Microbial synthesis of the forskolin precursor manoyl oxide in enantiomerically pure form

Morten T. Nielsen¹, Johan Andersen Ranberg², Ulla Christensen¹, Hanne Bjerre Christensen¹, Scott J. Harrison¹, Carl Erik Olsen², Björn Hamberger², Birger Lindberg Møller² and Morten H. H. Nørholm¹,³

¹Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, DK-2970 Hørsholm, Denmark
²Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
³Address correspondence to: MHHN morno@biosustain.dtu.dk. Phone: +45 217-99184 Fax: +45-353-33300

Keywords: Terpenes, cell factories, forskolin, Coleus forskohlii, medicinal natural products.

Total number of manuscript pages: 28
Total number of pages for Figures: 4
Total number of pages for Tables: 2
Total number of pages for supplemental material: 7
ABSTRACT

Forskolin is a promising medicinal compound belonging to the plethora of specialized plant metabolites that constitutes a rich source of bioactive high-value compounds. A major obstacle for exploitation of plant metabolites is that they often are produced in low amounts and in plants difficult to cultivate. This may result in insufficient and unreliable supply leading to fluctuating and high sales prices. Hence, substantial efforts and resources have been invested in developing sustainable and reliable supply routes based on microbial cell factories. Here, we report microbial synthesis of (13R)-manoyl oxide, a proposed intermediate in the biosynthesis of forskolin and other medically important labdane-type terpenoids. Process optimization enabled synthesis of enantiomerically pure (13R)-manoyl oxide as the sole metabolite providing a pure compound in just two steps with a yield of 10 mg/l. The work presented here demonstrates the value of a standardized bioengineering pipeline and the large potential of microbial cell factories as sources for sustainable synthesis of complex biochemicals.
INTRODUCTION

Terpenoids represent a highly diverse class of natural products with a multitude of applications in modern society ranging from renewable fuels to structurally complex medicinal compounds (1). For example they constitute the active compounds in several medicinal plant species including paclitaxel (*Taxus brevifolia*), forskolin (*Coleus forskohlii*) and artemisinin (*Artemisia annua*).

Although several terpenoid derived natural products are currently in or have passed clinical trial for treatment of a number of conditions (2), sufficient supply of the molecules remains an obstacle. Purification from naturally producing plants has proven infeasible in many cases due to low yield, while establishment of heterologous production systems have been hampered by insufficient insight into the biosynthetic routes. However, the advent and rapid development of whole genome and transcriptome sequencing have paved the way for several recent breakthroughs. Particularly tissue specific metabolomics combined with RNA deep sequencing is an exceptionally promising approach for gene discovery in non-model organisms (3, 4).

Our interest is focused on the diterpene forskolin produced by *C. forskohlii*. Forskolin is a promising pharmaceutical which has shown potential as treatment against a wide range of diseases (5-9). Forskolin is currently approved for treatment of glaucoma and heart-failure, while clinical trials against several forms of cancer are ongoing - the compound has been tested in over 800 independent assays and in 76 cases forskolin was reported to be potently active (< 2 micro molar) including activity as D3 dopamine and relaxin receptor agonist, as allosteric activator of Mas related G-protein coupled receptor X1 and as antagonist of the androgen receptor (PubChem, accessed August 2014 [http://pubchem.ncbi.nlm.nih.gov](http://pubchem.ncbi.nlm.nih.gov)). Finally, forskolin is extensively used in basic research as an inhibitor of protein kinases (more than 50,000 publications in Web of Science, Thompson Reuters). Recently, the combined use of metabolomics and RNA deep sequencing were successfully applied to *C. forskohlii* root cells resulting in identification of two diterpene synthases,
CfTPS2 and CfTPS3, that catalyze the formation of (13R)-manoyl oxide (13R-MO) from geranylgeranyl diphosphate (GGPP) (10). 13R-MO is proposed to be the first dedicated intermediate in forskolin biosynthesis (11). With this first biosynthetic knowledge in hand, establishment of a microbial system for further elucidation of the biosynthetic route as well as production of forskolin comes within reach.

From a microbial engineering perspective, terpenoids constitute an attractive class of compounds as they are all derived from a single precursor, isopentenyl pyrophosphate (IPP), and its isomer dimethylallyl pyrophosphate (DMAPP). Substantial efforts have therefore been made in increasing IPP-supply as a driver for efficient terpenoid synthesis in microbes (12). IPP can be synthesized through two biochemical routes; the mevalonate (MEV), and the 1-deoxy-D-xylulose 5-phosphate (DXP) pathways, with the MEV-pathway being prevalent in eukaryotes and archa, while the DXP-pathway dominates in eubacteria. While both routes have been subjected to numerous engineering efforts, the MEV-pathway has been most successfully applied in Escherichia coli (12, 13). One consensus minimal set of MEV-pathway enzymes required for boosting IPP/DMAPP supply can be formulated as mevalonate kinase (e.g. Saccharomyces cerevisiae [Sc] ERG12), phosphomevalonate kinase (e.g. ScERG8), mevalonate pyrophosphate decarboxylase (e.g. ScMVD1) and IPP isomerase (e.g. E. coli Idi2). Biosynthesis of diterpenes in E. coli faces the additional challenge that conversion of IPP to the general diterpene precursor, GGPP, is very inefficient in wild type E. coli strains and therefore requires introduction of a heterologous GGPP-synthase (14, 15). As previously shown for the C15 sesquiterpene amorphadiene, a substantial increase in product yield can be gained by combining a heterologous farnesyl diphosphate (FPP) synthase with the minimal consensus set of MEV-pathway genes (16). Likewise, high-level supply of GGPP in E. coli requires coordinated expression of a minimum of five genes: four for producing the precursor IPP/DMAPP (for example S. cerevisiae ERG12, ERG8, MVD1 and E. coli idi2), and one for a GGPP synthase.
As this consensus set is independent of the target diterpene, assembling all five genes on a single plasmid would substantially simplify experimental designs such as strain screenings. The Keasling group have reported the construction and verification of a single plasmid for synthesis of farnesyl diphosphate (13). Previous attempts of adaptation of this strategy for synthesis of GGPP unfortunately resulted in a non-functional plasmid (16). We have recently described a set of design principles for standardizing molecular cloning and formulation of a cloning pipeline to simplify multi-gene expression in *E. coli* (17). Here, we take advantage of this bioengineering pipeline to assemble and validate a single plasmid enabling GGPP synthesis from mevalonate, providing a versatile platform for diterpene biosynthesis in *E. coli*. By utilization of the platform, we achieve *in vivo* microbial synthesis of the proposed forskolin precursor 13R-MO. Through a limited series of iterative optimization steps, stereospecific synthesis of 13R-MO is achieved and yields are increased from trace amounts to 10 mg/l in micro-scale cultures. The final process yields 13R-MO in NMR-quality following a single, simple step for removal of impurities.
MATERIALS AND METHODS

Bacterial strains and media. Escherichia coli strain NEB5α was used for cloning and propagation of plasmids. BL21(DE3) was used for radioactive labeling studies and KRX (Promega, Madison, USA) for production of 13R-MO. Following transformations, bacteria were propagated on Luria-Bertoni (LB) agar plates supplemented with ampicillin (100 µg/ml), spectinomycin (50 µg/ml), kanamycin (50 µg/ml) or in combinations (ampicillin 50 µg/ml, spectinomycin 25 µg/ml, kanamycin 25 µg/ml) when required. Liquid cultures for plasmid preparation were based on 2xYT media, whereas pre-cultures for assays were prepared in LB media using antibiotic concentrations as described above. The PASM-5052 defined media used for expression assays were prepared as described by Studier (18) except for the exchange of Seleno methionine with methionine. For long term incubations (more than 24 h) ampicillin was replaced by carbencillin (50 µg/ml) as previously described (19).

PCR and uracil-excision cloning. Uracil-excision compatible PCR products were produced with 30 cycles in 50 µl reaction mixtures using proof-reading PfuX7 polymerase as previously described (20). Fragments of ScMVD1 (Gene ID: 855779), ScERG8 (Gene ID: 855260) and ScERG12 (Gene ID: 855248) were amplified from S. cerevisiae CEN.PK113-7D genomic DNA. Ecidi2 (NC_000913) fragments were amplified from E. coli K12 MG1655 genomic DNA. CfTPS2 (AWH04047.1) and CfTPS3 (AWH04048.1) were amplified from previously described constructs (10). AgGGPP (AF425235.2) was codon optimized for E. coli and synthetically synthesized (Genscript, Piscataway, USA). USER Fusion DNA assembly was performed essentially as described by Geu Flores et al (21). Purified PCR products of inserts and vector were mixed in a total insert to vector ratio of 6:1 (volume). When several DNA fragments were cloned simultaneously, use of equimolar amounts of each fragment was approximated by agarose gel.
The reaction mixture was buffered with 5x Phusion HF buffer (Invitrogen) and 1 U of USER™ enzyme mix (New England Biolabs, Ipswich, USA) was added. The reaction mixture was incubated for 20 min at 37°C, followed by 20 min at 25°C before transformation of chemically competent *E. coli* cells. All oligo nucleotides and synthetic genes used in this study are summarized in Table S1.

**Plasmid construction.** The principle behind the molecular cloning of open reading frames (ORFs) is illustrated in Fig. 1 and described previously in detail in (17). Briefly, ORFs were PCR amplified without start- or stop-codons and flanked by 9 base pair linkers mediating standardized cloning into a pETDuet-1 (Merck KGaA, Darmstadt, Germany) backbone entry vector. Transcriptional units containing a T7-promoter, a consensus Shine-Dalgarno and the entry clone ORFs were re-amplified using generic oligo nucleotides with complementary linkers facilitating multi-fragment assemblies. Vector-backbones used in multi-fragment assemblies were likewise amplified with generic oligo nucleotides. All plasmids containing ORFs were verified by sequencing.

**35S-methionine labeling of proteins.** Transcription and translation of ORFs used in this study was confirmed *in vivo* by the rifampicin blocking technique and 35S-labeling (22). Briefly, pre-cultures of BL21 (DE3) cells carrying relevant plasmids were sub-cultured (1:50) into 1 ml chemically defined rich medium lacking methionine (18). Following induction of T7 dependent transcription by isopropyl-beta-D-1-thiogalactopyranoside and addition of rifampicin, cells were lysed and samples applied on precast 4-20% SDS-PAGE (Expedeon, San Diego, USA). Upon completion of the electrophoresis, the gels were dried and the labeled proteins were visualized using a CyclonePl scanner (Perkin Elmer, Waltham, USA).
**Assay conditions.** For chromatographic assays, bacteria from single plate colonies were inoculated into pre-cultures of 500 µl LB and grown 16 h at 37°C, 300 rpm shaking. For assaying manoyl oxide production, 40 µl LB pre-culture were inoculated into 2 ml freshly prepared PASM-5052 defined media (1:50 dilution) and grown in 10 ml micro-titer plates at 37°C, 300 rpm to an OD of 0.3. At this point, synthesis of recombinant proteins was induced by addition of rhamnose (1 mM) and IPTG (0.4 mM). Rhamnose stimulates synthesis of the T7-RNA polymerase in KRX whereas IPTG alleviates the lac-repressor from the lac operator present in all protein expression plasmids. Mevalonolactone (9mM) was added at this point, when required. All metabolite concentrations are given as the final concentration in the assay. After induction, cultures were incubated at either 25°C, 150 rpm for 16 hours or 16°C, 150 rpm for 112 hours. Assays were performed with two independent colonies each analyzed in at least three replicates.

**Extraction procedure.** Metabolite extraction was performed by the following procedure; 500 µl of bacterial culture was transferred to a 2 ml HPLC vial (Mikrolab, Aarhus, Denmark), prior to application of an overlay of 500 µl hexane containing internal standard (1-eicosene, 1 mg/l, Sigma-Aldrich). Extractions were carried out at room temperature with shaking (240 rpm, table shaker (IKA-KS 130 basic)). After 1 or 16 h of extraction, 300 µl of the hexane phase were transferred to clean HPLC vials and subjected to gas chromatography analysis. Cultures for purification of 13R-MO were performed as above except that the total volume (2 ml) of 240 cultures were pooled in a 2 l Erlemeyer flask and extracted with 500 ml of hexane.

**Analytical method.** Gas chromatography mass spectrometry (GC-MS) analysis was performed as previously described (23). Extracts were analyzed on a Shimadzu GCMS-QP2010 Ultra instrument with an Agilent HP-5MS column (30 m x 0.250 mm i.d., 0.25 µm film thickness) installed. 1 µl of
extract was injected in the injection port set at 250°C in splitless mode. The gasflow was set to 2.2 ml/min and the purge flow to 1 ml/min, using helium as carrier gas. The GC program was set to: Hold for 1 min at 100°C, ramp 10°C/min to 250°C and a second ramp with 20°C/min to a final temperature of 310°C which was held for 2 min. Ion source and interface temperature were set to 230°C and 150°C, respectively, and the solvent cutoff was set for 8 min. Mass analyzer was set in scan mode from 50 Hz to 350 Hz and a scan rate of 625 u/sec. Data analysis was carried out with the GCMS solution Version 2.50 SU3 (Shimadzu Cooperation). Compound identification was done by comparison to reference spectra databases (Wiley Registry of Mass Spectral Data, 8th Edition, July 2006, John Wiley & Sons, ISBN: 978-0-470-04785-9) and comparison of retention indices found in (23). 13R-MO was quantified relatively by comparison of TIC peak area normalized to the peak area of internal standard (1mg/l 1-eicosene). Absolute quantification was performed by comparison of normalized peak areas as above with the standard curve containing purified 13R-MO. The standard curve was fitted with linear fit after log-log transformation.

Single step purification of 13R-MO. The hexane extract was concentrated by rotor evaporation (Buchi, Switzerland) set to 35°C and 220 mbar. The concentrated sample was subsequently purified by solid phase extraction (Dual Layer Florisil/Na$_2$SO$_4$ 6 ml PP SPE TUBE, Supelco Analytical, Sigma-Aldrich, St. Louis, USA) according to the manufacturers protocol. In brief, the sample was applied to the column, washed with hexane, before removal of residual solvent in a gentle nitrogen gas stream. The sample was eluted in deuterated chloroform (CDCl$_3$) and analysed by NMR for structural verification.

NMR analysis and quantification. NMR analysis was performed using a 400 MHz Bruker Avance 400 spectrometer, and the structure was confirmed by comparison of the observed $^1$H- and $^{13}$C-
NMR spectra with literature data (24) (see Table 1) (25). CDCl₃ was used as solvent. The amount of 13R-MO present in the NMR sample was determined by integration after adding 1.00 mg of dimethyl formamide (DMF). The integrals of one of the vinyl protons and one of the DMF methyl groups were used for the calculations. The NMR sample was subsequently used for creating a GCMS standard curve.
RESULTS

**Assembly of a two-plasmid system for producing 13R-MO from mevalonolactone.** Open reading frames (ORFs) of MEV-pathway genes from *S. cerevisiae (ScERG12, ScERG8 and ScMVD1)*, IPP isomerase from *E. coli (Ecidi2)*, GGPP synthase from *Abies grandis (AgGGPPs)* (26) and diterpene synthases from *C. forskohlii (CfTPS2 and CfTPS3)* (10) were initially cloned into an entry expression vector allowing assessment of proper expression and a standardized multi-gene assembly (Fig. 1A). All plant derived ORFs were codon optimized for *E. coli* and truncated to remove plastid-targeting signals according to signal peptide predictions by the ChloroP-algorithm (27). Single ORF expression plasmids were transformed into BL21 (DE3) and transcription and translation was verified by $^{35}$S-methionine labeling in the presence of rifampicin as previously described (28) (Fig. S1). From these entry clones, *ScMVD1, ScERG8, ScERG12, Ecidi2* and *AgGGPPs* were subsequently assembled into the plasmid pGGPP for GGPP synthesis, while *CfTPS2* and *CfTPS3* were assembled into pMO for conversion of GGPP into 13R-MO (Fig. 1B). In the setup employed here, the T7-promoter and a consensus Shine-Dalgarno sequence were maintained as transcriptional units in front of every ORF encoding the entry fragment in the assemblies (Fig. 1B).

**Co-transformation of pGGPP and pMO enables in vivo synthesis of 13R-MO.** In order to establish the system and test the functionality of the plasmids, pMO was co-transformed with either a plasmid harboring the AgGGPP synthase alone (pSyn), pGGPP or an empty vector control (pCDFDuet-1) into the *E. coli* strain KRX. Subsequently assays were performed in a defined rich media based on lactose and glycerol (18). In order to test for pGGPP-mediated conversion of mevalonolactone to GGPP, cultures were split in two and mevalonolactone (9 mM) was supplemented to one culture containing pGGPP at the time of induction. After 16 h induction at 25°C, non-polar metabolites were extracted with hexane and analyzed with gas chromatography.
coupled to mass spectrometry (GC-MS). Consistent with previous findings (15), introduction of the 222 AgGGPP synthase, either individually in pSyn or as part of pGGPP, enabled synthesis of trace amounts of two diterpene related compounds 1 and 2 (Fig. 2A and B). Supplementation of mevalonolactone to pGGPP containing cultures resulted in a modest, yet significant, increase in the yield of both compounds (Fig. 2C). The compounds were tentatively identified as 13R-MO and 13S-MO, respectively, by comparison of mass spectra with the Wiley Registry of Mass Spectral Data, (8th edition) (Fig. S2) as well as previously published data (23). Absolute identification of 13R-MO by NMR was achieved using the final process described below (Table 1 and Fig. S3). To our knowledge, this is the first report of in vivo biosynthesis of 13R-MO, the proposed precursor for forskolin, in microbial systems and therefore an important milestone in efforts towards establishment of a microbial cell factory for forskolin synthesis. Moreover, the mevalonate mediated increase in diterpene synthesis strongly indicates that all enzymes required for conversion of mevalonate to GGPP are functionally expressed from pGGPP and this plasmid might therefore serve as a future platform for diterpene synthesis in E. coli. The observed co-occurrence of 13S-MO with 13R-MO has previously been associated with insufficient activity of the class I diTPS, CfTPS3, (10) (see Fig. 2) and therefore indicated that the current process is sub-optimal.

**Process optimization facilitates increased yield** Several studies have demonstrated that substantial improvements of microbial production systems can be achieved using a simple set of process optimization steps including alteration of culture parameters and recovery procedures (13, 16, 29). Although no clear consensus protocol has emerged, several groups reported success with extended cultivation at low temperatures. Moreover, recovery by hexane extraction has previously been demonstrated to yield highly pure diterpene preparations from E. coli (16). A comparison between short incubation at medium temperature (16 h, 25°C) and longer incubation at low temperature (112 h, 16°C) as well as between short (1 h) and extended (16 h) hexane extraction was undertaken.
this experiment, *E. coli* KRX was transformed with pGGPP and pMO and all cultures were supplemented with mevalonolactone (9 mM) upon induction. Incubation as well as extraction time had a significant impact on diterpene yield when compared to the initial protocol contributing a 2-fold yield increase each (Fig. 3). These yield improvements proved to be cumulative when combined resulting in an approximately 5-fold total increase over the initial protocol. Under these experimental conditions, the process, however, still resulted in formation of both 13S- and 13R-MO in approximately equal amounts. At this stage, the expression level of *CfTPS3* from pMO was examined in detail. Remarkably, when pMO was transformed into BL21(DE3) and subjected to $^{35}$S-methionine labeling, only *CfTPS2* expression was detected (Fig. S1). Since it was already established that both *CfTPS2* and *CfTPS3* could be transcribed and translated in *E. coli* when expressed alone (Fig. S1), these findings indicated that transcription from the second T7 promoter is impeded in pMO. To test this hypothesis, we established a three-plasmid system consisting of pGGPP, p*CfTPS2* and p*CfTPS3*. Indeed, when the three plasmids were co-transformed into *E. coli* KRX cells and cultivated in consistence with the optimal parameters identified above, biosynthesis of enantiomerically pure 13R-MO was obtained (Fig. 4A). Moreover, the flux through the pathway was dramatically increased resulting in a more than two thousand fold yield increase over the initial protocol.

**Recovery, structure validation and absolute quantification.** An advantage of the *E. coli* based biosynthesis of diterpenes is high purity and therefore simple downstream processing. Indeed, with the hexane-based extraction utilized in this study, merely three compounds were observed with GC-MS in addition to 13R-MO. Two of these (4 and 5), could be attributed to impurities arising from the instrument (data not shown) and are thus experimental artifacts rather than genuine compounds present in the extract. The one remaining metabolite, 3, was identified as indole by comparison of retention time and mass spectra with an authentic standard (Fig. 4B). Indole is known to be a
signaling molecule in *E. coli* with substantial influence on metabolism and membrane potential (30, 31) and has been reported to inhibit terpene synthesis in *E. coli* (29). Thus elimination of indole formation might not only simplify recovery, but also increase yield. Recently, it was discovered that *E. coli* very efficiently converts extracellular tryptophan to indole (32). The defined media used in this study is supplemented with high concentrations (10 mM) of free amino acids, including tryptophan, to support high-level protein synthesis. Adjustment of the media composition by omission of tryptophan eliminated indole accumulation, resulting in an extract with 13R-MO as the sole metabolite observed by GC-MS analysis (Fig. 4A). Importantly, the overall yield of 13R-MO was unaffected. Following a simple removal of impurities by solid phase extraction, the recovered 13R-MO was sufficiently pure for absolute identification by NMR analysis (see Table 1). The yield of the established process was quantified absolutely to 10 mg/l, by GC-MS using a standard curve of purified 13R-MO as reference. Typical dephosphorylation products (and result of endogenous phosphatase activity) of the pathway intermediates GGPP and 8-hydroxy-colaplool diphosphate are geranyl-geraniol and dihydroxy-copalol. Under the conditions used, both compounds are detectable with our analytical setup, yet no traces were found, indicating that those pathway intermediates do not accumulate.
DISCUSSION

Biosynthesis of high-value terpenoids in microbial systems is emerging as an adequate and sustainable production platform, with several studies reporting yields of diterpene olefins in the scale of grams per liter (29, 33-35) (reviewed in (36)). A unifying feature of these studies is that yields in unoptimized micro-scale cultures are meagre, typical in the low milligram per liter range (0.15 mg/l, (33), 2.6 mg/l (16), 0.005 mg/l, (15), 4 mg/l, (35). A highly successful route for yield increase in micro-scale has been improvement of the pre-cursor pool by either introduction of the mevalonate pathway from yeast (13, 16, 35) or modulation of the endogenous MEP pathway by promoter engineering (29) or stacking of gene copies (33). Moreover, culture scale up into controlled bioreactors has proven essential obtaining gram scale yield (34, 35). Our aim is to use microbial systems for elucidation of biosynthetic pathways to high-value diterpenoids and simultaneously establish a microbial cell factory. Since gene discovery is likely to require a substantial screening effort, it was crucial for us to develop a process compatible with micro-scale cultures. We describe here the assembly of a single plasmid harboring five ORFs encoding enzymes sufficient for efficient GGPP synthesis from mevalonate in E. coli. Maintaining all these enzymatic functions on one plasmid substantially increases the flexibility in experimental design by serving as a platform to facilitate high throughput screening of novel diTPS candidates in different microbial strains. For our proof-of-principle, the process was optimized from affording trace amounts of two isomers, 13R-MO and 13S-MO, to obtaining 10 mg/l of enantiopure 13R-MO, corresponding to a more than two thousand fold increase. As such, our findings are comparable to previous findings regarding yield increase in micro-scale cultures (16, 33, 35), which hints that further engineering and culture scale up can substantially increase yield of 13R-MO, although the previously proposed cytotoxic effects (37-40) may represent a challenge. However, our interest in 13R-MO is exclusively as a pre-cursor for forskolin, thus maintaining the micro-scale format is crucial at this stage.
stage. Moreover, Ajikumar and co-workers observed that the strain developed for efficient production of the olefin taxadiene (1g/l) (29), merely produced 50mg/l of the next biosynthetic intermediate taxadien-5α-ol, thus indicating that careful balancing of enzyme activities might be most relevant once a complete pathway is assembled. We expect that the microbial synthesis process described here will serve as a foundation for full elucidation of the forskolin biosynthetic pathway by providing complex precursor molecules for \textit{in vitro} assays and facilitate high-throughput \textit{in vivo} screening. In due course, the microbial process may also be developed into a cell factory for sustainable production of forskolin and other labdane-type terpenes.
ACKNOWLEDGEMENTS

This work was supported by the Novo Nordisk Foundation, UNIK Synthetic Biology and an ERC Advanced grant: ERC-2012-AdG 323034 LightDriven P450s. We would like to thank Lone Riisberg for excellent technical assistance and Susanna Seppälä for critical reading of the manuscript. It is noted that parts of the discoveries presented are the subject of a patent application.
REFERENCES


FIGURES AND LEGENDS

**FIG 1** Schematic overview of the gene assembly procedure. (A) Open reading frames are cloned individually into standardized entry expression vectors. (B) Assembly of pGGPP and pMO. Transcriptional units consisting of a T7-promoter, a consensus Shine-Dalgaro and the open reading frame are PCR amplified with generic oligo nucleotides that anneal outside the different ORFs in the flanking regions common to all entry plasmids. The generic oligo nucleotides facilitate specific assembly through defined linker regions (A-D, Eu, Ed). Transcriptional units are assembled in a single cloning reaction by USER Fusion. Start and stop
The codons of open reading frames are indicated by green and red bars, respectively. T7 promoter and terminator are denoted pT7 and T7t, respectively.

**FIG 2** In vivo synthesis of two manoyl oxide isomers in *E. coli*. (A) Total ion GC-MS chromatograms (TIC) of hexane extract from *E. coli* transformed with either pGGPP and pMO (top) or empty vectors (bottom). IS denotes the internal standard, 1mg/l 1-eicosene. (B) Mass spectra of metabolites 1 and 2 identified as 13R- and 13S-MO, respectively, based on comparison to the Wiley Registry of Mass Spectral Data (Fig.S2). (C) Relative quantification of 13R- and 13S-MO synthesis in *E. coli*. Yields are normalized to a 1mg/l 1-Eicosene internal standard. N.d. indicates that neither 13R- nor 13S-MO were detectable. Numbers on the X-axis refer to the combination of plasmids present in the analyzed strain: 1 = pET-Duet/pCDF-Duet, 2 = pMO/pCDF-Duet, 3 = pMO/pSyn, 4 = pMO/pGGPP, 5 = pET-Duet/pGGPP, 4+mev = pMO/pGGPP supplemented with 9 mM mevalonolactone. Data represents the average of four independent replicates with standard deviations. (D) Biosynthetic route to 13R-MO. Solid
arrows indicate enzymatic reactions. The enzymes catalyzing each step are depicted next to the reaction arrow. The dashed arrows indicate the intrinsic degradation of labda-13-en-8-ol diphosphate into equimolar amounts of $13R$- and $13S$-MO occurring in the absence of CfTPS3. Phosphate groups in the chemical structures are abbreviated with P.

**FIG 3** Impact of iterative improvements of culture conditions and extraction time on manoyl oxide yields. *E. coli* transformed with pGGPP and pMO was cultivated at either 25°C for 16 h post-induction or at 16°C for 112 h post-induction. Cultures were extracted with hexane for either 1 or 16 h. Graphs depict the average yields of the two manoyl oxide isomers from two independent colonies each analyzed in four independent replicates. Error bars denote standard deviation.
FIG 4. Identification and elimination of indole in E. coli extracts.

Panel A displays Total Ion Chromatograms from GC-MS analysis of extracts from E. coli. The plasmid combination in question is listed in the top left corner of each chromatogram along with the initial concentration of tryptophan (trp) in the media. The metabolites observed were identified as 13R-MO (1), Indole (3) and impurities from the instrument (4 and 5). IS denotes the internal standard, 1mg/l 1-eicosene.

Panel B; Total Ion chromatogram of authentic indole standard (top) as well as mass spectra of compound 3 (middle) and indole standard (bottom). Based on comparison of retention time and mass spectra with authentic standard, compound 3 was identified as indole.
<table>
<thead>
<tr>
<th>#</th>
<th>C</th>
<th>ppm</th>
<th>ppm</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>39.12</td>
<td>39,018</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18.62</td>
<td>18,564</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>42.23</td>
<td>42,117</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>33.32</td>
<td>33,238</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>56.55</td>
<td>56,432</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>20.03</td>
<td>19,956</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>43.4</td>
<td>43,267</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>75.09</td>
<td>75,091</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>55.6</td>
<td>55,64</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>37.11</td>
<td>37,011</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>15.41</td>
<td>15,304</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>35.73</td>
<td>35,714</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>73.25</td>
<td>73,208</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>148.07</td>
<td>147,976</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>110.21</td>
<td>110,255</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>28.71</td>
<td>28,556</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>25.58</td>
<td>25,524</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>33.4</td>
<td>33,362</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>21.39</td>
<td>21,348</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>15.41</td>
<td>15,421</td>
<td>q</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 1

<table>
<thead>
<tr>
<th>Type</th>
<th>13C ppm</th>
<th>1H ppm</th>
<th>Coupling constants</th>
</tr>
</thead>
<tbody>
<tr>
<td>-CH3</td>
<td>0.83</td>
<td>0.806</td>
<td>s</td>
</tr>
<tr>
<td>-CH3</td>
<td>0.83</td>
<td>0.816</td>
<td>s</td>
</tr>
<tr>
<td>-CH3</td>
<td>0.9</td>
<td>0.877</td>
<td>s</td>
</tr>
<tr>
<td>-CH3</td>
<td>1.31</td>
<td>1.294</td>
<td>s</td>
</tr>
<tr>
<td>-CH3</td>
<td>1.33</td>
<td>1.313</td>
<td>s</td>
</tr>
<tr>
<td>=CH</td>
<td>4.94</td>
<td>4.938</td>
<td>q, J 11 &amp; 2Hz</td>
</tr>
<tr>
<td>=CH</td>
<td>5.16</td>
<td>5.164</td>
<td>q, J 17 &amp; 2Hz</td>
</tr>
<tr>
<td>=CH2</td>
<td>5.92</td>
<td>5.894</td>
<td>q, J 17 &amp; 11Hz</td>
</tr>
</tbody>
</table>

This work

Gonzales et al.

Coupling constants

| -CH3   | 0.83   | 0.806  | s |
| -CH3   | 0.83   | 0.816  | s |
| -CH3   | 0.9    | 0.877  | s |
| -CH3   | 1.31   | 1.294  | s |
| -CH3   | 1.33   | 1.313  | s |
| =CH    | 4.94   | 4.938  | q, J 11 & 2Hz      |
| =CH    | 5.16   | 5.164  | q, J 17 & 2Hz      |
| =CH2   | 5.92   | 5.894  | q, J 17 & 11Hz     |