Cloning and Characterization of Monacolin K Biosynthetic Gene Cluster from Monascus pilosus

YI-PEI CHEN,†,§ CHING-PING TSENG,§ LI-LING LIAW,† CHUN-LIN WANG,† I-CHING CHEN,† WEN-JUNG WU,† MING-DER WU,† and GWO-FANG YUAN*,†

Bioresource Collection and Research Center, Food Industry Research and Development Institute, and Department of Biological Science and Technology, National Chiao Tung University, HsinChu, Taiwan

Monacolin K is a secondary metabolite synthesized by polyketide synthases (PKS) from Monascus, and it has the same structure as lovastatin, which is mainly produced by Aspergillus terreus. In the present study, a bacterial artificial chromosome (BAC) clone, mps01, was screened from the BAC library constructed from Monascus pilosus BCRC38072 genomic DNA. The putative monacolin K biosynthetic gene cluster was found within a 42 kb region in the mps01 clone. The deduced amino acid sequences encoded by the nine genes designated as mokA–mokI, which share over 54% similarity with the lovastatin biosynthetic gene cluster in A. terreus, were assumed to be involved in monacolin K biosynthesis. A gene disruption construct designed to replace the central part of mokA, a polyketide synthase gene, in wild-type M. pilosus BCRC38072 with a hygromycin B resistance gene through homologous recombination, resulted in a mokA-disrupted strain. The disruptant did not produce monacolin K, indicating that mokA encoded the PKS responsible for monacolin K biosynthesis in M. pilosus BCRC38072.

KEYWORDS: Monacolin K; polyketide synthases; Monascus pilosus; bacterial artificial chromosome

INTRODUCTION

Monascus spp. are filamentous fungi that have been used in Chinese fermented foods for thousands of years. They are known as producers of various secondary metabolites with polyketide structures, including monacolins, pigments, and citrinin (1–4). Monacolin K, a cholesterol synthesis inhibitor, was first isolated from the medium of Monascus ruber (5), and the same substance was found in Aspergillus terreus as lovastatin (5). It belongs to polyketide synthesized by the iterative type I PKSs. The structure of monacolin K shares similarity with HMG-CoA; therefore, monacolin K competitively inhibits HMG-CoA reductase with HMG-CoA during cholesterol synthesis resulting in the reduction of cholesterol synthesis (6).

In previous studies, the lovastatin biosynthetic pathway was proposed in A. terreus. Two polyketide synthases (lovB and lovF), transerasesterase (lovD), enoyl reductase (ER) (lovC), and P450 monoxygenase (lovA) have been proven to be involved in the structural biosynthesis of lovastatin (7–9). Transformation of an extra copy of the lovE gene-encoded transcription factor into the wild-type strain resulted in a 7–10-fold overproduction of lovastatin (7). Although the lovastatin biosynthetic gene cluster in A. terreus has been characterized (10), the structural genes responsible for monacolin K (lovastatin) biosynthesis in Monascus are still unclear. In the present study, to explore the monacolin K biosynthetic gene cluster, construction of a bacterial artificial chromosome (BAC) library from M. pilosus BCRC38072 producing monacolin K was carried out. According to the conserved region of the lovB gene (lovastatin nonaketide synthase, LNKS), A. terreus was designed as a probe (11), and the mps01 clone containing the putative monacolin K biosynthetic gene cluster was isolated. Analysis of the disruption of the polyketide synthase gene (mokA) was conducted to identify the gene involved in monacolin K biosynthesis.

MATERIALS AND METHODS

Strain Used and Growth Conditions. M. pilosus BCRC38072, which is a monacolin K-producing strain isolated from red rice (anka), was collected from a local traditional market and used in this study. To identify the transcripts from monacolin K biosynthetic genes, the strain was incubated on YM (DIFCO 271120, Detroit, MI) agar for 1 week, and spore suspensions were obtained by washing cultured YM agar plates with distilled water. Mycelia were harvested after incubation for 12 days at 25 °C with constant agitation in liquid medium (7% glycerol, 3% glucose, 3% monosodium glutamate, 1.2% polyethylene, 0.2% NaNO3, and 0.1% MgSO4·7H2O).

BAC Library Construction. The methods of Peterson et al. (12) were used to construct the BAC library. Fragments of genomic DNA ranging in size from 150 to 300 kb were excised from pulse field gel electrophoresis (PFGE) and recovered by electroelution (BioRad, Hercules, CA). The eluted DNA was used for ligation. To perform the ligation, 100–200 ng of electroeluted DNA was mixed with 50 ng of linearized vector DNA (PIndigoBAC-5 HindIII Ready, Epicenter, Madison, WI), after which the ligation products were used to transform
Escherichia coli strain TransforMax EC 100 electrocompetent (Epicenter, Madison, WI) by electroporation. Transformed cells were cultured on LB agar plates supplemented with chloramphenicol (12.5 µg mL⁻¹), IPTG (100 µg mL⁻¹), and XGal (50 µg mL⁻¹). The resulting white bacterial colonies were harvested and transferred to 384 well plates for library screening or storage at −80 °C in freezing medium (2.5% [w/v] LB, 36 mM K₂HPO₄, 13.2 mM KH₂PO₄, 1.7 mM sodium citrate, 0.4 mM MgSO₄, 6.8 mM [NH₄]₂SO₄, and 4.4% v/v glycerol).

Library Screening, Sequencing, and Sequence Analysis. About 12000 clones from the BAC library were cultured on LB agar plates, after which they were transferred to nylon membranes and then subjected to alkali-sodium dodecyl sulfate lysis. The plasmid DNAs extracted from BAC clones were cross-linked to nylon membranes by UV irradiation. According to the conserved region of the los9b gene in A. terreus (10), the primer set (Mplovl, 5′-TCCACTGCGGTATTTAGTTC-3′; Mplovr, 5′-GATGGGGGTGAAGATGACGA-3′) was designed for the probe synthesis using a PCR DIG probe synthesis kit (Roche Diagnostics, Mannheim, Germany). The clones of membranes were screened to find a gene cluster involved in polyketide biosynthesis metabolism. Twenty-five positive BACs were identified in screening by colony hybridization. Two of these, mps01 and mps02, containing the longest inserted DNA, were sequenced. BAC mps01 was used to construct a complete monacolin K gene cluster instead of the incomplete mps02, with a reverse-phase C18 column (LichroCART 250-4, Rp-18e, 5 µm). Amplification products were sequenced and analyzed. The result of the PCR amplification using the PCR DIG probe synthesis kit (Roche Diagnostics, Mannheim, Germany). The primer sets of monacolin K biosynthetic genes were DIG-labeled by PCR transcription polymerase chain reaction (PCR) analyses. The probes of the monacolin K biosynthetic gene cluster were submitted to GenBank under the accession number DQ176595.

Cloning of Monacolin K Biosynthetic Gene Cluster. Studies on fungal polyketide biosynthetic genes indicate that metabolites are largely synthesized by iterative multifunctional polyketide synthase systems (17). Each PKS minimally carries a ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) domains to catalyze different modifications. To search for the genes related to monacolin K biosynthesis, degenerate primers designed according to the conserved region of the KS domain of the los9b gene in A. terreus (11), were used to amplify genomic DNA from M. pilosus BCRC38072. The candidate PCR products were sequenced and analyzed. The result of the

**RESULTS**

Cloning of Monacolin K Biosynthetic Gene Cluster. Studies on fungal polyketide biosynthetic genes indicate that metabolites are largely synthesized by iterative multifunctional polyketide synthase systems (17). Each PKS minimally carries a ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) domains to catalyze different modifications. To search for the genes related to monacolin K biosynthesis, degenerate primers designed according to the conserved region of the KS domain of the los9b gene in A. terreus (11), were used to amplify genomic DNA from M. pilosus BCRC38072. The candidate PCR products were sequenced and analyzed. The result of the
amplified DNA showed that the PCR product shared a 75% similarity with the KS domain of the lov gene in A. terreus (data not shown). This DNA fragment was further used to design a specific probe for cloning of the PKS gene.

A BAC library consisting of 12000 clones was constructed from the total DNA of M. pilosus BCRC38072. By screening the library with the specific PKS probe, 25 positive BACs were identified. To evaluate the sizes of the BACs, PFGE and Southern hybridization were carried out. The BAC designated as mps01 was selected for shotgun sequencing. It yielded a contig of approximately 160 kb. Database searches and open reading frame (ORF) prediction further provided information on the putative gene loci. The whole sequences of mps01 were retrieved from GenBank database. Table 1. Summary of Genes Identified in BAC mps01 Obtained from M. pilosus BCRC38072

<table>
<thead>
<tr>
<th>mok genes</th>
<th>amino acids</th>
<th>putative molecular mass</th>
<th>proposed function</th>
<th>homologous lov gene</th>
<th>protein similarity (%)</th>
<th>homologous mlc gene</th>
<th>protein similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mokA</td>
<td>3075</td>
<td>338.1</td>
<td>polyketide synthase</td>
<td>lovB</td>
<td>76</td>
<td>mlcA</td>
<td>66</td>
</tr>
<tr>
<td>mokB</td>
<td>2547</td>
<td>278.3</td>
<td></td>
<td>lovF</td>
<td>73</td>
<td>mlcB</td>
<td>61</td>
</tr>
<tr>
<td>mokC</td>
<td>524</td>
<td>60.6</td>
<td></td>
<td>lovA</td>
<td>85</td>
<td>mlcC</td>
<td>67</td>
</tr>
<tr>
<td>mokD</td>
<td>263</td>
<td>28.9</td>
<td></td>
<td>lovG</td>
<td>67</td>
<td>mlcF</td>
<td>53</td>
</tr>
<tr>
<td>mokE</td>
<td>360</td>
<td>38.9</td>
<td>dehydrogenase</td>
<td>lovC</td>
<td>81</td>
<td>mlcG</td>
<td>70</td>
</tr>
<tr>
<td>mokF</td>
<td>413</td>
<td>46.8</td>
<td>transesterase</td>
<td>lovD</td>
<td>74</td>
<td>mlcH</td>
<td>63</td>
</tr>
<tr>
<td>mokG</td>
<td>1052</td>
<td>113.0</td>
<td>HMG-CoA reductase</td>
<td>lovA</td>
<td>69</td>
<td>mlcD</td>
<td>39</td>
</tr>
<tr>
<td>mokH</td>
<td>455</td>
<td>49.4</td>
<td>transcription factor</td>
<td>lovE</td>
<td>54</td>
<td>mlcR</td>
<td>49</td>
</tr>
<tr>
<td>mokI</td>
<td>543</td>
<td>57.5</td>
<td>efflux pump</td>
<td>lovI</td>
<td>81</td>
<td>mlcE</td>
<td>68</td>
</tr>
</tbody>
</table>

*The deduced amino acid sequences were determined from cDNA sequences. The proposed gene functions were based on their homology to proteins in the GenBank database. The lovastatin biosynthetic gene cluster. Similarity was obtained by alignment using VectorNTI 9.0 (InforMax) software. The compactin biosynthetic gene cluster.*

Table 1. Several conserved domains were recognized in MokA and MokB by comparing their amino acid sequences with those of known PKSs. The domains of KS, AT, DH, MeT, KR, and ACP are included in both MokA and MokB. Additionally, MokB comprised an additional ER domain similar to the corresponding gene lovF of A. terreus, as shown in Figure 1C. The mokH gene-encoded transcription factor was suggested to be a positive regulatory protein for the production of monacolin K just like lovE, which is involved in lovastatin biosynthesis in A. terreus (Figure 2) (10). The arrangement of a cysteine-rich nucleotide-binding domain indicated that the consensus sequence CX2CX6CX11CX2CX6C represented a Zn2Cys6 type zinc finger (7, 10).

Polyketides are frequently synthesized from CoA thioesterified carboxylic acids, and the extent of keto-group processing varies from one condensation cycle to another (7). In this study, the phylogeny was further constructed according to the conserved domain of KS (Figure 3A). The result showed that the PKSs were divided into three clades. The clade of mokA and mokB was subdivided into two subclades belonging to the structural type of highly reduced polyketide. Because the MeT domain in mlcA of the compactin biosynthetic gene cluster is assumed to be inactive (18), the domain was compared among corresponding PKSs from M. pilosus, P. citrinum, and A. terreus, and our results showed that the amino acid residues of mlcA in consensus motifs were different from those of the other PKS, as boxed (Figure 3B).

Disruption of mokA Gene in M. pilosus BCRC38072. The mokA gene-encoded polyketide synthase was suggested to synthesize the nonaketide of monacolin K. In this study, the mokA gene was disrupted in M. pilosus BCRC38072 by homologous recombination to identify the gene involved in monacolin K biosynthesis. Plasmid pMkAko was linearized at the FspI sites and was used to transform strain BCRC38072 (Figure 4A). Forty-four transformants were isolated, one of which had a MokA gene genotype by Southern hybridization.

Southern hybridization analysis of NdeI-digested DNA indicated that instead of a 7.1 kb fragment corresponding to the mokA gene in wild-type BCRC38072, a 3.8 kb NdeI fragment was present in the disruptant BCRC38135 (Figure 4B,C). Thus, our result revealed that disruption of mokA gene had occurred. A precise gene replacement, in which the nonfunctional mokA gene construct (pMkAko) replaced the functional chromosomal mokA gene, yielded NdeI fragments of 3.8 kb. In addition, the amount and structure of monacolin K produced from M. pilosus BCRC38072 and BCRC38135 on the eighth day of cultivation were further determined by HPLC, mass, and 1H NMR spectroscopic analyses. Monacolin K was detectable in wild-type M. pilosus BCRC38072, as confirmed by UV light
absorption, mass, and $^1$H NMR spectra (Figure 5). The peak of monacolin K from M. pilosus BCRC38072 was identified by comparison to the monacolin K standard, which showed three maximum absorptions at ($\lambda_{\text{max}}$) 230, 237, and 246 nm (Figure 1. Identification of the monacolin K biosynthetic gene cluster of M. pilosus BCRC38072. (A) The genes involved in the biosynthesis of monacolin K. For proposed functions of assigned ORFs, see Table 1. The small black bars indicated the probes for Northern hybridization analysis. The lovastatin biosynthetic gene cluster in A. terreus was obtained from the GenBank database using the following accession numbers: AF141924, AF151722, and AF141925. The compactin biosynthetic gene cluster in P. citrinum was obtained from the GenBank database using the following accession number: AB072893. (B) Northern hybridization analyses. Total RNA isolated after 12 days of cultivation was blotted. Total RNA (6 $\mu$g per lane) was separated on 0.8% agarose gels by electrophoresis. The size of each transcript was estimated by comparison with markers of known size. (C) Arrangement of functional domains of genes encoding PKSs in the biosynthesis of monacolin K. DH, dehydratase; MeT, methyltransferase; KR, keto-reductase; and ACP, ACP.

Figure 2. Deduced amino acid sequences alignment of transcription factors from the mokH, lovE, and mlcR genes. The cysteine-rich nucleotide-binding domain represented a Zn$_2$Cys$_6$ type zinc finger with the consensus sequence CX$_2$CX$_6$CX$_{11}$CX$_2$CX$_6$C shown boxed.
Monacolin K Biosynthetic Gene Cluster


Figure 3. (A) Phylogenetic tree of PKSs from M. pilosus BCRC38072 and various organisms. The phylogeny of PKSs based on the conserved KS domains described by Kroken et al. (21) was constructed and rooted using KS domains of Saccharopolyspora erythraea DEBS (X56107 and X62569). Accession numbers for the polyketide synthase genes were used as follows: A. terreus lovF (AF141925), A. terreus lovB (AF151722), P. citrinum mokA, mokC (AB072893), Phoma sp. SQTKS (AY217789). The tree was constructed by the neighbor-joining method (23).

(B) Comparison of the MeT domain. The three MeT consensus motifs were shown boxed. The conserved residues of MeT are described by Kagan and Clarke (16)

5A). The mass spectrum of monacolin K revealed that the molecular weight was 404, which also agreed with the standard \((C_{22}H_{36}O_5)\) (Figure 5B). Moreover, the structure of monacolin K was further verified by \(^1H\) NMR spectrum \(^1H\) NMR (400 Hz, CDCl3): 6.98 (1H, d, J = 10.0 Hz, H-5), 5.76 (1H, dd, J = 6.4, 6.0 Hz, H-6), 5.50 (1H, d, J = 2.8 Hz, H-4), 5.36 (1H, dt, J = 4.8, 3.2 Hz, H-1), 4.60 (1H, m, H-5'), 4.33 (1H, m, H-3'), 2.72 (1H, dd, J = 5.2, 4.8 Hz, H-3-CH3), 2.62 (1H, ddd, J = 3.6, 3.2, 2.4, H-3″-CH3), 2.43 (2H, m, H-3), 2.38 (1H, m, H-7), 2.36 (1H, m, H-2′), 2.27 (1H, dd, J = 2.8, 2.8 Hz, H-8a), 1.98 (1H, m, H-9″-CH3), 1.45 (2H, m, H-2), 1.89 (1H, m, H-6), 1.72 (1H, m, H-8), 1.65 (1H, m, H-9″-CH3), 1.63 (1H, m, H-3″), 1.48 (1H, m, H-7), 1.42 (1H, m, H-3″), 1.38 (1H, m, H-7), 1.29 (1H, m, H-6), 1.11 (3H, d, J = 7.2 Hz, H-7′-CH3), 1.06 (3H, d, J = 7.2 Hz, H-3′-CH3), 0.88 (3H, d, J = 7.2 Hz, H-7′-CH3), 0.86 (3H, t, J = 7.6 Hz, H-4′)

DISCUSSION

Monacolin K, also known as lovastatin, is a polyketide used to reduce serum cholesterol levels in humans. Over the past years, it has become clear that polyketides are assembled in a variety of mechanistically complex ways (7). Studies on the lovastatin biosynthetic gene cluster of A. terreus have shown 18 putative ORFs based on the sequence alignment and characterization of genetically related fungal strains (10). Surprisingly, only nine genes in the BAC of M. pilosus BCRC38072 have revealed high homology to the genes involved in lovastatin biosynthetic gene cluster of A. terreus (Table 1). Moreover, the genomic arrangement of monacolin K biosynthetic genes in M. pilosus BCRC38072 has corresponded to the lovastatin biosynthetic genes in A. terreus (Figure 1A). The high homology between gene clusters of mok and lov implies that the mok gene cluster was responsible for monacolin K biosynthesis. To prove this, we disrupted the mokA gene-encoded polyketide synthase from wild-type M. pilosus BCRC38072. The phenotype of lost monacolin K productivity in the disruptant BCRC38135 indicates that the mokA gene was essential for monacolin K production (Figure 5C).

In particular, these genes also showed significant homology to genes identified in the compactin biosynthetic gene cluster of P. citrinum (18). However, the genomic arrangement of compactin biosynthetic genes was different from that of the monacolin K or lovastatin biosynthetic gene clusters (Figure 1A). The order and direction of P450 monoxygenase (mokC), polyketide synthase (mokA), oxidoreductase (mokD), dehydrogenase (mokE), and transesterase (mokF) was the same in M. pilosus, A. terreus, and P. citrinum, whereas the organization of other genes of the compactin biosynthetic gene cluster was different (10, 18). Furthermore, polyketide synthase (mokB), monoxygenase (mokC), oxidoreductase (mokD), dehydrogenase (mokE), and an efflux pump (mokF) appeared to have the same number of introns and similar intron positions among M. pilosus, A. terreus, and P. citrinum.

A. terreus and P. citrinum both belong to the family Trichocomaceae, but they are different from M. pilosus, which...
belongs to the family Monascaceae (19). Interestingly, lovastatin biosynthetic genes from \textit{A. terreus} revealed a higher homology with monacolin K biosynthetic genes from \textit{M. pilosus} than with compactin biosynthetic genes from \textit{P. citrinum}. Because polyketides play an ecological role in the environment regarding microbial competition, genetic differences might reflect extreme environmental stress and subsequent genetic changes in these species (20). In addition, many of the predicted PKSs in the PKS clade that produce highly reduced polyketides have divergent and presumably nonfunctional MeT domains (21). The structure of monacolin K differs from that of compactin, in which a methyl group derived from SAM is introduced at the C-6 position of the nonaketide-derived backbone (18). The MeT domain in \textit{mokA} and \textit{lovB} genes is assumed to be active instead of the \textit{mlcA} gene. The consensus motif of the MeT domain of \textit{mlcA} was found to be different from the relative PKSs (\textit{mokA} and \textit{lovB}) in some amino acid residues, A-L, G-I, QM-HL, and I-T (Figure 3B). Furthermore, there were two more introns located at the MeT

Figure 4. (A) Plasmid map of pMkAko for targeted gene disruption of \textit{mokA}. (B) Disruption of the \textit{mokA} gene in \textit{M. pilosus} BCRC38072. The strategy for disrupting \textit{mokA} gene was done by homologous recombination. A pMkAko vector containing the fusion protein of the hygromycin B resistance gene (HPH) with enhanced green fluorescent protein (EGFP) was flanked at the 5'-site by 2.0 kb (mokA2k) and at the 3'-site by 3.1 kb (mokA3k). pMkAko was linearized at the FspI sites and transformed into \textit{M. pilosus} BCRC38072. The homologous recombination event between the \textit{Monascus} genome and the \textit{FspI}-digested pMkAko fragment resulted in a truncated ORF for \textit{mokA} gene. (C) Southern hybridization analysis of disruption of \textit{mokA} gene in the genomes of \textit{M. pilosus} BCRC38072 (lane 1) and disruptant BCRC38135 (lane 2) hybridized with the probe indicated by a small black bar. The abbreviation F indicates an \textit{FspI} restriction enzyme site, and N indicates an \textit{Ndel} restriction enzyme site.
domain of mlcA, whereas mokA and lovB contained the same number of introns and similar intron positions. Therefore, the differences among amino acid residues could be the reason for the lack of MeT activity of mlcA in P. citrinum. These results could form the basis for the study of site-directed mutagenesis to understand the MeT activity of PKSs. Among these genes shown in Table 1, the transcription factor (mokH, lovE, and mlcR) and HMG-CoA reductase (mokG, lovA, and mlcD) were found to have fewer similarities to each other. The number and positions of introns were also different from one another. Nevertheless, the transcription factor and HMG-CoA reductase genes were assumed to be regulators responsible for up-regulation and down-regulation, respectively (7, 22). HMG-CoA reductase (mokG) could play a role in conferring resistance to monacolin K (22), and theoretically, there was no effect upon the structure of the polyketides (7).

The data for the monacolin K biosynthetic gene cluster can provide important information about the biosynthesis of monacolin K (lovastatin) between Monascus and Aspergillus. It is interesting that polyketide synthases between mok genes and lov genes are orthologues and also related in compactin biosynthesis by P. citrinum. Thus, there are three orthologous gene clusters in different fungi, which are useful to study evolution of genes for secondary metabolism. Moreover, this suggests that the structural variety of polyketides produced by fungi accompanies the enzymatic variation (7, 11). Studies on the regulation of fungal secondary metabolism and the development of novel polyketides have great potential for screening effective medications.

ACKNOWLEDGMENT

Support of the Ministry of Economic Affairs (Taiwan, ROC) (Grant 94-EC-17-A-17-R7-0563) to the Food Industry Research and Development Institute (FIRDI) is appreciated.

LITERATURE CITED


