FEBS Letters 488 (2001) 170–173 FEBS 24507

Identification of *gcpE* as a novel gene of the 2-*C*-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis in *Escherichia coli*

Narciso Campos[a], Manuel Rodríguez-Concepción[a], Myriam Seemann[b], Michel Rohmer[b], Albert Boronat[a,*]

[a] *Department de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, C/Martí i Franquès 1, 08028 Barcelona, Spain*
[b] *Université Louis Pasteur/CNRS, Institut Le Bel, 4 rue Blaise, 67070 Strasbourg Cedex, France*

Received 24 November 2000; accepted 19 December 2000

First published online 2 January 2001

Edited by Matti Saraste

Abstract The 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway for isoprenoid biosynthesis is essential in most eubacteria and plants and has remarkable biotechnological interest. However, only the first steps of this pathway have been determined. Using bioinformatic and genetic approaches, we have identified *gcpE* as a novel gene of the MEP pathway. The distribution of this gene in bacteria and plants strictly parallels that of the gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, which catalyses the first committed step of the MEP pathway. Our data demonstrate that the *gcpE* gene is essential for the MEP pathway in *Escherichia coli* and indicate that this gene is required for the trunk line of the isoprenoid biosynthetic route.

Key words: Isoprenoid biosynthesis; 2-*C*-methyl-D-erythritol 4-phosphate pathway; *GcpE*; Mevalonic acid

1. Introduction

Isoprenoids are the most structurally diverse family of compounds found in nature. More than 30 000 isoprenoid molecules are known to date [1]. Many isoprenoids have biotechnological applications as drugs, flavours, pigments, perfumes or agrochemicals. In living beings, isoprenoids have important roles in processes as diverse as electron transport, reproduction, growth regulation, signal transduction and defence. Isoprenoid biosynthesis is essential in all organisms. In *Escherichia coli*, for instance, the isoprenoids ubiquinone and bactoprenol are required, respectively, for respiration and cell wall biosynthesis. In spite of the structural and functional diversity of isoprenoids, all derive from the five-carbon isomers isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) which, altogether, constitute the basic building block of isoprenoids.

Two alternative pathways for the synthesis of IPP and DMAPP exist in nature: the well known mevalonate (MVA) pathway and the recently discovered 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway [2–4]. Most eubacteria, including *E. coli*, have only the MEP pathway. The initial step of this pathway involves the formation of 1-deoxy-D-xylulose 5-phosphate (DXP) from pyruvate and glyceraldehyde 3-phosphate (Fig. 1). This reaction is catalysed by 1-deoxy-D-xylulose 5-phosphate synthase (DXS), encoded by the *dxs* gene in *E. coli* [5,6]. DXP can be converted to isoprenoids, but also to thiamine and pyridoxal. Therefore, the following reaction, consisting in the conversion of DXP to MEP, is most probably the first committed step of the pathway. This reaction is catalysed by 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) which is encoded by the *dxr* gene [7]. In the next step, the *ygbP* gene product mediates the condensation of MEP with CTP to form 4-diphosphocytidyl-2-*C*-methyl-D-erythritol (CDP-ME) [8,9]. CDP-ME is then phosphorylated by the *ychB* gene product to 4-diphosphocytidyl-2-*C*-methyl-D-erythritol 2-phosphate (CDP-ME2P) [10,11]. The *ygbB* gene product catalyses the elimination of the CMP moiety from CDP-ME2P to yield 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cPP) [12,13] which is the last intermediate of the pathway known to date. Labelling studies indicated that, in *E. coli*, IPP and DMAPP are synthesised from MEP by routes that separate at some point [14–16]. Genetic studies confirmed the operation of separate routes for the synthesis of IPP and DMAPP in *E. coli* [17] and showed that the reactions catalysed by the *ygbP*, *ychB* and *ygbB* gene products belong to the trunk line of the pathway [18] (Fig. 1). These evidences led to conclude that the branching of the MEP pathway occurs after ME2,4-cPP [18].

To identify novel genes of the MEP pathway, we engineered *E. coli* for the synthesis of IPP and DMAPP from MVA [18] (Fig. 1). The presence of this engineered route for the production of isoprenoid units in the bacterium allows the rescue of otherwise lethal mutants of the MEP pathway. The growth of these mutants is dependent on the exogenous supply of MVA. In the present work, we use this system to demonstrate that the *gcpE* gene, which is broadly distributed in eubacteria and plants, is essential for the operation of the MEP pathway in *E. coli*.

*Corresponding author. Fax: (34)-93-4021219.
E-mail: aboronat@sun.bq.ub.es

Abbreviations: *CAT*, chloramphenicol acetyl transferase; CDP-ME, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol; CDP-ME2P, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol 2-phosphate; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; IPP, isopentenyl diphosphate; ME-2,4cPP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-*C*-methyl-D-erythritol 4-phosphate; MVA, mevalonate; ORF, open reading frame

0014-5793/01/$20.00
PII: S0014-5793(00)02420-0

2. Materials and methods

2.1. Deletion of the gcpE gene in strain EcAB3-1

Strain EcAB3-1 is a *recD* mutant host which allows efficient recombination of linear DNA with homologous sequences in the chromosome (Table 1). To delete the *gcpE* coding sequence in the *E. coli* genome, a recombination construct was prepared in which the *CAT* gene was cloned between the *gcpE* flanking regions (Fig. 2). Substitution of the *CAT* gene for the *gcpE* coding sequence in the genome was selected by chloramphenicol resistance.

Four PCR reactions were necessary to prepare the recombination construct. First, a genomic region of 3231 bp, encompassing the *gcpE* coding sequence (1116 bp), together with the corresponding flanking regions, was amplified using genomic DNA from strain MC4100 [19] as template and the oligonucleotides 1PE and 4PE as primers (Table 2). Next, the PCR product obtained was used as template to amplify the *gcpE* flanking regions. Primers 1PE and 22PE were used to amplify the 5′ flanking region. In this PCR, primer 22PE generated a *Sma*I restriction site. Primers 3PE and 4PE were used to amplify the 3′ flanking region. In this PCR, primer 3PE generated a *Pme*I restriction site. The PCR product corresponding to the 3′ flanking region (1061 bp) was cloned in the *Sma*I restriction site of plasmid pBluescript SK+. The plasmid obtained was named GC3. Subsequently, the PCR product corresponding to the 5′ flanking region (1102 bp) was cloned in the *Pme*I restriction site of plasmid GC3. The plasmid containing the two flanking regions was named GC53. The relative orientation of the 3′ and 5′ flanking regions in this plasmid was the same as in the *E. coli* genome. Finally, the *CAT* gene was amplified by PCR using plasmid pCAT19 as template [20] and the oligonucleotides CAT1 and CAT4 as primers. The PCR product (960 bp) was cloned in the *Sma*I restriction site of plasmid GC53. The construct obtained was named GC5CAT3. In this construct, the *CAT* gene had the same orientation as the coding sequence of the *gcpE* gene previously deleted. Since the *CAT* gene contained its own promoter elements and coding sequence, but did not contain any transcription termination signal, no polar effect downstream of the *gcpE* gene was expected.

Plasmid containing the construct GC5CAT3 was digested with *Hin*dIII, *Xba*I and *Xho*I restriction enzymes to release the recombination cassette (Fig. 2). This cassette was amplified by PCR using oligonucleotides 1PE and 4PE as primers. The PCR product was used to transform electrocompetent cells of strain EcAB3-1. The cells were plated on 2×TY medium [21] containing 17 μg/ml chloramphenicol, 6 μg/ml tetracycline, 25 μg/ml kanamycin, 0.2% (w/v) L-arabinose and 1 mM MVA. The presence of the *CAT* gene instead of the *gcpE* coding sequence in the genome of transformants was confirmed by PCR using oligonucleotides 0PE and 5PE as primers. The identity of the PCR product was verified by restriction analysis. Oligonucleotides 0PE and 5PE are complementary to sequences located outside of the region included in the deletion construct (Fig. 2). Therefore, the analysis of transformants confirmed both the absence of the original *gcpE* gene and the presence of the *CAT* gene. The novel strain lacking the *gcpE* gene was named EcAB3-3 (Table 1).

3. Results and discussion

3.1. Identification of the gcpE gene

To identify genes potentially involved in the MEP pathway, we followed a bioinformatic approach. Since bacterial genes with related functions are often organised in operons, we examined uncharacterised open reading frames (ORFs) that are beside known genes of the MEP pathway. An ORF of 1195 bp with unknown function was found just upstream of a DXS coding sequence of *Streptomyces coelicolor* (cosmid 6A5, accession number AL049485). This ORF is homologous to the *gcpE* gene of *E. coli* [22] and the *aarC* gene of *Providencia stuartii* [23]. Previous data indicate that *gcpE* and *aarC* are essential genes [22,23]. The *aarC* gene is required for the cell density dependent regulation of the 2′-*N*-acetyltransferase, an

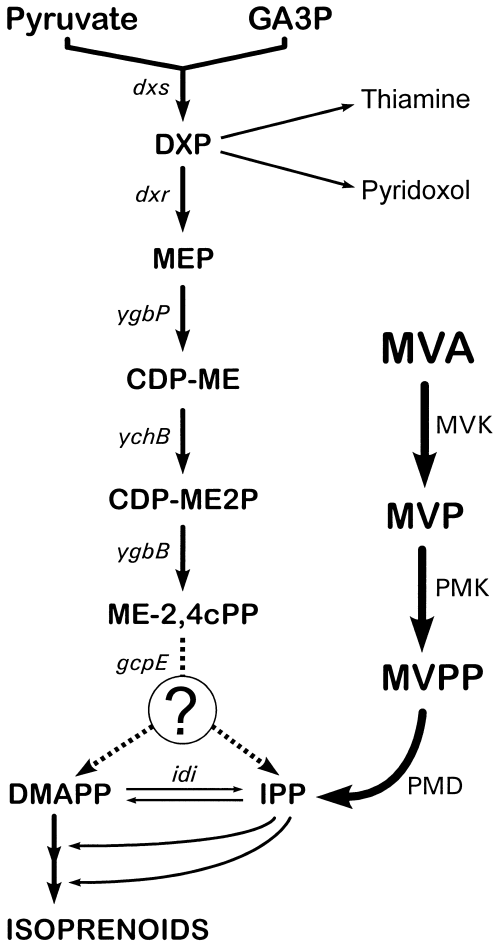

Fig. 1. The partial MVA pathway engineered in *E. coli* and the endogenous MEP pathway. The intermediates of the MEP pathway (on the left) are as follows: GA3P, glyceraldehyde 3-phosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-*C*-methyl-D-erythritol 4-phosphate; CDP-ME, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol; CDP-ME2P, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol 2-phosphate; ME-2,4cPP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate. The genes encoding the known enzymes of the MEP pathway and the *gcpE* gene are indicated. The intermediates of the engineered mevalonate pathway (on the right) are as follows: MVA, mevalonic acid; MVP, 5-phosphomevalonate; MVPP, 5-diphosphomevalonate. The enzymes of the mevalonate pathway are as follows: MVK, mevalonate kinase; PMK, 5-phosphomevalonate kinase; and PMD, 5-diphosphomevalonate decarboxylase. The *idi* gene encodes isopentenyl diphosphate isomerase.

Table 1
E. coli strains used in this study

Strain	Description
EcAB3-1	TE2680 MVA$^+$ [18]; parent strain for the generation of mutants of the MEP pathway
EcAB3-2	EcAB3-1 *dxs*::*CAT* [18]
EcAB3-3	EcAB3-1 *gcpE*::*CAT*

Table 2
Oligonucleotides used in this study

Name	Sequence
CAT1	5′-gagtccgaataaatacctgtg-3′
CAT4	5′-ccgaatttctgccattcatcc-3′
0PE	5′-tgggctttgtcacgagcacac-3′
1PE	5′-cgcggtgtgggtgagcatgatg-3′
22PE	5′-aaatctcccgggttacccgtctgttactgc-3′
3PE	5′-gcgtttaaactggacgaagcgcgtcgaattgac-3′
4PE	5′-tgcacgaccgcccagttgttcc-3′
5PE	5′-ggcccatagcaaaaccgacag-3′

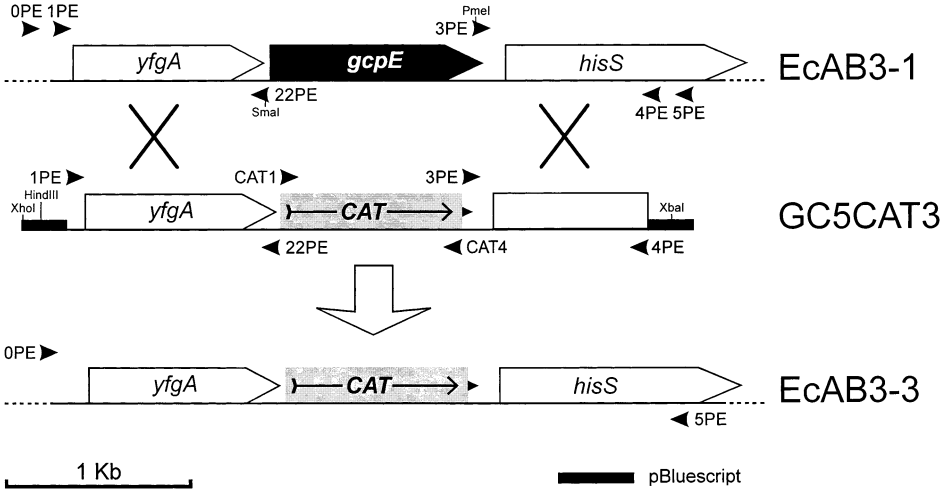

Fig. 2. Deletion of the *gcpE* gene in the *E. coli* genome. The *CAT* gene was substituted for the *gcpE* gene by homologous recombination. The construct GC5CAT3, containing the *CAT* gene surrounded by the *gcpE* flanking regions, was recombined with the *gcpE* genomic region of strain EcAB3-1. The resulting strain was named EcAB3-3. Arrowheads indicate the position of primers used to prepare the recombination construct and to analyse the mutated locus. Restriction enzyme sites are indicated.

enzyme involved in the O acetylation of peptidoglycans. However, no precise biochemical function has been assigned so far to the products of the *aarC* or the *gcpE* genes.

Genes with homology to *gcpE* are broadly distributed in eubacteria and plants. The occurrence of the *gcpE* gene in completely sequenced bacterial genomes strictly correlates with the occurrence of the *dxr* gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, which catalyses the first committed step in the MEP pathway (Fig. 1, Table 3). Nineteen out of 29 sequenced genomes contain both *dxr* and *gcpE* genes. Ten of the sequenced genomes do not contain *dxr* nor *gcpE*. This correlation suggested a close link between the *gcpE* gene and the MEP pathway. The identity between the *E. coli gcpE* gene product and the corresponding homologues of other bacteria ranges from 88.7% (*P. stuartii*) to 33.3% (*Xylella fastidiosa*). The similarity in the equivalent comparisons ranges from 95.2% (*P. stuartii*) to 54.0% (*X. fastidiosa*). Using the *E. coli gcpE* gene product as a query and the default parameters of the TBLASTN program accessible at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/), EST entries corresponding to *gcpE* homologues were detected in *Arabidopsis thaliana*, *Glycine max*, *Lycopersicon esculentum*, *Lycopersicon hirsutum*, *Medicago trunculata*, *Physcomitrella patens*, *Pinus taeda*, *Porphyra yezoensis*, *Secale cereale*, *Solanum tuberosum*, *Triticum aestivum* and *Zea mays* (data not shown). The presence of *gcpE* gene homologues in plants is consistent with a role of this gene in the MEP pathway. Also in agreement with this hypothesis, no homologue of the *gcpE* gene is present in the eukaryote *Saccharomyces cerevisiae*, in which all isoprenoids derive from MVA (Table 3).

3.2. Deletion of the gcpE gene in E. coli

To examine whether the *gcpE* gene of *E. coli* was indeed involved in the MEP pathway, we deleted this gene in strain EcAB3-1 (Fig. 2). Lethal mutants of the MEP pathway can be rescued in this strain by supplementation with MVA [18]. Conversion of MVA to IPP and DMAPP in strain EcAB3-1 and derivatives is mediated by a synthetic operon encoding yeast 5-diphosphomevalonate decarboxylase, human 5-phosphomevalonate kinase, yeast mevalonate kinase and *E. coli* isopentenyl diphosphate isomerase stably integrated in the chromosome. This operon is under control of the P_{BAD} promoter which can be induced with L-arabinose. Deletion of the *gcpE* gene was accomplished by homologous recombination between the genome and the construct GC5CAT3 in which a chloramphenicol resistance gene (*CAT*) is surrounded by the *gcpE* flanking regions (Fig. 2). Substitution of the *CAT* gene

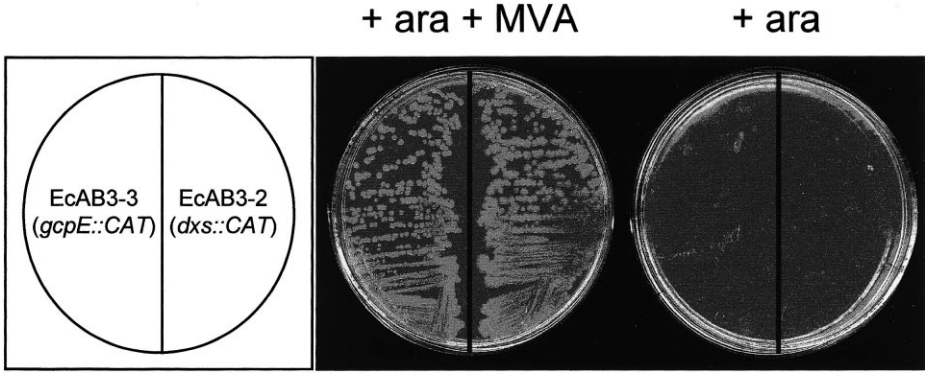

Fig. 3. Complementation of the *gcpE* mutant with MVA. Strains EcAB3-2 and EcAB3-3 were grown at 37°C in 2×TY medium [21] containing 1.5% agar (w/v), 6 μg/ml tetracycline, 25 μg/ml kanamycin, 17 μg/ml chloramphenicol and 0.2% (w/v) L-arabinose. MVA (1 mM) was present in the medium as indicated.

Table 3
Occurrence of *gcpE* and *dxr* genes in completely sequenced genomes[a]

Organism	*gcpE*	*dxr*
Eubacteria		
Aquifex aeolicus VF5	AE000745	AE000688
Bacillus halodurans C-125	AP001511	AP001515
Bacillus subtilis 168	Z99116	Z99112
Buchnera sp. APS	AP001118	A001118
Campylobacter jejuni NCTC 11168	AL139076	AL139078
Chlamydia pneumoniae CWL029	AE001621	AE001618
Chlamydia trachomatis D/UW-3/Cx	AE001280	AE001281
Deinococcus radiodurans R1	AE001898	AE001994
E. coli K-12 MG1655	AE000338	AE000126
Haemophilus influenzae Rd KW20	U32721	U32763
Helicobacter pylori 26695	AE000577	AE000541
Mycobacterium tuberculosis H37Rv	AL008883	Z74024
Neisseria meningitidis MC58	AE002479	AE002375
Pseudomonas aeruginosa PA01	AE004798	AE004785
Synechocystis sp. PCC 6803	D90908	D64000
Thermotoga maritima MSB8	AE001754	AE001754
Treponema pallidum Nichols	AE001221	AE001235
Vibrio cholerae N16961	AE004161	AE004297
X. fastidiosa 9a5c	AE004064	AE003942
Borrelia burgdorferi B31	–	–
Mycoplasma genitalium G-37	–	–
Mycoplasma pneumoniae M129	–	–
Rickettsia prowazekii Madrid E	–	–
Archea		
Aeropyrum pernix K1	–	–
Archaeoglobus fulgidus DSM4304	–	–
Methanobacterium thermoautotrophicum delta H	–	–
Methanococcus jannaschii DSM2661	–	–
Pyrococcus abyssi GE5	–	–
Pyrococcus horikoshii OT3	–	–
Eukaryotes		
S. cerevisiae S288C	–	–

The accession numbers at the NCBI database are indicated.
[a]Updated on 6th October, 2000.

for the *gcpE* coding sequence in the genome of transformants was confirmed by PCR and restriction analysis (see Section 2 for details). The *E. coli* strain lacking the *gcpE* gene was named EcAB3-3.

As shown in Fig. 3, the presence of MVA in the medium was required for growth of strain EcAB3-3. This metabolite also restored growth of the control strain EcAB3-2 in which the *dxs* gene, encoding the first enzyme of the MEP pathway, was disrupted (Table 1). The result demonstrates that the *gcpE* gene product is required for the operation of the MEP pathway. The *gcpE* gene is essential in *E. coli* and is likely to be essential as well in other bacteria in which the MEP pathway is the only alternative for the synthesis of IPP and DMAPP. The possibility that the *gcpE* gene is essential in other bacteria and plants where the MVA and the MEP pathways coexist remains to be investigated. As mentioned above, the MEP pathway is branched in *E. coli* (Fig. 1). It can be concluded that the *gcpE* gene product is involved in the trunk line of the MEP pathway. If the *gcpE* gene were required for one of the branches alone, deletion of this gene would not be lethal. A block produced in a single branch of the MEP pathway would be circumvented by isopentenyl diphosphate isomerase [18] (Fig. 1).

The biochemical function of the *gcpE* gene product is presently unknown. No significant homology to other characterised proteins was found in databank searches using the BLAST software accessible at the NCBI. In addition, no characteristic motif that might suggest a function was found with the PROSITE program accessible at the Expasy Molecular Biology Server (http://www.expasy.ch/). It is possible that the *gcpE* gene product has a novel catalytic activity, either alone or in combination with other protein subunits. Alternatively, the gcpE protein might have a regulatory role in the MEP pathway. However, this possibility appears unlikely because of the broad distribution of *gcpE* homologues in eubacteria and plants. Experiments are being conducted at present to determine the function of the *gcpE* gene.

Acknowledgements: We thank Dr. K. Poralla for bringing to our notice the presence of a DXS coding sequence in cosmid 6A5 from *S. coelicolor*. This work was supported by grants from the Spanish Ministerio de Educación y Cultura (DGICYT, PB96-0176), the Generalitat de Catalunya (CIRIT, 1999SGR00032) and the Institut Universitaire de France.

References

[1] (1996) Dictionary of Natural Products on CD-ROM (v. 5.1), Chapman and Hall, London.
[2] Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M.H. and Bacher, A. (1998) Chem. Biol. 5, R221–R233.
[3] Rohmer, M. (1999) Nat. Prod. Rep. 16, 565–574.
[4] Lichtenthaler, H.K. (1999) Annu. Rev. Plant Physiol. Plant Mol. Biol. 50, 47–65.
[5] Sprenger, G.A., Schorken, U., Wiegert, T., Grolle, S., de Graaf, A.A., Taylor, S.V., Begley, T.P., Bringer-Meyer, S. and Sahm, H. (1997) Proc. Natl. Acad. Sci. USA 94, 12857–12862.
[6] Lois, L.M., Campos, N., Rosa-Putra, S., Danielsen, K., Rohmer, M. and Boronat, A. (1998) Proc. Natl. Acad. Sci. USA 95, 2105–2110.
[7] Takahashi, S., Kuzuyama, T., Watanabe, H. and Seto, H. (1998) Proc. Natl. Acad. Sci. USA 95, 9879–9884.
[8] Rohdich, F., Wungsintaweekul, J., Fellermeier, M., Sagner, S., Herz, S., Kis, K., Eisenreich, W., Bacher, A. and Zenk, M.H. (1999) Proc. Natl. Acad. Sci. USA 96, 11758–11763.
[9] Kuzuyama, T., Takagi, M., Kaneda, K., Dairi, T. and Seto, H. (2000) Tetrahedron Lett. 41, 703–706.
[10] Lüttgen, H., Rohdich, F., Herz, S., Wungsintaweekul, J., Hecht, S., Schuhr, C.A., Fellermeier, M., Sagner, S., Zenk, M.H., Bacher, A. and Eisenreich, W. (2000) Proc. Natl. Acad. Sci. USA 97, 1062–1067.
[11] Kuzuyama, T., Takagi, M., Kaneda, K., Watanabe, H., Dairi, T. and Seto, H. (2000) Tetrahedron Lett. 41, 2925–2928.
[12] Herz, S., Wungsintaweekul, J., Schuhr, C.A., Hecht, S., Lüttgen, H., Sagner, S., Fellermeier, M., Eisenreich, W., Zenk, M.H., Bacher, A. and Rohdich, F. (2000) Proc. Natl. Acad. Sci. USA 97, 2486–2490.
[13] Takagi, M., Kuzuyama, T., Kaneda, K., Watanabe, H., Dairi, T. and Seto, H. (2000) Tetrahedron Lett. 41, 3395–3398.
[14] Giner, J.L., Jaun, B. and Arigoni, B. (1998) Chem. Commun. 17, 1857–1858.
[15] Leyes, A.E., Baker, J.A., Hahn, F.M. and Poulter, C.D. (1999) Chem. Commun. 8, 717–718.
[16] Charon, L., Hoeffler, J.F., Pale-Grosdemange, C., Lois, L.M., Campos, N., Boronat, A. and Rohmer, M. (2000) Biochem. J. 346, 737–742.
[17] Rodríguez-Concepción, M., Campos, N., Lois, L.M., Maldonado, C., Hoeffler, J.F., Grosdemange-Billiard, C., Rohmer, M. and Boronat, A. (2000) FEBS Lett. 473, 328–332.
[18] Campos, N., Rodríguez-Concepción, M., Sauret-Güeto, S., Gallego, F., Lois, L.M. and Boronat, A. (2000) Biochem. J., in press.
[19] Casadaban, M.J. (1976) J. Mol. Biol. 104, 541–555.
[20] Fuqua, W.C. (1992) BioTechniques 12, 223–225.
[21] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Short Protocols in Molecular Biology, Wiley, New York.
[22] Baker, J., Franklin, D.B. and Parker, J. (1992) FEMS Microbiol. Lett. 94, 175–180.
[23] Rather, P.N., Solinsky, K.A., Paradise, M.R. and Parojcic, M.M. (1997) J. Bacteriol. 179, 2267–2273.