Functional analysis of the *Arabidopsis thaliana* GCPE protein involved in plastid isoprenoid biosynthesis

Jordi Querol, Narciso Campos, Santiago Imperial, Albert Boronat, Manuel Rodríguez-Concepción*

Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1–7, 08028 Barcelona, Spain

Received 23 January 2002; revised 31 January 2002; accepted 31 January 2002

First published online 20 February 2002

Edited by Ulf-Ingo Flügge

Abstract Plastid isoprenoids are synthesized via the 2-*C*-methyl-D-erythritol 4-phosphate pathway. A few years after its discovery, most of the *Escherichia coli* genes involved in the pathway have been identified, including *gcpE*. In this work, we have identified an *Arabidopsis thaliana* protein with homology to the product of this gene. The plant polypeptide, GCPE, contains two structural domains that are absent in the *E. coli* protein: an N-terminal extension and a central domain of 30 kDa. We demonstrate that the N-terminal region targets the *Arabidopsis* protein to chloroplasts in vivo, consistent with its role in plastid isoprenoid biosynthesis. Although the presence of the internal extra domain may have an effect on activity, the *Arabidopsis* mature GCPE was able to complement a *gcpE*-defective *E. coli* strain, indicating the plant protein is a true functional homologue of the bacterial *gcpE* gene product.

Key words: GCPE; Isoprenoid biosynthesis; 2-*C*-Methyl-D-erythritol 4-phosphate pathway; Plastid targeting; *Arabidopsis*

1. Introduction

Isoprenoids (also called terpenoids) are present in all organisms but are especially abundant and diverse in plants [1–3]. Tens of thousands of plant isoprenoids have been reported to date, many of which have biotechnological interest as flavors, pigments, drugs, or agrochemicals. In spite of their diversity of functions and structures, all isoprenoids derive from the common building units isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which can be interconverted by the enzyme IPP isomerase (Fig. 1). Isoprenoid biosynthesis in plants is unique because two independent biosynthetic pathways coexist within the cell [4–6]. Mitochondrial and cytosolic isoprenoids (including ubiquinone, sterols, and the hormones cytokinins and brassinosteroids) derive from IPP synthesized from mevalonic acid (MVA). By contrast, the biosynthesis of plastidial IPP and derived isoprenoids (such as carotenoids, monoterpenes, the side chain of chlorophylls, plastoquinone, and tocopherol, and the hormones abscisic acid and gibberellins) is accomplished through a completely different pathway [4–6]. Despite its essential role in plant biology, the plastid pathway was completely overlooked until very recently. This now uncovered MVA-independent pathway, the only one present in most eubacteria, is best characterized in *Escherichia coli* (reviewed in [4,7]) (Fig. 1). The initial reaction forms 1-deoxy-D-xylulose 5-phosphate (DXP) by condensation of D-glyceraldehyde 3-phosphate with (hydroxyethyl)thiamine derived from pyruvate in a reaction catalyzed by DXP synthase (DXS). The enzyme DXP reductoisomerase (DXR) then synthesizes 2-*C*-methyl-D-erythritol 4-phosphate (MEP) by intramolecular rearrangement and reduction of DXP. Since DXP can be used for the synthesis of vitamins B_1 (thiamine) and B_6 (pyridoxol), MEP is actually the first committed intermediate of the pathway (Fig. 1) and it is therefore accepted to name the pathway accordingly. In the following reactions of the MEP pathway, the enzymes encoded by the *E. coli* genes *ygbP*/*ispD*, *ychB*/*ispE* and *ygbB*/*ispF* mediate the sequential formation of 4-diphosphocytidyl-2-*C*-methyl-D-erythritol (CDP-ME), CDP-ME 2-phosphate and 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate (ME-cPP), respectively (reviewed in [4,7]). Although the remaining enzymatic reactions leading to the synthesis of IPP and DMAPP remain to be fully elucidated, there is strong evidence for the involvement of two more genes in the pathway: *gcpE* [8,9] and *lytB* [10–12]. The data reported to date suggest that the product of the *gcpE* gene participates in the conversion of ME-cPP into hydroxymethylbutenyl 4-diphosphate [13–15], a compound that accumulates in *lytB*-deficient *E. coli* cells [16] (Fig. 1).

In spite of the impressive progress on the elucidation of the MEP pathway in *E. coli*, further work is also required to identify the components in other organisms, including plants. The best characterized MEP pathway enzymes in plants are DXS and DXR [17–29], but proteins from *Arabidopsis thaliana* and tomato (*Lycopersicon esculentum*) with homology to those encoded by the *E. coli ygbP* and *ychB* genes, respectively, have also been shown to have the predicted enzymatic activities [30,31]. With more and more plant 'expressed sequence tags' (ESTs) deposited in the databases and the *Arabidopsis* genome completely sequenced, it is becoming easier to identify putative plant homologues to the reported *E. coli* genes involved in the MEP pathway. However, only function-

*Corresponding author: Fax: (34)-93-402 1219.
E-mail address: mrodrigu@sun.bq.ub.es (M. Rodríguez-Concepción).

Abbreviations: DXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-*C*-methyl-D-erythritol 4-phosphate; CDP-ME, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol; ME-cPP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; MVA, mevalonic acid; GcpE, *Escherichia coli gcpE* gene product; GCPE, *Arabidopsis* homologue of GcpE

0014-5793/02/$22.00
PII: S0014-5793(02)02402-X

al analysis of each proposed plant protein can ascertain its role in the biosynthesis of plastid isoprenoids. This work addresses the identification and functional analysis of the *Arabidopsis* GCPE protein, the first reported plant homologue of the bacterial *gcpE* gene product.

2. Materials and methods

2.1. Sequence analysis

The databases available on the National Center for Biotechnology Information (NCBI) web page (www.ncbi.nlm.nih.gov) were searched using the sequence from the *E. coli gcpE* gene product (accession number P27433) as a query and the TBLASTN algorithm. EST clone 135H1T7 (T46582) was obtained from the *A. thaliana* DNA Stock Center at the *Arabidopsis* Biological Resource Center (ABRC, Ohio State University, OH, USA). Both strands of the cDNA, cloned into the *Sal*I–*Not*I sites of the pZL1 vector (Gibco-BRL), were sequenced with the ABI PRISM BigDye DNA Sequencing kit (PE Biosystems) using vector primers and gene specific primers. Sequence analyses were performed using the GCG 9.0 software package (Genetics Computer Group Inc.) and the available NCBI web programs. Plastid targeting signals were predicted with ChloroP (www.cbs.dtu.dk/services/chlorop).

2.2. Microbombardment assay

A 0.3 kb *Eco*RI fragment from the 5′ region of the cDNA in clone 135H1T7 was cloned into pGFP-MRC [32]. In the resulting construct, the CaMV *35S* promoter directed the expression of the fusion protein N49–green fluorescent protein (GFP), in which the 5′ untranslated region and the sequence encoding the first 49 amino acid residues of the *Arabidopsis* GCPE protein were fused in frame to the N-terminus of a synthetic GFP. Subcellular localization of the N49–GFP fusion protein in *Arabidopsis* leaf cells was detected by green fluorescence after microbombardment using a Biolistic PDS-1000/He system (Bio-Rad) as described [23].

2.3. E. coli complementation

Plasmid pQE-AGm was created by subcloning a *Bgl*II–*Sph*I cDNA fragment from clone 135H1T7 (encoding the predicted *Arabidopsis* mature GCPE protein) into the *Bam*HI–*Sph*I sites of pQE30 (Qiagen). The complementation assay was carried out using the *E. coli* strain EcAB3-3 [9]. After electroporation of competent cells with pQE-AGm and plasmid pUBS520-argU [33], transformants were selected on LB plates containing 17 μg/ml chloramphenicol, 100 μg/ml ampicillin, 0.4 mM IPTG, 100 μg/ml kanamycin, 0.04% arabinose, and 0.5 mM MVA. The ability of the cloned protein to rescue the lethal disruption of the *gcpE* gene was tested by streaking independent colonies on plates without MVA.

3. Results and discussion

To identify plant sequences with homology to the *E. coli* protein encoded by the *gcpE* gene (accession number AE000338), a similarity search was carried out in the databases available on the NCBI web page using the TBLASTN algorithm. This search retrieved a hypothetical protein predicted by an *A. thaliana* chromosome V genomic sequence (AB005246) and 21 ESTs. Sequencing of one of the EST clones found, 135H1T7 (T46582), showed that it contained a 2520 bp cDNA sequence with an ORF of 2220 bp encoding a protein of 740 amino acid residues (Fig. 2A). Comparison of the complete cDNA sequence with that of the corresponding chromosomal gene revealed 20 exons and 19 introns in the region from position 2227 to 6625 in the AB005246 genomic sequence entry (Fig. 2B). A polypeptide sequence (BAB09833) had been predicted from this genomic region earlier by computer analysis in the context of whole genome sequencing. However, the first coding exon had not been predicted correctly, since the N-terminal 25 residues of the protein (from the first methionine residue to the methionine at position 26, deduced to be the translation start point in the database entry) were missing. The sequence data deduced from EST and genomic entries are consistent with a single gene encoding this protein in *Arabidopsis*.

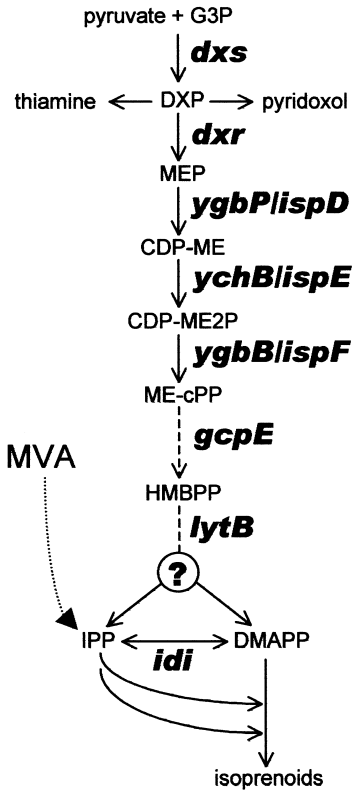

Fig. 1. The MEP pathway as described in *E. coli*. The indicated genes (in italics) encode the following enzymes: *dxs*, DXP synthase; *dxr*, DXP reductoisomerase; *ygbP*/*ispD*, CDP-ME synthase; *ychB*/*ispE*, CDP-ME kinase; *ygbB*/*ispF*, ME-cPP synthase; *idi*, IPP isomerase. The *E. coli* strain used in this work was engineered to utilize exogenously supplied MVA for the biosynthesis of IPP (dotted line).

Alignment of the deduced *Arabidopsis* sequence (GCPE) with that predicted from the *E. coli gcpE* gene showed that the plant protein contains an N-terminal region of 75 residues that is absent in the bacterial counterpart (Fig. 2A). N-terminal extensions are also found in all the previously described plant enzymes of the MEP pathway and are proposed to comprise a signal sequence for plastid import, consistent with the subcellular localization of the MEP pathway in plants. However, plastid targeting has only been demonstrated for DXS [17,23]. The hypothetical cleavage site predicted by the ChloroP program is indicated in Fig. 2A. To determine whether this domain was functional in targeting the protein to plastids, we fused the N-terminal 49 residues of the *Arabidopsis* protein to GFP for expression of the corresponding fusion protein (N49–GFP) in plant cells. *Arabidopsis* leaves were microbombarded with constructs to transiently express either GFP or N49–GFP under the control of the CaMV *35S* promoter. While green fluorescence corresponding to GFP was localized in the cytoplasm and the nucleus, fluorescence from N49–GFP colocalized with chlorophyll autofluorescence (Fig. 3). This result shows that the N-terminal peptide fused to GFP is functional in targeting the protein to chloroplasts in

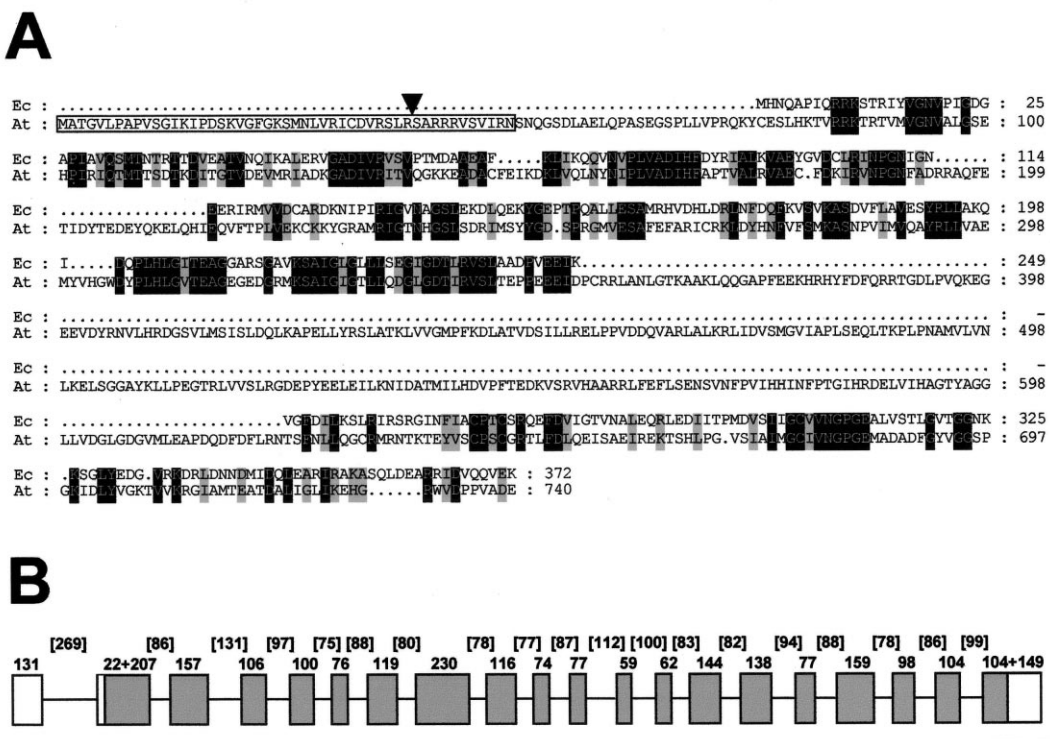

Fig. 2. Sequence analysis of GCPE proteins. A: Alignment of sequences from *E. coli* GcpE (P27433) and *A. thaliana* GCPE (AF434673). Identical residues are highlighted in black boxes and conservative changes in gray boxes. Arrowhead indicates the position predicted with the ChloroP algorithm for the cleavage of the plastid targeting peptide from the *Arabidopsis* protein. The peptide fused to GFP to create N49–GFP is boxed. B: Intron/exon topology of the *Arabidopsis* gene encoding GCPE. Exons are indicated with boxes (white for untranslated sequences and gray for coding sequences) and introns with a line. Numbers indicate length in nucleotides (those corresponding to introns are between brackets).

vivo, in agreement with a proposed role of GCPE in the MEP pathway for the biosynthesis of plastidic isoprenoids.

Besides the N-terminal region, the alignment of the predicted *Arabidopsis* and *E. coli* proteins showed the presence of a large additional domain of 268 residues (ca. 30 kDa) in the plant protein (Fig. 2A). The sequence deduced from this domain showed no significant homology with any other known protein. Interestingly, GCPE is the only protein of the MEP pathway containing a large additional domain in the mature plant protein that is absent in the *E. coli* counterpart. In spite of the overall similarity between *Arabidopsis* and *E. coli* sequences (Fig. 2A), the presence of an additional large domain in the plant protein makes a functional analysis especially relevant to ascertain its putative GcpE activity. Since the biochemical activity of the GcpE protein remains to be well established, a complementation assay was carried out using the *gcpE*-defective *E. coli* strain EcAB3-3 [9]. In this strain, which was engineered to synthesize IPP and DMAPP from exogenously supplied MVA (Fig. 1), the chromosomal *gcpE* gene is substituted by the *CAT* marker conferring chloramphenicol resistance. Since the deletion of *gcpE* prevents IPP and DMAPP biosynthesis via the MEP pathway, mutant EcAB3-3 cells cannot grow unless MVA is supplied [9].

For the complementation assay, plasmid pQE-AGm was constructed by subcloning the cDNA fragment encoding the predicted *Arabidopsis* mature protein (lacking the predicted plastid signal peptide) into the pQE30 expression vector. Careful examination of the *Arabidopsis* sequence revealed the presence of abundant arginine residues, especially in the N-terminal region (Fig. 2A). Since the AGA and AGG codons used for arginine in *Arabidopsis* are not common in *E. coli*, it is possible that problems in reading such codons (especially in the regions where they are clustered, like the 5′ region of the cloned cDNA) may block the synthesis of a complete protein from the cloned *Arabidopsis* cDNA. To address this possibility, we cotransformed EcAB3-3 cells with both pQE-AGm and pUBS520-argU, a plasmid encoding the *E. coli* minor tRNA(Arg) species that decodes the triplets AGA and AGG [33], to create strain EcAB3-3(pQE-AGm). Similarly, strain EcAB3-3(pQE) was created by transforming

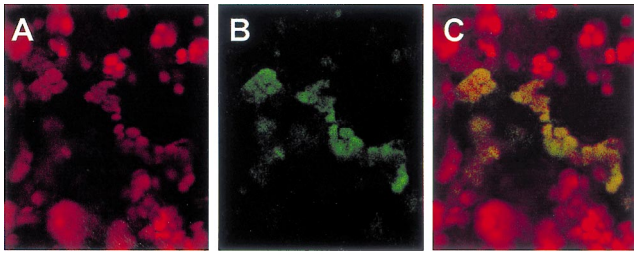

Fig. 3. Plastid targeting of the N49–GFP fusion protein in *Arabidopsis*. Leaves were microbombarded with a plasmid to express N49–GFP (with the N-terminal 49 residues of the putative *Arabidopsis* GCPE protein fused to GFP; Fig. 2A) under the CaMV *35S* promoter. Red autofluorescence of chlorophyll (A) and green fluorescence from the fusion protein (B) were overlapped to show colocalization of N49–GFP fluorescence and chloroplasts (C). The lesser amount of chlorophyll fluorescence in the chloroplasts expressing N49–GFP is due to the absorption of excitation energy by GFP.

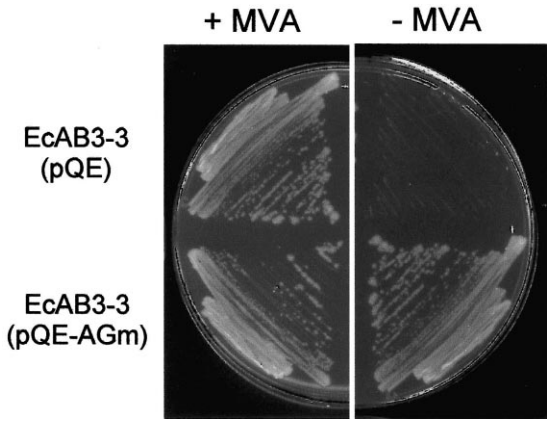

Fig. 4. Complementation of *gcpE*-deficient *E. coli* with the *Arabidopsis* mature GCPE protein. Cells of the *gcpE*-deficient EcAB3-3 strain were transformed with constructs to express the predicted *Arabidopsis* mature GCPE protein (pQE-AGm), or with the original pQE30 vector as a control. After recovering the generated strains EcAB3-3(pQE-AGm) and EcAB3-3(pQE) on plates supplemented with 1 mM MVA (to rescue the lethal deletion of the chromosomal *gcpE* gene), colonies were streaked on new plates with (+) or without (−) MVA and incubated at 37°C.

EcAB3-3 cells with plasmid pQE30 as a control. After selection on plates containing MVA, separate colonies from the generated strains were streaked on plates without MVA (Fig. 4). As expected, EcAB3-3(pQE) cells were not able to grow when streaked on plates without MVA but EcAB3-3(pQE-AGm) cells did overcome MVA auxotrophy (Fig. 4). Our results show that the cloned *Arabidopsis* cDNA encodes a protein capable of complementing the lethal *gcpE*-defective *E. coli* mutant. Together, we have demonstrated that the *Arabidopsis* GCPE protein is the functional homologue of the *E. coli gcpE* gene product, despite the differences in their amino acid sequence such as the presence of an additional internal domain in the plant protein. Although it is possible that because of this extra domain the plant GCPE protein could have distinct regulatory or catalytic functions, these remain to be established.

Acknowledgements: We thank the DNA Stock Center at ABRC, for the *Arabidopsis* 135H1T7 EST clone. The help and advice of the staff of the Universitat de Barcelona Serveis Cientificotècnics (confocal microscopy) and Serveis de Camps Experimentals (greenhouse facilities) are greatly appreciated. We also thank Drs. M. Rohmer, M.A. Phillips and J.F. Martínez-García for critical reading of the manuscript, A. Orozco for valuable technical assistance, and the laboratory members for their collaboration. This work was supported by grants from the Spanish Ministerio Ciencia y Tecnología (BIO1999-0503-C02-01), and Generalitat de Catalunya (CIRIT, 1999SGR-00032).

References

[1] Chappell, J. (1995) Annu. Rev. Plant Physiol. Plant Mol. Biol. 46, 521–547.
[2] Goldstein, J.L. and Brown, M.S. (1990) Nature 343, 425–430.
[3] McGarvey, D.J. and Croteau, R. (1995) Plant Cell 7, 1015–1026.
[4] Eisenreich, W., Rohdich, F. and Bacher, A. (2001) Trends Plant Sci. 6, 78–84.
[5] Lichtenthaler, H.K. (1999) Annu. Rev. Plant Physiol. Plant Mol. Biol. 50, 47–65.
[6] Rohmer, M. (1999) Nat. Prod. Rep. 16, 565–574.
[7] Rohdich, F., Kis, K., Bacher, A. and Eisenreich, W. (2001) Curr. Opin. Chem. Biol. 5, 535–540.
[8] Altincicek, B., Kollas, A.K., Sanderbrand, S., Wiesner, J., Hintz, M., Beck, E. and Jomaa, H. (2001) J. Bacteriol. 183, 2411–2416.
[9] Campos, N., Rodríguez-Concepción, M., Seemann, M., Rohmer, M. and Boronat, A. (2001) FEBS Lett. 488, 170–173.
[10] Altincicek, B., Kollas, A.K., Eberl, M., Wiesner, J., Sanderbrand, S., Hintz, M., Beck, E. and Jomaa, H. (2001) FEBS Lett. 499, 37–40.
[11] Cunningham, F.X., Lafond, T.P. and Gantt, E. (2000) J. Bacteriol. 182, 5841–5848.
[12] McAteer, S., Coulson, A., McLennan, N. and Masters, M. (2001) J. Bacteriol. 183, 7403–7407.
[13] Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A., Arigoni, D. and Rohdich, F. (2001) Proc. Natl. Acad. Sci. USA 26, 14837–14842.
[14] Seemann, M., Campos, N., Rodríguez-Concepción, M., Hoeffler, J.F., Grosdemange-Billiard, C., Boronat, A. and Rohmer, M. (2002) Tetrahedron Lett. 43, 775–778.
[15] Seemann, M., Campos, N., Rodríguez-Concepción, M., Ibáñez, E., Duvold, T., Tritsch, D., Boronat, A. and Rohmer, M. (2002) Tetrahedron Lett. 43, 1413–1415.
[16] Hintz, M., Reichenberg, A., Altincicek, B., Bahr, U., Gschwind, R.M., Kollas, A.K., Beck, E., Wiesner, J., Eberl, M. and Jomaa, H. (2001) FEBS Lett. 509, 317–322.
[17] Araki, N., Kusumi, K., Masamoto, K., Niwa, Y. and Iba, K. (2000) Physiol. Plant 108, 19–24.
[18] Bouvier, F., d'Harlingue, A., Suire, C., Backhaus, R.A. and Camara, B. (1998) Plant Physiol. 117, 1423–1431.
[19] Estévez, J.M., Cantero, A., Romero, C., Kawaide, H., Jiménez, L.F., Kuzuyama, T., Seto, H., Kamiya, Y. and Leon, P. (2000) Plant Physiol. 124, 95–103.
[20] Estévez, J.M., Cantero, A., Reindl, A., Reichler, S. and Leon, P. (2001) J. Biol. Chem. 25, 22901–22909.
[21] Lange, B.M., Wildung, M.R., McCaskill, D. and Croteau, R. (1998) Proc. Natl. Acad. Sci. USA 95, 2100–2104.
[22] Lange, B.M. and Croteau, R. (1999) Arch. Biochem. Biophys. 365, 170–174.
[23] Lois, L.M., Rodríguez-Concepción, M., Gallego, F., Campos, N. and Boronat, A. (2000) Plant J. 22, 503–513.
[24] Mahmoud, S.S. and Croteau, R.B. (2001) Proc. Natl. Acad. Sci. USA 98, 8915–8920.
[25] Mandel, M.A., Feldmann, K.A., Herrera-Estrella, L., Rocha-Sosa, M. and León, P. (1996) Plant J. 9, 649–658.
[26] Rodríguez-Concepción, M., Ahumada, I., Diez-Juez, E., Sauret-Güeto, S., Lois, L.M., Gallego, F., Carretero-Paulet, L., Campos, N. and Boronat, A. (2001) Plant J. 27, 213–222.
[27] Schwender, J., Müller, C., Zeidler, J. and Lichtenthaler, H.K. (1999) FEBS Lett. 455, 140–144.
[28] Veau, B.M.C., Oudin, A., Chénieux, J.C., Rideau, M. and Clastre, M. (2000) Biochim. Biophys. Acta 1517, 159–163.
[29] Walter, M.H., Fester, T. and Strack, D. (2000) Plant J. 21, 571–578.
[30] Rohdich, F., Wungsintaweekul, J., Eisenreich, W., Richter, G., Schuhr, C.A., Hecht, S., Zenk, M.H. and Bacher, A. (2000) Proc. Natl. Acad. Sci. USA 97, 6451–6456.
[31] Rohdich, F., Wungsintaweekul, J., Lüttgen, H., Fischer, M., Eisenreich, W., Schuhr, C.A., Fellermeier, M., Schramek, N., Zenk, M.H. and Bacher, A. (2000) Proc. Natl. Acad. Sci. USA 97, 8251–8256.
[32] Rodríguez-Concepción, M., Yalovsky, Y., Zik, M., Fromm, H. and Gruissem, W. (1999) EMBO J. 18, 1996–2007.
[33] Brinkmann, U., Mattes, R.E. and Buckel, P. (1989) Gene 85, 109–114.