Massetolide A Biosynthesis in *Pseudomonas fluorescens*[∇]

I. de Bruijn,¹ M. J. D. de Kock,¹ P. de Waard,² T. A. van Beek,³ and J. M. Raaijmakers^{1*}

Laboratory of Phytopathology, Wageningen University, Wageningen, The Netherlands¹; Wageningen NMR Centre, Wageningen University, Wageningen, The Netherlands²; and Natural Products Chemistry Group, Laboratory of Organic Chemistry, Wageningen University, Wageningen, The Netherlands³

Received 28 September 2007/Accepted 29 October 2007

Massetolide A is a cyclic lipopeptide (CLP) antibiotic produced by various *Pseudomonas* strains from diverse environments. Cloning, sequencing, site-directed mutagenesis, and complementation showed that massetolide A biosynthesis in *P. fluorescens* SS101 is governed by three nonribosomal peptide synthetase (NRPS) genes, designated *massA*, *massB*, and *massC*, spanning approximately 30 kb. Prediction of the nature and configuration of the amino acids by *in silico* analysis of adenylation and condensation domains of the NRPSs was consistent with the chemically determined structure of the peptide moiety of massetolide A. Structural analysis of massetolide A derivatives produced by SS101 indicated that most of the variations in the peptide moiety occur at amino acid positions 4 and 9. Regions flanking the *mass* genes contained several genes found in other *Pseudomonas* CLP biosynthesis clusters, which encode LuxR-type transcriptional regulators, ABC transporters, and an RND-like outer membrane protein. In contrast to most *Pseudomonas* CLP gene clusters known to date, the *mass* genes are not physically linked but are organized in two separate clusters, with *massA* disconnected from *massB* and *massC*. Quantitative real-time PCR analysis indicated that transcription of *massC* is strongly reduced when *massB* is mutated, suggesting that these two genes function in an operon, whereas transcription of *massA* is independent of *massBC* and vice versa. Massetolide A is produced in the early exponential growth phase, and biosynthesis appears not to be regulated by *N*-acylhomoserine lactone-based quorum sensing. Massetolide A production is essential in swarming motility of *P. fluorescens* SS101 and plays an important role in biofilm formation.

The cyclic lipopeptide (CLP) surfactant massetolide A consists of a 9-amino-acid cyclic oligopeptide linked to 3-hydroxydecanoic acid and was first identified in cultures of a marine *Pseudomonas* sp. isolated from the surface of a leafy red alga collected in Masset Inlet, British Columbia, Canada (15). Massetolide A was subsequently identified in *Pseudomonas fluorescens* SS101, a biocontrol strain isolated from the wheat rhizosphere (10), and later described for *Pseudomonas* sp. strain MF-30, a strain that inhibits the growth of the fungal pathogens *Fusarium oxysporum* and *Drechslera teres* (26). Like several other CLPs produced by *Pseudomonas* and *Bacillus* species (35, 43), massetolide A has potent surfactant and broad-spectrum antimicrobial activities: it inhibits the growth of *Mycobacterium tuberculosis* and *Mycobacterium avium-intracellulare* (15) and has destructive effects on zoospores of multiple *Oomycete* plant pathogens, including *Pythium* and *Phytophthora* species (8, 10). Massetolide A is an important determinant of the activity of *P. fluorescens* SS101 against *Phytophthora infestans*, the causal agent of late blight disease of tomato, and contributes to the rhizosphere competence of strain SS101 (55). The activity of massetolide A against the late blight pathogen was attributed, at least in part, to its zoosporicidal activity and to the induction of a systemic resistance response in tomato plants (55). In spite of its potential for the control of plant and human pathogenic microorganisms, little

is known to date about the genes involved in massetolide A biosynthesis in *Pseudomonas*.

Biosynthesis of CLPs is generally governed by multifunctional nonribosomal peptide synthetases (NRPS) (43). NRPS consist of several modules, each having a specific function in the biosynthesis of CLPs and other peptide antibiotics (14, 16, 51). The number of NRPS modules is in most cases consistent with the number of amino acids in the peptide moiety (“colinearity rule”). The modules can be further subdivided into initiation and elongation modules. Initiation modules typically consist of an adenylation (A) domain, responsible for amino acid selection and activation, and a thiolation (T) domain, responsible for thioesterification of the activated amino acid (13, 14). For CLP biosynthesis, however, the initiation module also contains a condensation (C) domain, which is postulated to catalyze N acylation of the first amino acid in the peptide chain (27, 46). Elongation modules contain A, T, and C domains, in which the C domain is responsible for peptide bond formation between two neighboring substrates to elongate the peptide chain. Collectively, these domains generate a linear lipopeptide which is cleaved at the end of the assembly line by a thioesterase (TE) domain, resulting in the release of a linear product or a cyclic molecule via an intramolecular cyclization reaction (5, 14, 25, 47, 51).

In this study, we describe the identification and characterization of the massetolide A biosynthesis genes from *P. fluorescens* SS101. Transposon mutagenesis, bacterial artificial chromosome (BAC) cloning, sequence analyses, site-directed mutagenesis, and complementation revealed that massetolide A biosynthesis is governed by three large NRPS genes, designated *massA*, *massB*, and *massC*. Sequence analysis of the

* Corresponding author. Mailing address: Laboratory of Phytopathology, Binnenhaven 5, 6709 PD Wageningen, The Netherlands. Phone: 31 317 483427. Fax: 31 317 483412. E-mail: jos.raaijmakers@wur.nl.

[∇] Published ahead of print on 9 November 2007.

regions flanking the *mass* genes was performed to identify the presence of genes conserved in other CLP biosynthesis clusters. Quantitative real-time PCR (Q-PCR) analysis was performed to investigate the expression of each of the three *mass* genes. The dynamics of massetolide A production by *P. fluorescens* SS101, the identity of massetolide A derivatives produced by strain SS101, and the role of massetolide A in surface motility and biofilm formation are presented.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *P. fluorescens* SS101 was grown on *Pseudomonas* agar F (Difco) plates or in liquid King's medium B (KB) at 25°C. The transposon mutants were obtained as described by De Souza et al. (10), and plasposon mutants were obtained with plasmid pTnModOKm (9). *Escherichia coli* strain DH5 α and EPI3000 were used as hosts for the plasmids for site-directed mutagenesis, complementation, and construction of the BAC library. *E. coli* strains were grown on Luria-Bertani (LB) plates or in LB broth amended with the appropriate antibiotics.

Site-directed mutagenesis. Site-directed mutagenesis of the *mass* genes was performed with the pKnockout-G suicide vector (57). Fragments of the genes of interest were amplified by PCR with the primers 5'-CATTCTGGCGTTGGCTGG-3' (*massA* forward primer), 5'-TGCAGCATTCTCCAGCC TG-3' (*massA* reverse primer), 5'-AAATTCACGGGCGCTGGCAT-3' (*massB* forward primer), 5'-ACATGCCTCGTTGTCCTGG-3' (*massB* reverse primer), 5'-TCCTGGCGTTGATGGAAGG-3' (*massC* forward primer), and 5'-AACGACAGGTGCAACTTGGC-3' (*massC* reverse primer) and first cloned into pGEM-T Easy vector Systems I (Promega) according to the manufacturer's instructions. Inserts were subcloned by *Apa*I/*Sac*I digestions into pKnockout-G and transferred into SS101 by triparental mating with helper strain *E. coli* HB101 carrying plasmid pRK2013. Transformants were selected on KB agar plates supplemented with rifampin (100 μ g/ml) and gentamicin (75 μ g/ml). Integrations in the target gene were verified by PCR using one primer specific for the insert and one primer specific for the gene fragment flanking the pKnockout insertion site.

Construction of BAC library. *P. fluorescens* SS101 cells from a 25-ml overnight culture grown at 25°C were washed twice with sterile demineralized water and embedded in 1% low-melting-point agarose (Invitrogen) dissolved in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0). The cell-agarose mixture was taken up into a 1-ml syringe and cooled down to 4°C to solidify. The cell-agarose worm was extruded from the syringe and incubated with 10 ml of lysis buffer (10 mM Tris, 50 mM NaCl, 0.2 M EDTA, 1% Sarkosyl, 0.2% sodium deoxycholate, 1 mg/ml lysozyme, pH 8.0) with gentle agitation for 3 h at 37°C. The agarose worms were subsequently transferred to 40 ml of 1% Sarkosyl and proteinase K (1 mg/ml) and incubated with gentle agitation for 16 h at 55°C. After refreshing of the buffer and incubation for 1 hour, agarose worms were washed three times with 50 ml T₁₀E₁ buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and subsequently incubated for 1 hour at room temperature in T₁₀E₁ buffer supplemented with 1 mM phenylmethylsulfonyl fluoride. The agarose worms were washed three times with T₁₀E₁ buffer and incubated overnight in storage buffer (10 mM Tris, 50 mM EDTA, pH 8.0). Pulse-field gel electrophoresis was performed at 6 V/cm for 16 h with a 5- to 15-s switch time at a 120° angle to analyze DNA yield and quality. Plugs were preincubated with *Sau*3AI digestion buffer (New England Biolabs) and subsequently treated with *Sau*3AI. Partially digested DNA was fractionated by pulse-field gel electrophoresis, and fragments of 50 to 250 kb were isolated from agarose by Gelase (Epicenter) treatment according to manufacturer's instructions; 225 ng of DNA fragments was ligated to 25 ng of *Bam*HI-digested and dephosphorylated pCC1BAC vector DNA (Epicenter) according to the supplier's protocol and electroporated into *E. coli* EPI3000 cells (Epicenter). Cells were plated on LB with chloramphenicol (12.5 μ g/ml).

Identification and sequencing of the *mass* genes. Library clones were blotted onto Hybond N+ membranes (Amersham) and hybridized with ³²P-labeled probes amplified by PCR with the same primers as described for the site-directed mutagenesis (see above). Hybridization was performed overnight, and membranes were washed at 65°C with 0.5 $\times$ SSC (75 mM NaCl, 7.5 mM sodium citrate)-0.1% sodium dodecyl sulfate. Hybridization-positive clones were subjected to detailed restriction digestion and hybridization analysis using *Pst*I, *Eco*RV, and previously described probes. Contigs were constructed by cluster analysis of these experimental data by the unweighted-pair group method using average linkages. Clones 2H12, containing *massA*, and 7B4, containing *massB*

and *massC*, were sent for shotgun sequencing (Macrogen, Seoul, Korea). Sequence gaps were closed by primer walking and by sequencing the PCR products overlapping the gaps. Bacterial operons and genes were subsequently predicted by the Softberry FGENESB program (Softberry, Inc., Mount Kisco, NY), and the identified open reading frames (ORFs) were analyzed using Blastx in the NCBI database and Pseudodb (http://xbase.bham.ac.uk/pseudodb/). Putative promoter sequences were identified by the Softberry BPROM program, and putative terminator sequences were identified by the RNA secondary structure prediction program of Genebee (http://www.genebee.msu.su/). Specific domains in the deduced protein sequences of the *mass* genes were analyzed with PFAM (http://pfam.sanger.ac.uk/search?tab=searchSequenceBlock). Protein sequences of specific domains were aligned in ClustalX (version 1.81). Trees were inferred by neighbor joining using 1,000 bootstrap replicates. Identification of the flanking genes of *massA*, *massB*, and *massC* was performed by Blastx analysis in NCBI, Pseudomonas.com (http://v2.pseudomonas.com/), or Pseudodb and by comparison with genes flanking the known CLP biosynthesis clusters for syringomycin, syringopeptin, viscosin, orfamide, and arthofactin.

Construction of pME6031-based vectors for complementation. A 7.7-kb fragment containing the *massA* gene, including the promoter and terminator, was obtained by PCR (forward primer, CAGACAAATCCTTCTTACC; reverse primer, GCGAGCTGCTGGATAACCCA) with Phusion DNA polymerase (Finnzymes). This PCR fragment was subcloned in pGEM-T Easy vector systems I (Promega) according to the manufacturer's instructions, and the obtained plasmid was digested with *Eco*RI. Restriction analysis of the BAC clone containing *massB* and *massC* revealed unique restriction sites for *Bam*HI. Digestion with this enzyme resulted in a fragment of approximately 30 kb containing *massB*, *massC*, and homologs of *macA* and *macB*. The fragments containing the *mass* genes were obtained by excising the fragments from gel and isolating the DNA with the NucleoTrap kit (Macherey-Nagel). These fragments were cloned into the shuttle vector pME6031 (19), which was digested, dephosphorylated (shrimp alkaline phosphatase; Promega), and purified with the NucleoTrap kit according to the manufacturer's instructions. *E. coli* DH5 α was transformed with the obtained plasmids pME6031-*massA* and pME6031-*massBC* by heat shock transformation (20), and transformed colonies were selected on LB agar plates supplemented with tetracycline (25 μ g/ml). Integration of the inserts was verified by PCR analysis and restriction analysis of isolated plasmids. The correct pME6031-*massA* and pME6031-*massBC* constructs were subsequently electroporated into the massetolide-deficient Δ *massA*, Δ *massB*, and Δ *massC* transposon mutants. Electrocompetent cells were obtained by washing the cells three times with 1 mM MOPS (morpholinepropanesulfonic acid) and 15% glycerol from a 5-ml overnight culture and finally dissolving the cells in 100 μ l of the washing buffer. Cells were stored at -80°C for at least 1 hour prior to electroporation. Electroporation occurred at 2.4 kV and 200 μ F, and after incubation in SOC medium (2% Bacto tryptone [Difco], 0.5% Bacto yeast extract [Difco], 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose [pH 7]) for 2 h at 25°C, cells were plated on KB supplemented with tetracycline (25 μ g/ml). Verification of transformation was performed by PCR analysis using one primer specific for the insert and one primer specific for the pME6031 vector. Massetolide A production in the complemented mutants was tested with a drop collapse assay followed by high-pressure liquid chromatography (HPLC) analysis and liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Chemical identification of massetolide A derivatives produced by *P. fluorescens* SS101. Analytical HPLC separations were carried out on Alltech end-capped 5- μ m C₁₈ columns 250 mm in length and 4.6 mm (UV [210 nm] detection and evaporative light-scattering detection) or 2.1 mm (LC-MS) in diameter at a flow rate of 1.2 or 0.20 ml/min, respectively. For the separations on the 4.6-mm column, a Midas autosampler (Spark), two Spectroflow 400 pumps (Kratos), a Spectroflow UV detector 785 (Applied Biosystems), a Sedex 55 evaporative light-scattering detector (Sedere) and Dionex Chromeleon software were used. Preparative separations were performed with an Alltech end-capped 5- μ m C₁₈ column (250 by 22 mm) at 23 ml/min with methyl cyanide-methanol-H₂O (3,800:1,925:4,275) containing 0.1% trifluoroacetic acid as the solvent on a Shimadzu autopreparative system. Two hundred milligrams of crude surfactant extract of strain SS101, obtained as described by De Souza et al. (10), was dissolved in 6 ml dimethyl sulfoxide. After membrane filtration, five injections of 1,150 μ l each were carried out. The fractions eluting at 47.2, 51.3, 57.0, 62.9, 65.0, 68.7, 79.4, and 91.3 min were collected on the basis of the UV signal in round-bottom flasks, and the eluent was removed with a rotary evaporator (Büchi) in vacuo. All fractions were investigated by means of analytical HPLC and infusion (+)-electrospray ionization (ESI)-MS/MS. The fractions at 57.0 and 65.0 min were shown to contain a second, minor compound. Sufficient amounts of the fractions at 62.9, 68.7, 79.4, and 91.3 min were collected to perform two-dimensional nuclear magnetic resonance (NMR) experiments. The MS system consisted of a

Finnigan LCQ ion trap mass spectrometer equipped with a Finnigan ESI interface. Data were processed with the Finnigan Xcalibur software system (ThermoQuest, Breda, The Netherlands). For off-line MS studies, all peaks of the preparative separation were introduced by continuous infusion using a syringe pump (Hamilton, NV) at a flow rate of 5 μ l/min. Spectra were recorded in positive mode over a period of 2 min and averaged. The scan range was m/z 310 to 1200 at a scan rate of 0.20 s. For the MS/MS experiments, helium was used as the collision gas. Only a single parent ion was kept in resonance (isolation width m/z 3); all other ions were ejected from the trap without mass analysis. The ion was then agitated and allowed to fragment by collision-induced dissociation. A collision energy of 22% was used to give >90% yield of fragmentation. For NMR, fractions were dissolved in CD₃OD (99.9 atom% D; Acros) and transferred to a standard 5-mm NMR tube. NMR spectra were recorded at a probe temperature of 25°C on a Bruker DPX-400 (1H, correlation [COSY], total correlation [TOCSY], nuclear Overhauser effect [NOESY], ¹³C, distortionless enhancement by polarization transfer [DEPT], and heteronuclear multiple bond correlation [HMBC] spectroscopy) or a Bruker AMX 500 (1H, heteronuclear multiple quantum coherence [HMQC] spectroscopy) spectrometer, both from the Wageningen NMR Centre. Chemical shifts are expressed in ppm relative to dimethyl sulfoxide (δ ¹H 2.50, δ ¹³C 39.52). One- and two-dimensional double quantum filtered COSY, TOCSY, NOESY, HMBC, and HMQC spectra were acquired using standard pulse sequences delivered by Bruker. The mixing time for the TOCSY was 80 ms, and that for the NOESY was 200 ms.

The analysis of the D/L configurations of the constituting amino acids was carried out as described by Gerard et al. (15) with some modifications. In short, 2 mg of purified lipopeptide was dissolved in 4 ml of 6 M HCl and heated at 110°C for 24 h in a closed glass vial. The HCl solution was removed in a fume hood by blowing a stream of N₂ over the solution at 45°C. The residue was dissolved in 250 μ l of isopropanol saturated with HCl gas and then heated in a closed vial at 110°C for 45 min. After cooling the solvent was removed with N₂ as before. The residue was dissolved in 250 μ l of CH₂Cl₂, and 100 μ l of pentafluoropropionyl anhydride was added. After the vial was closed, it was heated at 110°C for 15 min. The reagent solution was removed with N₂ as before and redissolved in 200 μ l of CH₂Cl₂, and 1 μ l was injected (split 1:100) onto a 25-m ChiralSil-Val HeliFlex gas chromatography (GC) column installed in an Agilent 6890 GC equipped with both a mass selective detector and an flame ionization detector. Other parameters were as follows: injector, cooled injection system starting at 100°C and going to 240°C at 10°C/s; temperature program, 60°C (hold for 2 min) to 170°C at 4°C/min. Helium was used as a carrier gas. Amino acids were identified by comparing the retention time and mass spectrum with authentic reference D- and L-amino acids.

Massetolide A production and transcription of *massA*, *massB*, and *massC*. Cells were grown in a 24-well plate with 1.25 ml KB broth per well and shaking at 220 rpm at 25°C. At specific time points, 1 ml of cell culture was collected and spun down. The cells were frozen in liquid N₂ and stored at -80°C. For the RNA isolations and cDNA synthesis, four biological replicates were used for each time point. Massetolide A production was measured qualitatively by the drop collapse assay and quantitatively by tensiometric analysis of the cell-free supernatant (K6 tensiometer; Krüss GmbH, Hamburg, Germany) at room temperature. To get sufficient volume for the tensiometric analysis, the supernatants of four biological replicates were collected and pooled for each time point. The surface tension of each sample was measured in triplicate. RNA was isolated from the frozen bacterial cells with Trizol reagent (Invitrogen), followed by DNase I (GE Healthcare) treatment. One microgram of RNA was used for cDNA synthesis with Superscript III (Invitrogen) according to the manufacturer's protocol. For the Q-PCR, conducted with the 7300SDS system from Applied Biosystems, the SYBR green core kit (Eurogentec) with a final concentration of 3.5 mM MgCl₂ was used according to the manufacturer's protocol. The concentrations of the primers were optimized (400 nM final concentration for all), and a dissociation curve was performed to check the specificity of the primers. The primers used for the Q-PCR were as follows: for *massA*, 5'-GCTGTACAACATTGGCGGCT-3' (forward) and 5'-GGTATGCAGTTGAGTGCCTAGC-3' (reverse); for *massB*, 5'-AACAACGACCGGAGATGCC-3' (forward) and 5'-AAGGTGTGCAGCAAGTGATGG-3' (reverse); for *massC*, 5'-GTCGACCCCTCAACGCGTCT-3' (forward) and 5'-CCACCGACAGTTGGTCAAGC-3' (reverse); for the 16S rRNA gene, 5'-GCGCAACCCTTGCTCTTAGTT-3' (forward) and 5'-TGTGTAGCCAGGCCGTAA-3' (reverse); and for *rpoD*, 5'-GCAGCTCTGTGTCCGTGATG-3' (forward) and 5'-TCTACTTCGTTGCCAGGGAATT-3' (reverse). To correct for small differences in template concentration, the 16S rRNA gene and *rpoD* were used as housekeeping genes. The cycle where the SYBR green fluorescence crosses a manually set threshold cycle (C_T) was used to determine transcript levels. For each gene the threshold was fixed based on the exponential segment of the PCR curve. The C_T value for *massA* was corrected

for the housekeeping gene as follows: $\Delta C_T = C_{T(massA)} - C_{T(rpoD)}$. The same formula was used for *massB* and *massC*. The relative quantification (RQ) values were calculated by the formula $RQ = 2^{[\Delta C_T(mutant) - \Delta C_T(wild\ type)]}$. If there is no difference in transcript level between mutant and wild type, then $RQ = 1$ (2^0) and $\log RQ = 0$. Q-PCR analysis was performed in duplicate (technical replicates) on four independent RNA isolations (biological replicates). Statistically significant differences were determined for log-transformed RQ values by analysis of variance ($P < 0.05$) followed by the Bonferroni and Dunnett post hoc multiple comparisons.

Motility and biofilm. The motilities of wild-type strain SS101 and the massetolide A-deficient mutants were assessed on soft (0.6% agar, wt/vol) standard succinate medium (SSM) [32.8 mM K₂HPO₄, 22 mM KH₂PO₄, 7.6 mM (NH₄)₂SO₄, 0.8 mM MgSO₄, 34 mM succinic acid, adjusted pH to 7 with NaOH]. Overnight cultures of SS101 and the mutants were washed three times, and 5 μ l of a cell suspension (1×10^{10} cells/ml) was spotted in the center of the soft SSM agar plate and incubated for 48 to 72 h at 25°C. Biofilm formation was assessed according to the method described by O'Toole et al. (37) using flat-bottom 96-well plates made of transparent polystyrene (Greiner) with 200 μ l KB broth per well. Statistically significant differences were determined with Student's *t* test ($P < 0.05$).

Nucleotide sequence accession numbers. The sequences of the contig containing *massA* and the contig containing *massB* and *massC* have been deposited in GenBank under accession numbers EU199080 and EU199081, respectively.

RESULTS AND DISCUSSION

Cloning and sequencing of the *massA*, *massB*, and *massC* genes. Five mutants, designated 10.24, 17.18, 1G12, 9.26, and 11.17, deficient in massetolide A production were obtained by random mutagenesis. HPLC analysis confirmed that each of these five mutants did not produce massetolide A or any of the other massetolide derivatives produced by wild-type strain SS101 (data not shown) (10). Southern hybridization showed that each of the five mutants contained a single transposon integration. Subsequent cloning and sequencing of the regions flanking the transposon revealed that the transposon had integrated in NRPS genes. A BAC library was constructed from the SS101 genome to clone and sequence the complete gene cluster for massetolide A biosynthesis. The clones in the BAC library had an average insert size of 55 kb and covered 7.5 genome equivalents. The library was screened by hybridization with probes corresponding to specific sequences flanking the transposon insertions, resulting in seven positive BAC clones (Fig. 1A). Based on restriction analysis, hybridization patterns, and BAC end sequencing, the seven clones aligned in two contigs, of which clones 2H12 and 7B4 were completely sequenced (Fig. 1A). On the first contig (64 kb), one large ORF of 6,270 bp, designated *massA*, was identified and predicted to encode an NRPS of 2,089 amino acids (aa) with a mass of 230 kDa. On the second contig (80 kb), two large ORFs were identified: one ORF of 12,924 bp, designated *massB*, was predicted to encode an NRPS protein of 4,307 aa with a mass of 469 kDa; the second ORF, of 11,328 bp and designated *massC*, was predicted to encode an NRPS protein of 3,775 aa with a mass of 410 kDa. No overlap was found between the two contigs, indicating that *massA* is disconnected from *massB* and *massC* (Fig. 1A). Collectively, *massA*, *massB*, and *massC* span a region of 30.5 kb (Fig. 1). Site-directed mutagenesis of each of these three NRPS genes with pKnockout-G (57) generated mutants that were deficient in massetolide A biosynthesis. In addition, complementation of the transposon mutants 10.24 (Δ *massA*), 17.18 (Δ *massB*), and 9.26 (Δ *massC*) with pME6031-*massA* and pME6031-*massBC*, respectively, restored massetolide A biosynthesis, which was confirmed by a drop collapse

FIG. 1. (A) Representation of the contig assembly of BAC clones 2F11, 3G1, 2H12, 7F11, 6F12, 8G2, and 7B4. The first contig (64 kb) harbors the *massA* gene, and the second contig (80 kb) contains the *massB* and *massC* genes. (B) Organization of the CLP gene cluster identified in *P. fluorescens* SS101 by sequence analysis of BAC clones 2H12 and 7B4. The three genes designated *massA*, *massB*, and *massC* are responsible for massetolide A biosynthesis. In the genome of strain SS101, *massA* is disconnected from *massB* and *massC*. Underneath the genes are the module and domain organization of MassA, MassB, and MassC. The domains are as follows: C, condensation; A, adenylation; T, thiolation; and TE, thioesterification. Underneath the domains are the amino acids that are incorporated into the CLP peptide moiety. The number associated with the amino acid refers to the position of the amino acid in the CLP peptide chain. Triangles represent the positions of the single transposon disruptions in the *massABC* genes obtained by random mutagenesis. The arrows indicate putative promoter sequences, and the closed circles represent putative terminator sequences. For each of the three *mass* genes, double arrows indicate the locations of the primers used for Q-PCR.

assay and HPLC analysis. These results confirm the role of *massA*, *massB*, and *massC* in massetolide A biosynthesis.

Putative promoter and terminator sequences were identified for each of the three *mass* genes (Fig. 1B). For *massA*, the -35 (TTGATG) and -10 (TTATAAAAT) putative promoter regions were identified at 221 and 213 bp upstream of *massA*, respectively. A putative terminator sequence was identified by RNA secondary structure analysis; sense (TGTAGGAGCGA GCTTGCTCGCGAAAA) and antisense stem-loops were identified with a ΔG of 51.6 kcal/mol at 500 and 574 bp downstream of *massA*, respectively. For *massB*, the -35 (TTACCA) and -10 (CGGCAGACT) putative promoter regions were identified at 412 and 394 bp upstream of *massB*, respectively. Sense (GCCTGGCGC) and antisense stem-loops were identified with a ΔG of 23 kcal/mol at 520 and 548 bp downstream of *massB*, respectively. For *massC*, the -35 (CTCACT) and -10 (CTATGTGAT) putative promoter regions were identified at 1,290 and 1,310 bp upstream of *massC*, respectively, and are located in the 3' region of *massB*. Sense (GCCCCACCA CTCGGCACCTCGCCTAGGCTCGGTGTGCCCCG) and antisense stem-loops with a ΔG of 94.9 kcal/mol were identified at 144 and 567 bp downstream of *massC*.

Characteristics of massetolide A synthetases. Analysis of the deduced NRPS amino acid sequences revealed two modules in MassA, four in MassB, and three in MassC. Each module consists of C, A, and T domains, and in MassC two TE

domains also were identified (Fig. 1B). The N-terminal C domain in MassA clusters closely with C1 domains of other NRPS involved in CLP biosynthesis (Fig. 2) and is presumably involved in N acylation of the first amino acid of the CLP molecule (8, 46). In silico analysis of the substrate specificity of the nine A domains and subsequent prediction of the amino acids in the CLP peptide moiety based on signature sequences (6, 8, 53) were consistent with the chemically determined structure of the peptide moiety of massetolide A (Fig. 1B). These results indicate that *massA* is the first and *massC* the last gene involved in massetolide A biosynthesis and that the nine modules in the massetolide biosynthetic template are colinear with the number of amino acids in massetolide A.

Chiral GC analysis performed in this study confirmed that the peptide moiety of massetolide A produced by *P. fluorescens* SS101 contains 4 aa in the L configuration (aa 1, 5, 7, and 9) and 5 amino acids in the D configuration (aa 2, 3, 4, 6, and 8) (Fig. 1B). In contrast to CLP biosynthetic templates in *Bacillus* species (40, 51), no epimerization (E) domains were found in the massetolide A synthetases (Fig. 1B) or in any of the other six CLP biosynthetic templates described to date for *Pseudomonas* (4, 8, 17, 18, 39, 45, 48, 59). Roongsawang et al. (45) initially postulated that external racemases may be responsible for the D configuration of the amino acids in arthrofactin and that sequence differences downstream of a conserved core motif [FFELGGHSLA(V/M)] in the T domains might reflect

FIG. 2. Phylogenetic analysis of amino acid sequences of 51 C domains identified in the known CLP biosynthesis clusters for massetolide A (mass), arthrofactin (arf), syringomycin (syr), and syringopeptin (syp). C domains predicted to have both condensation and epimerase activities are referred to as dual C/E domains and are indicated above the dotted line, whereas C domains predicted to function only in condensation (i.e., conventional C domains) are found below the dotted line. The C domains of the massetolide A biosynthesis cluster are boxed. Exceptions in this classification of C domains are indicated with an arrow followed by the upstream donor substrate. Dab, 2,4-diaminobutyric acid. C domains that are presumably involved in linking the fatty acid to the first amino acid of the CLP peptide moiety are the so-called C1 domains. The numbers at the nodes indicate the level of bootstrap support higher than 500, based on neighbor-joining analysis of 1,000 resampled data sets. The bar indicates the relative number of substitutions per site. The abbreviations in the tree indicate the gene and the C domain number, which refers to the module number in the CLP cluster.

the recognition by these external racemases. However, when sequences of the T domains of massetolide A were aligned, alone or together with T domains of other CLP biosynthesis clusters, including arthrfactin, no relationship could be established between this sequence motif and the amino acid configuration (data not shown). Subsequent studies on arthrfactin biosynthesis by Balibar et al. (2) had indicated that the D configuration of the amino acids is generated by specific C domains that have dual catalytic activities, i.e., condensation and epimerization. In their study, they showed that this subclass of C domains, referred to as C/E domains, is involved in epimerization of the amino acid that is loaded onto the T domain of the preceding module. Given that aa 2, 3, 4, 6, and 8 have the D configuration in massetolide A (Fig. 1B), this would suggest that the C domains of the third, fourth, fifth, seventh, and ninth modules should fall within this subclass of C/E domains. Subsequent alignment of the primary sequence of each of the individual C domains from NRPS genes involved in massetolide A, arthrfactin, syringomycin, and syringopeptin biosynthesis showed that C domains 3, 4, 5, 7, and 9 from massetolide A indeed cluster with the C/E domains of arthrfactin, syringomycin, and syringopeptin (Fig. 2). There were, however, two exceptions: based on the alignment, the second and sixth C domains from massetolide A were also predicted to have dual catalytic activity but follow an L-Leu residue (aa 1 and A5) (Fig. 2). Also Balibar et al. (2) also found three exceptions for syringopeptin (i.e., C5, C13, and C22 of syringopeptin synthetase [Fig. 1B]) and suggested that these C/E domains could also function as dual condensation/dehydration domains with or without prior epimerization.

Subsequent analysis of the primary sequence of the proposed C/E domains for massetolide A biosynthesis further revealed that they harbor the elongated His motif HHI/LxxxGD in the N-terminal sequence (Fig. 3A), as was described for arthrfactin (2). This elongated His motif is present in addition to the conventional His motif found more downstream in all C domains (data not shown). Moreover, the second His and terminal Asp in this elongated His motif have been shown to be critical for catalysis in both condensation and epimerase domains (2, 3, 54). These essential amino acids of the elongated His motif are present in the predicted C/E domains C3, C4, C5, C7, and C9 of the massetolide A synthetases (Fig. 3A). The second and sixth domains in the massetolide A synthetases also have an elongated His motif, but in the C2 domain the Gly in front of the Asp is missing, and in the C6 domain the second His is replaced by a Tyr (Fig. 3A). Whether these deficiencies affect or eliminate epimerase activity is not known.

MassC terminates with two TE domains of approximately 250 aa, each containing the conserved GxSxG sequence motif (Fig. 3B) (45, 47). Both TE domains of MassC contain the residues Ser⁸⁰, Asp¹⁰⁷, and His²⁰⁷ (Fig. 3B), which form a catalytic triad in the TE domain of SrfA to -C, the synthetases involved in surfactin biosynthesis in *Bacillus* (5). These results suggest that both TE domains in MassC are likely to be active in massetolide A biosynthesis and may function in tandem to enhance the rate of product release from the NRPS assembly line, as was shown for the two TE domains in arthrfactin biosynthesis (45, 47).

Organization of the massetolide biosynthesis genes and identification of flanking regions. Compared with other *Pseudomonas* CLP biosynthesis genes described to date (4, 8, 17, 18, 39, 45, 48, 59), the *massA*, *massB*, and *massC* genes showed highest similarity (81 to 84% identity) to *viscA*, *viscB*, and *viscC* of *P. fluorescens* SBW25, respectively (Fig. 4). Upstream of *massA*, two additional ORFs were found (Fig. 4). The first ORF, of 1,424, bp showed 75% identity with the outer membrane protein NodT (44) of *P. fluorescens* Pf0-1, belonging to the family of resistance nodulation and cell division (RND) efflux systems. The tripartite RND efflux system PseABC, identified at the left border of the *syrr-syp* genomic island in *Pseudomonas syringae* pv. *syringae*, encodes an outer membrane protein (PseA), a periplasmic membrane fusion protein (PseB), and a cytoplasmic membrane protein (PseC) (21). Mutations in each of the *pseABC* genes resulted in a significant decrease (40 to 60%) in syringomycin and syringopeptin production (21). Interestingly, the predicted RND-like outer membrane protein flanking *massA* showed only 30% identity to PseA (Psynr_2620) but 69% identity to another outer membrane protein found in *P. syringae* pv. *syringae*. This other membrane protein, designated Psyr_2606, is also located close to the *syrr-syp* genomic island (49), but its function in transport of syringomycin and syringopeptin is, to our knowledge, not yet known. Given that homologs of the RND-like outer membrane protein upstream of *massA* are also found upstream of the biosynthesis clusters for viscosin, orfamide, and syringomycin/syringopeptin (Fig. 4), we postulate that this gene plays a role in transport of massetolide A and other CLPs produced by *Pseudomonas*.

The predicted gene product of ORF 2, located upstream of *massA*, belongs to the group of LuxR-type transcriptional regulators which are also found upstream of other *Pseudomonas* CLP biosynthesis clusters (Fig. 4). This ORF shows 49% identity to *syrrF*, which is involved in regulation of syringomycin biosynthesis (30, 56). The role of this LuxR-type regulator in massetolide A biosynthesis is yet unknown, but preliminary results showed that overexpression of this gene resulted in an increased production of massetolide A in strain SS101 (data not shown). Downstream of *massA*, ORFs 3, 4, and 5 were identified, with 76% identity to mannosyltransferase, 69% identity to the melittin resistance protein PqaB, and 50% identity to the Ais protein (aluminum-induced protein), respectively (Fig. 4). These three ORFs are also found downstream of *viscA* in *P. fluorescens* SBW25, but their role in massetolide A or viscosin biosynthesis, if any, has not been resolved. Located even more downstream of *massA* are two ABC transporters (ORFs 6 and 7) and a hypothetical protein (ORF 8). Although close homologs of ORFs 6 through 8 were identified downstream of *viscA* in *P. fluorescens* SBW25 and in the genomes of other CLP-producing *Pseudomonas* strains, their role in CLP biosynthesis remains elusive. The ABC transporters flanking *massA* have relatively low identity (<25%) to SyrD, an ABC transporter involved in virulence and lipopeptide transport in *Pseudomonas syringae* pv. *Syringae* (42).

Upstream of *massB* and *massC*, an 11.3-kb ORF showing 61% identity to an outer membrane transport barrel of *P. fluorescens* Pf0-1 was identified. The presence of this gene close to the CLP biosynthesis cluster seems to be unique for strain SS101, since it is not found near any of the other CLP biosyn-

FIG. 3. (A) Alignment of the amino acid sequences of the 51 (C) domains identified in the known CLP biosynthesis clusters encoding the synthetases of massetolide A (mass), arthofactin (arf), syringomycin (syr), and syringopeptin (syp). The C domains of the massetolide A biosynthesis cluster are boxed. The conserved motif (HHI/LxxxxGD) for C domains with dual catalytic activity for condensation

thesis clusters described to date. Downstream of *massB* and *massC*, two ORFs with 84 to 85% identity to the ABC-type macrolide efflux proteins MacA and MacB of *P. fluorescens* Pf-5 were identified. In *E. coli*, this transport system confers resistance against the macrolides erythromycin and azithromycin (24). The presence of these genes downstream of the CLP biosynthesis cluster is conserved among CLP-producing pseudomonads (Fig. 4), suggesting that the MacA and MacB homologs may play a role in CLP transport. ORF 4, located downstream of the *macA* and *macB* genes, has 63% identity with a LuxR-type transcriptional regulator of *P. fluorescens* Pf0-1. Homologs of this gene are also found downstream of the viscosin and orfamide biosynthesis clusters in *P. fluorescens* strains SBW25 and Pf-5, respectively, but show relatively low identity (39%) to *salA*, an important LuxR-type transcriptional regulator of syringomycin and syringopeptin biosynthesis (22, 23, 30, 31). Site-directed mutagenesis of the LuxR-type regulators of the massetolide A biosynthesis cluster, as performed for strain DC3000 (4), as well as expression analyses should be conducted to more conclusively assess their functions in massetolide A biosynthesis in *P. fluorescens*. Collectively, these results indicate that the massetolide A biosynthesis cluster, including flanking genes, is most closely related to the viscosin biosynthesis cluster in *P. fluorescens* SBW25 and harbors specific features found in other known CLP biosynthesis clusters.

Identification of massetolide A derivatives in *P. fluorescens* SS101. Previous results obtained by De Souza et al. (10) indicated that *P. fluorescens* SS101 produces at least four other CLPs in addition to massetolide A. In this study, conditions to separate these putative CLPs were optimized and resulted in nine peaks corresponding to compounds having molecular masses ranging from 1,112 to 1,158 Da (Fig. 5). ESI-MS/MS and NMR analyses confirmed that the main peak, with a retention time of 91.3 min, is massetolide A. The peaks with retention times of 62.9 and 79.4 min were identified by ESI-MS/MS and NMR as viscosin and massetolide D, respectively (Fig. 5). Viscosin differs from massetolide A by the replacement of the *allo*-isoleucine at aa 4 with a valine, and massetolide D differs from massetolide A by the replacement of the isoleucine at aa 9 with a leucine (Fig. 5). Based on ESI-MS/MS studies, the small peaks at 47.2 and 54.7 min were tentatively identified as massetolides E and F, respectively. Massetolides E and F resemble viscosin, with the only difference being that at aa 9 a valine replaces the isoleucine in massetolide E and a leucine replaces the isoleucine in massetolide F. The molecular

and epimerization is indicated by arrows. The abbreviations in the alignment indicate the gene and the C domain number, which refers to the module number in the CLP cluster. (B) Alignment of the amino acid sequences of 11 TE domains identified in known CLP biosynthesis clusters encoding the synthetases of massetolide A (mass), viscosin (visc), orfamide (orf), arthofactin (arf), syringomycin (syr), and syringopeptin (syp). The TE domains of the massetolide A biosynthesis cluster are boxed. The conserved motif (GxSxG) for TE domains is indicated by arrows. Also indicated are Ser⁸⁰, Asp¹⁰⁷, which form a catalytic triad of the TE domain in the NPRS encoding surfactin biosynthesis. The abbreviations in the alignment indicate the gene and the TE domain number, referring to the first or second TE domain.

fractions, the lipid tail was identified as 3-hydroxydecanoic acid.

None of the five massetolide A biosynthesis mutants (Fig. 1B) produces any of the derivatives of massetolide A, indicating that these derivatives are the result of flexibility in amino acid selection and activation by the A domains of the massetolide A synthetases, and in particular A domains 4 and 9. Substrate flexibility of A domains is a common phenomenon in nonribosomal peptide synthesis, resulting in the production of a range of structural analogs that may have different biological functions or activities. For example, Gerard et al. (15) showed that for the massetolides produced by a marine *Pseudomonas* strain, variations at the fourth and ninth amino acid positions

FIG. 5. (A) HPLC profile of a crude surfactant extract of *P. fluorescens* SS101. The main peak (peak 9) represents massetolide A. The identities of the other eight peaks are given in panel B. (B) Peak numbers, retention times, names, and masses (*m/z*) of the pseudomolecular ions and amino acid sequence of the peptide moiety for each of the peaks in the chromatogram shown in panel A.

resulted in a significant change in antibacterial activity. Although massetolide A is the most predominant CLP produced by *P. fluorescens* SS101, culture conditions may have a significant effect on the production of these structural analogs. For example, carbon sources and the nature of amino acids added to the culture medium may greatly affect the production levels and structural diversity of CLPs (11) and may also be exploited in precursor-directed biosynthesis to generate structurally novel derivatives with different activities (15).

Relationship between cell density, massetolide A production, and transcription of *mass* genes. Cell density is an important feature in the regulation of CLP biosynthesis in several *Pseudomonas* strains (43). Production of viscosinamide, tensin, and amphisin occurs in the late exponential growth phase or stationary phase (34). Also, in *Pseudomonas putida* PCL1445, putisolvin production occurs at the end of the exponential growth phase (28). For *P. fluorescens* SS101, tensiometric analysis of cell-free culture supernatant showed a significant drop in the surface tension already after 12 h of growth, i.e., during the early exponential growth phase (Fig. 6A). The growth rate

of each of the three *mass* mutants was similar to that of wild-type strain SS101, but no reduction in surface tension of the culture supernatant was observed (Fig. 6A). To investigate the relationship between massetolide A production and expression of each of the three *mass* genes, RNA was isolated from samples taken at specific time points (8, 12, 16, and 24 h) in the growth curve and cDNA was subjected to Q-PCR with different primers specific for each of the three *mass* genes (positions of the primers are indicated in Fig. 1B). Transcript levels were determined in four independent RNA isolations and related to transcript levels of the housekeeping gene *rpoD* to correct for small differences in template concentration; correction with 16S rRNA gene transcript levels gave similar results (data not shown). During growth of wild-type strain SS101, transcript levels of *massA*, *massB*, and *massC* increased over time, reaching a maximum after 16 h of growth (Fig. 6B). Analysis of transcript levels of the *mass* genes in each of the three mutants was performed after 12 and 16 h of growth. The results show that at both time points, *massC* transcript levels were significantly and consistently decreased in the Δ *massB* mutant but

FIG. 6. (A) Growth of SS101 and the $\Delta massA$, $\Delta massB$, and $\Delta massC$ mutants at 25°C in liquid KB medium. Circles, SS101; triangles, $\Delta massA$; diamonds, $\Delta massB$; squares, $\Delta massC$. Closed symbols correspond to cell density, and open symbols correspond to the surface tension of the cell-free culture supernatant. Error bars represent standard deviations. OD600, optical density at 600 nm. (B) Transcript levels of *massA*, *massB*, and *massC* corrected for the transcript levels of the housekeeping gene *rpoD* [$\Delta Ct = C_{T(mass\ gene)} - C_{T(rpoD)}$]. (C and D) Transcript levels of *massA*, *massB*, and *massC* in SS101 and in the $\Delta massA$, $\Delta massB$, and $\Delta massC$ mutants after 12 h (C) and 16 h (D) of growth. For each time point, transcript levels are presented relative to the transcript level in wild-type SS101 (log RQ), with $RQ = 2^{[\Delta Ct(mutant) - \Delta Ct(wild\ type)]}$. For each time point, mean values for four biological replicates are given. Error bars represent the standard errors of the means.

were not affected in the $\Delta massA$ mutant (Fig. 6C and D). Transcript levels of *massA* and *massB* were variable between the two time points but were not substantially and consistently changed in all three mutants, indicating that initiation of transcription was not affected (Fig. 6C and D). These results show that transcript levels of *massA*, *massB*, and *massC* follow the same dynamics as the growth and biosurfactant production by strain SS101. A mutation in *massB* strongly reduces *massC* expression, suggesting that these two genes function in an operon.

Collectively, these results indicate that under the culture conditions used in this study, massetolide A biosynthesis is initiated in the early exponential growth phase. Although the results obtained with strain SS101 do not point to cell density-dependent regulation of massetolide A production, various methods were adopted to investigate whether *N*-acylhomoserine lactone (*N*-AHL)-mediated quorum sensing plays a role. When strain SS101 was coinoculated with the *N*-AHL reporter strains *Chromobacterium violaceum* O26 (32), *Pseudomonas aureofaciens* 30-841 (41), *Agrobacterium tumefaciens* NTLR4 (32, 52), *E. coli* pSCR1 (1), and *E. coli* pSB401 (58), no response of these *N*-AHL reporter strains was observed (data not shown). Furthermore, to separate putative metabolites

produced by SS101 that may inhibit the induction of the *N*-AHL reporters, cell-free supernatant from stationary-phase cultures of SS101 were extracted with ethyl acetate and separated by HPLC, followed by overlay thin-layer chromatography analysis with *C. violaceum* CV026 as the reporter strain (29, 50). This approach also did not provide any indications of *N*-AHL production by strain SS101 (data not shown). Taken together, these results strongly suggest that, under the conditions tested, *N*-AHL-dependent quorum sensing does not appear to play a role in massetolide A biosynthesis in *P. fluorescens* SS101. Similar results were found for viscosin biosynthesis in *P. fluorescens* SBW25 (data not shown). In the biosynthesis of amphisin and syringomycin, *N*-AHL-mediated quorum sensing also does not appear to play a role, even though these CLPs are produced in the late exponential and stationary growth phases (34). The fact that *N*-AHL-mediated quorum sensing does play a role in viscosin biosynthesis in the plant pathogenic *P. fluorescens* strain 5064 (7) and in putisolvin biosynthesis in *P. putida* strain PCL1445 (12) indicates that this type of regulation is strain dependent.

Role of massetolide A in swarming and biofilm formation. The role of CLP production in surface motility and biofilm formation is well established for other *Pseudomonas* strains (8,

FIG. 7. Role of massetolide A in biofilm formation on an artificial surface by *P. fluorescens* SS101, its mutants, and the mutants complemented with *massA* or *massBC*. Wells of microtiter plates were filled with 200 μ l of KB broth and inoculated with strain SS101 or its mutants at an initial density of 1×10^8 cells/ml. After incubation for 24 h at 25 $^{\circ}$ C, cells were stained with crystal violet, wells were washed, and cells attached to the walls of the microtiter wells were quantified spectrophotometrically (optical density at 600 nm [OD600]). Asterisks indicate statistically significant ($P < 0.05$) differences relative to wild-type SS101. Error bars represent the standard errors of the means.

17, 28, 43, 45). Soft agar assays performed in this study showed that the $\Delta massA$, $\Delta massB$, and $\Delta massC$ mutants also were completely impaired in surface motility, as was shown previously for viscosin-deficient mutants of *P. fluorescens* SBW25 (8). Microtiter plate assays showed that the $\Delta massA$, $\Delta massB$, and $\Delta massC$ mutants produced significantly less biofilm than wild-type strain SS101 (Fig. 7). The biofilm formed by SS101 was located mostly at the air-liquid interface. The deficiency in biofilm formation of the mutants was restored by complementation (Fig. 7). The complemented mutants produced even more biofilm than wild-type strain SS101, which is most likely due to the copy number ($n = 5$ to 7 [19]) of the vector used to reintroduce *massA* and *massBC* in the mutants. The role of CLPs in biofilm formation can differ considerably between different strains. For example, arthrfactin and putisolvin were shown to adversely affect biofilm formation, since mutants deficient in the biosynthesis of these CLPs produced more and differently structured biofilms than their respective parental strains (28, 45). How CLPs influence biofilm formation is still unclear, but their effect on cell surface hydrophobicity may play an important role in this process. Hydrophobic interactions and surface-active compounds, including CLPs, have been widely suggested to play a role in the adherence of cells to surfaces (33, 36, 38). More specifically, biosurfactants may be oriented with the hydrophilic part to the cell surface, thereby exposing the hydrophobic part to the outside and facilitating attachment to hydrophobic surfaces; when the orientation is the other way around, i.e., when the hydrophobic part

of the biosurfactant is anchored in the outer layers of the cell surface, the cell can interact with a hydrophilic surface but not with a hydrophobic interface (33). Given the diversity in structures and hydrophobicities of various CLPs produced by *Pseudomonas* strains, we postulate that depending on the cell surface of the producing strain as well as the structure and hydrophobicity of the CLP produced, the role in biofilm formation may be entirely different.

Conclusion. CLPs are produced by a variety of microorganisms and have activity against a wide range of plant and human pathogenic microorganisms, including fungi, oomycetes, enveloped viruses, mycoplasmas, trypanosomes, and gram-positive bacteria (43). Insight into the biosynthesis, regulation, and transport of these versatile compounds can ultimately be exploited in combinatorial biosynthesis to generate new derivatives with more specific or different activities (14, 43). The mutagenesis, BAC cloning, sequencing, complementation, and in silico analysis performed in this study revealed that massetolide A biosynthesis is governed by three large NRPS genes. Although other CLP biosynthesis genes have been identified previously for *Pseudomonas* species, the complete massetolide A biosynthesis cluster was unknown. Furthermore, in contrast to other *Pseudomonas* CLP biosynthesis clusters, the *mass* genes are not physically linked in the SS101 genome but are organized in two separate clusters that are transcribed independently. The only *Pseudomonas* CLP biosynthesis cluster known to date, for which the genes are also physically disconnected, is the viscosin biosynthesis cluster in *P. fluorescens*

SBW25, where *viscA* is separated from *viscBC* by 1.6 Mb on the physical map of the draft genome sequence (8). Several flanking genes of the massetolide A biosynthesis cluster are conserved among other *Pseudomonas* CLP biosynthesis clusters and include LuxR-type transcriptional regulators, ABC transport carriers, and an RND-like outer membrane protein. Although their function in massetolide A biosynthesis needs to be assessed, the conserved positioning of these genes suggests an important role in CLP biosynthesis. Interestingly, no genes involved in the biosynthesis of the lipid side chain or in acyl transfer were found up- or downstream of the massetolide biosynthesis genes. Similar observations were made for the orfamide biosynthesis cluster (17) and other CLP gene clusters (43). Gross et al. (17) postulated that the hydroxy fatty acids composing the lipid side chains of these CLPs may be produced by the primary metabolism, i.e., the type II fatty acid synthase systems.

ACKNOWLEDGMENTS

We are grateful to Marc Lemmers, Elbert van der Klift, and Frank Claasen for their assistance with reverse-phase HPLC, LC-MS/MS, and chiral GC analysis and to Andreas Untergasser for his help with the pME6031-*massBC* construct. The *N*-AHL reporter strain *P. aureofaciens* 30-841 was kindly provided by Leland S. Pierson III; *A. tumefaciens* NTLR4 was kindly provided by Yves Dessaux; and *E. coli* pSCR1, *E. coli* pSB401, and *C. violaceum* CV026 were kindly provided by Paul Williams and Vitorrio Venturi.

This work was funded by the Dutch Technology Foundation (STW), the applied science division of NWO.

REFERENCES

- Aguilar, C., I. Bertani, and V. Venturi. 2003. Quorum-sensing system and stationary-phase sigma factor (RpoS) of the onion pathogen *Burkholderia cepacia* genomovar I type strain, ATCC 25416. *Appl. Environ. Microbiol.* **69**:1739–1747.
- Balibar, C. J., F. H. Vaillancourt, and C. T. Walsh. 2005. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* **12**:1189–1200.
- Bergendahl, V., U. Linne, and M. A. Marahiel. 2002. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **269**:620–629.
- Berti, A. D., N. J. Greve, Q. H. Christensen, and M. G. Thomas. 2007. Identification of a biosynthetic gene cluster and the six associated lipopeptides involved in swarming motility of *Pseudomonas syringae* pv. tomato DC3000. *J. Bacteriol.* **189**:6312–6323.
- Bruner, S. D., T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, and M. T. Stubbs. 2002. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**:301–310.
- Challis, G. L., J. Ravel, and C. A. Townsend. 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**:211–224.
- Cui, X., R. Harling, P. Mutch, and D. Darling. 2005. Identification of *N*-3-hydroxyoctanoyl-homoserine lactone production in *Pseudomonas fluorescens* 5064, pathogenic to broccoli, and controlling biosurfactant production by quorum sensing. *Eur. J. Plant Pathol.* **111**:297–308.
- De Bruijn, I., M. J. D. de Kock, M. Yang, P. de Waard, T. A. van Beek, and J. M. Raaijmakers. 2007. Genome-based discovery, structure prediction and functional analysis of cyclic lipopeptide antibiotics in *Pseudomonas* species. *Mol. Microbiol.* **63**:417–428.
- Dennis, J. J., and G. J. Zylstra. 1998. Plasmidons: modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. *Appl. Environ. Microbiol.* **64**:2710–2715.
- De Souza, J. T., M. De Boer, P. De Waard, T. A. Van Beek, and J. M. Raaijmakers. 2003. Biochemical, genetic, and zoosporicidal properties of cyclic lipopeptide surfactants produced by *Pseudomonas fluorescens*. *Appl. Environ. Microbiol.* **69**:7161–7172.
- Dubern, J. F., and G. V. Bloemberg. 2006. Influence of environmental conditions on putisolins I and II production in *Pseudomonas putida* strain PCL1445. *FEMS Microbiol. Lett.* **263**:169–175.
- Dubern, J. F., B. J. J. Lugtenberg, and G. V. Bloemberg. 2006. The PpuL-RsaL-PpuR quorum-sensing system regulates biofilm formation of *Pseudomonas putida* PCL1445 by controlling biosynthesis of the cyclic lipopeptides putisolins I and II. *J. Bacteriol.* **188**:2898–2906.
- Finking, R., and M. A. Marahiel. 2004. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**:453–488.
- Fischbach, M. A., and C. T. Walsh. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**:3468–3496.
- Gerard, J., R. Lloyd, T. Barsby, P. Haden, M. T. Kelly, and R. J. Andersen. 1997. Massetolides A-H, antimycobacterial cyclic depsipeptides produced by two pseudomonads isolated from marine habitats. *J. Nat. Prod.* **60**:223–229.
- Gewolb, J. 2002. Bioengineering. Working outside the protein-synthesis rules. *Science* **295**:2205–2207.
- Gross, H., V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper, and W. H. Gerwick. 2007. The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem. Biol.* **14**:53–63.
- Guenzi, E., G. Galli, I. Grgurina, D. C. Gross, and G. Grandi. 1998. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**:32857–32863.
- Heeb, S., Y. Itoh, T. Nishijyo, U. Schnider, C. Keel, J. Wade, U. Walsh, F. O'Gara, and D. Haas. 2000. Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. *Mol. Plant-Microbe Interact.* **13**:232–237.
- Inoue, H., H. Nojima, and H. Okayama. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**:23–28.
- Kang, H., and D. C. Gross. 2005. Characterization of a resistance-nodulation-cell division transporter system associated with the *syr-syp* genomic island of *Pseudomonas syringae* pv. *syringae*. *Appl. Environ. Microbiol.* **71**:5056–5065.
- Kinscherf, T. G., and D. K. Willis. 2002. Global regulation by GidA in *Pseudomonas syringae*. *J. Bacteriol.* **184**:2281–2286.
- Kitten, T., T. G. Kinscherf, J. L. McEvoy, and D. K. Willis. 1998. A newly identified regulator is required for virulence and toxin production in *Pseudomonas syringae*. *Mol. Microbiol.* **28**:917–929.
- Kobayashi, N., K. Nishino, and A. Yamaguchi. 2001. Novel macrolide-specific ABC-type efflux transporter in *Escherichia coli*. *J. Bacteriol.* **183**:5639–5644.
- Kohli, R. M., C. T. Walsh, and M. D. Burkart. 2002. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**:658–661.
- Konnova, E. V., R. Hedman, C. J. Welch, and B. Gerhardson. 2004. Character of massetolide A production by biocontrol strain *Pseudomonas* sp. MF-30, p. 565–568. In I. A. Tikhonovich, B. J. J. Lugtenberg, and N. A. Provov (ed.), *Biology of plant-microbe interactions*, vol. 4. International Society for Plant-Microbe Interactions, St. Paul, MN.
- Konz, D., S. Doekel, and M. A. Marahiel. 1999. Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J. Bacteriol.* **181**:133–140.
- Kuiper, I., E. L. Lagendijk, R. Pickford, J. P. Derrick, G. E. M. Lamers, J. E. Thomas-Oates, B. J. J. Lugtenberg, and G. V. Bloemberg. 2004. Characterization of two *Pseudomonas putida* lipopeptide biosurfactants, putisolins I and II, which inhibit biofilm formation and break down existing biofilms. *Mol. Microbiol.* **51**:97–113.
- Laue, R. E., Y. Jiang, S. R. Chhabra, S. Jacob, G. S. A. B. Stewart, A. Hardman, J. A. Downie, F. O'Gara, and P. Williams. 2000. The biocontrol strain *Pseudomonas fluorescens* F113 produces the *Rhizobium* small bacteriocin, *N*-(3-hydroxy-7-cis-tetradecenyl)homoserine lactone, via HdtS, a putative novel *N*-acylhomoserine lactone synthase. *Microbiology* **146**:2469–2480.
- Lu, S. E., B. K. Scholz-Schroeder, and D. C. Gross. 2002. Characterization of the *salA*, *syrF*, and *syrG* regulatory genes located at the right border of the syringomycin gene cluster of *Pseudomonas syringae* pv. *syringae*. *Mol. Plant-Microbe Interact.* **15**:43–53.
- Lu, S. E., N. Wang, J. Wang, Z. J. Chen, and D. C. Gross. 2005. Oligonucleotide microarray analysis of the *salA* regulon controlling phytotoxin production by *Pseudomonas syringae* pv. *syringae*. *Mol. Plant-Microbe Interact.* **18**:324–333.
- McClellan, K. H., M. K. Winson, L. Fish, A. Taylor, S. R. Chhabra, M. Camara, M. Daykin, J. H. Lamb, S. Swift, B. W. Bycroft, G. S. A. B. Stewart, and P. Williams. 1997. Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of *N*-acylhomoserine lactones. *Microbiology* **143**:3703–3711.
- Neu, T. R. 1996. Significance of bacterial surface-active compounds in interaction of bacteria with interfaces. *Microbiol. Rev.* **60**:151–166.
- Nybroe, O., and J. Sorensen. 2004. Production of cyclic lipopeptides by fluorescent pseudomonads, p. 147–172. In J.-L. Ramos (ed.), *Pseudomonas*, biosynthesis of macromolecules and molecular metabolism, vol. 3. Kluwer Academic/Plenum Publishers, New York, NY.
- Ongena, M., E. Jourdan, A. Adam, M. Paquot, A. Brans, B. Joris, J. L. Arpigny, and P. Thonart. 2007. Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. *Environ. Microbiol.* **9**:1084–1090.
- O'Toole, G., H. B. Kaplan, and R. Kolter. 2000. Biofilm formation as microbial development. *Annu. Rev. Microbiol.* **54**:49–79.

37. O'Toole, G. A., L. A. Pratt, P. I. Watnick, D. K. Newman, V. B. Weaver, and R. Kolter. 1999. Genetic approaches to study of biofilms. *Methods Enzymol.* **310**:91–109.
38. Palmer, J., S. Flint, and J. Brooks. 2007. Bacterial cell attachment, the beginning of a biofilm. *J. Ind. Microbiol. Biotechnol.* **34**:577–588.
39. Paulsen, I. T., C. M. Press, J. Ravel, D. Y. Kobayashi, G. S. A. Myers, D. V. Mavrodi, R. T. DeBoy, R. Seshadri, Q. Ren, R. Madupu, R. J. Dodson, A. S. Durkin, L. M. Brinkac, S. C. Daugherty, S. A. Sullivan, M. J. Rosovitz, M. L. Gwinn, L. Zhou, D. J. Schneider, S. W. Cartinhour, W. C. Nelson, J. Weidman, K. Watkins, K. Tran, H. Khouri, E. A. Pierson, L. S. Pierson III, L. S. Thomashow, and J. E. Loper. 2005. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* **23**:873–878.
40. Peypoux, F., J. M. Bonmatin, and J. Wallach. 1999. Recent trends in the biochemistry of surfactin. *Appl. Microbiol. Biotechnol.* **51**:553–563.
41. Pierson, L. S., III, V. D. Keppenne, and D. W. Wood. 1994. Phenazine antibiotic biosynthesis in *Pseudomonas aureofaciens* 30-84 is regulated by PhzR in response to cell density. *J. Bacteriol.* **176**:3966–3974.
42. Quigley, N. B., Y. Y. Mo, and D. C. Gross. 1993. SyrD is required for syringomycin production by *Pseudomonas syringae* pathovar *syringae* and is related to a family of ATP-binding secretion proteins. *Mol. Microbiol.* **9**:787–801.
43. Raaijmakers, J. M., I. De Bruijn, and M. J. D. De Kock. 2006. Cyclic lipopeptide production by plant-associated *Pseudomonas* spp.: diversity, activity, biosynthesis, and regulation. *Mol. Plant-Microbe Interact.* **19**:699–710.
44. Rivilla, R., J. M. Sutton, and J. A. Downie. 1995. *Rhizobium leguminosarum* NodT is related to a family of outer-membrane transport proteins that includes TolC, PrtF, CyaE and AprF. *Gene*. **161**:27–31.
45. Roongsawang, N., K. Hase, M. Haruki, T. Imanaka, M. Morikawa, and S. Kanaya. 2003. Cloning and characterization of the gene cluster encoding arthrofacin synthetase from *Pseudomonas* sp. MIS38. *Chem. Biol.* **10**:869–880.
46. Roongsawang, N., S. P. Lim, K. Washio, K. Takano, S. Kanaya, and M. Morikawa. 2005. Phylogenetic analysis of condensation domains in the non-ribosomal peptide synthetases. *FEMS Microbiol. Lett.* **252**:143–151.
47. Roongsawang, N., K. Washio, and M. Morikawa. 2007. *In vivo* characterization of tandem C-terminal thioesterase domains in arthrofacin synthetase. *Chem Biochem.* **8**:501–512.
48. Scholz-Schroeder, B. K., J. D. Soule, and D. C. Gross. 2003. The *sypA*, *sypB*, and *sypC* synthetase genes encode twenty-two modules involved in the non-ribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. *Mol. Plant-Microbe Interact.* **16**:271–280.
49. Scholz-Schroeder, B. K., J. D. Soule, S. E. Lu, I. Grgurina, and D. C. Gross. 2001. A physical map of the syringomycin and syringopeptin gene clusters localized to an approximately 145-kb DNA region of *Pseudomonas syringae* pv. *syringae* strain B301D. *Mol. Plant-Microbe Interact.* **14**:1426–1435.
50. Shaw, P. D., G. Ping, S. L. Daly, C. Cha, J. E. Cronan, Jr., K. L. Rinehart, and S. K. Farrand. 1997. Detecting and characterizing *N*-acyl-homoserine lactone signal molecules by thin-layer chromatography. *Proc. Natl. Acad. Sci. USA* **94**:6036–6041.
51. Sieber, S. A., and M. A. Marahiel. 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**:715–738.
52. Smadja, B., X. Latour, D. Faure, S. Chevalier, Y. Dessaux, and N. Orange. 2004. Involvement of *N*-acylhomoserine lactones throughout plant infection by *Erwinia carotovora* subsp. *atroseptica* (*Pectobacterium atrosepticum*). *Mol. Plant-Microbe Interact.* **17**:1269–1278.
53. Stachelhaus, T., H. D. Mootz, and M. A. Marahiel. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**:493–505.
54. Stachelhaus, T., and C. T. Walsh. 2000. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **39**:5775–5787.
55. Tran, H., A. Ficke, T. Asimwe, M. Hofte, and J. M. Raaijmakers. 2007. Role of the cyclic lipopeptide massetolide A in biological control of *Phytophthora infestans* and in colonization of tomato plants by *Pseudomonas fluorescens*. *New Phytol.* **175**:731–742.
56. Wang, N., S. E. Lu, A. R. Records, and D. C. Gross. 2006. Characterization of the transcriptional activators SalA and SyrF, which are required for syringomycin and syringopeptin production by *Pseudomonas syringae* pv. *syringae*. *J. Bacteriol.* **188**:3290–3298.
57. Windgassen, M., A. Urban, and K. E. Jaeger. 2000. Rapid gene inactivation in *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* **193**:201–205.
58. Winson, M. K., S. Swift, L. Fish, J. P. Throup, F. Jorgensen, S. R. Chhabra, B. W. Bycroft, P. Williams, and G. S. A. B. Stewart. 1998. Construction and analysis of *luxCDABE*-based plasmid sensors for investigating *N*-acyl homoserine lactone-mediated quorum sensing. *FEMS Microbiol. Lett.* **163**:185–192.
59. Zhang, J. H., N. B. Quigley, and D. C. Gross. 1995. Analysis of the *syrB* and *syrC* genes of *Pseudomonas syringae* pv. *syringae* indicates that syringomycin is synthesized by a thiotemplate mechanism. *J. Bacteriol.* **177**:4009–4020.