Molecular Basis for Mycophenolic Acid Biosynthesis in *Penicillium brevicompactum*\(^\dagger\)

Torsten Bak Regueira,\(^\dagger\)‡ Kanchana Rueksomtawin Kildegaard,\(^1\) Bjarne Gram Hansen,\(^1\) Uffe H. Mortensen,\(^1\) Christian Hertweck,\(^3\) and Jens Nielsen\(^1,2^*\)

Department of Systems Biology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark; Department of Chemical and Biological Engineering, Systems Biology Group, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden; and Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans Knöll Institute, Beutenbergstr. 11a, 07745 Jena, Germany

Received 23 December 2010/Accepted 28 February 2011

Mycophenolic acid (MPA) is the active ingredient in the increasingly important immunosuppressive pharmaceuticals CellCept (Roche) and Myfortic (Novartis). Despite the long history of MPA, the molecular basis for its biosynthesis has remained enigmatic. Here we report the discovery of a polyketide synthase (PKS), MpaC, which we successfully characterized and identified as responsible for MPA production in *Penicillium brevicompactum*. *mpaC* resides in what most likely is a 25-kb gene cluster in the genome of *Penicillium brevicompactum*. The gene cluster was successfully localized by targeting putative resistance genes, in this case an additional copy of the gene encoding IMP dehydrogenase (IMPDH). We report the cloning, sequencing, and the functional characterization of the MPA biosynthesis gene cluster by deletion of the polyketide synthase gene *mpaC* of *P. brevicompactum* and bioinformatic analyses. As expected, the gene deletion completely abolished MPA production as well as production of several other metabolites derived from the MPA biosynthesis pathway of *P. brevicompactum*. Our work sets the stage for engineering the production of MPA and analogues through metabolic engineering.

Mycophenolic acid (MPA) is a fungal metabolite that was initially discovered by Bartolomeo Gosio in 1893 as an antibiotic against anthrax bacillus, *Bacillus anthracis* (see the excellent review by R. Bentley for details [4]). MPA has also been reported to possess antiviral (8, 17), antifungal (35), antibacterial (28, 48), and antipsoriasis (19) activities. Most importantly, it is being used as an immunosuppressant in kidney, heart, and liver transplantation patients and is marketed under the brands CellCept (mycophenolate mofetil; Roche) and Myfortic (mycophenolate sodium; Novartis). Mycophenolate, the active component in both drugs, inhibits IMP dehydrogenase (IMPDH). This enzyme is the rate-controlling enzyme in GMP biosynthesis (12, 47). The proliferation of B and T lymphocytes is inhibited in the presence of MPA, because these cell types rely entirely on the IMPDH dependent de novo pathway for purine biosynthesis. Unlike B and T lymphocytes, most other cell types express the IMPDH-independent salvage pathway, which allows purine production despite inhibition of IMPDH by MPA. This explains why MPA has found excellent use as an immunosuppressive pharmaceutical (2).

MPA is a meroterpenoid consisting of an acetate-derived phthalide nucleus and a terpene-derived side chain (6). The acetate origin of the phthalide identifies this part of the molecule as a polyketide, which refers to an enormously diverse group of bioactive compounds (16). Polyketide biosynthesis is catalyzed by polyketide synthases (PKSs), which are structurally and mechanistically closely related to fatty acid synthases (FASs) (16). Several different types of PKSs have been identified in nature, and among the fungal PKSs are large multifunctional enzymes with multiple active domains that are used iteratively during polyketide biosynthesis (3, 16).

MPA biosynthesis has been investigated extensively at the chemical level by using labeled substrates and by feeding cell cultures with analogues. This has provided the first insights into the reaction steps of MPA biosynthesis (5–7, 10, 13, 33, 44) and resulted in the model shown in Fig. 1.

The production of the phthalide moiety of MPA is initiated by the assembly of a nonreduced, methylated tetraketide (Fig. 1, MpaC). Hence, the MPA PKS as a minimum contains the starter unit acyl carrier protein transacylase (SAT), β-ketoacyl synthase (KS), acyltransferase (AT), product template (PT), acyl carrier protein (ACP), and methyltransferase (MT) domains. The fact that no reductive steps occur during biosynthesis indicates that the PKS likely does not contain any reductive domains. Moreover, the proposed MT domain could be involved in the methylation of the C-6 position, as shown in Fig. 1.

Gene clusters of biologically active metabolites typically harbor genes which confer the necessary tolerance to the compounds. As described, MPA is a known inhibitor of IMPDH, which converts IMP to XMP in the de novo pathway of GMP biosynthesis (47). This is an important reaction in almost all living organisms. Sequence analyses of IMPDH genes from different organisms show that this gene is highly conserved among different species.

\(^\dagger\) Corresponding author. Mailing address: Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. Phone: 46317723804. Fax: 46317723801. E-mail: nienjen@chalmers.se.

\(^\dagger\) Present address: Novozymes A/S, Krogshoejvej 36, DK-2880 Bagsvaerd, Denmark.

\(^\dagger\) Supplemental material for this article may be found at http://aem.asm.org/.

\(^\dagger\) Published ahead of print on 11 March 2011.
To advance the understanding of MPA biosynthesis, we set out to identify the gene cluster that is responsible for production of this important compound. Since only very few fungal PKSs have been characterized that produce methylated, nonreduced polyketides, it is difficult to isolate a gene encoding such a PKS simply by using DNA sequence information from close PKS homologues. We therefore took another approach, one which was based on the assumption that an organism is often resistant to the secondary metabolites it produces. For example, gene clusters responsible for production of lovastatin and compactin have been reported to contain homologues of the 3-hydroxy-3-methylglutaryl-coenzyme A (CoA) reductase, which is a target for these PKSs. In this way, the tolerance to these statins is increased. Similarly, we hypothesized that the MPA biosynthesis gene cluster may therefore contain a gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44). Using this rationale, we indeed identified a putative MPA biosynthesis gene encoding an IMPDH homologue (43, 44).

Materials and Methods

Strains and plasmids. *P. brevicompactum* IBT23078 was obtained from the strain collection at the Center for Microbial Biotechnology at the Technical University of Denmark and used as the source for genomic DNA and fungal transformation. Plasmid pAN7-1 (GenBank accession number Z32698.1) harboring the hygromycin resistance cassette under the control of the *gpdA* promoter from *Aspergillus nidulans* was used as a template for constructing the gene-targeting cassette. Manipulation of plasmid DNA and introduction of plasmid was performed by Amplicon
Express, Pullman, WA; PBBAC consisted of 3,072 clones with average inserts of 110 kb. pEBCBAC1 contains the chloramphenicol resistance marker.

Screening of PBBAC. Probing of PBBAC was performed using PCR products and RapidHyb hybridization buffer (Amersham-Pharma). The washing temperature was 65°C. Isolation of bacterial artificial chromosomes (BACs) from PBBAC was performed using Jetstar 2.0 (Genomed, Löhne, Germany).

Southern blot hybridization of BACs. In order to verify that correct BACs were identified and picked from PBBAC, Southern blots of enzymatically (HindIII and XbaI)-digested BACs were constructed using standard techniques and probed with the original probe. Approximately 2 µg of each BAC was digested and blotted onto a membrane as described by Radner (41). RapidHyb hybridization buffer (Amersham-Pharma) was used for probing, and the washing temperature was 65°C. The insert lengths of the isolated BACs were tested by pulsed-field gel electrophoresis.

Degenerate PCR. Degenerate PCR was used to obtain all probes for screening of PBBAC. The technique was used routinely to amplify fragments of PKS genes in a two-step procedure: first, an initial PCR with genomic DNA as template, followed by a second PCR with 2 µl of the product from the first reaction mixture as template. The amplifications were obtained using Taq polymerase (New England Biolabs) under the following standard conditions in 50-µl reaction volumes: 1× PCR buffer (10 mM Tris-Cl, 1.5 mM MgCl₂, 50 mM KCl, pH 8.0), 200 µM deoxynucleoside triphosphates, 2.5 U of enzyme, and approximately 100 ng of genomic DNA or 2 µl of the first PCR product as templates. Samples were run with hot start, an initial denaturation step of 2 min at 94°C, and the following cycle parameters (for 30 cycles): 94°C for 30 s, 60°C for 30 s (with a 0.5°C decrease in annealing temperature per cycle with genomic DNA as template, or at a fixed 50°C in the second-round PCR), and 72°C for 2 min.

Isolation of genomic DNA for PCR. Genomic DNA from P. brevicompactum was isolated by spooling high-molecular-weight DNA onto a sheep’s crook as described in standard laboratory manuals (41).

Sequencing and annotation of genes on BACs. The BACs were sequenced by Macrogen Inc., Seoul, South Korea. The annotation of sequenced BACs was carried out using several bioinformatic tools. Primarily, a Blastx similarity search located open reading frames (ORFs) with similarities to proteins submitted to GenBank, EMBL, DDBJ, and PDB. When the genes were identified, the exon-intron structures were determined by running analyses using GENSCAN (http://genes.mit.edu/GENSCAN.html) (9), HMHMGene (http://www.cbs.dtu.dk/services/HMMGene/) (29), NetAspGene 1.0 (http://www.cbs.dtu.dk/services/NetAspGene/) (50), and Genesplicer (http://www.cbcb.umd.edu/software/Genesplicer/gene_splt.html) (40). Blastx and Blastp similarity searches were performed using the online web interface at http://blast.ncbi.nlm.nih.gov/.

Bioinformatic characterization of predicted proteins: detection of transmembrane regions. All predicted protein sequences were submitted to the Phobius prediction server at http://phobius.cbg.ki.se/ (27), which determines transmembrane regions in the proteins. Signal peptides were predicted using SignalP3.0 (18) as well as PSORT II (http://wolfsort.org/).

Functional analysis of MpaC. (i) Oligonucleotides, PCR, and sequencing analysis. Degenerate PCR was performed to generate bipartite PCR fragments and to investigate the targeting pattern were synthesized by Sigma-Aldrich (United States) and were used for gene targeting. Specifically, the primer pairs used to detect the insertion of the hph gene and the other either in the upstream or in the downstream region used for gene targeting. Similarly, the presence of the wild-type mpaC locus was determined using primer pairs in which one of the primers was located in the mpaC gene and the other either in the upstream or in the downstream region used for gene targeting. Specifically, the primer pairs used to detect the insertion of the lphp cassette were BHGA687/Upst-HygR-N and Dwst-HygF-N/BGHA688. The primers used to detect the presence of mpaC were BHGA687/BHGA471 and BHGA1483/BHGA688.

Southern blotting was carried out using the Biotin chromogenic detection kit (Fermentas, Germany) and a biotin DecaLabel DNA labeling kit (Fermentas, Germany) following the manufacturer’s instructions. EcoRI-digested genomic DNAs from wild-type P. brevicompactum and the mpaC::3ph allele strain were used for Southern blotting. The probe was PCR amplified using primers BHGA696 and BHGA697.

Nucleotide sequence accession number. The nucleotide sequence of the MPA biosynthesis gene cluster can be found in the GenBank database (http://www.ncbi.nlm.nih.gov/GenBank/index.html) under accession number HQ731031.

RESULTS

Identification of the MPA biosynthesis gene cluster. A part of the IMPDH gene from P. brevicompactum was cloned using primers IMP_FW and IMP_RV, which were both designed based on conserved regions in five putative IMPDH gene sequences from the following fungi (Uniprot IDs shown in parentheses): Pneumocystis carinii (Q9UVLO), Aspergillus

...
fungi (O4WHZ9), Neurospora crassa (Q7SFX7), Kluyveromyces lactis (Q6CWA8), Candida glabrata (Q6FM59). A 1,115-bp PCR product was obtained and cloned into the pGEMT vector system, and subsequent sequence analyses showed that the amplified product was 92% identical in 202 amino acid residues to the putative IMPDH from A. fumigatus. The obtained PCR product was then used for screening a DNA library (referred to as PBBAC) of P. brevica\ldash pactum.

PBBAC was constructed with the aim of obtaining a DNA library with a statistically determined coverage of approximately 10-fold, calculated based on a genome size of approximately 30 Mbp. Other Penicillium spp. have genomes reported to range from 26 Mbp (P. mamefei) to 34 Mbp (P. chrysogenum). PBBAC was screened with the obtained PCR product of the IMPDH gene. As the PCR product was not subcloned prior to screening PBBAC, the probe may have contained two or more amplicons. The screening resulted in a total of 24 hybridization signals, which led us to suspect that two IMPDH homologues were present. To investigate whether the hybridizations really were the result of the presence of two IMPDH homologues, five BACs were randomly selected, digested with HindIII and XbaI, and used for a Southern blot assay and hybridized to the IMPDH probe. The results clearly showed that the BACs hybridized in two distinct patterns that grouped the BACs into two groups (data not shown): the first group (1-B12, 1-B16, and 1-H11) had hybridization signals at 1 kb, 3 kb, and >20 kb. The second group (1-C23 and 1-E13) shared hybridization signals at 2 kb and 5 kb. The only fungus known to have more than one IMPDH gene is the distantly related P. brevicompactum, like Aspergillus nidulans, Aspergillus terreus, Magnaporthe grisea, and N. crassa, only contain a single IMPDH gene (data not shown). This strongly suggested that P. brevicompactum, unlike closely related fungi, contains two genes encoding IMPDH. This was consistent with, but not direct evidence for, our hypothesis that one IMPDH gene resides in the MPA biosynthesis gene cluster.

As probing for MPA biosynthesis-related genes other than IMPDH had previously proven unsuccessful, it seemed that one way to identify the group of BACs most likely to harbor the MPA biosynthesis gene cluster was through process of elimination. If we could identify the group of BACs containing the IMPDH gene that does not reside in the MPA biosynthesis gene cluster, the other group of BACs would be candidates for carrying the MPA biosynthesis genes. There were several assumptions here: first, that it would be possible to identify the IMPDH gene based on the genetic context and synteny with related non-MPA-producing fungi; second, that the other IMPDH gene in P. brevicompactum did not just occur from an unrelated gene duplication event but rather resided in the MPA biosynthesis gene cluster. Based on these assumptions, the BACs analyzed above (1-B12, 1-C23, 1-E13, 1-B16, and 1-H11) were subjected to further analyses.

We investigated whether partial synteny in non-MPA-producing fungi could be detected in the regions surrounding IMPDH genes. Hence, genomic regions around IMPDH genes from several closely related fungal genomes were aligned, and a conserved gene encoding a “rasGTPase-activating protein” was located in all fungi at a distance between 1.5 and 4 kb from the IMPDH gene. It is uncertain whether this gene in fact is involved in GMP biosynthesis; however, as the alignment of this region in all cases showed the presence of this gene, it was assumed that if this gene was located on some of the BACs identified with the IMPDH probe, these BACs would most likely harbor the original IMPDH gene. Two degenerate primers (GTP_FW/RV) were designed based on conserved regions of the identified “rasGTPase-activating protein” sequences. Degenerate PCR was carried out with the hybridizing BACs 1-C23, 1-E13, 1-B12, 1-B16, and 1-H11 as templates. Three of these (1-B12, 1-B16, and 1-H11) resulted in a PCR product of the expected length (approximately 2 kb), whereas the remaining two (1-C23 and 1-E13) did not (data not shown). These findings corresponded exactly to the grouping observed in the Southern blot results, and sequencing of the PCR product verified that the amplified region was 78% identical on the amino acid level to the rasGTPase-activating proteins of A. terreus (GenBank accession number XP_001218147) and Aspergillus oryzae (GenBank accession number BA62830).

These results showed that the BACs 1-C23 and 1-E13 harbored an IMPDH gene situated in a genetic context different from the IMPDH gene present in most fungi. After verifying the length of the DNA inserts of BACs 1-C23 and 1-E13 to >100 kb each, BAC 1-C23 was sequenced and annotated. This revealed a putative IMPDH gene, mpaF (Fig. 2), flanked by the polyketide biosynthesis genes mpaD, mpaE, mpaG, and...
Oxidative cleavage. However, the PKS as well as other genes necessary for MPA production were missing on the BAC. Subsequent BAC walking resulted in the identification of several overlapping BACs, of which one, 1-E13, turned out to have an insert that contained the missing region. Sequencing and sequence analyses of both 1-C23 and 1-E13 revealed the putative gene pattern depicted in Fig. 2.

The DNA region upstream of mpaA has no similarity to any known proteins. Furthermore, the nearest identified ORFs upstream of mpaA have similarities to ribosome biogenesis protein Rsa4 (ORF30), DNA mismatch repair protein Msh4 from Coccidioides posadasii (ORF29), and a C6H2Zn finger-carrying protein (ORF28) with weak similarity to ubiquitin ligase E3 protein, which may also be involved in DNA repair. The nearest putative genes downstream of mpaH are a DAHP synthase homologue (88% identical; ORF39), two hypothetical proteins (ORF40 and ORF41) with no characterized homologues, and a glycoside hydrolase (57% identical; ORF42).

Hence, neither upstream nor downstream of the proposed MPA biosynthesis gene cluster are the putative genes predicted to be involved in MPA biosynthesis. Upstream, the gene products are predicted to be involved in certain DNA binding or repair reactions, whereas downstream, the predicted gene products may be involved in central metabolism, such as aromatic amino acid biosynthesis and sugar degradation reactions. Hence, the borders of the MPA biosynthesis gene cluster are predicted to be mpaA and mpaH.

**Bioinformatic characterization of the MPA gene cluster.**

The annotation of the sequenced BACs revealed a 25-kb DNA region containing several predicted ORFs similar to polyketide biosynthetic genes. The region is flanked by 4-kb and 7-kb DNA regions with no predicted genes, and a Blastx analysis confirmed that these regions only have very low similarities to genes in the GenBank database. A Blastx analysis of the gene cluster identified eight regions with high similarities to genes in the nonredundant (nr) GenBank database, and these are denoted mpaA to mpaH in Fig. 2. A GENSCAN analysis detected all eight regions as ORFs, but the GENSCAN algorithm combined several of these regions, resulting in a total of only five predicted genes in the gene cluster. Submission to the hidden Markov model genefinder, HMMGene, and to the exon-intron probability calculator NetAspGene 1.0, resulted in the eight genes listed in Table 1. For one gene (mpaH), the algorithm Genesplicer predicted an additional intron which was not detected by any of the other algorithms. Discrepancies between the predictions were expected, as the algorithms were developed for predictions in organisms as different as Aspergillus (NetAspGene 1.0), Caenorhabditis elegans (HMMGene), Arabidopsis thaliana (GENSCAN), and humans (Genesplicer).

In order to propose functions of the putative MPA genes, they were subjected to bioinformatic analyses such as Blastx, Blastp, and signal protein and transmembrane predictions, as well as functional domain predictions using CDD (20). A brief overview of these results is presented in Table 1.

<table>
<thead>
<tr>
<th>Putative activity</th>
<th>Size (aa)</th>
<th>Predicted domains and features</th>
<th>Closest characterized homologue</th>
<th>% similarity (no. of aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prenyltransferase</td>
<td>315</td>
<td>Seven transmembrane helices; CDD UbA prenyltransferase family</td>
<td>Neosartorya fischeri (XP_001261885)</td>
<td>42 (322)</td>
</tr>
<tr>
<td>Unknown function</td>
<td>392</td>
<td>Type III reverse signal membrane anchor; CDD none</td>
<td>Putative dephospho-CoA kinase (ZP_01083610)</td>
<td>28 (202)</td>
</tr>
<tr>
<td>Polyketide synthase</td>
<td>2448</td>
<td>SAT domain; CDD KS, AT, PP, MT, esterase</td>
<td>Citrinin PKS (dbj BAD44749.1)</td>
<td>33 (2,189)</td>
</tr>
<tr>
<td>P450 monoxygenase</td>
<td>396</td>
<td>Possible membrane anchor</td>
<td>Pisetin demethylase (P450) (gb AAC01762.1)</td>
<td>35 (346)</td>
</tr>
<tr>
<td>Zn-dependent hydrolase</td>
<td>261</td>
<td>CDD metallo-beta-lactamase superfamilly II</td>
<td>AhID (Zn-dependent hydrolase) (gb AAP57766.1)</td>
<td>32 (84)</td>
</tr>
<tr>
<td>IMP dehydrogenase (MPA tolerance)</td>
<td>527</td>
<td>CDD IMPDH, TIM phosphate binding</td>
<td>Candida albicans (gb AAW65380.1)</td>
<td>62 (524)</td>
</tr>
<tr>
<td>O-Methyltransferase</td>
<td>398</td>
<td>CDD O-MT; SAM-binding motif and catalytic residues</td>
<td>Hypocrea viros (gb ABE60721.1)</td>
<td>32 (375)</td>
</tr>
<tr>
<td>Oxidative cleavage</td>
<td>421</td>
<td>CDD M-factor, weak similarity to α/β-hydrolase fold 1, C-terminal peroxisomal targeting signal (GKL)</td>
<td>Akt2, AK toxin synthesis (dbj BAA36589.1)</td>
<td>20 (255)</td>
</tr>
</tbody>
</table>

aa, amino acids. Only characterized sequences (as of December 2010) are included as closest characterized homologues here, as sequences with no characterization add no extra information about the function of the MPA biosynthesis genes. dbj, DNA Data Bank of Japan; gb, GenBank.
bers with the following domain structure: KS, AT, PP (phosphopantetheine attachment site), optional PP, MT, and an optional cyclase domain. Kroken et al. identified only putative genes with this domain structure in *Botryotinia fuckeliana*, *Gibberela zeae*, and *Cochliobolus heterostrophus*. Since then, the citrinin PksCT has been described, which has a similar domain structure except for the terminal thioesterase domain (46). The determined domain structure of MpaC is presented in Fig. 3.

When comparing the sequence obtained after subcloning the IMPDH DNA probe (impA) with *mpaF* from the MPA gene cluster, several differences were observed over the entire length of the probe. In order to analyze the phylogenetic relationship between MpaF and its homologues, an alignment of MpaF and ImpA as well as of several IMPDH enzymes from filamentous fungi and two other eukaryotes was carried out. The result was used to calculate distances using the nearest-neighbor method, and the result is the unrooted phylogenetic tree presented in Fig. 4. Interestingly, the similarity between MpaF and ImpA was lower than between MpaF and IMPDH from *P. chrysogenum*. We also compared MpaF with Chinese hamster (*Cricetulus griseus*) type 2 IMPDH, as this IMPDH has been analyzed in detail with regard to residues that are essential for activity. We found that residues that have been found essential for activity are conserved in MpaF, which points toward MpaF as indeed a functional IMPDH.

**Functional analysis of MpaC.** To functionally characterize the identified gene cluster, a gene deletion strategy for *mpaC* was undertaken. In order to construct the *mpaC* deletion strain, the bipartite gene-targeting method (38) was used based on the fungal selectable marker *hph*, which confers hygromycin resistance to the host. A total of 20 independent transformants were picked randomly, streak purified, and subjected to PCR analysis in order to investigate the gene-targeting events. PCR tests were performed by using primer pairs in which one of the primers was located within the hygromycin cassette (*hph*) and the other was located either upstream or downstream of the homologous region. Of 20 transformants, 9 transformants were found to have the correct *mpaC* deletion, based on the expected PCR fragment length. The identified deletion mutants were MPA1.1, -1.3, -1.8, -2.3 to -2.7, and 2.9. The remaining 11 transformants were most likely due to nonhomologous random integration. However, a PCR test of the presence of *mpaC* in the nine deletion mutants simultaneously showed that *mpaC* was still present in all nine transformants. This led us to believe that the strains were not adequately purified. Hence, from all nine transformants, in total 198 single spores were started on YES plates, and 49 germinating spores were transferred to new YES plates as three point stabs and subsequently analyzed for MPA production by thin-layer chromatography. MPA was produced in all strains but one (strain 2.3:1). Liquid chromatography-mass spectrometry (LC/MS) analysis of extracts from the wild type and mutant strain 2.3:1 verified that the mutant strain showed no evidence of MPA production. The purified

![FIG. 3. Map of the gene encoding MpaC. The Conserved Domain Database (CDD) at the National Center for Biotechnology Information (NCBI) was used to predict the PKS domains. SAT, starter unit acyl carrier protein transacylase; KS, β-ketoacyl synthase; AT, acyltransferase; PP, phosphopantetheine attachment site; MT, methyltransferase; esterase, esterase domain, similar to Aes of *E. coli* (31a). Gaps indicate predicted introns. The scale indicates the number of nucleotides.](http://aem.asm.org/)

![FIG. 4. Unrooted phylogenetic tree based on alignment of several IMPDH proteins. The tree was bootstraped 100 times, and values (as percentages) are indicated at each branch point. ImpA_Pbr, IMPDH of *P. brevicompactum* not located in the MPA biosynthesis gene cluster; Pchr, *Penicillium chrysogenum*; Pbr, *Penicillium brevicompactum*; Pma, *Penicillium marneffei*; Aor, *Aspergillus oryzae*; Afu, *Aspergillus fumigatus*; Ate, *Aspergillus terreus*; Ncr, *Neurospora crassa*; Kla, *Kluyveromyces lactis*; Cgl, *Candida glabrata*.](http://aem.asm.org/)
strain contained a clean \textit{mpaC::hph} allele, as demonstrated by four diagnostic PCRs as well as by Southern blotting (data not shown). Sequencing of the PCR fragments of four of the originally isolated \textit{mpaC} deletion strains as well as the repurified deletion strain confirmed that \textit{mpaC} had been deleted as intended.

LC/MS analyses of extracts from the wild type and the \textit{mpaC::hph} mutant strain verified that the mutant strain lost its ability to produce MPA (Fig. 5 and 1; see also Fig. S1 in the supplemental material). In addition to MPA, the MPA-related compounds MPA diol lactone, 4-\textit{O}-mycophenolate, and deacetylpebrolide are absent in the \textit{mpaC::hph} strain (Fig. 5). Reference standards of MPA were coanalyzed. Additional target compounds were identified based on the UV spectrum, retention time, accurate masses, and relative intensities of [M+H] and [M+Na+]. Extracted ion chromatograms for all target peaks were analyzed for all extracts in order to exclude that MPA, MPA diol lactone, 4-\textit{O}-mycophenolate, or deacetylpebrolide was produced by the \textit{mpaC} deletion mutant (data not shown).


discussion

The isolation of novel polyketide biosynthesis gene clusters is important for the continued progress in understanding the biosynthetic mechanisms of one of today’s industrially most important bioactive group of compounds. The size of the polyketide biosynthesis gene cluster varies for different polyketides. For example, among the reported biosynthetic gene clusters are the compactin biosynthesis gene cluster from \textit{Penicillium citrinum} (38 kb; AB072893), aurofusarin from \textit{Gibberella zeae} (25 kb) (21), aflatoxin from \textit{Aspergillus parasiticus} (82 kb; AY371490), bikaverin from \textit{Gibberella fujikuroi} (15 kb) (51), and zearalenone from \textit{Gibberella zeae} (15 kb) (23); several others have also been reported (11, 32). The biosynthesis of these polyketides is dependent on PKSs, transcription factors, transporters, and modifying enzymes encoded by genes residing in the clusters. Hence, in the hunt for the MPA biosynthesis gene cluster, identification of any of the involved genes in the cluster followed by sequencing of up- and downstream regions from this gene would result in the identification of the complete gene cluster.

The identification of an IMPDH homologue located in the proposed MPA biosynthesis gene cluster enabled us to sequence the gene locus. In addition, involvement of \textit{mpaC}, the gene residing in the proposed gene cluster (\textit{mpaA} to \textit{mpaH}) in the biosynthesis of MPA, was demonstrated by deleting \textit{mpaC}. The \textit{mpaC} deletion mutant lost the ability to produce MPA and several MPA-related compounds, such as MPA diol lactone, 4-\textit{O}-mycophenolate, and deacetylpebrolide. The deletion was confirmed with PCR and Southern blotting.

Notably, \textit{MpaC} belongs to the group of the few described, nonreducing, fungal PKSs with a functional MT domain. In addition, \textit{MpaC} contains a starter unit acyl carrier protein transacylase (SAT) domain, which could transfer the starter acetyl-CoA unit to the ACP domain (15). There is a predicted cyclase/thioesterase domain at the C-terminal end of the protein, which could catalyze the cyclization and release of the polyketide from the PKS. Thioesterases, which belong to the
same family of proteins, have previously been reported to be involved in chain length determination, cyclization, and lactonization (22). Recently, the long-sought-after PKS required for the biosynthesis of orsellinic acid was discovered (45). Although orsellinic acid and 5-methylorsellinic acid are structurally very similar, the respective amino acid sequences of the synthases orxA (AN7909) and mpaf are only 29% identical globally and 40% identical based on the KS domains. This is in part due to the fact that the orsellinic acid synthase lacks the MT domain present in MpaC.

For lactonization to occur for 5-methylorsellinic acid, the C-3 methyl group must be oxidized to the alcohol, which is a reaction often catalyzed by P450 monooxygenases. Of the MPA biosynthesis enzymes, only MpaD has similarity to a P450 monooxygenase, and thus MpaD is predicted to catalyze this reaction. Raistrick et al. observed that the 3,5-dihydroxyphalic acid was produced by P. brevicaffectum, which is probably derived from orsellinic acid (42). As the structures of 5-methylorsellinic acid and orsellinic acid are almost identical, oxidations of the C-6 methyl group of these two compounds could both be catalyzed by MpaD.

Another enzyme that strongly links the gene cluster to MPA biosynthesis is the predicted prenyltransferase MpaA. The putative prenyltransferase could catalyze prenylation of the phthalide intermediate (Fig. 1) and is predicted to be membrane bound, with seven transmembrane hydrophobic regions.

The step following prenylation in MPA biosynthesis is an oxidative cleavage of the central double bond of the farnesyl chain. The mechanism has been reported to include epoxidation of the double bond, followed by hydrolysis, rearrangement to a ketone, α-hydroxylation, and Woodward reaction of the α-hydroxyketone with C-C bond cleavage and formation of the acid (13). It is uncertain which enzyme in the gene cluster could catalyze this reaction, but MpaH has similarity to Akt2, which is involved in AK toxin biosynthesis in Alternaria alternata. It was recently shown that Akt1, Akt2, and Akt3 are localized in peroxisomes, which are involved in fatty acid β-oxidation (25). A PSORT II analysis (34) of the amino acid sequence of MpaH identified a PTS1-like tripeptide GKL at the C-terminal end of this protein, suggesting that MpaH may be located in the peroxisomes. Hence, MpaH could have a similar oxidative role in MPA biosynthesis and, with localization in the peroxisomes, cleave the farnesyl chain.

Demethylmycophenolic acid is converted to MPA by methylation of the 5-hydroxyl group, which could be catalyzed by MpaG, the only O-methyltransferase encoded in the MPA gene cluster (Table 1).

No genes for transcription factors, like MlcR, which is found in the compactin gene cluster (1), could be identified within the MPA biosynthesis gene cluster. Initial studies of MPA production have demonstrated that MPA is produced during the early growth phase on several types of complex media (data not shown), and not only as a response to poor nutrient supply during the stationary phase, when most other secondary metabolites are produced (39). Thus, the question is if there are any conditions under which the strain does not produce MPA, or if the MPA gene cluster is constitutively expressed in P. brevicaffectum.

The bioinformatic analysis of the MPA gene cluster does not readily assign catalytic functions to the deduced proteins MpaB, MpaE, and MpaH. Future functional analyses will elucidate the functions of these proteins in MPA biosynthesis.

The discovery of the MPA gene cluster has provided new insights into the biosynthesis of MPA, a very important immunosuppressive drug of today. In summary, the performed experiments show that mpaC is a key gene involved in biosynthesis of MPA by P. brevicaffectum, and our results set the groundwork for future experiments to elucidate the remaining intriguing biosynthetic steps in MPA biosynthesis. The accompanying manuscript by Hansen et al. (24) conclusively shows that the polyketide synthase MpaC catalyzes the production of 5-methylorsellinic acid. A thorough study of several of the catalytic steps in MPA biosynthesis, like the prenyl transfer, heterocyclization, and chain cleavage steps, may result in a deeper understanding of the biosynthesis of this valuable immunosuppressant. Furthermore, construction of higher-yielding mutants would have immediate applicability for the pharmaceutical industry. The knowledge gained through this and future work may lead to the construction of mutants capable of producing interesting new phthalide analogues with fine-tuned biological activities and may have great potential as pharmaceutical drugs.

ACKNOWLEDGMENTS

Part of this work was financed by the Danish Research Council for Technology and Production. We also acknowledge the Chalmers Foundation for financial support. Furthermore, the work by B. G. Hansen and U. H. Mortensen was supported by grant numbers 09-064240 and 09-064967 from the Danish Council for Independent Research, Technology and Product Sciences.

We thank Jens C. Frisvad (Technical University of Denmark) for his insightful help with the MPA analyses. We also thank Martin Engelhard Kornholt for his valuable technical assistance in the laboratory.

REFERENCES


3042 REGUEIRA ET AL. APPL. ENVIRON. MICROBIOL.