JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Toward a Biosynthetic Route to Sclareol and Amber Odorants

Michel Schalk,* Laurence Pastore, Marco A. Mirata,# Samretthy Khim, Marina Schouwey, Fabienne Deguerry, Virginia Pineda, Letizia Rocci, and Laurent Daviet*

Biotechnology Department, Corporate R&D Division, FIRMENICH SA, Geneva CH-1211, Switzerland

Supporting Information

ABSTRACT: Ambergris, a waxy substance excreted by the intestinal tract of the sperm whale, has been a highly prized fragrance ingredient for millenia. Because of supply shortage and price inflation, a number of ambergris substitutes have been developed by the fragrance industry. One of the key olfactory components and most appreciated substitutes of ambergris, Ambrox is produced industrially by semisynthesis from sclareol, a diterpene-diol isolated from Clary sage. In the present study, we report the cloning and functional characterization of the enzymes responsible for the biosynthesis of sclareol. Furthermore, we reconstructed the sclareol biosynthetic pathway in genetically engineered *Escherichia coli* and reached sclareol titers of ~1.5 g/L in high-cell-density fermentation. Our work provides a basis for the development of an alternative, sustainable, and cost-efficient route to sclareol and other diterpene analogues.

Sclareol (**4**) is a naturally occurring diterpene alcohol used extensively as starting material for the synthesis of fragrance molecules with ambergris notes, such as Ambrox (**5**). This semisynthetic route was originally developed to provide an alternative to ambergris, a biliary secretion of the sperm whale.[1] Ambergris is highly appreciated for its pleasant sweet and earthy scent and has been historically used as a perfume ingredient. Due to the increasing demand for ambergris, coupled with its short supply, the chemical synthesis of ambergris constituents and molecules with ambergris character has been the focus of extensive research in the flavor and fragrance industry. Among these molecules, Ambrox (**5**) is one of the most appreciated substitutes for ambergris.[2] More recently, sclareol (**4**) has also been reported to have anticancer effects, inducing apoptotic death in several human cancer cell lines.[3]

Currently, the main sources of sclareol are the flowers and leaves of *Salvia sclarea* (Clary sage), a biennial herb native of Southern Europe and belonging to the *Lamiaceae* family. In plants, diterpenes are derived from geranylgeranyl diphosphate (GGPP) (**1**) through cyclization reactions catalyzed by diterpene synthases (diTPS).[4] This genetically diverse family of enzyme is responsible for the large diversity of carbon skeletons found in the diterpene family. Two types of cyclization mechanism occur in nature and are attributed to two types of diTPS: the class I and class II diTPS.[5] Class I diTPS catalyze a cyclization/rearrangement of GGPP initiated by ionization of the diphosphate ester function, while in class II diTPS, the cyclization of GGPP is initiated by the protonation of the terminal double bond of GGPP and leads to a cyclic diterpene diphosphate intermediate. Often class II diTPS work in combination with a class I diTPS in two successive enzymatic reactions.[6] Finally, examples of diterpenes produced from GGPP by a single enzyme catalyzing successively the two types of reaction have been reported.[7,8] The biosynthesis of sclareol has previously been studied using crude extracts or partially purified protein fractions from sclareol-producing plants.[9,10] These biochemical experiments suggested that a single bifunctional synthase rather than two distinct diTPS is responsible for sclareol biosynthesis. However, the corresponding gene has so far not been isolated.

In the present study, we investigate at the molecular level the biosynthesis of sclareol in its native host, *S. sclarea*. We report the cloning and functional characterization of two diTPS that act sequentially in sclareol biosynthesis from the universal GGPP precursor. Furthermore, we heterologously expressed optimized versions of these two synthases into bacterial platform strains engineered to overproduce GGPP. Peak titers of ~400 mg/L and 1.5 g/L were obtained by cultivation in shake flasks and bench-scale bioreactors, respectively.

A polymerase chain reaction (PCR)-based approach was first used to isolate diTPS genes from a Clary sage flower complementary DNA (cDNA) library. Degenerate oligonucleotides were designed from conserved motifs in the amino acid sequences of a selection of class I and II diTPS from plant origin. This approach provided two variant cDNAs, SsTpsSa3 and SsTpsSa9, encoding for 785 and 784 amino acid proteins, respectively. The sequences of these two proteins differed by five amino acids and one residue insertion and showed strong homology with diTPS and namely with copalyl diphosphate synthase. The DxDD motif, involved in protonation-initiated cyclization, was present in these amino acid sequences, whereas the DDxxD motif, characteristic of ionization-dependent terpene synthases, was not found, suggesting that these enzymes were monofunctional class II diTPS. The recombinant diTPS were produced in *Escherichia coli*, purified, and evaluated *in vitro* for the conversion of GGPP. Consistent with the expected presence of a plastid targeting sequence, deletions of amino-terminal portions resulted in a strong increase of the levels of heterologous expression (Figure S4).[11] No diterpene product was detected by direct analysis of the *in vitro* assay. An alkaline phosphatase treatment followed by extraction and GC-MS analysis allowed the identification of labdenediol ((13*E*)-labdene-8α,15-diol (**3**)), the hydrolysis product of (13*E*)-8α-

Received: August 2, 2012
Published: October 31, 2012

dx.doi.org/10.1021/ja307404u | J. Am. Chem. Soc. 2012, 134, 18900−18903

hydroxylabden-15-yl diphosphate (LDPP) (**2**) (Figure 1 and Supporting Information). This showed that the diTPS encoded

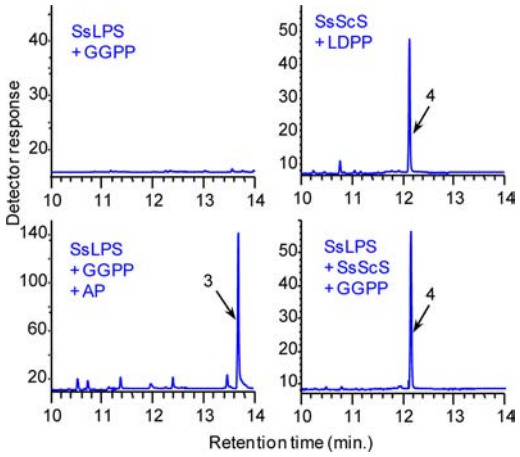
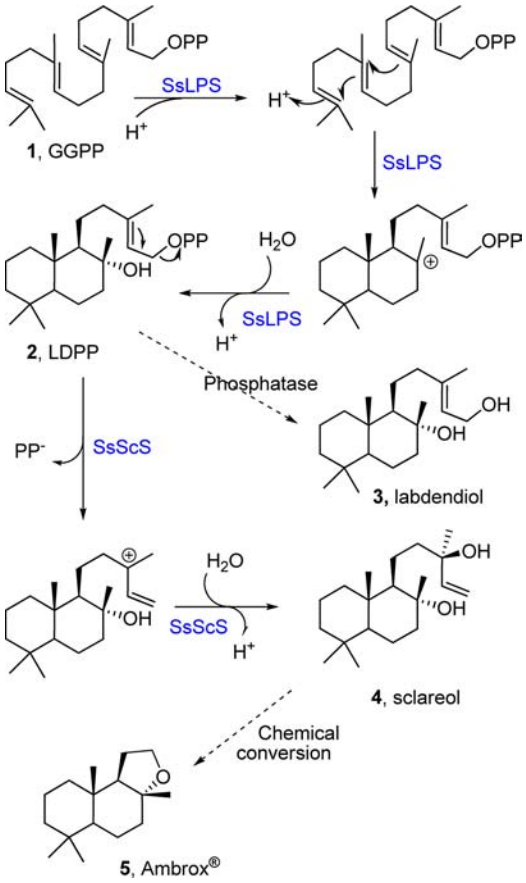


Figure 1. GC analysis of the *in vitro* assays with the recombinant *S. sclarea* diTPS. Incubation of the SsLPS protein encoded by the SsTpsSa9 or SsTpsSa3 cDNAs with GGPP (top left) and with GGPP followed by alkaline phosphatase (AP) treatment (bottom left). Incubation of the SsScS protein encoded by the SsTps1132 cDNA with LDPP (top right) and co-incubation of the SsLPS and SsScS proteins with GGPP (bottom right). Labdenediol (**3**) and sclareol (**4**) were identified by matching the retention times and mass spectra with authentic standards by GC-MS analysis (Figure S12). Sclareol (**4**) was also characterized by ^{13}C and ^{1}H NMR spectroscopy, which reveal the presence of an isomer of sclareol, 13-*epi*-sclareol. The 85:15 epimeric ratio of the mixture was similar to that of the plant-extracted products (Figure S15).

by SsTpsSa3 and SsTpsSa9 catalyze only the protonation-initiated cyclization of GGPP to LDPP (Scheme 1). We observed that LDPP was unstable, especially in acidic conditions, and that this property can lead to non-enzymatic formation of sclareol and therefore mischaracterization of the enzyme function (Figures S8 and S9). A class II diTPS with the same biochemical activity as SsTpsSa3 and SsTpsSa9 has recently been identified in the angiosperm *Cistus creticus*.[12] In view of our results, earlier findings of a single protein carrying the two-step GGPP-to-sclareol conversion should be re-evaluated, although the presence of a bifunctional sclareol synthase in other plant species is not excluded.

Following the discovery of a monofunctional class II diTPS involved in the first step of the conversion of GGPP to sclareol, we explored the *S. sclarea* flower transcriptome for class I diTPS. A homology-based PCR cloning strategy was attempted to isolate cDNAs encoding for class I diterpene synthases from the *S. sclarea* cDNA library. Degenerate oligonucleotides were deduced from conserved regions in the amino acid sequences of class I diTPS from several plants. Two regions in the N-terminal ends were used to design the forward oligonucleotides, and the DDxxD motif region located in the C-terminal part was used to design the reverse oligonucleotides. These primers were found to be ineffective for the amplification of diTPS-encoding cDNA fragments from the cDNA library. We therefore undertook a whole transcriptome sequencing approach exploiting the "next generation" sequencing technology developed by Illumina.[13] It is noteworthy that this sequencing was performed in 2007 and was one of the very first *de novo* transcriptome sequencing attempts using this technology. Compared to the performance of newer versions of the

Scheme 1. Successive Steps in the Conversion of GGPP (1) to Sclareol (4) by the Two Diterpene Synthases from *Salvia sclarea* (SsLPS and SsScS), and Structures of Ambrox (5) Derived Chemically from Sclareol, and of Labdenediol (3) Obtained after Hydrolysis of the Pyrophosphate Group of LDPP (2)



technology,[13] the sequencing resulted in low read numbers (1.9 million), short read lengths (35 bases), and low coverage of the transcriptome. Nevertheless, following *de novo* assembly, 2050 contigs of a maximum length of 1330 bases were obtained. Highly represented transcripts were almost fully re-assembled, and 8% of the SaTpsSa3 cDNA was reconstituted. The data could be exploited to recover several fragments of new diTPS genes, including one fragment containing the DDxxD motif characteristic of class I terpene synthases. Starting from these fragments, we assembled a full-length cDNA (SsTps1132) encoding for a novel diTPS. The protein sequence encoded by this cDNA was substantially shorter than any other known plant diTPS except a miltiradiene synthase from another *Salvia* specie.[14,15] The N-terminal region was shorter by 230−280 amino acids, and this domain was replaced by a 70-amino-acid sequence predicted as a plastid targeting sequence (Figure S2). This unusual feature and the low overall sequence homology with published sequences explain the failure to isolate this cDNA by PCR-based cloning. Besides the *Salvia* miltiradiene synthase, SsTps1132 is phylogenetically related to a heterogeneous group consisting of diterpene, sesquiterpene, and monoterpene synthases (Figure S10).

The SsTps1132 encoded protein was functionally expressed in *E. coli* cells as full-length or amino-terminal deleted proteins (to remove the predicted plastid-targeting sequence), and *in*

dx.doi.org/10.1021/ja307404u | *J. Am. Chem. Soc.* 2012, 134, 18900−18903

vitro enzymatic assays were performed. With GGPP as substrate, no enzymatic activity was observed (Supporting Information). Using LDPP as substrate, we observed the formation of sclareol as the unique product with both the full-length and the truncated enzymes. Enzymatic assays were also performed by co-incubation of the SsTps1132 and the SsTpsSa3 proteins with GGPP as substrate, in these conditions, the conversion of GGPP to sclareol was observed (Figures 1, S11, and S13). This series of experiments confirms that SaTps1132 is encoding for a sclareol synthase (SsScS) using LDPP as substrate.

Thus, using homology-based cloning and whole transcriptome sequencing combined with heterologous expression in *E. coli*, we isolated and functionally characterized two novel diTPS from *S. sclarea*, a LDPP synthase (SsLPS) and a sclareol synthase (SsScS).[16] Moreover, we showed that the biochemical conversion of GGPP to sclareol is a two-enzyme step process in this plant (Scheme 1). Given the high value of sclareol, a biotechnological route to this molecule using a microbial-based process could provide a cost-effective source of sclareol as an alternative or complement to the plant-based production. We therefore evaluated the potential of these genes for the production of sclareol from a cheap carbon source using *E. coli*, a widely used platform host for the production of biochemicals.

Microbial production of sclareol in high titers may be restricted by its potential toxicity to the producing host. This is particularly relevant for sclareol that has been shown to have a strong fungitoxic activity.[17] When added exogenously to the growth medium, sclareol shows low to no toxicity to *E. coli*. An analysis of the maximum specific growth rate μ_{max} showed that *E. coli* cell growth is comparable at 0 g/L and up to 2.0 g/L of exogenously added sclareol (Figure S16). The lack of toxicity to the producing host should enable the production of sclareol in high titers. Our initial attempt to synthesize sclareol from *E. coli* by co-expressing genes encoding for a geranylgeranyl diphosphate synthase (GGPPS) from *Pantoea agglomerans* (*CrtE* gene; Genbank accession AAA24819) together with truncated and codon-optimized versions of SsLPS and SsScS resulted in only a small amount of sclareol, ~4 mg/L in shake flask (Figure 2). To improve *in vivo* productivity, we undertook a metabolic engineering approach via overexpression of a heterologous mevalonate pathway into *E. coli*. We conceptually divided the mevalonate pathway into two portions: an upper pathway that converts acetyl-CoA into mevalonate, and a bottom pathway that converts mevalonate into isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), much as previously reported by the Keasling group.[18] As an upper pathway, we created a synthetic operon consisting of an acetoacetyl-CoA thiolase from *E. coli* encoded by *atoB* and a HMG-CoA synthase and a HMG-CoA reductase from *Staphylococcus aureus* encoded by *mvaA* and *mvaS*, respectively. A similar operon has previously been shown to sustain high-level production of mevalonate into *E. coli*.[19] As a lower mevalonate pathway, we selected a natural operon from the Gram-negative bacteria, *Streptococcus pneumoniae*, encoding a mevalonate kinase (*mvaK1*), a phosphomevalonate kinase (*mvaK2*), a phosphomevalonate decarboxylase (*mvaD*), and an isopentenyl diphosphate isomerase (*fni*).[20,21] Finally, a *Saccharomyces cerevisiae* farnesyl pyrophosphate synthase gene (*ERG20*) was introduced at the 3′-end of the upper pathway operon to convert IPP and DMAPP into FPP. Each operon was subcloned into one of the multiple-cloning sites of a low-copy expression plasmid under the control of a bacteriophage T7 promoter (pACYCDuet-1, Invitrogen). The FPP-overproducing strain was then adapted to the production of sclareol by co-expressing *CrtE* and the two cDNA-encoding SsLPS and SsScS. The three genes were co-expressed from a medium-copy plasmid (pETDuet-1, Invitrogen) under the control of strong T7 promoters. The resulting strain was then tested for sclareol production under shake flask conditions. To prevent air stripping of secreted sclareol, this study was carried out as liquid−liquid two-phase cultivation using a 10% (v/v) dodecane overlay. A peak titer of 418 ± 41 mg/L of sclareol was reached after 80 h of growth (72 h post-induction), which is more than 100-fold improvement over the basal, non-engineered strain (Figure 2). Interestingly, we detected a significant amount of the dephosphorylated form of LDPP, the product of SsLPS. We hypothesized that the presence of this intermediate may be due to its intracellular accumulation, leading to its dephosphorylation and subsequent excretion by the producing cell. In order to limit the diffusion of LDPP and favor substrate channeling between SsLPS and SsScS, we fused genetically the two synthases and expressed the chimeric construct in the GGPP-overproducing platform.[22] Surprisingly, the resulting strain accumulates extremely low titers of sclareol (29 ± 0.07 mg/L), suggesting that the hybrid enzyme did not express properly in the host (Figure 2).

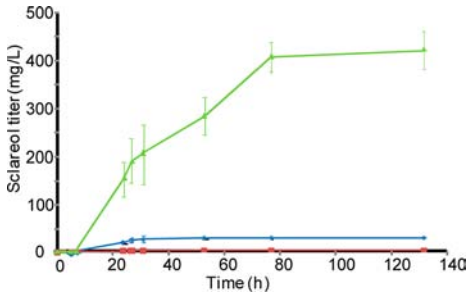


Figure 2. Time course profiles of sclareol accumulation in shake flask experiments. *E. coli* strain co-expressing the *CrtE* GGPPS, the LDPP synthase (SsLPS), and the sclareol synthase (SsScS) (red). Pathway-engineered *E. coli* strains co-expressing the *CrtE* GGPPS and either a chimeric SsLPS-SsScS synthase (blue) or the SsLPS and SsScS synthases as individual polypeptides (green). Data are mean ± SD of experiments performed in triplicate.

In order to further investigate the performance of the more productive *E. coli* strain, we carried out fed-batch cultivation experiments in laboratory-scale bioreactors under controlled conditions. Similarly to the shake flask experiments, the fed-batch cultivation studies were carried out using an organic solvent overlay.[23] In defined media with controlled glycerol feeding, sclareol titer reached a maximum of 1.46 ± 0.17 g/L in almost 2 days and ~20 h after the induction of sclareol production (Figure 3). The application of an indirect feedback control scheme that couples feeding with the concentration of dissolved oxygen allowed an automatic supply of glycerol at a restricted rate favoring sclareol formation and keeping acetate <1 g/L throughout the cultivation (Figure S17). Maximum cell densities >150 OD_{600nm} (equivalent to 60 g/L dry cell weight) were consistently achieved during these high cell density fermentations (Figure 3).

In conclusion, we have identified, in a commercially exploited species of Salvia, the terpene synthases responsible for the production of sclareol, an intermediate to a key fragrance

dx.doi.org/10.1021/ja307404u | J. Am. Chem. Soc. 2012, 134, 18900−18903

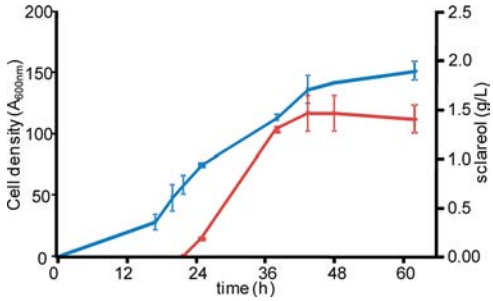


Figure 3. High cell density fermentation of mevalonate pathway engineered *E. coli* strains co-expressing the CrtE GGPPS, and the SsLPS and SsScS diTPS. Time courses of sclareol accumulation (red) and cell growth (blue) are shown. Data are mean ± SD of experiments performed in duplicate.

compound and potential anticancer molecule. Heterologous expression of optimized versions of these enzymes in a microbial strain engineered to overproduce GGPP led to a dramatic increase in sclareol titers reaching ~1.5 g/L in high cell density fermentation. To the best of our knowledge, this is the highest titer reported so far for a diterpene in a microbial host.[22,24−27] A second source of sclareol to increase the supply and stabilize the price of this key fragrance intermediate is highly desirable. Our work describes progress toward this goal by providing an alternative route to the conventional plant extraction process.

■ ASSOCIATED CONTENT

Supporting Information

Full experimental procedure, enzyme characterization data, and additional figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

Michel.Schalk@firmenich.com; Laurent.Daviet@firmenich.com

Present Address

[#]LCRDN, Lonza AG, Visp CH-3930, Switzerland

Notes

The authors declare the following competing financial interest(s): The authors declare an interest in FIRMENICH SA as employees of the R&D division. FIRMENICH SA is engaged in the production and commercialization of high-value natural and synthetic ingredients for the flavor and fragrance industry.

■ ACKNOWLEDGMENTS

We thank Christopher Dean and Andreas Taglieber (Firmenich) for their critical reading of the manuscript, and Wolfgang Fieber, Horst Sommer, Joëlle Porret, Robert Brauchli (Firmenich), and Guillaume Marti (University of Geneva) for their help with the GC-MS and NMR analyses of LDPP and sclareol.

■ REFERENCES

(1) Decorzant, R.; Vial, C.; Näf, F.; Whitesides, G. M. *Tetrahedron* **1987**, *43*, 1871−1879.

(2) Chauffat, C.; Morris, A. *Perfum. Flavor.* **2004**, *29*, 34−41.

(3) Mahaira, L. G.; Tsimplouli, C.; Sakellaridis, N.; Alevizopoulos, K.; Demetzos, C.; Han, Z.; Pantazis, P.; Dimas, K. *Eur. J. Pharmacol.* **2011**, *666*, 173−182.

(4) Dewick, P. M. *Nat. Prod. Rep.* **2002**, *19*, 181−222.

(5) Wendt, K. U.; Schultz, G. E. *Structure* **1994**, *6*, 127−133.

(6) Peters, R. J. *Nat. Prod. Rep.* **2010**, *27*, 1521−1530.

(7) Vogel, B. S.; Wildung, M. R.; Vogel, G.; Croteau, R. *J. Biol. Chem.* **1996**, *271*, 23262−23268.

(8) Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. *Plant J.* **2011**, *66*, 212−229.

(9) Banthorpe, D. V.; Brown, J. T.; Morris, G. S. *Phytochemistry* **1992**, *31*, 3391−3395.

(10) Guo, Z.; Severson, R. F.; Wagner, G. J. *Arch. Biochem. Biophys.* **1994**, *308*, 103−108.

(11) Prisic, S.; Peters, R. J. *Plant. Physiol.* **2007**, *144*, 445−454.

(12) Falara, V.; Pichersky, E.; Kanellis, A. K. *Plant. Physiol.* **2010**, *154*, 301−310.

(13) Bennett, S. *Pharmacogenomics* **2004**, *4*, 433−438. A sequencing of the same plant material was performed in 2012 with the Illumina technology, and over 23 million of 2 × 100 bases paired-end reads were obtained. The *de novo* assembly resulted in 45 241 unique transcript sequences with an average length of 1300 bases (data not shown).

(14) Gao, W.; Hillwig, M. L.; Huang, L.; Cui, G.; Wang, X.; Kong, J.; Yang, B.; Peters, R. J. *Org. Lett.* **2009**, *11*, 5170−5173.

(15) Hillwig, M. L.; Xu, M.; Toyomasu, T.; Tiernan, M. S.; Wei, G.; Cui, G.; Huang, L.; Peters, R. J. *Plant J.* **2010**, *68*, 1051−1060.

(16) cDNA sequences described in this paper have been submitted to GenBank with accession numbers JN133923 (SsTpsSa3), JN133924 (SsTpsSa9), and JN133922 (SsTps1132).

(17) van den Brûle, S.; Müller, A.; Fleming, A. J.; Smart, C. C. *Plant J.* **2002**, *30*, 649−662.

(18) Martin, V. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. *Nat. Biotechnol.* **2003**, *21*, 796−802.

(19) Tsuruta, H.; Paddon, C. J.; Eng, D.; Lenihan, J. R.; Horning, T.; Anthony, L. C.; Regentin, R.; Keasling, J. D.; Renninger, N. S.; Newman, J. D. *PLoS One* **2009**, *4*, e4489.

(20) Wilding, E. I.; Brown, J. R.; Bryant, A. P.; Chalker, A. F.; Holmes, D. J.; Ingraham, K. A.; Iordanescu, S.; So, C. Y.; Rosenberg, M.; Gwynn, M. N. *J. Bacteriol.* **2000**, *182*, 4319−4327.

(21) Yoon, S. H.; Lee, S. H.; Das, A.; Ryu, H. K.; Jang, H. J.; Kim, J. Y.; Oh, D. K.; Keasling, J. D.; Kim, S. W. *J. Biotechnol.* **2009**, *140*, 218−226.

(22) Zhou, Y. J.; Gao, W.; Rong, Q.; Jin, G.; Chu, H.; Liu, W.; Yang, W.; Zhu, Z.; Li, G.; Zhu, G.; Huang, L.; Zhao, Z. K. *J. Am. Chem. Soc.* **2012**, *134*, 3234−3241.

(23) Newman, J. D.; Marshall, J.; Chang, M.; Nowroozi, F.; Paradise, E.; Pitera, D.; Newman, K. L.; Keasling, J. D. *Biotechnol. Bioeng.* **2006**, *95*, 684−691.

(24) Leonard, E.; Ajikumar, P. K.; Thayer, K.; Xiao, W. H.; Mo, J. D.; Tidor, B.; Stephanopoulos, G.; Prather, K. L. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13654−13659.

(25) Ajikumar, P. K.; Xiao, W. H.; Tyo, K. E.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. *Science* **2010**, *330*, 70−74.

(26) Cyr, A.; Wilderman, P. R.; Determan, M.; Peters, R. J. *J. Am. Chem. Soc.* **2007**, *129*, 6684−6685.

(27) Morrone, D.; Lowry, L.; Determan, M. K.; Hershey, D. M.; Xu, M.; Peters, R. J. *Appl. Microbiol. Biotechnol.* **2010**, *85*, 1893−1906.

■ NOTE ADDED AFTER ASAP PUBLICATION

After publishing ASAP on November 7, 2012, a graphic citation and a reference citation were corrected, this reposted November 13, 2012.