Squalene-hopene cyclase from *Bradyrhizobium japonicum*: cloning, expression, sequence analysis and comparison to other triterpenoid cyclases

Michael Perzl,¹ Peter Müller,² Karl Poralla¹ and Elmar L. Kannenberg^{1,3}

Author for correspondence: Elmar L. Kannenberg. Tel: +49 7071 2972080 or +1 706 542 4401. Fax: +49 7071 295002 or +1 706 542 4412. e-mail: poralla@uni-tuebingen.de or elmark@ccrc.uga.edu

¹ Mikrobiologie/
Biotechnologie, Universität
Tübingen, Auf der
Morgenstelle 1, D-72076
Tübingen, Germany

² Fachbereich Biologie-
Botanik, Universität
Marburg, Lahnberge, D-
35043 Marburg, Germany

³ Complex Carbohydrate
Research Center and
Department of
Biochemistry and
Molecular Biology,
University of Georgia, 220
Riverbend Road, Athens,
GA 30602-4712, USA

With the help of a PCR-based screening method, the gene encoding squalene-hopene cyclase (SHC) of *Bradyrhizobium japonicum* USDA 110 was isolated from a cosmid library. The SHC catalyses the cyclization of squalene to hopanoids, a class of triterpenoid lipids recently discovered in nitrogen-fixing, root-nodule-forming *Bradyrhizobium* bacteria. Hybridization experiments showed that the gene is present in bacteria of all *Bradyrhizobium* strains tested and in photosynthetic bacteria forming stem nodules on tropical legumes of the genus *Aeschynomene*. The *Bradyrhizobium shc* gene is 1983 bp in length and encodes a protein of 660 amino acid residues with a calculated molecular mass of 73 671 Da. Comparison of the deduced amino acid sequence with the sequences of other SHCs revealed highest similarity (70%) to the SHC from the Gram-negative *Zymomonas mobilis* and lower similarity (48%) to the SHCs from the Gram-positive *Alicyclobacillus acidocaldarius* and *Alicyclobacillus acidoterrestris*. *Bradyrhizobium* SHC also showed similarity (38–43%) to eukaryotic oxidosqualene cyclases. The *B. japonicum shc* gene was expressed in *Escherichia coli*. The recombinant SHC catalysed the cyclization of squalene to the hopanoids hopene and diplopterol *in vitro*. However, the formation of the gammacerane derivative tetrahymanol, which is produced in addition to hopanoids in *B. japonicum* strains *in vivo*, could not be detected *in vitro*. Therefore, the presence of a second squalene cyclase in *B. japonicum* can be assumed. Sequence analysis of 0.5 kb upstream from the *shc* gene identified a partial ORF with significant similarity to the C-terminus of an ORF located immediately upstream from the *shc* gene in *Z. mobilis*. Both ORFs also showed similarity to phytoene desaturases from cyanobacteria and plants. The 3'-end of this ORF from *B. japonicum* overlaps with 13 bp at the 5'-end of *shc*. The close proximity of this ORF to *shc* suggests that *shc* and this ORF may be part of an operon.

Keywords: *Bradyrhizobium japonicum*, triterpenoids, hopanoids, squalene-hopene cyclase

INTRODUCTION

Hopanoids are a class of pentacyclic triterpenoid lipids occurring in a wide range of Gram-positive and Gram-

Abbreviations: OSC, oxidosqualene cyclase; SHC, squalene-hopene cyclase; USDA, United States Department of Agriculture.

The EMBL accession number for the nucleotide sequence reported in this paper is X86552.

negative bacteria (Sahm *et al.*, 1993). Among aerobic bacteria with the ability to fix atmospheric nitrogen, hopanoids were found in free-living *Azotobacter* and *Beijerinckia* (Vilcheze *et al.*, 1994), and in the symbiotic actinomycete *Frankia* (Berry *et al.*, 1991). Recently, hopanoids were also discovered in nitrogen-fixing *Bradyrhizobium* bacteria (Kannenberg *et al.*, 1995) that form root nodules in symbiosis with various legume plants, and in photosynthetic bacteria forming stem

Fig. 1. Enzymic cyclization of squalene. I, Squalene; II, hopene; III, diplopterol; IV, tetrahymanol. Hopene and diplopterol are formed in prokaryotes from squalene by SHC activity. Elimination of a proton leads to hopene; addition of a hydroxyl group derived from water leads to formation of diplopterol. In some cases such as *B. japonicum*, *R. palustris* or the ciliate *Tetrahymena*, addition of a hydroxyl group can lead to formation of the gammacerane derivative tetrahymanol. ?, Not known if this reaction is catalysed by SHC.

nodules on tropical legumes of the genus *Aeschynomene* (Kannenberget al., 1996). Besides hopanoids, *Bradyrhizobium* contains another pentacyclic triterpenoid, the gammacerane derivative tetrahymanol (unpublished results). The only prokaryote in which tetrahymanol has been discovered in addition to hopanoids is the purple non-sulphur bacterium *Rhodospseudomonas palustris* (Kleemann et al., 1990). Phylogenetic studies based on 16S rRNA sequence analysis revealed that *R. palustris* is a close relative of *Bradyrhizobium* (Jarvis et al., 1986). Neither hopanoids nor tetrahymanol have been detected so far in *Rhizobium* species or in additional members of the *Rhizobiaceae* (Kannenberget al., 1995, 1996).

Hopanoids and tetrahymanol are reinforcers of cellular membranes (Kannenberget al., 1980; Ourisson et al., 1987), condensing membranes in a similar way to cholesterol in eukaryotic membranes. However, hopanoids and tetrahymanol may have additional functions besides membrane reinforcement that have not yet been detected. In the case of the nitrogen-fixing *Frankia*, it has been suggested that hopanoids are involved in oxygen protection of the nitrogenase complex by forming a diffusion barrier (Berry et al., 1993). This may also be true in the case of *Bradyrhizobium* which, unlike *Rhizobium*, is able to fix nitrogen non-symbiotically in the free-living state (Wilcockson & Werner, 1976).

The key enzyme in hopanoid biosynthesis is the squalene-hopene cyclase (SHC). This enzyme catalyses the cyclization of the linear triterpenoid squalene to hopene and diplopterol (Fig. 1). These reaction products are the presumed precursors of elongated hopanoids with a great diversity of polar side chains. Mechanistically, this cyclization reaction is not well understood and the cyclization products can differ depending on the

organism involved; for example, in the ciliate *Tetrahymena* an analogous reaction leads predominantly from squalene to the formation of tetrahymanol (Caspi et al., 1968; Abe et al., 1993).

SHCs from several bacteria (*Alicyclobacillus acidocaldarius*, *Zymomonas mobilis* and *R. palustris*) have been purified and characterized recently (Ochs et al., 1992; Tappe, 1993; Kleemann et al., 1994). The genes of the SHCs from *A. acidocaldarius*, *Alicyclobacillus acidoterrestris* and *Z. mobilis* have been isolated and their nucleotide sequences determined (Ochs et al., 1992; EMBL accession no. X89854; Reipen et al., 1995). Prokaryotic SHCs exhibit similarity to oxidosqualene cyclases (OSCs) from plants, animals and fungi, supporting the hypothesis that both types of cyclases have evolved from a common ancestor cyclase (Ourisson et al., 1987).

We chose *Bradyrhizobium* to start an investigation on the molecular biology of hopanoid biosynthesis for several reasons. First, bradyrhizobia live and thrive in two environments: as soil colonizers and as symbiotic endophytes. This characteristic allows us to address the function of hopanoids in these different habitats and as part of plant-microbe interactions. Second, bradyrhizobia contain a range of hopanoid derivatives and structurally related compounds (e.g. tetrahymanol) whose biological function has not been explained. Since bradyrhizobia are genetically accessible, they are ideal for addressing the functions of these different compounds. Third, the ease of genetically manipulating the amino acid sequence of the SHC protein will allow us to address questions about the cyclization reaction mechanism and important features in the SHC enzyme structure.

In this paper, we describe the cloning, DNA sequencing, expression and sequence analysis of the *shc* gene from *B. japonicum* as the first step in an examination of hopanoid biosynthesis and *in vivo* function.

METHODS

Strains, plasmids and growth conditions. *B. japonicum* USDA 110spc4 (Regensburger & Hennecke, 1983), *Bradyrhizobium* sp. (*Lupinus*) (Deutsche Sammlung von Mikroorganismen und Zellkulturen 30140) and the photosynthetic *Aeschynomene* stem-nodulating strain MKAa3 (USDA 4088; Wong et al., 1994) were cultivated in YM medium (Vincent, 1970) at 28 °C. *Escherichia coli* strains (S17-1, XL-1 Blue and DH5 α) were grown in LB medium (Gibco) routinely at 37 °C. *E. coli* strain S17-1 (Simon et al., 1983) was used for construction of the cosmid library, and strains XL-1 Blue and DH5 α were used for cloning procedures. Cloning and subcloning were carried out with plasmid vectors pUC18 and pUC19 (Vieira & Messing, 1982), respectively. For construction of the cosmid library, cosmid vector pVK100 (Knauf & Nester, 1982) was used. The concentrations of antibiotics were 100 mg ampicillin l⁻¹, 10 mg tetracycline l⁻¹ and 30 mg kanamycin l⁻¹.

Standard DNA manipulation procedures. Genomic DNA of *B. japonicum* USDA 110 was isolated according to the procedure of Somasegaran & Hoben (1994). Digestion with restriction

endonucleases, ligation and transformation of *E. coli* were carried out as described in Sambrook *et al.* (1989). Large-scale isolation of plasmid DNA was performed with a Macherey and Nagel Nucleobond AX-100 kit. Small-scale isolation of plasmid DNA was done using the alkaline lysis method (Birnboim & Doly, 1979).

Southern hybridization was performed with digoxigenin-labelled probes using the random primed method according to the protocol of the manufacturer (Boehringer Mannheim). Digested DNA was separated on agarose gels and blotted with a Vakublot apparatus (Pharmacia) on positively charged nylon membranes (Boehringer Mannheim). Hybridization was done at 65 °C (2 h prehybridization, at least 12 h hybridization) in a hybridization oven (Hybaid). Washing was performed twice with $2 \times$ SSC (0.3 M NaCl; 0.03 M tri-sodium citrate, pH 7.4), 0.1% SDS at room temperature, followed by washing twice with $0.1 \times$ SSC, 0.1% SDS at 65 °C. Detection was carried out with anti-digoxigenin-antibody conjugates (anti-digoxigenin-alkaline phosphatase Fab fragments) and the chemiluminescence substrate CSPD (Boehringer Mannheim) according to the instructions of the manufacturer.

Construction of the cosmid library. Genomic DNA of *B. japonicum* USDA 110 was partially digested with *EcoRI* and separated by gel electrophoresis. Fragments of 20–25 kb were cloned in the single *EcoRI* site of cosmid vector pVK100, packaged *in vitro* in phage λ , and propagated in *E. coli* S17-1 (Parniske *et al.*, 1993).

PCR screening of the cosmid library. Clones of the cosmid library from *B. japonicum* USDA 110 harbouring the *shc* gene were identified with the help of a PCR-based screening method modifying a method described by Griffin *et al.* (1993). The PCR conditions were as follows: denaturation at 94 °C for 1 min; annealing at 48 °C for 1 min; and elongation at 73 °C for 1 min. A 50 μ l reaction contained 1 U vent DNA polymerase (Biolabs), 5 μ l $10 \times$ polymerase buffer (Biolabs), 1 mM each dNTP, 0.5 mM $MgSO_4$, 10 pmol each primer and 1 μ g genomic DNA or 1 μ g cosmid DNA. The reaction mixture was overlaid with 50 μ l mineral oil.

Conserved regions of known SHCs and OSCs (Fig. 2) were used to design a pair of degenerate oligonucleotide primers (P1/P2 in Fig. 2). All oligonucleotides were custom-synthesized by Eurogentec.

P1 (5'-TGGTTC/TGGC/TCGC/TTGGGG-3') and P2 (5'-GCCCCAG/ACCG/ACCG/ATCCGGTGGTC-3') amplified a 150 bp fragment of the *shc* gene from genomic DNA of *B. japonicum* USDA 110. Homologous primers B1 (5'-CATCTATGGAACCTGGTC-3') and B2 (5'-TCGAGG-CCAGCCAGTCGG-3') were synthesized according to the nucleotide sequence of the 150 bp PCR fragment. PCR with these two primers was expected to result in amplification of a 98 bp PCR fragment (for more details see Results and Discussion and Fig. 2). The conditions in both PCRs were identical with the exception of an annealing temperature of 52 °C in the case of primers B1 and B2. These primers were used to PCR-screen a *B. japonicum* cosmid library for clones harbouring the *shc* gene. The cosmid library was divided into pools of 100 clones that were grown on agar plates. The clones were washed off with 5 ml LB medium, and the cosmid DNA of these clones was collectively isolated. An aliquot of the isolated cosmid DNA was used as PCR template. Only those pools that allowed amplification of the expected 98 bp fragment were submitted to the next screening step. Positive pools were subdivided into pools containing 10 clones, cultivated in 1 ml LB medium, and the isolated cosmid DNA

was screened by PCR. The indicative 98 bp fragment could be amplified from two of these pools. The remaining 20 clones were now individually assayed by PCR.

Sequence determination. Double-stranded DNA was sequenced with the AutoRead Sequencing kit and an ALF DNA Sequencer (Pharmacia). As sequencing primers, fluorescein-labelled M13 universal and reverse primers or fluorescein-labelled *shc*-specific primers (custom-synthesized by Pharmacia) were used. Sequence analysis was done with the PCGENE program package (IntelliGenetics). Database searches were performed with the BLASTN and BLASTP programs of NCBI (Altschul *et al.*, 1990).

Enzyme activity assay. *E. coli* DH5 α cells, containing plasmid DNA with or without the *B. japonicum shc* gene, were grown overnight at 28 °C in liquid LB medium, harvested by centrifugation, and washed in a 0.9% NaCl solution. The cells were resuspended in buffer (100 mM Tris/HCl, 250 mM saccharose, 20 mM ascorbic acid, pH 8). The frozen cells (–20 °C) were passed five times through an X-press homogenator for cell rupture. The homogenate was thawed, lysozyme and DNase were added, and debris and unbroken cells were removed by centrifugation. Appropriate aliquots (between 100 and 400 μ l) of the different homogenates were added to individual assay solutions containing the substrate squalene (100 μ l of a squalene stock solution of 500 μ M squalene; 0.1% Triton X-100; 100 mM sodium citrate, pH 6.5), 100 μ l 1 M sodium citrate (pH 6.5) and water to make up a final volume of 1 ml. The mixture was incubated for 2 h at 28 °C and then extracted with 2 ml hexane/2-propanol (3:2, v/v). The organic phase was collected and evaporated, and the remaining lipids were redissolved in 30 μ l hexane/2-propanol for analysis.

The reaction products were analysed by GLC analysis with a Shimadzu GC-9 gas chromatograph on a capillary column (DB-1, methylsilicon; 0.32 mm $\times$ 20 m). The oven temperature was set from 250 °C to 320 °C with 4 °C min^{–1} increase followed by 10 min at 320 °C. Detection was made with a flame ionization detector at 330 °C. Triterpenoids were identified either by comparison with the retention times of standard compounds or by GLC-MS fragmentation analysis.

RESULTS AND DISCUSSION

Isolation of the *shc* gene

The importance of hopanoid lipids for bradyrhizobia in the soil and as symbiotic endophytes has yet to be elucidated. As a first step in the analysis of hopanoid function and biosynthesis, we attempted to isolate the gene encoding one of the key enzymes in biosynthesis, the SHC.

For isolation of the *shc* gene from *B. japonicum* USDA 110, a two-step PCR-based screening strategy was used. In the first step, genomic DNA was used to amplify a 150 bp DNA fragment with homology to the *shc* gene. The degenerate primers P1 and P2 were designed according to two of the most conserved regions in known SHCs and OSCs (Fig. 2). The deduced amino acid sequence of the amplified DNA fragment from strain USDA 110 had significant identity (57% in the case of *Z. mobilis* SHC) to the corresponding region of other SHCs and OSCs (Fig. 2), indicating that we had amplified a DNA fragment that should be part of a

Fig. 2. Alignment of the amino acid sequence deduced from a 150 bp PCR fragment amplified from genomic DNA of *B. japonicum* USDA 110 with the homologous sequences of four cyclases already characterized. Partial amino acid sequences of the OSCs of *Arabidopsis thaliana* (A. t.; Corey *et al.*, 1993) and *Candida albicans* (C. a.; Buntel & Griffin, 1992; Roessner *et al.*, 1993), and of the SHCs of *A. acidocaldarius* (A. a.; Ochs *et al.*, 1992) and *Z. mobilis* (Z. m.; Reipen *et al.*, 1995) are shown and aligned with the amino acid sequence of the *B. japonicum* (B. j.) fragment. The position of the first and last amino acid residues of the fragment sequences in the proteins are indicated as numbers. Gaps in the alignment are indicated by a dash. Conserved amino acid residues and residues that have been conservatively exchanged are in bold type. The 150 bp PCR fragment was obtained with degenerate PCR primers P1/P2 (their location is indicated by arrows). The homologous primers B1/B2 were used to PCR-screen in a cosmid library of *B. japonicum* USDA 110 for a 98 bp fragment, indicative for cosmids carrying the cloned *shc* gene (for details see Methods and Results and Discussion).

squalene cyclase gene. Southern hybridization of the digoxigenin-labelled 150 bp fragment with digested genomic DNA of strain USDA 110 under high stringency resulted in hybridization signals with a 5 kb *Eco*RI and a 1.7 kb *Bam*HI fragment (data not shown). This result confirmed the origin of the amplified DNA fragment from *Bradyrhizobium* DNA.

In a second step, a set of homologous primers (B1 and B2) amplifying a 98 bp fragment from the *shc* gene was designed according to the nucleotide sequence of the 150 bp fragment. These primers were used to identify, by PCR screening, cosmid clones harbouring the *shc* gene. To avoid the screening of individual clones, the cosmid library of strain USDA 110 was divided into increasingly smaller sublibraries. Those sublibraries that did not allow the amplification of the expected 98 bp fragment were omitted from the PCR screening. This screening method led finally to the isolation of two cosmid clones (pBJ64 and pBJ81). Restriction fragment analysis revealed that these clones overlapped, having a 5, 6 and 12 kb *Eco*RI fragment in common. The orientation of these three fragments is shown in Fig. 3 and was determined by restriction fragment analysis and Southern hybridization. Southern hybridization with the 150 bp PCR fragment resulted in hybridization signals with the 5 kb *Eco*RI fragment and, in the case of digesting the cosmid DNA with *Bam*HI, in a 1.7 kb *Bam*HI fragment, indicating the same hybridization pattern as with genomic DNA from strain USDA 110. Both fragments were subcloned (pMPE2 contains the 5 kb *Eco*RI fragment and pMPB2 contains the 1.7 kb *Bam*HI fragment) and further analysed. The 1.7 kb

*Bam*HI fragment was found to be entirely part of the 5 kb *Eco*RI fragment (Fig. 3).

Hybridization signals with the 150 bp PCR fragment were also found with genomic DNA of other *Bradyrhizobium* strains (data not shown), e.g. *Bradyrhizobium* sp. (*Lupinus*) and also MKAa3 (USDA 4088), a stem-nodulating, photosynthetic isolate from the tropical legume *Aeschynomene* (Wong *et al.*, 1994). These strains are grouped in one cluster together with *R. palustris*, which also contains hopanoids and tetrahymanol (Kleemann *et al.*, 1990), on several recently suggested phylogenetic trees for the α -subdivision of proteobacteria (e.g. Young *et al.*, 1991; Martinez-Romero & Caballero-Mellado, 1996). Our findings about the distribution of the *shc* gene confirm the phylogenetic relatedness of these strains.

Gene and protein structure

Sequencing of almost the entire 5 kb *Eco*RI fragment led to the identification of an ORF that was designated the *shc* gene. The *shc* gene is 1983 bp long and encodes a protein of 660 amino acids with a calculated molecular mass of 73 671 Da (Fig. 4). These values agree with those of known SHCs, which have a molecular mass ranging between 71 and 74 kDa and a length of between 631 and 658 amino acids.

A putative ribosome-binding site (GGAGTAA) is located at position -16 to -10 upstream of the ATG start codon of *shc*. Downstream of the *shc* gene a region capable of forming a hairpin [ΔG^0 -20 kcal mol⁻¹ (-84 kJ mol⁻¹)] was found representing a putative termination sequence. The *shc* gene has a GC ratio of 64.8%, which is in the range of the GC ratio of *B. japonicum* (61–65%; Jordan, 1982). The sequence of the 150 bp PCR fragment amplified from genomic DNA was found to be identical with the sequence obtained from the isolated cosmid clones with the exception of base changes in the P1/P2 primer regions.

An alignment of the deduced amino acid sequence of the *Bradyrhizobium* SHC with other SHCs and OSCs revealed significant overall homology. The highest amount of homology (70% similarity; 59% identity) was found with the SHC from *Z. mobilis*, which belongs like *B. japonicum* to the α -subgroup of proteobacteria. The homology to the SHC from the Gram-positive *A. acidocaldarius* and *A. acidoterrestis* is significantly lower (48% similarity; 38% identity). Among the SHCs, 165 amino acid residues are strictly conserved. The eukaryotic OSCs show about 38–43% similarity and 25–27% identity to the *B. japonicum* SHC. Overall homology of SHC to other terpenoid cyclases, e.g. mono-, sesqui- or diterpene cyclases, is negligible, indicating that triterpenoid cyclases form a separate group of enzymes.

Interestingly, database searches revealed an amino acid sequence from an ORF from the cytochrome P-450 gene cluster of *B. japonicum* USDA 110 (Tully & Keister, 1993; EMBL accession no. BJ12678) that has some

Fig. 3. (a) Orientation of three *EcoRI* fragments of approximately 5, 6 and 12 kb derived from a cosmid (pBJ81) isolated from a *B. japonicum* USDA 110 cosmid library by PCR screening. (b) Restriction map of the 5 kb *EcoRI* fragment (cloned as pMPE2) harbouring the *shc* gene encoding SHC. A 1.7 kb *Bam*HI fragment (cloned as pMPB2) from the cosmid clone also showed hybridization with the 150 bp PCR fragment obtained in the first step of the PCR screening for the *shc* gene. ORF1' indicates a partial ORF with similarity to an ORF located immediately upstream of the *shc* gene of *Z. mobilis*. The parts of the *EcoRI* and *Bam*HI fragments that were sequenced are hatched. The black box indicates the location of the 150 bp fragment. E, *EcoRI*; S, *SacI*; B, *Bam*HI; K, *KpnI*.

Fig. 4. Nucleotide sequence and deduced amino acid sequence of a partial ORF and of the *shc* gene from *B. japonicum* USDA 110. A putative ribosome-binding site upstream of the start codon of the *shc* gene and a termination signal downstream of *shc* in the nucleotide sequence are underlined. In the amino acid sequence, the QW motifs are in bold type and underlined. The DV/LDDTA motif is surrounded by a box. Amino acid residues that are identical in all known SHCs are marked with an asterisk. The deduced amino acid residues of the partial ORF1' (start of sequence indicated by an arrow) of *B. japonicum* that are identical with amino acid residues from an ORF located upstream of the *Z. mobilis* *shc* gene are in bold type. The base numbers are shown on the right side of the nucleotide sequence.

Fig. 5. *In vitro* activity of the *shc* gene of *B. japonicum* USDA 110 cloned and expressed in *E. coli*. *E. coli* cells harbouring the cloned *shc* gene [DH5 α (pMPE2)] were grown overnight at 28 °C, harvested, and homogenized for release of the cyclase (see Methods for details). Cyclase activity was tested by adding squalene (Sq) and identifying the reaction products of the cyclase [hopene (H) and diplopterol (D)] by GLC analysis. (a) Assay stopped immediately after addition of cell homogenate; (b) assay stopped after 2 h incubation time at 28 °C. The arrow indicates the position where tetrahymanol should appear as determined with authentic tetrahymanol.

similarity to SHCs and OSCs. The deduced amino acid sequence of this ORF of length 1550 bp contains some amino acid stretches that resemble a characteristic motif of triterpenoid cyclases (the QW motif; see below). However, the function of this ORF has not been determined so far.

The chemistry and enzymology of the complex ring-forming reaction fostered by cyclases has attracted a great deal of attention, and ideas have been put forward about some of the molecular properties of these enzymes and their relatedness. Basically, it is expected that these enzymes derive from a common ancestor and that they show common molecular features necessary for the cyclization reaction (Poralla, 1994). *B. japonicum* SHC shares two striking amino acid motifs with other SHCs. First, a characteristic feature of SHCs and also of OSCs is a repetitive non-tandem motif (QW motif) with the consensus sequence (K/R)(G/A)X₂₋₃(F/Y/W)LX₃QX₂₋₅GXW. The QW motif occurs seven to eight times in SHCs (Fig. 4) and seven times in OSCs in alignable positions (Poralla *et al.*, 1994). This motif contains highly conserved residues of aromatic amino acids. It is noteworthy that the total amount of aromatic amino acids in all known triterpenoid cyclases is unusually high (ranging from 10.4% in the case of *B. japonicum* to 13.5% in the case of *Saccharomyces cerevisiae* OSC; Corey *et al.*, 1994) compared to the normal amount of aromatic residues

found in proteins, which is in the range of 8.4% (Robinson & Robinson, 1991). A high proportion of the aromatic amino acids are strictly conserved throughout the SHCs (30 of 69 aromatic amino acid residues from *B. japonicum* SHC).

Second, the *B. japonicum* SHC contains an aspartate-rich amino acid motif (Fig. 4) with the consensus sequence DV/LDDTA which is also highly conserved among SHCs. Similar aspartate-rich motifs are found in other enzymes of isoprenoid biosynthesis. In these enzymes, the aspartate-rich motifs may be involved in binding and/or stabilization of the diphosphate moieties of the allylic substrates via divalent cations (Mg²⁺ or Mn²⁺; Ashby *et al.*, 1990). However, SHC activity does not depend on the presence of divalent cations, and squalene does not possess a diphosphate group.

One possible function of these conserved negatively charged amino acids and also of the conserved aromatic amino acid residues could be stabilization of intermediate carbocations that occur during the cyclization process (Abe *et al.*, 1993). Recently, site-directed mutagenesis of the second and third aspartate residue of the DV/LDDTA motif of the SHC from *A. acidocaldarius* led to complete inactivation of the SHC, indicating an important role for these residues in enzymic activity (Feil *et al.*, 1996). In total, 11 aspartate and 5 glutamate residues are strictly conserved throughout the known SHCs.

In vitro cyclization of squalene

The cloned *shc* from *B. japonicum* was expressed in *E. coli*. *E. coli* produces neither squalene nor hopanoids and is therefore a suitable system for assaying SHC activity. Homogenates of *E. coli* cells harbouring plasmid pMPE2 were assayed for their ability to catalyse the cyclization of squalene *in vitro*. *shc* in plasmid pMPE2 is located in correct orientation of the *lacZ* promoter. Fig. 5 reveals that after an incubation time of 2 h, squalene was converted partially (to about 20%) to two new compounds with retention times identical to authentic hopene and diplopterol. Both new compounds were further identified by GLC-MS analysis as hopene (*m/z* 410 with a base peak at *m/z* 191) and diplopterol (*m/z* 428 and a base peak at *m/z* 191). The amount of hopene formed was approximately five times that of diplopterol. However, the formation of tetrahymanol from squalene *in vitro* by the *shc* gene product could not be detected. Even a change in assay conditions (e.g. longer incubation times up to 16 h, a change to lower or higher incubation temperatures or a change of the detergent used in the assay) did not lead to the formation of tetrahymanol. This finding agrees with results obtained from *R. palustris* (Kleemann *et al.*, 1994), where the purified SHC converted squalene to hopene and diplopterol but was unable to produce tetrahymanol *in vitro*. It remains unclear if there is a second squalene cyclase in *B. japonicum* and *R. palustris* responsible for tetrahymanol biosynthesis. Partial sequencing of the downstream region of the *shc* gene on the 5 kb *EcoRI*

fragment (Fig. 3) did not lead to identification of sequence homologies to genes encoding triterpenoid cyclases, indicating that a second squalene cyclase gene is not located immediately downstream of *shc*. On the basis of the reaction mechanism and biochemical data from the purified cyclase from the ciliate *Tetrahymena thermophila* (Saar *et al.*, 1991; Abe *et al.*, 1993), one can assume that a squalene-tetrahymanol cyclase should resemble the SHCs in its primary structure. Another possible explanation for the lack of tetrahymanol biosynthesis could be the need of the cyclase for a special membrane environment in the form of certain lipids or proteins to enable the squalene cyclase from *B. japonicum* and *R. palustris* to convert squalene to tetrahymanol. In that case, it is possible that tetrahymanol could only be synthesized under physiological conditions.

Analysis of the upstream region of *shc*

The sequence analysis of 0.5 kb upstream of the *shc* gene from *B. japonicum* on plasmid pMPE2 revealed significant homology between this DNA region of *B. japonicum* and the upstream region of the *shc* gene from *Z. mobilis*. A part of a potential ORF (ORF1'; 163 amino acids; Fig. 4) lacking the 5'-end was discovered that has 64% similarity and 51% identity to the C-terminal 140 amino acids (of 415) of an ORF located immediately upstream of *shc* from *Z. mobilis* (Reipen *et al.*, 1995). Both ORFs have similarity to a stretch of amino acids found in phytoene desaturases, enzymes of the carotenoid biosynthesis pathway, of cyanobacteria and plants (Sandmann, 1994). *Bradyrhizobium* and *Zymomonas* are usually non-pigmented, indicating that this ORF in these two organisms might not be involved in biosynthesis of carotenoids. In the case of *B. japonicum*, this ORF overlaps with *shc* for 13 bp. The close proximity of ORF1' to *shc* may indicate that these two ORFs are members of an operon. Future genetic analysis of this identified gene region, harbouring the *shc* gene and the neighbouring ORFs, will allow the function of these gene regions in the biosynthesis of hopanoids in *Bradyrhizobium* and the role of hopanoids in general in this and other bacteria to be addressed.

ACKNOWLEDGEMENTS

We thank Michel Rohmer for performing GLC-MS analysis and Thomas Härtner for technical assistance. Strain USDA 4088 was kindly provided by Peter van Berkum (United States Department of Agriculture). We also thank Rosemary Nuri for her help in preparing the manuscript. Michael Perzl was supported by the Graduiertenkolleg Mikrobiologie of the University of Tübingen. Grants were also given by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 323 and Po 117/16-1) and the University of Tübingen.

REFERENCES

Abe, I., Rohmer, M. & Prestwich, G. D. (1993). Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chem Rev* 93, 2189–2206.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.

Ashby, M. N., Spear, D. H. & Edwards, P. A. (1990). Prenyltransferases. From yeast to man. In *Molecular Biology of Atherosclerosis*, pp. 27–34. Edited by A. D. Attie. Amsterdam: Elsevier.

Berry, A. M., Moreau, R. A. & Jones, A. D. (1991). Bacteriohopantetrol: abundant lipid in *Frankia* cells and in nitrogen-fixing tissue. *Plant Physiol* 95, 111–115.

Berry, A. M., Harriott, O. T., Moreau, R. A., Osman, S. F., Benson, D. R. & Jones, A. D. (1993). Hopanoid lipids compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proc Natl Acad Sci USA* 90, 6091–6094.

Birnboim, C. & Doly, J. (1979). Rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7, 1513–1523.

Buntel, C. J. & Griffin, J. H. (1992). Nucleotide and deduced amino acid sequences of the oxidosqualene cyclase from *Candida albicans*. *J Am Chem Soc* 114, 9711–9713.

Caspi, E., Zander, J. M., Greig, J. B., Mallory, F. B., Conner, R. L. & Landrey, J. R. (1968). Evidence for a nonoxidative cyclization of squalene in the biosynthesis of tetrahymanol. *J Am Chem Soc* 90, 3563–3564.

Corey, E. J., Matsuda, S. P. T. & Bartel, B. (1993). Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. *Proc Natl Acad Sci USA* 90, 11628–11632.

Corey, E. J., Matsuda, S. P. T. & Bartel, B. (1994). Molecular cloning, characterization, and overexpression of *ERG7*, the *Saccharomyces cerevisiae* gene encoding lanosterol synthase. *Proc Natl Acad Sci USA* 91, 2211–2215.

Feil, C., Süßmuth, R., Jung, G. & Poralla, K. (1996). Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase. *Eur J Biochem* 242, 51–55.

Griffin, H. G., l'Anson, K. J. & Gasson, M. J. (1993). Rapid isolation of genes from bacterial lambda libraries by direct polymerase chain reaction screening. *FEMS Microbiol Lett* 112, 49–54.

Jarvis, B. D. W., Gillis, M. & De Ley, J. (1986). Intra- and intergeneric similarities between the ribosomal ribonucleic acid cistrons of *Rhizobium* and *Bradyrhizobium* species and some related bacteria. *Int J Syst Bacteriol* 36, 129–138.

Jordan, D. C. (1982). Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *Int J Syst Bacteriol* 32, 136–139.

Kannenbergh, E., Poralla, K. & Blume, A. (1980). A hopanoid from the thermoacidophilic *Bacillus acidocaldarius* condenses membranes. *Naturwissenschaften* 67, 458–459.

Kannenbergh, E. L., Perzl, M. & Härtner, T. (1995). The occurrence of hopanoid lipids in *Bradyrhizobium* bacteria. *FEMS Microbiol Lett* 127, 255–262.

Kannenbergh, E. L., Perzl, M., Müller, P., Härtner, T. & Poralla, K. (1996). Hopanoid lipids in *Bradyrhizobium* and other plant associated bacteria and cloning of the *Bradyrhizobium* squalene-hopene cyclase gene. *Plant Soil* 186, 107–112.

Kleemann, G., Poralla, K., Englert, G., Kjösen, H., Liaaen-Jensen, S., Neunlist, S. & Rohmer, M. (1990). Tetrahymanol from the phototrophic bacterium *Rhodospseudomonas palustris*: first report of a gammacerane triterpene from a prokaryote. *J Gen Microbiol* 136, 2551–2553.

Kleemann, G., Kellner, R. & Poralla, K. (1994). Purification and

- properties of the squalene-hopene cyclase from *Rhodospseudomonas palustris*, a purple non-sulfur bacterium producing hopanoids and tetrahymanol. *Biochim Biophys Acta* **1210**, 317–320.
- Knauf, V. C. & Nester, E. W. (1982). Wide host range cloning vectors: a cosmid clone bank of an *Agrobacterium* Ti plasmid. *Plasmid* **8**, 45–54.
- Martinez-Romero, E. & Caballero-Mellado, J. (1996). *Rhizobium* phylogenies and bacterial genetic diversity. *Crit Rev Plant Sci* **15**, 113–140.
- Ochs, D., Kaletta, C., Entian, K.-D., Beck-Sickinger, A. & Poralla, K. (1992). Cloning, expression, and sequencing of squalene-hopene cyclase, a key enzyme in triterpenoid metabolism. *J Bacteriol* **174**, 298–302.
- Ourisson, G., Rohmer, M. & Poralla, K. (1987). Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annu Rev Microbiol* **41**, 301–333.
- Parniske, M., Kosch, K., Werner, D. & Müller, P. (1993). ExoB mutants of *Bradyrhizobium japonicum* with reduced competitiveness for nodulation of *Glycine max*. *Mol Plant-Microbe Interact* **6**, 99–106.
- Poralla, K. (1994). The possible role of a repetitive amino acid motif in evolution of triterpenoid cyclases. *Bioorg Med Chem Lett* **4**, 285–290.
- Poralla, K., Hewelt, A., Prestwich, G. D., Abe, I., Reipen, I. & Sprenger, G. (1994). A specific amino acid repeat in squalene and oxidosqualene cyclases. *Trends Biochem Sci* **19**, 157–158.
- Regensburger, B. & Hennecke, H. (1983). RNA polymerase from *Rhizobium japonicum*. *Arch Microbiol* **135**, 103–109.
- Reipen, I. G., Poralla, K., Sahm, H. & Sprenger, G. A. (1995). *Zymomonas mobilis* squalene-hopene cyclase gene (*shc*): cloning, DNA sequence analysis, and expression in *Escherichia coli*. *Microbiology* **141**, 155–161.
- Robinson, A. B. & Robinson, L. R. (1991). Distribution of glutamine and asparagine residues and their near neighbours in peptides and proteins. *Proc Natl Acad Sci USA* **88**, 8880–8884.
- Roessner, C. A., Min, C., Hardin, S. H., Harris-Haller, L. W., McCollum, J. C. & Scott, A. I. (1993). Sequence of the *Candida albicans* *erg7* gene. *Gene* **127**, 149–150.
- Saar, J., Kader, J.-C., Poralla, K. & Ourisson, G. (1991). Purification and some properties of the squalene-tetrahymanol cyclase from *Tetrahymena thermophila*. *Biochim Biophys Acta* **1075**, 93–101.
- Sahm, H., Rohmer, M., Bringer-Meyer, S., Sprenger, G. A. & Welle, R. (1993). Biochemistry and physiology of hopanoids in bacteria. *Adv Microb Physiol* **35**, 247–273.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sandmann, G. (1994). Carotenoid biosynthesis in microorganisms and plants. *Eur J Biochem* **223**, 7–24.
- Simon, R., Priefer, U. & Pühler, A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Biotechnology* **1**, 784–791.
- Somasegaran, P. & Hoben, H. J. (1994). *Handbook for Rhizobia: Methods in Legume-Rhizobium Technology*. New York: Springer.
- Tappe, C. H. (1993). *Squalen-Hopen-Cyclasen: Reinigung, Charakterisierung und Inhibitor-Experimente*. PhD thesis, University of Tuebingen, Germany.
- Tully, R. E. & Keister, D. L. (1993). Cloning and mutagenesis of a cytochrome P-450 locus from *Bradyrhizobium japonicum* that is expressed anaerobically and symbiotically. *Appl Environ Microbiol* **59**, 4136–4142.
- Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259–268.
- Vilcheze, C., Llopiz, P., Neunlist, S., Poralla, K. & Rohmer, M. (1994). Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria *Azotobacter vinelandii*, *Beijerinckia indica* and *Beijerinckia mobilis*. *Microbiology* **140**, 2749–2753.
- Vincent, J. M. (1970). *A Manual for the Practical Study of Root-Nodule Bacteria*. IBP Handbook no. 15. Oxford: Blackwell Scientific.
- Wilcockson, J. & Werner, D. (1976). Nitrogenase activity by *Klebsiella* and *Rhizobium* on solid substrata exposed to air. *Ber Dtsch Bot Ges* **89**, 587–607.
- Wong, F. Y. K., Stackebrandt, E., Ladha, J. K., Fleischman, D. E., Date, R. A. & Fuerst, J. A. (1994). Phylogenetic analysis of *Bradyrhizobium japonicum* and photosynthetic stem-nodulating bacteria from *Aeschynomene* species grown in separated geographical regions. *Appl Environ Microbiol* **60**, 940–946.
- Young, J. P. W., Downer, H. L. & Eardly, B. D. (1991). Phylogeny of the phototrophic *Rhizobium* strain BTAi1 by polymerase chain reaction-based sequencing of a 16S rRNA gene segment. *J Bacteriol* **173**, 2271–2277.

Received 20 June 1996; revised 21 October 1996; accepted 28 October 1996.