  V  P  V  A  G  E     949

4021 CGACTATGGTGTTCACCACCCAAATGGATGACCAGACCAAGCCTTTCCCAGGAAGCCATC  4080
 950 T  T  M  V  F  T  T  Q  M  D  D  Q  T  K  P  F  P  G  S  H     969

4081 CACTGCACGGCTCTGAGATTGTTCCGGCTGCTGCCCTTGTCAACACTTTCTTGCATGCCA  4140
 970 P  L  H  G  S  E  I  V  P  A  A  A  L  V  N  T  F  L  H  A     989

4141 CCAGGGCTACCACCCTTTCCAACATTACTCTTCGCGTCCCAGTGGCCATCAGCCAGCCGC  4200
 990 T  R  A  T  T  L  S  N  I  T  L  R  V  P  V  A  I  S  Q  P    1009
```

```
4201 GCGACATCCAGGTGGTGGTGTCACAGAATCAAATCAAGATCTGCTCCCGTCTCACTCAGA  4260
1010 R  D  I  Q  V  V  V  S  Q  N  Q  I  K  I  C  S  R  L  T  Q   1029

4261 AGGCGGATTCTGGGGCAGACGAAGGTTCCTGGCTGACACACACTACGGGTCAGTGGGAAG  4320
1030 K  A  D  S  G  A  D  E  G  S  W  L  T  H  T  T  G  Q  W  E   1049

4321 CTGGTGGAAGCAAGAACCCCCCGGCGCAACTCGATATTGCTGCTATCAAGGCTCGTCTCG  4380
1050 A  G  G  S  K  N  P  P  A  Q  L  D  I  A  A  I  K  A  R  L   1068

4381 CTAATAACAAGTTGGCGGACAACTTCTCCATCGACTATTTGGACAAGGTTGGCGTTTCGG  4440
1070 A  N  N  K  L  A  D  N  F  S  I  D  Y  L  D  K  V  G  V  S   1089

4441 CCATGGGCTTCCCTTGGGCAGTTACAGAGCACTACGGCACCCTGCAGGAGATGATCGCTC  4500
1090 A  M  G  F  P  W  A  V  T  E  H  Y  G  T  L  Q  E  M  I  A   1109

4501 GCGTTGATGTCGCGCCAGACGTCCCCGCGACCAGTCCACTCCCCTGGGATGCTGCCTCTT  4560
1110 R  V  D  V  A  P  D  V  P  A  T  S  P  L  P  W  D  A  A  S   1129

4561 GGGCCCCGATCCTCGATGCGGCCACCTCAGTGGGATCCACGCTCTTTTTCGATCAGCCTC  4620
1130 W  A  P  I  L  D  A  A  T  S  V  G  S  T  L  F  F  D  Q  P   1149

4621 GCCTGCGCATGCCGGCTCACATTCACGGGGTTCAAGTCTACACCACGCAGCCGCCTCCCA  4680
1150 R  L  R  M  P  A  H  I  H  G  V  Q  V  Y  T  T  Q  P  P  P   1169

4681 AGGTGGGTTACCTGTACGTGGAAAAGGCTGGCGATCGGGATCTGGCGGTGCATGTCAGTG  4740
1170 K  V  G  Y  L  Y  V  E  K  A  G  D  R  D  L  A  V  H  V  S   1189

4741 TCTGCGACGAGCTCGGAACCGTCTTAGCTCGATTCGAATCCATGCGCTTTTCCGAGATCG  4800
1190 V  C  D  E  L  G  T  V  L  A  R  F  E  S  M  R  F  S  E  I   1209

4801 AAGGCACGCCGGGCAGTAACGGCAGCGAGGAGAGTCTTGTCCACCAGCTCGCATGGCCTC  4860
1210 E  G  T  P  G  S  N  G  S  E  E  S  L  V  H  Q  L  A  W  P   1229

4861 CCGCCACCTACAGCGAGAAGCCGCTGACAATCAACAATGTCGTCCTCATTTCCCGGGATA  4920
1230 P  A  T  Y  S  E  K  P  L  T  I  N  N  V  V  L  I  S  R  D   1249

4921 GGAACGTCGCAGATCTCTACTGTGGGTCTTTGAAAGATCGTGTGTCATCTATCACGGTGC  4980
1250 R  N  V  A  D  L  Y  C  G  S  L  K  D  R  V  S  S  I  T  V   1269

4981 TGGATGCTGCTGCCGACCTGCTTTCCCTTTCGCAGGATCCCAGCAGTGTCTTGCAAGCAA  5040
1270 L  D  A  A  A  D  L  L  S  L  S  Q  D  P  S  S  V  L  Q  A   1289

5041 AGGATACAGCGGTGGTGTACGTGCCCGGTCCCCTCCACAGCGCGGATTCTATCCCGACTG  5100
1290 K  D  T  A  V  V  Y  V  P  G  P  L  H  S  A  D  S  I  P  T   1309

5101 CGGCCCATTCCTTCCTCATGGAATTGCTCCTCCTGGTCAAAATCATTGTCAATGGCTCTT  5160
1310 A  A  H  S  F  L  M  E  L  L  L  L  V  K  I  I  V  N  G  S   1329

5161 TGCCCACCAAGGTCTTTGTCCTTACGGACCGCGTCTGCGAGAGTGAGTCTGCGACGGCTC  5220
1330 L  P  T  K  V  F  V  L  T  D  R  V  C  E  S  E  S  A  T  A   1349

5221 TCGCTCAGTCTCCGATCCACGGTGTCTCCCGCATCATTGCTTCGGAGCACCCAGATCAAT  5280
1350 L  A  Q  S  P  I  H  G  V  S  R  I  I  A  S  E  H  P  D  Q   1369

5281 GGGGCGGACTGATTGACGTCGAAACGCCGGGCCAGTTCCCACTCGAGACGATGAAGTATG  5340
1370 W  G  G  L  I  D  V  E  T  P  G  Q  F  P  L  E  T  M  K  Y   1389

5341 TGCAGGAGGCGGACAACATCCGCATCTCGGATGGCATACCCAGAATTGCTCGTCTGCGCC  5400
1390 V  Q  E  A  D  N  I  R  I  S  D  G  I  P  R  I  A  R  L  R   1409

5401 CGCTTCCTCGCGACAAGCTCCTACCGCCTAGCAAGCAGACTTCCCTGCTTCCCCGGCCCG  5460
1410 P  L  P  R  D  K  L  L  P  P  S  K  Q  T  S  L  L  P  R  P   1429

5461 AAGGTACCTACTTAATTACGGGTGGACTGGGCGCTCTGGGGTTGGAGGTCGCACAGTTCC  5520
1430 E  G  T  Y  L  I  T  G  G  L  G  A  L  G  L  E  V  A  Q  F   1449

5521 TGGTGGAAAAGGGTGCTCGTCGATTGATCCTCGTTTCTCGGCGTGCCTTGCCTCCGCGCC  5580
1450 L  V  E  K  G  A  R  R  L  I  L  V  S  R  R  A  L  P  P  R   1469

5581 GGGAGTGGGCAGACATCCTTGCTGATCCATCGTCCTCGCTGGCGCCGGCGCTGGAGACAA  5640
1470 R  E  W  A  D  I  L  A  D  P  S  S  S  L  A  P  A  L  E  T   1489

5641 TCCAGGCCCTTGAGACACAGGGAGCCACTGTCCACACCCTCGCAGTGGACATTTCCTCTC  5700
1490 I  Q  A  L  E  T  Q  G  A  T  V  H  T  L  A  V  D  I  S  S   1509

5701 CTGACGCAGCGCCTCAACTGGCAGTCGCCATTGATGCTCTGTCGCTACCTCCAGTCCGCG  5760
1510 P  D  A  A  P  Q  L  A  V  A  I  D  A  L  S  L  P  P  V  R   1529

5761 GCGTGGTCCACGCAGCAGGCGTTCTGGACAGCCAGCTGGTCCTCTCCGCCACGTCAGACT  5820
1530 G  V  V  H  A  A  G  V  L  D  S  Q  L  V  L  S  A  T  S  D   1549

5821 CTGTCGAGCGCGTGCTGGCGCCCAAGATCACCGGAGCGCTGGTCCTTGGCACCGTCTTCC  5880
1550 S  V  E  R  V  L  A  P  K  I  T  G  A  L  V  L  G  T  V  F   1569

5881 CCCCCAAGGCACTCGATTTCTTCATGCTATTCTCCTCATGCGGACAGATACTAGGCTTCC  5940
1570 P  P  K  A  L  D  F  F  M  L  F  S  S  C  G  Q  I  L  G  F   1589

5941 CAGGTCAAGCATCCTACGCGTCCGGAAACGCGTTCCTTGATGCATTCGCAACATCGCGCC  6000
1590 P  G  Q  A  S  Y  A  S  G  N  A  F  L  D  A  F  A  T  S  R   1609

6001 GACACCAAGGAGACAACGCTGTCGCCGTGCAGTGGACCAGCTGGCGCTCCCTCGGCATGG  6060
1610 R  H  Q  G  D  N  A  V  A  V  Q  W  T  S  W  R  S  L  G  M   1629

6061 CAGCCAGTACCGACTTCATCAACGCTGAGCTAGCCAGCAAGGGCATCACTGACATCACTC  6120
1630 A  A  S  T  D  F  I  N  A  E  L  A  S  K  G  I  T  D  I  T   1649

6121 GCGACGAGGGGTTCCGCGCGTGGATGCATATTTCCAAATATGATATCGACCAGGCCGCCG  6180
1650 R  D  E  G  F  R  A  W  M  H  I  S  K  Y  D  I  D  Q  A  A   1669

6181 TCTTGCGCAGTCTGGCCTTCGAGGCCGATGAACCCCTCCCCACCCCTATCCTTACGGATA  6240
1670 V  L  R  S  L  A  F  E  A  D  E  P  L  P  T  P  I  L  T  D   1689

6241 TTGCCGTCCGCAAGGCTGGCTCCGCCTCCTCCGCTGATGCTCCCTCTGCTGCACCGAAAG  6300
1690 I  A  V  R  K  A  G  S  A  S  S  A  D  A  P  S  A  A  P  K   1709

6301 AGACGAACGAAATGCCGGAATCGATCCCGGAGCGTCGTACCTGGTTGGATGAACGAATCC  6360
1710 E  T  N  E  M  P  E  S  I  P  E  R  R  T  W  L  D  E  R  I   1729

6361 GTGATTGTGTGGCCCGCGTGCTTCAGCTGGGGAGCAGCGATGAGGTTGATTCCAAGGCCG  6420
1730 R  D  C  V  A  R  V  L  Q  L  G  S  S  D  E  V  D  S  K  A   1749

6421 CTCTGAGTGATCTGGGAGTCGACAGCGTCATGACCGTTAGCTTGAGAGGTCAGCTGCAGA  6480
1750 A  L  S  D  L  G  V  D  S  V  M  T  V  S  L  R  G  Q  L  Q   1769

6481 AGACGTTGGGGGTCAAGGTGCCACCCACACTGACCTGGAGTTGCCCGACGGTGTCACATC  6540
1770 K  T  L  G  V  K  V  P  P  T  L  T  W  S  C  P  T  V  S  H   1789

6541 TGGTGGGATGGTTTTTGGAAAAGATGGGAAATTGATTAGAGCTGATGGTTTCTCCTTGTG  6600
1790 L  V  G  W  F  L  E  K  M  G  N  *                             1800

6601 TGTTCTTCATTTGATATATATTTGTTGTTTCGTCTCTGGTTTCCCCCTTTGCCTGTTCCC  6660

6661 CTGAATCAATTTGCCAAGACTGTGATGCCAACTGAACCTGACGAGATTGTATATGTCACA  6720

6721 TGATATTAATTTGTATGAGGATATTAGCAATAGACGAGTCATGTTTTTGGATATATATAT  6780

6781 ACTCTCCCTGCGCGAAAGGAGGCAGGAATGAAGCTGGAGAAGAGGCGATGAATTGGATAG  6840

6841 GCTTTCTGCAGACGCCTCTGGAGTAGGGGATGTGCTAGGCCTAAACGGGCAGGCTCACGG  6900

6901 CCTGAGGCTCCAGCGTTCCCAATTGTTTCCCTATCAAGTCAAGGGTGTTGGGGACGGAGA  6960

6961 GCTTTCTCTTGCGCAATAAAGAATAGTCGATTTAGTTTCTTGAACGTGCACATACCGCAG  7020

7021 CATTGTAGGAATTGGCTCGTCAATTGACATAATTCGGATGTATCTCAACATCATCTGTAG  7080

7081 AGCATCGTGTGTGAAATATATATCATACCAACTTGAATACCATCAAGAAGACATAGGTTC  7140

7141 TAGTTACTGAGATAGTTGCATTAAACAGCTCGTATCAACTATGCGTATGCTTCGCTAAAT  7200

7201 ACTGTAGCATGCAAAGATACCACACTCACTGAAAGTTGATCTGAGATGAGAACTGCAAGA  7260

7261 GGCAGTTTTCACCGAGTCAGGCTGTCCTATCTTCTGGTGTCTAGACAACTTCCCACTGCT  7320

7321 ACGTATATTCATCGACTAAGGTGTGGACAGTCTAGTACATTATCTGTAAGCTACTTCAGG  7380

7381 CAGTACATGCGACGTGTGCCAGAGATAGTCTCGGATGCAGTATAAGACATCGTGGCAATG  7440

7441 AAAAAGATCATCTAAAACCGGAGCTTGCTTATAATAACCTGACAGGATAACCACTTATAT  7500

7501 TTGCACTCCGAACCCCGGAGTGTATGCTCCAATACAGCAGTAGCCCATCACCTCAGGAAA  7560

7561 CCTCATAAATTCACTTACTCAAGATATC                                  7588
```

Fig. 4 Nucleotide and deduced amino acid sequence of the *A. terreus atX* gene. The intron is indicated in *lower case*. A potential TATA box and CAAT box are *underlined*

With the *P. patulum* MSAS probe, no hybridizing signal was detected in the genome of orsellinic acid-producing *P. cyclopium* even at low stringency conditions, indicating a significant difference between orsellinic acid synthase and MSAS at the genetic level, although these PKSs catalyze the formation of single aromatic ring compounds that differ by only one hydroxyl group.

An MKS-homologous gene, *atX*, was cloned from *A. terreus* IMI 16043, which produces (+)-geodin as a main metabolite. (+)-Geodin is a seco-anthraquinone compound derived from octaketide emodinanthrone. Thus, this MKS-homologous gene from *A. terreus* was expected to encode emodinanthrone synthase. However, no transcript of *atX* was detected during the growth of *A. terreus* by Northern blot analysis. The dot matrix comparison of the amino acid sequences of ATX and the MSAS of *P. patulum* showed high homology of

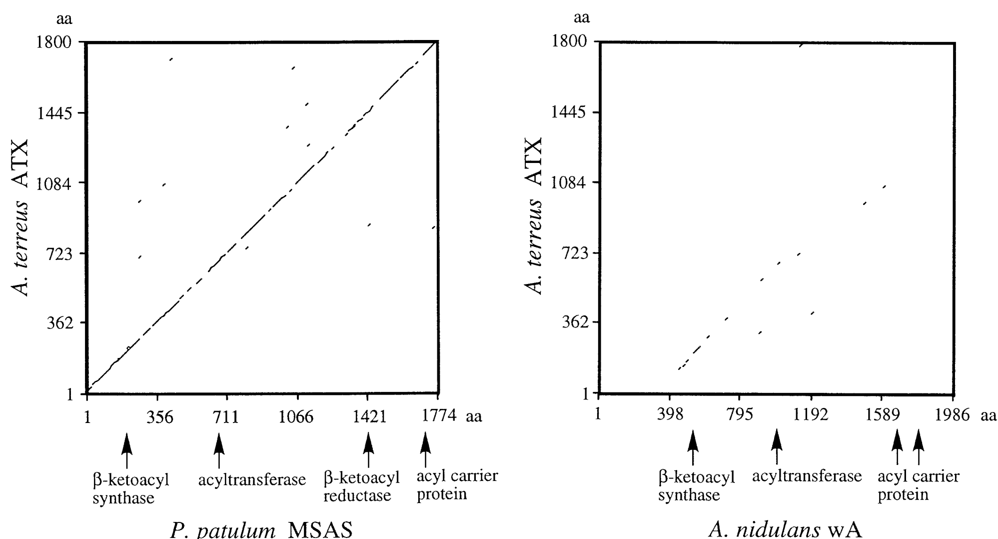

Fig. 5 Dot matrix comparison of ATX with fungal polyketide synthases. The ATX polypeptide was compared with the *Penicillium patulum* MSAS polypeptide (Beck et al. 1990) and the *A. nidulans wA* polypeptide (Mayorga and Timberlake 1992). Comparison was carried out with a window size of 10 and a stringency of 6

A β-ketoacyl synthase

RISYHLNLMGPSTAVDAA**C**ASSLVAIHHGRGAILQGE	*A. terreus* ATX
RISYHLNLMGPSTAVDAA**C**ASSLVAIHHGVQAIRLGE	*P. patulum* MSAS
RINYYFKFSGPSVSVDTA**C**SSSLAAIHLACNSIWRND	*A. nidulans wA*
RINYHFGFSGPSLNVDTA**C**SSSAAALNVACNSLWQKD	*C. lagenarium PKS* 1
RISYHLNLMGPSTAVDAA**C**ASSLVAIHHGVQAIRLGE	*S. erythraea ery*AIII
EVAWEAGAEGPVTVVSTG**C**TSGLDAVGYGTELIRDGR	*S. glaucescens tcm*K

B Acyltransferase

GLEPQAVIGH**S**VGEIAASVAAGCLTAEEGALIVT	*A. terreus* ATX
GITPQAVIGH**S**VGEIAASVVAGALSPAEGALIVT	*P. patulum* MSAS
GITPSFVLGH**S**LGDFAAMNAAGVLSTSDTIYACG	*A. nidulans wA*
GIRPSAVMGH**S**LGEYAALNAAGVLSASDTIYLVG	*C. lagenarium PKS* 1
GVEPAAVVGH**S**QGEIAAAHVAGALTLEDAAKLVA	*S. erythraea ery*AIII

C β-ketoacyl reductase

G T Y L I T G **G** L **G** A L **G** L E V A	*A. terreus* ATX
G T Y L I T G **G** L **G** V L **G** L E V A	*P. patulum* MSAS
G T A L V T G **G** T **G** A L **G** G H V A	*S. erythraea ery*AIII KR2
V N N A G R P **G** G **G** A T **A** E L A D	*S. coelicolor act*III

D Acyl carrier protein

S D L G V D **S** V M T V S L R	*A. terreus* ATX
A D L G V D **S** V M T V T L R	*P. patulum* MSAS
N E L G M D **S** L L S L T V L	*A. nidulans wA* ORF 1
A D Y G V D **S** L L S L T V T	*A. nidulans wA* ORF 2
T D L G C D **S** L M A L T V S	*C. lagenarium PKS* 1 ORF 1
A A L G M D **S** L M S L S I L	*C. lagenarium PKS* 1 ORF 2
T E L G F D **S** L T A V G L R	*S. erythraea ery*AIII ACP2
Q D L G Y D **S** I A L L E I S	*S. glaucescens tcm* M

Fig. 6A–D Alignment of the ATX amino acid sequence with the active site regions of polyketide synthases. **A** β-ketoacyl synthase motif; **B** acyltransferase motif; **C** β-ketoacyl reductase motif; **D** acyl carrier protein motif (Hallam et al. 1988; Bibb et al. 1989; Beck et al. 1990; Cortés et al. 1990; Bevitt et al. 1992; Mayorga and Timberlake 1992; Takano et al. 1995) *Bold letters* indicate the active site cysteine residue (**A**), the active site serine residue (**B**), The NADPH-binding motif (**C**), and the pantetheine-binding serine residue (**D**)

the complete sequences from N-terminus to C-terminus. In particular, high homology was observed in their active sites, e.g., nearly 100% identity between the β-ketoacyl synthase regions (Fig. 6). These results strongly suggested that *atX* is the gene coding for MSAS or a closely related PKS of *A. terreus.*

Identification of ATX as the MSAS of *A. terreus* was carried out by expression of the *atX* gene in *A. nidulans*, which is the genetically most well characterized fungus and a suitable host for the functional analysis of cloned fungal genes. The pTA-ATX transformant of *A. nidulans* produced quite high amounts of 6-methylsalicylic acid, which was identified by physicochemical analysis. This successful expression also demonstrated that the intervening sequence of the *atX* gene was correctly processed in *A. nidulans.*

Several other fungal PKS genes have been cloned and their nucleotide sequences reported, e.g., *A. nidulans wA* (Mayorga and Timberlake 1992), *A. nidulans* PKS^{ST} (Yu and Leonard 1995), and *C. lagenarium PKS1* (Takano et al. 1995). Of these, the PKS^{ST} and *PKS1* products are assumed to be norsolorinic acid anthrone and tetrahydroxynaphthalene, respectively. However, no direct identification of these PKS products, including the *wA* product has been carried out. Interestingly, fairly high homology exists between these three fungal PKS sequences, including tandemly duplicated active sites for acyl carrier protein characteristic of fungal PKSs except MSAS. Structural identification of these PKS products is the first necessary step in dissecting functional mechanisms of these types of fungal PKS reactions. Our expression system with the pTAex3 vector and *A. nidulans* will be a useful tool for this purpose and future functional analysis by genetic manipulation such as site-directed mutagenesis.

Recently, Khosla and colleagues reported the successful expression of the MSAS gene in the bacterial host *Streptomyces coelicolor*, using their expression vector (Bedford et al. 1995). In their case, an intervening sequence had to be removed to construct the expres-

sion plasmid and the production level was comparatively low.

In fungi, DNAs introduced by transformation are usually integrated into host genomes and the transformants are mitotically highly stable (Fincham 1989). Once the desired transformant has been selected, a further selection force, such as antibiotic addition, is usually unnecessary. Therefore, the fungal system is suitable for large-scale expression and product analysis.

Acknowledgements We thank Professor E. Schweizer of the University of Erlangen-Nürnberg for providing the 6-methylsalicylic acid synthase gene of *Penicillium patulum.* We also thank the Human Genome Center (Institute of Medical Science, The University of Tokyo) for the use of their program package for *nucleotide and protein sequence analysis.*

References

Beck J, Ripka S, Signer A, Schiltz E, Schweizer E (1990) The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum.* Eur J Biochem 192 : 487–498

Bedford DJ, Schweizer E, Hopwood DA, Khosla C (1995) Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2). J Bacteriol 177 : 4544–4548

Bevitt DJ, Cortés J, Haydock SF, Leadlay PF (1992) 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea.* Eur J Biochem 204 : 39–49

Bibb MJ, Biró S, Motamedi H, Collins JF, Hutchinson CR (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. EMBO J 8 : 2727–2736

Biel SW, Parrish FW (1986) Isolation DNA from fungal mycelia and sclerotia without the use of density gradient ultracentrifugation. Anal Biochem 154 : 21–25

Chargwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18 : 5294–5299

Chen ZG, Fujii I, Ebizuka Y, Sankawa U (1992) Emodin O-methyltransferase from *Aspergillus terreus.* Arch Microbiol 158 : 29–34

Chen ZG, Fujii I, Ebizuka Y, Sankawa U (1995) Emodinanthrone oxygenase from *Aspergillus terreus.* Phytochemistry 38 : 299–305

Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF (1990) An unusually large multifunctional polypeptide in the erythromycin-polyketide synthase of *Saccharopolyspora erythrea.* Nature 346 : 176–178

Dimroth P, Walter H, Lynen F (1970) Biosynthese von 6-methylsalicylsäure. Eur J Biochem 13 : 98–110

Donadio S, Katz L (1992) Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin biosynthesis in *Saccharopolyspora erythrea.* Gene 111 : 51–60

Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252 : 675–679

Fincham JRS (1989) Transformation in fungi. Microbiol Rev 53 : 148–170

Fujii I, Ebizuka Y, Sankawa U (1982) Partial purification and some properties of emodin-O-methyltransferase from (+)-geodin producing strain of *Aspergillus terreus.* Chem Pharm Bull (Tokyo) 30 : 2283–2286

Fujii I, Iijima H, Ebizuka Y, Sankawa U (1983) Purification and characterization of dihydrogeodin oxidase from a fungal strain of *Aspergillus terreus* producing (+)-geodin. Chem Pharm Bull (Tokyo) 31 : 337–340

Fujii I, Iijima H, Tsukita S, Ebizuka Y, Sankawa U (1987) Purification and properties of dihydrogeodin oxidase from *Aspergillus terreus.* J Biochem 101 : 11–18

Fujii I, Ebizuka Y, Sankawa U (1988) A novel anthraquinone ring cleavage enzyme from *Aspergillus terreus.* J Biochem 103 : 878–883

Fujii I, Chen ZG, Ebizuka Y, Sankawa U (1991) Identification of emodinanthrone oxygenase in fungus *Aspergillus terreus.* Biochem Int 25 : 1043–1049

Fujii T, Yamaoka H, Gomi K, Kitamoto K, Kumagai C (1995) Cloning and nucleotide sequence of the ribonuclease T_1 gene (*rntA*) from *Aspergillus oryzae* and its expression in *Saccharomyces cerevisiae* and *Aspergillus oryzae.* Biosci Biotech Biochem 59 : 1869–1874

Gomi K, Iimura Y, Hara S (1987) Integrative transformation of *Aspergillus oryzae* with a plasmid containing the *Aspergillus nidulans argB* gene. Agric Biol Chem 51 : 2549–2555

Gurr S, Unkles SE, Kinghorn JR (1987) The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn (ed) Gene structure in eukaryotic microbes. IRL Press, Oxford, pp 93–139

Hallam SE, Malpartida F, Hopwood DA (1988) DNA sequence, transcription and deduced function of a gene involved in polyketide antibiotic biosynthesis in *Streptomyces coelicolor*. Gene 74 : 305–320

Hopwood DA, Khosla C (1992) Genes for polyketide secondary metabolic pathways in microorganisms and plants. Ciba Found Symp 17 : 88–112

Hopwood DA, Sherman DH (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu Rev Genet 24 : 37–66

Huang KX, Fujii I, Ebizuka Y, Gomi K, Sankawa U (1995a) Molecular cloning and heterologous expression of the gene encoding dihydrogeodin oxidase, a multicopper blue enzyme from *Aspergillus terreus.* J Biol Chem 270 : 21495–21502

Huang KX, Iwakami N, Fujii I, Ebizuka Y, Sankawa U (1995b) Transformation of *Penicillium islandicum* and *Penicilliium frequentans* that produce anthraquinone-related compounds. Curr Genet 28 : 580–584

Hutchinson CR, Fujii I (1995) Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. Annu Rev Microbiol 49 : 201–238

Malpartida F, Hallam SE, Kieser HM, Motamedi H, Hutchinson CR, Butler MJ, Sugden DA, Warren M, McKillop C, Bailey CR, Humphrey GO, Hopwood DA (1987) Homology between *Streptomyces* genes coding for synthesis of different polyketide used to clone antibiotic biosynthesis genes. Nature 325 : 818–821

Mayorga ME, Timberlake WE (1990) Isolation and molecular characterization of the *Aspergillus nidulans wA* gene. Genetics 126 : 73–79

Mayorga ME, Timberlake WE (1992) The developmentally regulated *Aspergillus nidulans wA* gene encodes a polypeptide homologous to polyketide and fatty acid synthases. Mol Gen Genet 235 : 205–212

O'Hagen D (1991) The polyketide metabolites. Ellis Horwood, NY

O'Hagen D (1992) Biosynthesis of polyketide metabolites. Nat Prod Rep 9 : 447–479

O'Hagen D (1993) Biosynthesis of fatty acid and polyketide metabolites. Nat Prod Rep 10 : 593–624

Pieper R, Luo G, Cane DE, Khosla C (1995) Cell-free synthesis of polyketides by recombinant erythromycin polyketide synthases. Nature 378 : 263–266

Pontecorvo P, Roper JA, Hammons LM, Macdonald KD, Bufton AW (1953) The genetics of *Aspergillus nidulans.* Adv Genet 5 : 141–238

Shen B, Hutchinson CR (1993) Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. Science 262 : 1535–1540

Spencer JB, Jordan PM (1992a) Investigation of the mechanism and steric course of the reaction catalyzed by 6-methylsalicylic acid

synthase from *Penicillium patulum* using (*R*)-[1-^{13}C; 2-^{2}H]- and (*S*)-[1-^{13}C; 2-^{2}H] malonates. Biochemistry 31 : 9107–9116

Spencer JB, Jordan PM (1992b) Purification and properties of 6-methylsalicylic acid synthase from *Penicillium patulum*. Biochem J 288 : 839–846

Tada S, Gomi K, Kitamoto K, Takahashi K, Tamura G, Hara S (1991) Construction of a fusion gene comprising the Taka-amylase A promoter and the *Escherichia coli β*-glucuronidase gene and analysis of its expression in *Aspergillus oryzae*. Mol Gen Genet 229 : 301–306

Takano Y, Kubo Y, Shimizu K, Mise K, Okuno T, Furusawa I (1995) Structural analysis of *PKS1*, a polyketide synthase gene involved in melanin biosynthesis of *Colletotrichum lagenarium*. Mol Gen Genet 249 : 162–167

Tilburn J, Scazzocchia C, Taylor GG, Zabichy-Zissman JH, Lockington RA, Davies RW (1983) Transformation by integration in *Aspergillus nidulans*. Gene 26 : 205–221

Tsukamoto N, Fujii I, Ebizuka Y, Sankawa U (1992) Cloning of aklavinone biosynthesis genes from *Streptomyces galilaeus*. J Antibiot 45 : 1286–1294

Tsukamoto N, Fujii I, Ebizuka Y, Sankawa U (1994) Nucleotide sequence of the *aknA* region of the aklavinone biosynthesis gene cluster of *Streptomyces galilaeus*. J Bacteriol 176 : 2473–2475

Turner WB, Aldridge (1983) Fungal metabolites II. Academic Press, London

Woo ER, Fujii I, Ebizuka Y, Sankawa U, Kawaguchi A, Beale JM, Shibuya M, Mocek U, Floss HG (1989) Nonstereospecific proton removal in the enzymatic formation of orsellinic acid from chiral malonate. J Am Chem Soc 111 : 5498–5500

Yu JH, Leonard TJ (1995) Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type I polyketide synthase. J Bacteriol 177 : 4792–4800