

University of Groningen

## Identification of 3-Ketosteroid 9-Alfa-hydroxylase genes and microorganisms bloked in 3-Ketosteroid 9-Alfa-hydroxylase activity

van der Geize, Robert; van der Meijden, Peter ; Hessels, Gerda; Dijkhuizen, Lubbert

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

### *Document Version*

Publisher's PDF, also known as Version of record

### *Publication date:*

2003

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

van der Geize, R., van der Meijden, P., Hessels, G., & Dijkhuizen, L. (2003). Identification of 3-Ketosteroid 9-Alfa-hydroxylase genes and microorganisms bloked in 3-Ketosteroid 9-Alfa-hydroxylase activity. (Patent No. WO03070925).

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 August 2003 (28.08.2003)

PCT

(10) International Publication Number  
**WO 03/070925 A2**

- (51) International Patent Classification<sup>7</sup>: **C12N 9/02**
- (21) International Application Number: PCT/EP03/50025
- (22) International Filing Date: 19 February 2003 (19.02.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
02075723.3      21 February 2002 (21.02.2002)      EP
- (71) Applicant (for all designated States except US): **AKZO NOBEL N.V.** [NL/NL]; Velperweg 76, NL-6824 BM Arnhem (NL).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **VAN DER GEIZE, Robert** [NL/NL]; Zaagmuldersweg 586, NL-9713 LZ Groningen (NL). **VAN DER MEIJDEN, Peter** [NL/NL]; DIOSYNTH B.V., P.O. Box 20, NL-5340 BH Oss (NL). **HESELS, Gerda** [NL/NL]; Kraaienest 131, NL-9733 HK Groningen (NL). **DIJKHUIZEN, Lubbert** [NL/NL]; Ter Borch 28, NL-9474 RB Zuidlaren (NL).
- (74) Agent: **Van Wezenbeek, P.M.G.F.**; P.O. Box 20, NL-5340 BH Oss (NL).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



**WO 03/070925 A2**

(54) Title: IDENTIFICATION OF 3-KETOSTEROID 9-ALFA-HYDROXYLASE GENES AND MICROORGANISMS BLOCKED IN 3-KETOSTEROID 9-ALFA-HYDROXYLASE ACTIVITY

(57) Abstract: The invention relates to an isolated polynucleotide sequence comprising a nucleic acid sequence encoding the amino acid sequence of KshA protein or of KshB protein, encoded by nucleotides 499-1695 of SEQ ID NO: 1 or by nucleotides 387-1427 of SEQ ID NO:2, respectively, and functional homologues thereof. The polynucleotides of the invention can be used to construct genetically modified microorganisms blocked in 3-ketosteroid 9 $\alpha$ -hydroxylase activity, which are useful in the microbial degradation of steroids to accumulate certain steroid products.

## IDENTIFICATION OF 3-KETOSTEROID 9-ALFA-HYDROXYLASE GENES AND MICROORGANISMS BLOCKED IN 3-KETOSTEROID 9-ALFA-HYDROXYLASE ACTIVITY

- 5 The invention relates to isolated polynucleotide sequences encoding 3-ketosteroid 9 $\alpha$ -hydroxylase components, to microorganisms blocked in 3-ketosteroid 9 $\alpha$ -hydroxylase activity, to a method for the preparation of said microorganisms, and to the use of such microorganisms in steroid  $\Delta^1$ -dehydrogenation.
- 10 To date very limited knowledge is available on 3-ketosteroid 9 $\alpha$ -hydroxylase (KSH), the enzyme performing the 9 $\alpha$ -hydroxylation of 4-androstene-3,17-dione (AD) and 1,4-androstadiene-3,17-dione (ADD) in microbial sterol/steroid degradation. No nucleotide sequences of the genes encoding KSH components have been reported. Furthermore, difficulties are faced during enzyme purification procedures (Chang, F. N. *et al.* *Biochemistry* (1964) **3**:1551-1557; Strijewski, A. *Eur. J. Biochem.* (1982) **128**:125-135). A three-component monooxygenase with KSH activity has been partially purified from *Nocardia* sp. M117 and was found to constitute a three-component enzyme system, composed of a flavoprotein reductase and two ferredoxin proteins (Strijewski, A. *Eur. J. Biochem.* (1982) **128**:125-135). In *Arthrobacter oxydans* 317, 9 $\alpha$ -hydroxylation of the steroid poly-cyclic ring structure appeared plasmid-borne (Dutta, R.K. *et al.* *J. Basic Microbiol.* (1992) **32**:317-324).
- 20 Nucleotide sequence analysis of the plasmid, however, was not reported.

The lack of genetic data has hampered the construction of molecularly defined mutant strains with desired properties (i.e. blocked 9 $\alpha$ -hydroxylation of steroids) by genetic engineering. Mutants have been isolated by classical mutagenesis, but these strains usually are inadequate in industrial processes mostly due to genetic instability and/or low bioconversion efficiencies. Molecularly defined mutants have advantages compared to mutants generated by classical mutagenesis. The constructed mutants are genetically stable and the introduced mutations are well-defined genetic modifications. Construction of genetically engineered strains make the widespread use of chemical agents to block 9 $\alpha$ -hydroxylation (e.g.  $\alpha,\alpha$ -dipyridyl, *o*-phenanthroline) obsolete. Chemical agents used to block KSH activity mostly are not reaction specific and inhibit other important enzymatic reactions (e.g. sterol 26-hydroxylation in sterol side chain degradation), which may have negative effects on sterol bioconversion efficiency. The use of defined mutants by genetic engineering overcomes these problems.

- 35 3-Ketosteroid 9 $\alpha$ -hydroxylase (KSH) is a key-enzyme in the microbial steroid ring B-opening pathway. KSH catalyzes the conversion of AD into 9 $\alpha$ -hydroxy-4-androstene-3,17-dione (9OHAD) and ADD into the chemically unstable compound [9OHADD]. KSH activity has been found in many

bacterial genera (Martin, C.K.A. Adv. Appl. Microbiol. (1977) **22**: 29-58; Kieslich, K. J Basic Microbiol. (1985) **25**: 461-474; Mahato, S.B. *et al.* Steroids (1997) **62**: 332-345): *e.g.* *Rhodococcus* (Datcheva, V.K. *et al.* Steroids (1989) **54**:271-286; Van der Geize *et al.* FEMS Microbiol. Lett. (2001) **205**: 197-202, *Nocardia* (Strijewski, A. Eur. J. Biochem. (1982) **128**:125-135), *Arthrobacter* (Dutta, R.K. *et al.* J. Basic Microbiol. (1992) **32**:317-324) and *Mycobacterium* (Wovcha, M.G. *et al.* Biochim Biophys Acta (1978) **531**:308-321). Bacterial strains lacking KSH activity are being considered important in sterol/steroid biotransformation. Mutants blocked in KSH activity will be able to perform only the KSTD (3-ketosteroid  $\Delta^1$ -dehydrogenase) reaction, thereby allowing selective  $\Delta^1$ -dehydrogenation of steroid compounds. Examples are the cortisol biotransformation into prednisolone and the AD biotransformation into ADD. Sterol bioconversion by mutants blocked at the level of steroid  $9\alpha$ -hydroxylation may also carry out a selective degradation of the sterol side chain thereby accumulating AD and/or ADD which are excellent precursors for the synthesis of bioactive steroid hormones.

15 According to one aspect of the present invention, the isolated polynucleotide sequences of two genes, designated *kshA* and *kshB* of *Rhodococcus erythropolis* are now provided: SEQ ID NO:1 and SEQ ID NO:2, respectively. KshA protein is encoded by nucleotides 499-1695 of SEQ ID NO:1 and KshB protein by nucleotides 387-1427 of SEQ ID NO:2. Thus, in particular preferred are polynucleotides comprising the complete coding DNA sequences of the nucleotides 499-1695 of SEQ ID NO:1 and of the nucleotides 387-1427 of SEQ ID NO:2, respectively.

20 Furthermore, to accommodate codon variability the invention also includes sequences coding for the same amino acid sequences of the KshA protein and the KshB protein. Also portions of the coding sequences coding for individual domains of the expressed protein are part of the invention as well as allelic and species variations thereof. Sometimes, a gene is expressed as a splicing variant, resulting in the inclusion of an additional exon sequence, or the exclusion of an exon. Also a partial exon sequence may be included or excluded. A gene may also be transcribed from alternative promoters that are located at different positions within a gene, resulting in transcripts with different 5' ends. Transcription may also terminate at different sites, resulting in different 3' ends of the transcript. These sequences as well as the proteins encoded by these sequences all are expected to perform the same or similar functions and form also part of the invention. The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequence disclosed herein can be readily used to isolate the complete genes which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

The present invention further relates to polynucleotides having slight variations or having polymorphic sites. Polynucleotides having slight variations encode polypeptides which retain the same biological function or activity as the natural, mature protein.

- 5 The DNA according to the invention may be obtained from cDNA using suitable probes derived from SEQ ID NO:1 or SEQ ID NO:2. Alternatively, the coding sequence might be genomic DNA, or prepared using DNA synthesis techniques. The polynucleotide may also be in the form of RNA. If the polynucleotide is DNA, it may be in single stranded or double stranded form. The single strand might be the coding strand or the non-coding (anti-sense) strand.
- 10 The present invention further relates to polynucleotides which have at least 70%, preferably 80%, more preferably 90%, even more preferred 95%, and highly preferably 98% and most preferred at least 99% identity with the entire DNA sequence of the nucleotides 499-1695 of SEQ ID NO:1 and of the nucleotides 387-1427 of SEQ ID NO:2, respectively. Such polynucleotides encode polypeptides which retain the same biological function or activity as the natural, mature protein.
- 15 Alternatively, also fragments of the above mentioned polynucleotides which code for domains of the protein which still are capable of binding to substrates are embodied in the invention.

The percentage of identity between two sequences can be determined with programs such as Clustal W 1.7 (Thompson J. D., *et al.* Nucleic Acids Res.(1994) **22**: 4673-4680: "CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix.") used in default settings. The percentage identity generally is defined by the number of identical residues between the two sequences divided by the total number of residues of the known sequence.

20 Similarity is defined as a combination of identity together with all semi-conserved amino acid residues in the alignment according to the groups as defined in ClustalW 1.7:

'\*' = identity = indicates positions which have a single, fully conserved residue

':' = semi-conserved = indicates that one of the following 'strong' groups is fully conserved. STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

30 '.' = semi-conserved = indicates that one of the following 'weaker' groups is fully conserved. CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY.

These are all the positively scoring groups that occur in the Gonnet Pam250 matrix.

Also within the scope of this invention are functional homologues of the new genes *e.g.* in the family of Actinomycetales (*e.g.* *Rhodococcus*, *Nocardia*, *Arthrobacter*, *Corynebacterium* and *Mycobacterium*).

35

In order to identify such genes with similar action in other microorganisms, any method for detection of (poly)nucleotides known in the art for such purpose is included herewith. For example, nucleotide elongation methods / amplification methods may be considered, but also, such method may comprise the steps of: hybridizing to a sample a probe specific for a polynucleotide encoding an amino acid sequence of KshA or KshB under conditions effective for said probe to hybridize specifically to said polynucleotide and determining the hybridization of said probe to polynucleotides in said sample. The term "specific" in this respect means that the majority of hybridization takes place with a polynucleotide of this invention. Preferably, said probe comprises at least 25 of the nucleotides of SEQ ID NO:1 or SEQ ID NO:2. More preferred, the probe comprises 50, and in particular preferred more than 100, nucleotides of SEQ ID NO:1 or SEQ ID NO:2. Most preferred, the probe consists of a polynucleotide of nucleotides selected from the nucleotides 499-1695 of SEQ ID NO:1 and of the nucleotides 387-1427 of SEQ ID NO:2, respectively. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45 °C, followed by a wash of 2.0 x SSC at 50 °C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from low stringency of about 2.0 x SSC at 50 °C to a high stringency of about 0.2 x SSC at 50 °C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22 °C, to high stringency conditions at about 65 °C.

Alternatively, the polynucleotides of this invention may also be used for targeting specific genes, *e.g.* for the purpose of gene disruption in other species (see for example WO 01/31050 and references cited therein).

The sequence of the newly identified polynucleotide of the present invention, SEQ ID NO:1 and SEQ ID NO:2 may also be used in the preparation of vector molecules for the expression of the encoded protein in suitable host cells. A wide variety of host cell and cloning vehicle combinations may be usefully employed in cloning the nucleic acid sequences coding for the proteins KshA or KshB or parts thereof. For example, useful cloning vehicles may include chromosomal, non-chromosomal and synthetic DNA sequences such as various known bacterial plasmids and wider host range plasmids and vectors derived from combinations of plasmids and phage or virus DNA. Vehicles for use in expression of the polynucleotides of the present invention or a part thereof comprising a functional domain will further comprise control sequences operably linked to the nucleic acid sequence coding for the protein. Such control sequences generally comprise a promoter sequence and sequences which regulate and/or enhance expression levels. Of course control and other sequences can vary depending on the host cell selected.

Suitable expression vectors are for example bacterial or yeast plasmids, wide host range plasmids and vectors derived from combinations of plasmid and phage or virus DNA. Vectors derived from chromosomal DNA are also included. Furthermore an origin of replication and/or a dominant selection marker can be present in the vector according to the invention. The vectors according to the invention are suitable for transforming a host cell. Integrative vectors may also be regarded as suitable expression vehicles.

Recombinant expression vectors comprising DNA of the invention as well as cells transformed with said DNA or said expression vector also form part of the present invention.

Suitable host cells according to the invention are bacterial host cells, yeast and other fungi, insect, plant or animal host cells such as Chinese Hamster Ovary cells or monkey cells or human cell lines. Thus, a host cell which comprises DNA or expression vector according to the invention is also within the scope of the invention. The engineered host cells can be cultured in conventional nutrient media which can be modified e.g. for appropriate selection, amplification or induction of transcription. The culture conditions such as temperature, pH, nutrients etc. are well known to those ordinary skilled in the art.

The techniques for the preparation of DNA or the vector according to the invention as well as the transformation or transfection of a host cell with said DNA or vector are standard and well known in the art, see for instance Sambrook et al., *Molecular Cloning: A laboratory Manual*. 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

In another aspect of the invention, there is provided for a protein comprising the amino acid sequence encoded by any of the above described DNA molecules. Preferably, the protein according to the invention comprises an amino acid sequence encoded by the nucleotides 499-1695 of SEQ ID NO:1 or of the nucleotides 387-1427 of SEQ ID NO:2, respectively. Also part of the invention are proteins resulting from post translational processing, which proteins are encoded by the polynucleotide of this invention.

Also functional equivalents, that is proteins homologous to amino acid sequences of KshA and KshB or parts thereof having variations of the sequence while still maintaining functional characteristics, are included in the invention.

The variations that can occur in a sequence may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. Amino acid substitutions that are expected not to essentially alter biological activities, have been described. Amino acid replacements between related amino acids or replacements which have occurred frequently in evolution are, inter alia Ser/Ala, Ser/Gly, Asp/Gly,

Asp/Asn, Ile/Val (see Dayhof, M.D., Atlas of protein sequence and structure, Nat. Biomed. Res. Found., Washington D.C., 1978, vol. 5, suppl. 3). Based on this information Lipman and Pearson developed a method for rapid and sensitive protein comparison (Science, 1985, 227, 1435-1441) and determining the functional similarity between homologous polypeptides. It will be clear that also  
5 polynucleotides coding for such variants are part of the invention.

The polypeptides according to the present invention also include polypeptides comprising an amino acid sequence encoded by the nucleotides 499-1695 of SEQ ID NO:1 or of the nucleotides 387-1427 of SEQ ID NO:2, respectively, but further polypeptides with a similarity of at least 70%, preferably 80%, more preferably 90%, and even more preferred 95%. Also portions of such polypeptides still  
10 capable of conferring biological effects are included.

Another aspect of the present invention are genetically modified microorganisms. For the construction of mutant strains incapable of 9 $\alpha$ -hydroxylation, the genes encoding KSH activity must be identified and their nucleotide sequence must be known. The two genes of this invention,  
15 designated *kshA* and *kshB*, were identified in *Rhodococcus erythropolis* SQ1 to encode KSH. These genes were cloned by functional complementation of two UV mutants, designated strains RG1-UV26 and RG1-UV39, both impaired in 9 $\alpha$ -hydroxylation of AD(D). They were isolated via an extensive screening of UV irradiated cells of *R. erythropolis* strain RG1 (van der Geize, R. *et al.* FEMS Microbiol. Lett. Submitted 2001), a *kstD* (encoding 3-ketosteroid  $\Delta^1$ -dehydrogenase =  
20 KSTD1) gene deletion mutant of strain SQ1. Strains RG1-UV26 and RG1-UV39 were unable to grow on AD and ADD, but grew normally on 9OHAD, indicating deficient KSH activity.

For functional complementation of the KSH deficient mutant strains RG1-UV26 and RG1-UV39 and cloning of the *kshA* and *kshB* genes, a genomic library of *R. erythropolis* RG1 was constructed using  
25 *Rhodococcus-E. coli* shuttle vector pRESQ (Fig. 1). *Sau3A* digested chromosomal DNA of *R. erythropolis* RG1 was sized by sucrose gradient to 6-10 kb and ligated into *Bgl*III digested pRESQ. Transformation of *E. coli* Top10F' (Invitrogen Corp.) with this ligation mixture generated a gene library of approximately 15,000 transformants in which approximately 90% of the constructs contained insert. An average insert size of 6 kb was estimated. No complications with stability or  
30 rearrangements were apparent. The gene library represents the complete genome (p>0.99) assuming a genome size of approximately 6 Mb.

Introducing the *R. erythropolis* RG1 genomic library into strains RG1-UV39 and RG1-UV26 and subsequent screening for complementation of KSH deficiency, resulted in the cloning of two  
35 independent DNA fragments containing the *kshA* gene and the *kshB* gene, respectively (Fig. 2).



Analysis of these genes revealed that *kshA* encodes a 398 amino acid protein (KshA). KshA showed high similarity (58% identity; 84% similarity) to a hypothetical protein encoded by gene Rv3526 (DDBJ/EMBL/GenBank accession no. CAB05051) in *Mycobacterium tuberculosis* (Cole, S.T. *et al.* Nature (1998) **393**: 537-544). Rv3526 is thus expected to be the homologue of *kshA* in *M. tuberculosis*. Comparison of the obtained nucleotide sequence of *kshA* to databases further revealed that *kshA* is identical (97%) to a hypothetical gene (ORF12), found by Maeda, M. *et al.* (Appl. Environ. Microbiol. (1995) **61**:549-555) in *R. erythropolis* strain TA421 (DDBJ/EMBL/GenBank accession no. D88013) upstream of *bphC1* (Fig. 3). In analogy with the molecular organization found in strain TA421, a hypothetical ORF11, identified in this strain downstream of ORF12, was also identified downstream of *kshA* in strain SQ1. The nucleotide sequences of the DNA fragments of strain SQ1 and strain TA421 were therefore merged and the resulting theoretical nucleotide sequence was used for the successful construction of plasmid pKSH126 (Fig. 3) used for the introduction of an unmarked in-frame *kshA* gene deletion in *R. erythropolis* strain SQ1 (rendering strain RG2) and strain RG8 (rendering strain RG9).

The *kshB* gene encodes a 346 amino acid protein (KshB). Database similarity searches revealed that KshB showed high similarity to ferredoxin reductase components of multi-component oxygenases. Highest similarity (56% identity; 85% similarity) was found with Rv3571 of *M. tuberculosis* (DDBJ/EMBL/GenBank accession no.A70606).

Inactivation of *kshA* or *kshB* by unmarked gene deletion renders molecularly defined and genetically stable mutant strains capable of selective  $\Delta^1$ -dehydrogenation of AD producing ADD that is not further metabolized due to absence of KSH activity (see WO 01/31050). Using the *sacB* counter selection system (described in WO 01/31050) three unmarked gene deletion mutant strains were constructed: a *kshA* mutant *R. erythropolis* RG2 using pKSH126 (Fig. 3), a *kshB* mutant *R. erythropolis* RG4 using pKSH212 (Fig. 4) and a *kstD kstD2 kshA* mutant *R. erythropolis* RG9. Strains RG2 and RG4 are derived from *R. erythropolis* SQ1. Strain RG9 is derived from *kstD kstD2* mutant *R. erythropolis* RG8 using pKSH126 (Fig. 3). Strain RG8 lacks both 3-ketosteroid  $\Delta^1$ -dehydrogenase isoenzymes (KSTD1 and KSTD2; described in WO 01/31050).

Thus, another aspect of this invention is a microorganism blocked in 3-ketosteroid 9 $\alpha$ -hydroxylase activity characterized in that it is a genetically modified microorganism, in particular of the family of Actinomycetales, preferably of the *Rhodococcus* genus and most preferred of *Rhodococcus erythropolis*. Also preferred is a strain, wherein at least one gene encoding 3-ketosteroid  $\Delta^1$ -dehydrogenase activity is inactivated, preferably by unmarked gene deletion. In particular preferred are the strains RG2, RG4 and RG9.

Also an aspect of the present invention is a method to construct a genetically modified strain of a

steroid degrading microorganism lacking the ability to degrade the steroid nucleus, the method comprising inactivation of the genes encoding KSH-activity, preferably the gene *kshA* and/or the gene *kshB*. Preferably, the inactivation of the gene(s) is accomplished by targeted, preferably unmarked, gene deletion.

5

A further aspect of the present invention is the use of a genetically modified microorganism in steroid  $\Delta^1$ -dehydrogenation, in particular in the preparation of 1,4-androstadiene-3,17-dione and prednisolone. Preferably, the microorganism for such use has been obtained by targeted gene inactivation, preferably unmarked gene deletion, of the genes encoding KSH-activity in a  
10 microorganism of the family of Actinomycetales, preferably the gene *kshA* and/or the gene *kshB*. Preferred microorganism for this use is selected from the genetically modified strains RG2, RG4 and RG9.

The micro-organism strains *Rhodococcus erythropolis* RG2, RG4 and RG9 have been deposited at  
15 the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Mascheroder Weg 1b, D-38124 Braunschweig, Germany under the accession numbers DSM 14544, DSM 14545 and DSM 14546, respectively. These deposits have been made under the terms of the Budapest Treaty.

Methods to construct vehicles to be used in the mutagenesis protocol are well known (Sambrook *et al.*, Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, latest edition). Furthermore, techniques for site directed mutagenesis, ligation of additional  
20 sequences, PCR, sequencing of DNA and construction of suitable expression systems are all, by now, well known in the art. Portions or all of the DNA encoding the desired protein can be constructed synthetically using standard solid phase techniques, preferably to include restriction sites  
25 for ease of ligation.

Modifications and variations of the method for introducing disrupted gene mutations, targeted gene inactivation, and in particular unmarked gene deletion as well as transformation and conjugation will be obvious to those skilled in the art from the detailed description of the invention. Such modifications and variations are intended to come within the scope of present application.

30 A person skilled in the art will understand how to use the methods and materials described and referred to in this document in order to construct micro-organisms lacking KSH activity.

The following examples are illustrative for the invention and should in no way be interpreted as limiting the scope of the invention.

35

## Legends to figures

**Figure 1.** The pZErO-2.1 (curved thin line) based *Rhodococcus-E. coli* shuttle vector pRESQ used for constructing a genomic library of *R. erythropolis* RG1. rep: 2.5 kb region of pMVS301 coding for autonomous replication in *Rhodococcus* sp (thick curved line). *lacZ-ccdB*: marker for positive selection in *E. coli*. *aphII*: kanamycin resistance marker for selection in *Rhodococcus* and *E. coli*.

**Figure 2.** Strategy for the separate cloning and identification of the *kshA* and *kshB* genes encoding KSH activity in *R. erythropolis* SQ1 by functional complementation of UV mutant strains RG1-UV39 (A) and RG1-UV26 (B), respectively, using several pRESQ derived constructs.

**Figure 3. (A)** Overview of the 2.6 kb DNA fragment of *R. erythropolis* SQ1 encoding *kshA* and its 4.1 kb counterpart in *R. erythropolis* TA421. The grey bar of ORF12, together with the black arrow (proposed size of ORF12 by Maeda et al. (1995)), represents the actual size of ORF12 in *R. erythropolis* TA421, which is identical (97%) to *kshA*. The “X” indicates the point of merger of the two sequences. **(B)** Scheme of the theoretically merged nucleotide sequences of the DNA fragments of strain SQ1 and strain TA421, and its use in the construction of plasmid pKSH126 for unmarked in-frame *kshA* gene deletion in strain SQ1 and strain RG8. Numbers 1-4 indicate the primers used to obtain PCR products 1 and 2 for the construction of plasmid pKSH126 used for *kshA* unmarked gene deletion in parent strain SQ1 and *kstD kstD2* mutant strain RG8.

**Figure 4.** Cloning scheme for the construction of plasmid pKSH212 used in the construction of *kshB* unmarked gene deletion mutant strain RG4 from parent strain SQ1.

## Examples

### General

#### Construction of the pRESQ shuttle vector.

A pZErO-2.1 (Invitrogen Corp. San Diego, Calif.) derivative was constructed in which the *Bam*HI site was replaced by a *Bg*III site (Fig. 1). A *Sac*I-*Stu*I fragment of pZErO-2.1, containing the *lacZ-ccdB* gene, was duplicated by PCR using a mutagenic forward primer (5' ACCGAGCTCAGATCTACTAGTAACGGC 3', SEQ ID NO:3), containing the desired *Bg*III restriction site (double underlined) and a *Sac*I restriction site (underlined), and reverse primer (5' ATTCAGGCCTGACATTTATATTCCCC 3', SEQ ID NO:4) with a *Stu*I restriction site (underlined). The obtained PCR product was digested with restriction enzymes *Sac*I and *Stu*I and

ligated in *SacI-StuI* digested pZErO-2.1 (pRES14). The *aphII* gene from pWJ5 (Jäger, W.A. et al. (1992). J Bacteriol 174:5462-5465) was cloned as blunted *HindIII-BamHI* fragment (Klenow fill in) into *EcoRV* digested pBlueScript (II) KS to construct pBsKm1. The unique *BglIII* site in pBsKm1 was destroyed by *BglIII* digestion followed by Klenow fill in (pRES4). The ampicillin cassette present in pRES4 was removed by self ligation following *BspHI* digestion (pRES9). A 2.5 kb *BglIII-XbaI* fragment of pMVS301 (Vogt-Singer, M.F. et al., (1988) J. Bacteriol. 170:638-645), containing the region for autonomous replication in *Rhodococcus* sp., was subsequently ligated into *BamHI-XbaI* digested pRES9 to construct pRES11. Construction of pRESQ (6.55 kb) was completed by ligating a 4.15 kb *BspHI/NcoI* fragment of pRES11 into *BspHI-NcoI* digested pRES14.

10

### Example 1

Inactivation of steroid 9 $\alpha$ -hydroxylase activity by UV mutagenesis.

15

Late exponential phase *R. erythropolis* RG1 (= *kstD* mutant) cells ( $2 \cdot 10^8$  CFUs $\cdot$ ml $^{-1}$ ) grown in 10 mM glucose mineral medium were sonicated for a short period of time to obtain single cells. Diluted ( $10^4$ ) samples were spread on glucose mineral agar medium and irradiated for 15-20 sec with a UV lamp (Philips TAW 15W) at a distance of 27 cm, on average resulting in 95% killing of cells. After 4 days of incubation, colonies that had appeared were replica plated on AD (0.5 g $\cdot$ l $^{-1}$  solubilized in DMSO (50 mg $\cdot$ ml $^{-1}$ )) mineral agar medium. A screening for AD(D) growth deficient mutants of *R. erythropolis* RG1 able to grow on 9OHAD mineral medium yielded 2 mutants that were clearly impaired in the KSH reaction. These mutants, designated strain RG1-UV26 and strain RG1-UV39, showed no growth after 3-4 days with either AD or ADD as sole carbon and energy source, while growth on 9OHAD mineral agar medium was normal.

20

### Example 2

Cloning and molecular characterization of *kshA* and *kshB*.

30

The *R. erythropolis* strain RG1 gene library was introduced into strain RG1-UV39 by electrotransformation to complement its mutant phenotype (Fig. 2). A clone was isolated containing a 6 kb insert (pKSH101) that was able to restore growth of strain RG1-UV39 on AD mineral agar medium. Restriction enzyme mapping analysis, subcloning in pRESQ and subsequent complementation experiments resulted in identification of a 1.8 kb *BamHI-Sau3A* DNA fragment (pKSH106) that was still able to complement strain RG1-UV39 (Fig. 2). This 1.8 kb insert was cloned into pBlueScript (II) KS and its nucleotide sequence determined. Nucleotide sequence analysis revealed a single 1,197 nt ORF (*kshA*, 499-1695 of SEQ ID NO:1) encoding a putative

35

protein of 398 aa (KshA).

Complementation of *R. erythropolis* RG1-UV26 with the strain RG1 gene library resulted in isolation of clone pKSH200 able to restore growth of strain RG1-UV26 on AD mineral agar medium (Fig. 2). By subsequent restriction mapping analysis, subcloning and complementation experiments of pKSH200 we identified a 2.8 kb *Bgl*III-*Sau*3A fragment (pKSH202) which was still able to restore the mutant phenotype of strain RG1-UV26 (Fig. 2). This fragment was subcloned into pBlueScript (II) KS and its nucleotide sequence was determined. The ORF responsible for complementing the RG1-UV26 mutant phenotype was identified from a subsequent complementation experiment. An *Asp*718 restriction site mapped on the 2.8 kb fragment was used to construct pKSH205, which could no longer complement the mutant phenotype (Fig. 2). The *Asp*718 restriction enzyme thus is located within the ORF responsible for complementation. The identified ORF of 1,041 nt was designated *kshB* (387-1427 of SEQ ID NO:2; GC content, 62.3%) encoding a putative protein of 346 amino acids with a calculated molecular weight of 37.1 kDa (KshB).

### 15 Example 3

Unmarked gene deletion of *kshA* in *R. erythropolis* SQ1.

For unmarked in-frame gene deletion of *kshA* ( $\Delta kshA$ ) pKSH126 was constructed. A 1.3 kb fragment (PCR product 1) was obtained from pKSH101 using a primer (Fig. 3 primer 1) annealing to sequences upstream of *kshA* (5' CGCGGGCCCATCGAGAGCACGTT 3', SEQ ID NO:5), and a primer (Fig. 3 primer 2) annealing to the 5'-end of the *kshA* gene (5' GCGCCCGGGTCCGAGTGCCATGTCTTC 3', SEQ ID NO:6) containing a *Sma*I site (underlined). Primers for PCR product 2 were developed using the nucleotide sequence of the merged sequences of strain SQ1 and strain TA421 (DDBJ/EMBL/GenBank accession no. D88013). PCR product 2 (840 bp) was obtained from SQ1 chromosomal DNA using forward primer (Fig. 3 primer 3) annealing to the 3' end of the *kshA* gene (5' GCGCCCGGGACAACCTCCTGATTCGCAGTC 3', SEQ ID NO:7), including a *Sma*I restriction site (underlined), and reverse primer (Fig. 3 primer 4) annealing to ORF11 (5' GCGTCTAGAGTGGAAGAGCATTCCTCGCA 3', SEQ ID NO:8), including a *Xba*I restriction site (underlined). The *Sma*I restriction site was introduced to give an in-frame deletion of *kshA* following ligation of this PCR fragment behind the 5'-truncated *kshA* gene from PCR product 1. Finally, a 2.05 kb *Sal*I-*Xba*I fragment from pKSH125 was ligated into pK18mobsacB vector (pKSH126).

Unmarked in-frame deletion of the *kshA* gene was obtained by introducing the mutagenic vector pKSH126 in strain SQ1 followed by *sacB* counter selection (WO 01/31050). Wild type *kshA* gene was reduced to an ORF ( $\Delta kshA$ ) of 30 nt, encoding only 9 amino acids (MALGPGTTS). Gene deletion of *kshA* was confirmed by Southern analysis of *Bam*HI digested chromosomal DNA using

the 2 kb insert of pKSH126 as a probe: a 2.05 kb wild type *Bam*HI DNA fragment was reduced to 0.88 kb in the gene deletion mutant strains. The resulting strain is designated *R. erythropolis* RG2.

#### Example 4

##### 5 Unmarked gene deletion of *kshB* in *R. erythropolis* SQ1.

For unmarked in-frame gene deletion of *kshB* construct pKSH212 was made (Fig. 4). For this purpose pKSH201 was fused to a PCR product (1,275 bp) obtained from pKSH202 as a template. PCR product was obtained with forward primer 5' GCGGGTACCGATCGCCTGAAGATCGAGT  
10 3' (SEQ ID NO:9) and reverse primer 5' GCGAAGCTTGCCGGCGTCGCAGCTCTGTG 3' (SEQ ID NO:10) and ligated (blunt) into *Eco*RV digested pBlueScript(II) KS (pKSH210). At the 5'-terminal end of this PCR product an *Asp*718 restriction site, preceding the stop codon of *kshB* (see forward primer), was introduced to ensure proper in-frame deletion of *kshB* after ligation with the *Asp*718 restriction site of pKSH201. At the 3'-terminal end a *Hind*III restriction site (see reverse  
15 primer) was added, compatible to the *Hind*III restriction site of pKSH201, for cloning purposes. PCR product was isolated from pKSH210 by *Asp*718-*Hind*III digestion (1,263 bp) and ligated into *Asp*718-*Hind*III digested pKSH201 (Fig. 2) (pKSH211), thereby introducing the desired *kshB* in-frame deletion (Fig. 4). Finally, a *Bam*HI-*Hind*III DNA fragment (2.2 kb) containing the *kshB* deletion was ligated into pK18mobsacB (see WO 01/31050) digested with *Bam*HI and *Hind*III  
20 (pKSH212).

Plasmid pKSH212 was introduced into *R. erythropolis* SQ1 by conjugation using *Escherichia coli* S17-1. Unmarked *kshB* gene deletion was obtained using the *sacB* counter selectable system (WO 01/31050). Potential *kshB* mutants were screened by replica plating on AD mineral agar plates, which enabled us to isolate *kshB* mutants unable to grow on AD. Southern analysis was performed  
25 on *Asp*718 digested chromosomal DNA of wild type and three AD growth deficient mutants. Hybridization with the complete *kshB* gene showed that *kshB* was not present in the genome of the putative *kshB* mutants. A clear hybridization signal (4.3 kb fragment) was exclusively found with wild type chromosomal DNA. Additional Southern analysis with an alternative probe, being the 2.2 kb insert of pKSH212 comprising both flanking regions of *kshB*, furthermore confirmed *kshB* gene  
30 deletion: a 4.3 kb *Asp*718 wild type DNA fragment containing the *kshB* gene was reduced to 3.3 kb in a *kshB* mutant, demonstrating replacement of the 1,041 bp *kshB* gene by a *kshB* in-frame remnant of 30 nt (encoding MTTVEVPIA). The resulting strain is designated *R. erythropolis* RG4.

#### Example 5

35 Use of a genetically modified strains RG1-UV26, RG1-UV39, RG2, RG4 and RG9 in steroid  $\Delta^1$ -dehydrogenation.

Strains RG2 and RG4 were plated on mineral agar media containing AD, ADD or 9OHAD as sole carbon and energy source. Both strains showed no growth on AD(D), whereas growth on 9OHAD was comparable to strain SQ1. These phenotypes are in agreement with those found with UV mutant  
5 strains RG1-UV26 and RG1-UV39. Bioconversion of AD ( $1 \text{ g}\cdot\text{l}^{-1}$ ) with strain SQ1 results in AD utilization but not in accumulation of ADD or other metabolites. Bioconversion of AD ( $1 \text{ g}\cdot\text{l}^{-1}$ ) by strain RG2 or strain RG4 resulted in comparable accumulation levels of ADD (varying between  $0.3\text{-}0.5 \text{ g}\cdot\text{l}^{-1}$  after 168 h). AD(D)  $9\alpha$ -hydroxylation thus is blocked by inactivation of either *kshA* or *kshB*, demonstrating the essential role of both KshA and KshB in KSH activity in *R. erythropolis* SQ1. In  
10 AD bioconversion experiments with strain RG9, neither a decline in the initial AD concentration nor 9OHAD formation was observed. Mutant strain RG9 thus confirms that *kshA* encodes AD  $9\alpha$ -hydroxylase activity and that, contrary to the KSTD isoenzymes, no further KSH isoenzymes are present in *R. erythropolis* SQ1.

15

## Claims

- 5 1. An isolated polynucleotide sequence comprising a nucleic acid sequence encoding the amino acid sequence of KshA protein or of KshB protein, encoded by nucleotides 499-1695 of SEQ ID NO:1 or by nucleotides 387-1427 of SEQ ID NO:2, respectively, and functional homologues thereof.
- 10 2. The polynucleotide sequence of claim 1 comprising the nucleic acid sequence SEQ ID NO:1, preferably comprising the nucleotides 499-1695.
3. The polynucleotide sequence of claims 1 or 2 consisting of the nucleic acid sequence SEQ ID NO:1.
- 15 4. The polynucleotide sequence of claim 1 comprising the nucleic acid sequence SEQ ID NO:2, preferably comprising comprising the nucleotides 387-1427 of SEQ ID NO:2.
- 20 5. The polynucleotide sequence of any one of claims 1 or 4 consisting of the nucleic acid sequence SEQ ID NO:2.
6. A polypeptide sequence encoded by nucleotides 499-1695 of SEQ ID NO:1 or by nucleotides 387-1427 of SEQ ID NO:2.
- 25 7. A microorganism blocked in 3-ketosteroid 9 $\alpha$ -hydroxylase activity characterized in that it is a genetically modified microorganism, preferably of the family of Actinomycetales.
8. The microorganism of claim 7, being of the *Rhodococcus* genus.
- 30 9. The microorganism according to claim any one of claims 7 or 8, obtainable by targeted, preferably unmarked, gene inactivation of the genes encoding KSH-activity, preferably the gene *kshA* and/or the gene *kshB*.
- 35 10. The microorganism according to any one of claims 7-9, wherein also at least one gene encoding 3-ketosteroid  $\Delta^1$ -dehydrogenase activity is inactivated, preferably by unmarked gene deletion.



11. The microorganism according to any one of claims 7-10, being selected from the genetically modified microorganisms RG2, RG4 and RG9.
- 5 12. A method to construct a genetically modified strain of a steroid degrading microorganism lacking the ability to degrade the steroid nucleus, the method comprising inactivation of the genes encoding KSH-activity, preferably the gene *kshA* and/or the gene *kshB*.
- 10 13. The method of claim 12, wherein the inactivation of the gene(s) is accomplished by targeted, preferably unmarked, gene deletion.
14. A use of a genetically modified microorganism in steroid  $\Delta^1$ -dehydrogenation, preferably in the preparation of 1,4-androstadiene-3,17-dione and prednisolone.
- 15 15. The use of claim 14, wherein the microorganism has been obtained by targeted gene inactivation, preferably unmarked gene deletion, of the genes encoding KSH-activity in a microorganism of the family of Actinomycetales, preferably the gene *kshA* and/or the gene *kshB*.
- 20 16. The use of claims 14 or 15, wherein the microorganism is selected from the genetically modified strains RG2, RG4 and RG9.

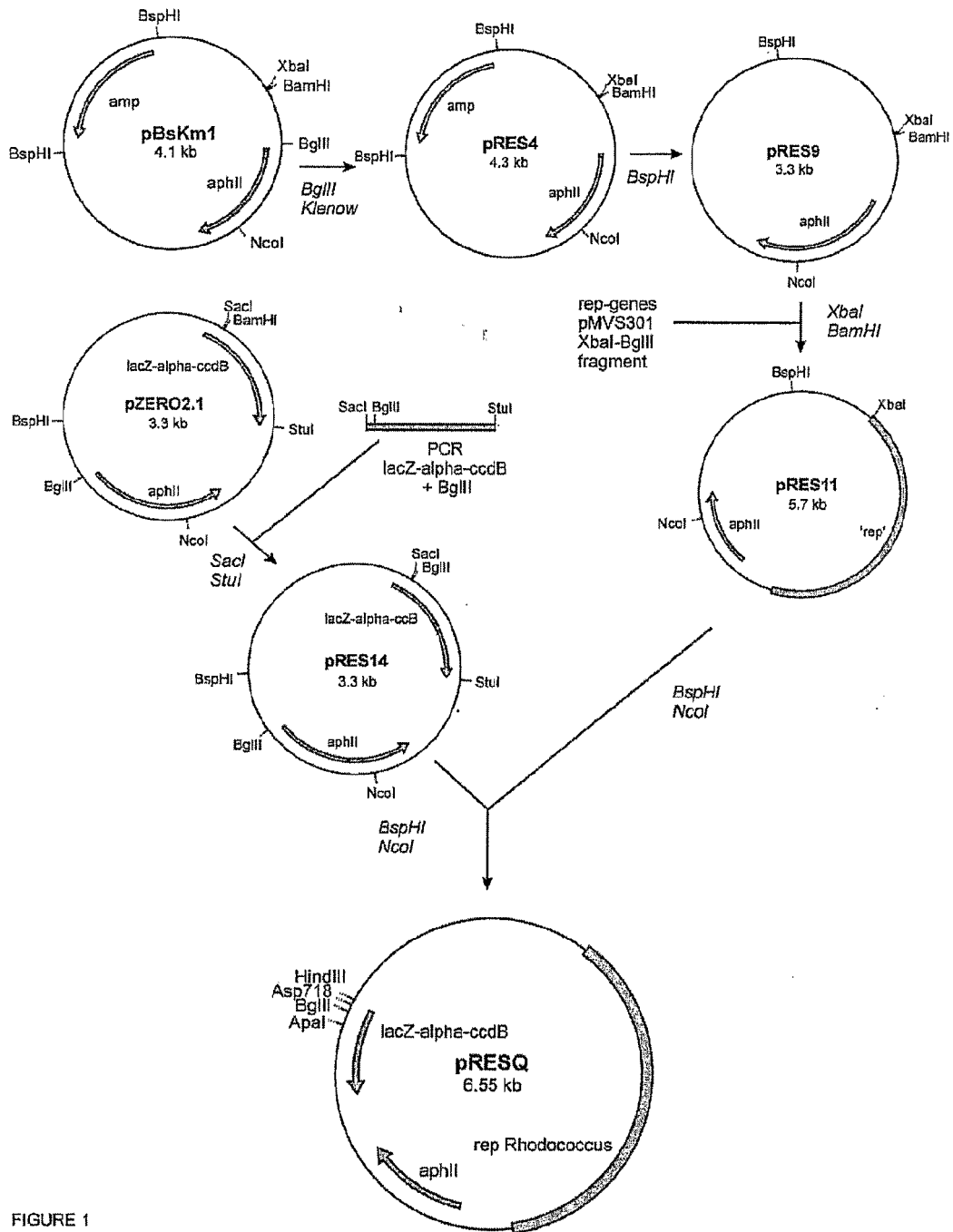


FIGURE 1

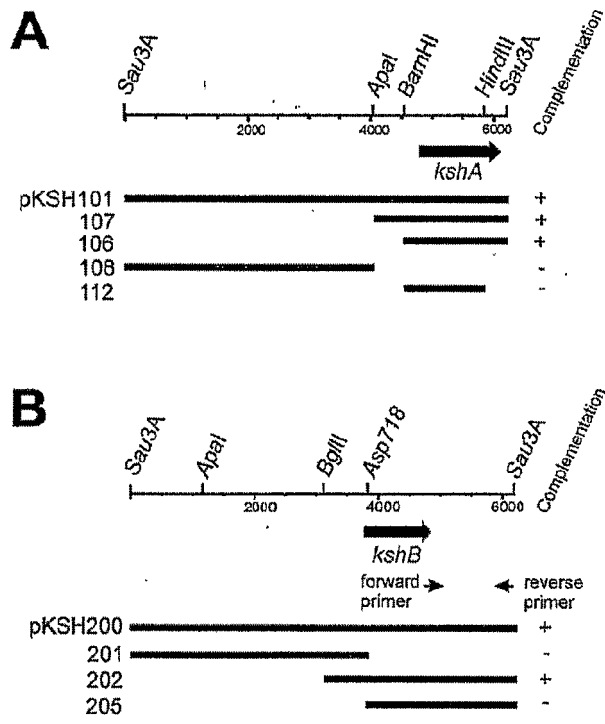


FIGURE 2

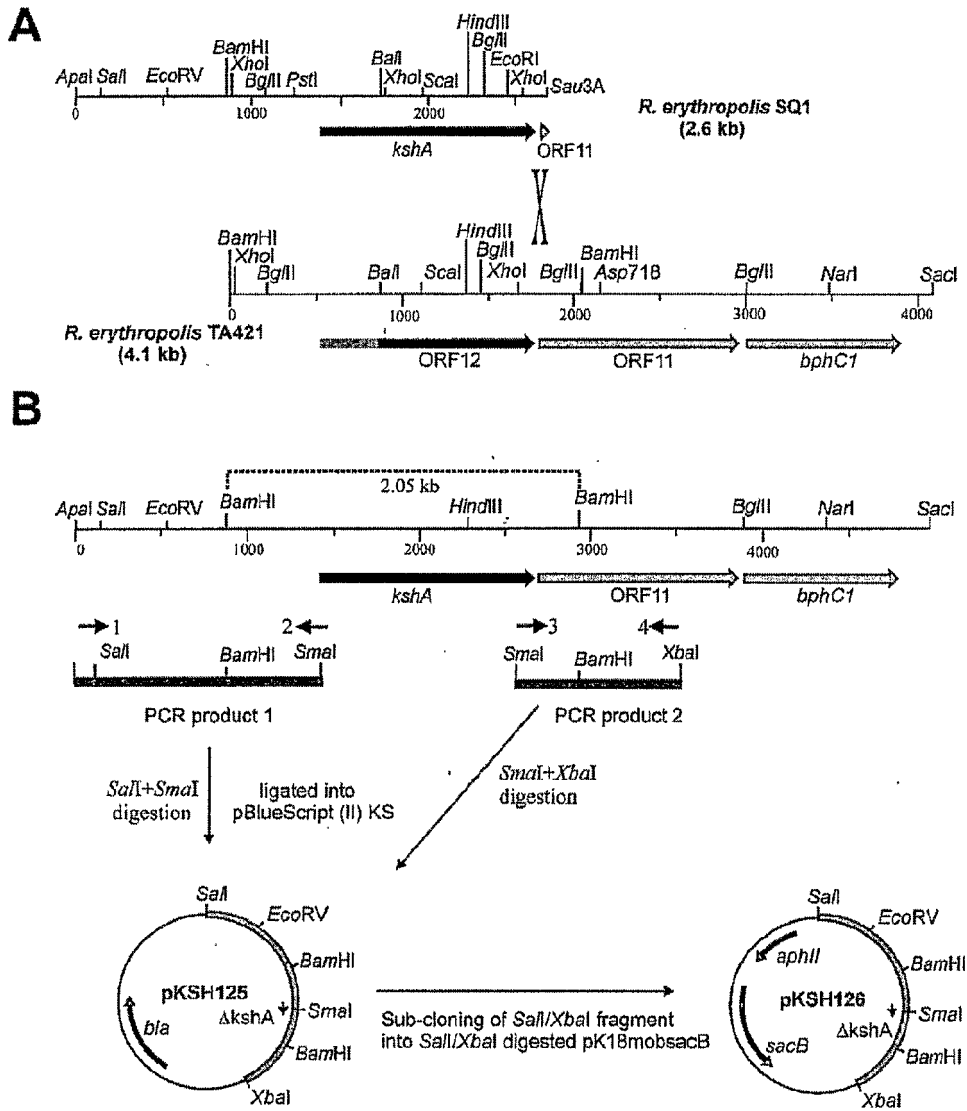


FIGURE 3

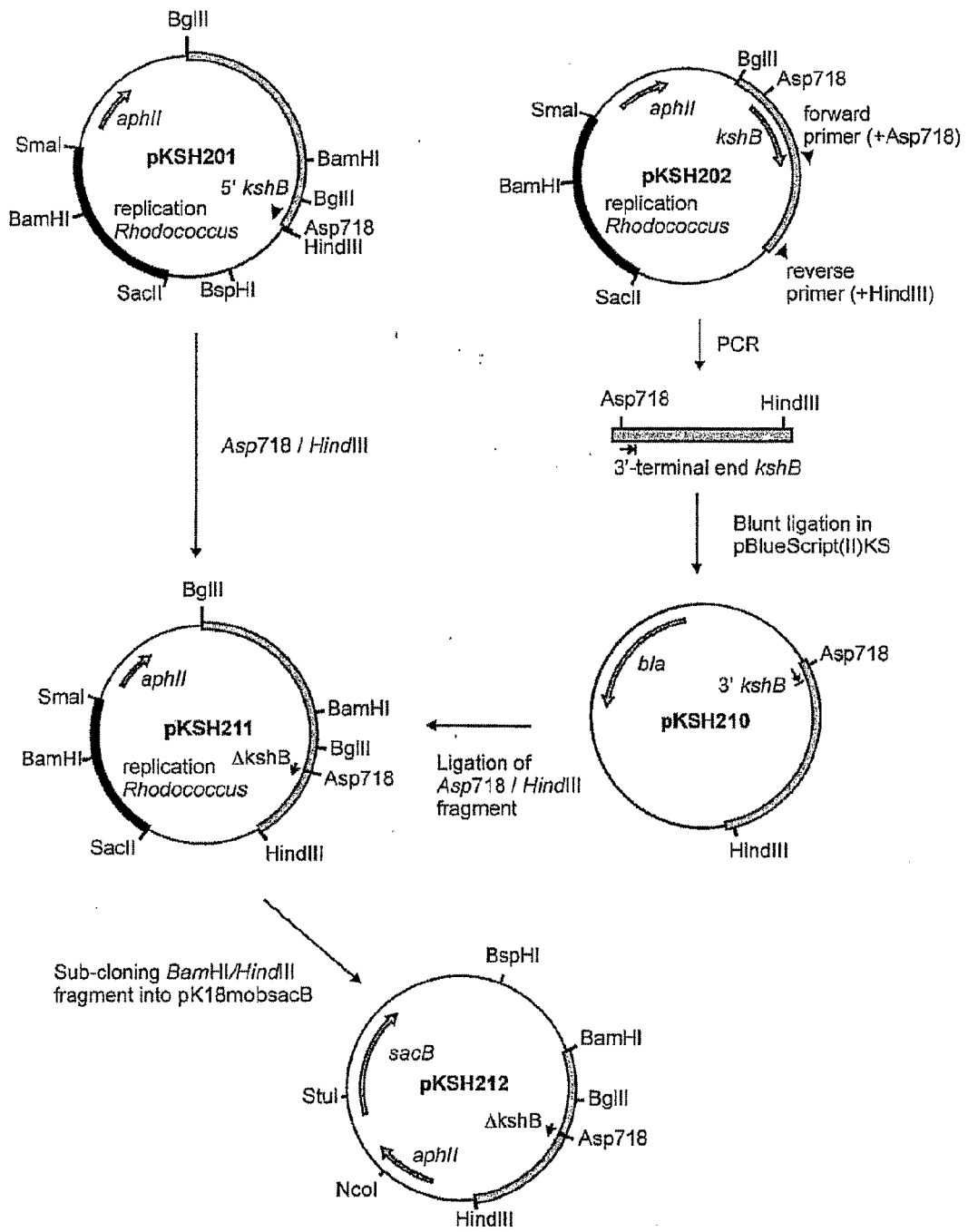


FIGURE 4

## SEQUENCE LISTING

<110>  
 5 <120> Microbial delta-1-dehydrogenation of steroids  
 <130>  
 <160> 10  
 <170> PatentIn version 3.1

10 SEQ ID NO:1  
 <210> 1  
 <211> 1770  
 <212> DNA  
 15 <213> Rhodococcus erythropolis SQ1  
 <400> 1

ggatccgtga ggccgctggg ctcgactctc gagcggatca ccgaatacca cgaccagacc 60  
 20 gggcatctgg ttgccgaggt cttcgtccgg gaagccccga cagcgggcac cggaaccgat 120  
 tccgttgctg cccttgccga tttcgccgct cgtcgttacg gctcggcgtc gctacgacgg 180  
 ttccgcggac gttgcaccgc aacttccgcy ggagcagatc tgaactgctc cgtgacggtg 240  
 25 ctcgctacgt acgagaagga ccacgagagc gtcgcccagc tagctctggg cgccaccgat 300  
 cccaccgggg tcgtcaccgc tcgggcctgg gcgacttacg cgatcgacga tgtggcgaca 360  
 30 ctcgcgagct gaatctctgc aggtcaccga gcggtgttc catccagcgg gaactgccgt 420  
 atctccagag gacgcttcgt cgttgatga gacggatcac atcaactggag cgtcgcaaga 480  
 cgcggctcgg aggaagacat ggcactcggg accggcccac tcaccactac ggacgcctcg 540  
 35 acgcagtcgg gcgcccggca ggaaattcga gaaatcgaag cggcagcggc cccggcggcg 600  
 ttcgcgcgctg gatggcactg cctcggactg agcaatacct atcgcgacgg tcagccccac 660  
 40 cagatcgagg cattcggcac cagccttgtc gtcttcgccg acagcaaggg cgacatcaag 720  
 attctcgacg cctactgccg ccacatgggt ggcaacctcg ctacggaac cgtcaagggc 780  
 gattcgatcg cctgcccgtt ccacgactgg cgttggggcg gtaacggaaa atgtacggcg 840  
 45 atcccctacg cacgacgcgt tccgcccgtg gccaaagcgc gggcgtggac cacactcgag 900  
 aagaacggtc agctgttcgt ctggcacgat ccccagggga atccccgcc cgccgaggtc 960  
 50 acaattcccg acatcgaggg cttcggaagc gatgagtgga gcgactggag ctggaatacg 1020  
 ctccaccatcg aaggctctca ttgccgagc atcgtggaca acgtcgtgga catggcccac 1080  
 ttcttctacg ttcactactc gttcccccaag tacttcaaga acatcttoga aggtcacgta 1140  
 55 gcgagccagt acatggaatc ggtgggacgc gaggacatca tcagcggcac caactacggc 1200  
 gatcccaacg ctgttcttcg ttcggatgct tcgtatttcg gtcgctcgta catgatcgac 1260

	tggatcaaga	gccaagccaa	cgggcagatc	atcgaaactg	tattgatcaa	ctgccactac	1320
5	cccatctcga	acaatgcctt	cgtggtgcag	tacggcgcga	tggccaagaa	gcttcccggg	1380
	atggacgacg	agtccactgc	cgcgatggcc	gcgcagttca	ccgagggcgt	cgagatggga	1440
	ttccttcagg	acgtggagat	ctggaagaac	aaggccccga	tcgacaacc	gctgctctcg	1500
10	gaggaggacg	gcccggttta	ccagttgcgg	cgctggtaca	accagttcta	caccgatgtc	1560
	gagaacgtca	cggaagacat	gacgaagcgc	ttcgaattcg	agatcgacac	taccgcgcga	1620
15	gtggagagct	ggaagcggga	agtcgccgaa	aacgttgccg	cacgtgacgc	ccaggcgctc	1680
	gagacaacct	cctgattcgc	agtcgaaaac	actgctgaga	caacctgact	gaacgcaagg	1740
	aacaacatga	cgtacgaagt	gctcgaacag				1770
20	SEQ ID NO:2						
	<210>	2					
	<211>	1500					
25	<212>	DNA					
	<213>	Rhodococcus erythropolis SQ1					
	<400>	2					
30	agatctgggt	gagctcgata	gccaatcggt	tctcctcatg	gatgtgcacg	aatgtgactg	60
	cgggagggat	tcgagccgat	tctttgtgac	cggcgtctca	cctggttccc	actgtgtacc	120
	gacaaacacc	ccgtgcacag	acttcggcag	ctcaaccgcc	cggacgcgcg	ccgcgaaccg	180
35	cccgttcggg	caggtcgtcg	ccaccccgga	ccgccaggaa	ggccggatca	gactgcogtg	240
	gttcccgctg	atcgggaacg	tgttgctcgg	cgactacgat	cccagtagc	ttttcatcac	300
40	tgctaacatt	gaactagaac	acgttctaga	tgttgactt	cactgacctg	cgctttacc	360
	agcttcacga	gaggactagc	ttcgcgatga	cgacagtcga	ggtaccgcac	agctcgcgat	420
	cggcgggtact	caccgtgtcc	ggagtcatcg	aagagacttc	cgacgcccgg	tcgctgggtg	480
45	tcgagattcc	cgcagaactg	aaggacaagt	tcgactaaa	gccaggtcaa	ttcctgacc	540
	tgcgcatccc	gagtgatcaa	accggctcgg	tcgcacgctg	ctactcctg	gccagttcac	600
50	ccttcaccga	cgacgcccc	aaggtgaccg	tcaagcgaac	cgtcgacggc	tacggctcca	660
	actggttgtg	cgacaagctc	caggtcggcg	acaccatcga	agtgctgccg	ccgtcaggtg	720
	tcttcacgcc	caaatecctc	gatcacgact	tcctgctctt	cggtgcaggc	agcggaatca	780
55	ctccgggtgat	ctcgattctc	aagtccgcc	tcacacaggg	cagcggcaat	gtcgtgctga	840
	tctacgcgaa	tcgcgacgag	aagtccgtca	tcttcggcgc	cgaactgcgc	gagctggcag	900

cgcaacaccc cggccgcctc accgtggtcc actggatcga aaccgtgcag ggactccccg 960  
 5 ctgtctcaca actggcgacg ctcgccaagc cttttgtcgc gtacgaggcc ttcattgtgcg 1020  
 gtccaggccc tttcatggac gccgtccaca aggcactggc agaagcgggc atgccccgca 1080  
 cgcaggttca cgccgaggtc ttcaattcac tcgccggcga cccgttccgt gacgtcgagg 1140  
 10 tagcagaggt ctccgacgag gaagctgccg acgccgcaac cgtcgaggtc gaactggacg 1200  
 gtgaaaacca caccctcgtc tggccgcgca agcagacgct cgtcgacatc atgctcgcaa 1260  
 agggctctga cgttccgtac tcttgcaagg aaggctgagtc cgggtcgtgt gcctgcaccg 1320  
 15 tcaccgaagg cgaagttcag atggacaact ccgagatcct cgacgccgag gacgtggcga 1380  
 acggctacat cctcggttgc aagcgaagcc gatcaccgat cgctgaaga tcgagtctga 1440  
 20 accctgcttc gatgcatgac ggccccttga ttccctgctg aatcgggggc tttgtcatct 1500

SEQ ID NO:3

25 <210> 3  
 <211> 27  
 <212> DNA  
 <213> artificial: primer  
 <400> 3  
 30 accgagctca gatctactag taacggc 27

SEQ ID NO:4

35 <210> 4  
 <211> 26  
 <212> DNA  
 <213> artificial: primer  
 40 <400> 4  
 attcaggcct gacatttata ttcccc 26

SEQ ID NO:5

45 <210> 5  
 <211> 23  
 <212> DNA  
 50 <213> artificial: primer  
 <400> 5  
 cgcgggccca tcgagagcac gtt 23

SEQ ID NO:6

55



<210> 6  
 <211> 27  
 <212> DNA  
 <213> artificial: primer  
 5 <400> 6  
  
 gcgcccgggt ccgagtgcc tgtcttc 27  
  
 10 SEQ ID NO:7  
  
 <210> 7  
 <211> 30  
 <212> DNA  
 15 <213> artificial: primer  
 <400> 7  
  
 gcgcccggga caacctcctg attcgcagtc 30  
  
 20 SEQ ID NO:8  
  
 <210> 8  
 <211> 30  
 25 <212> DNA  
 <213> artificial: primer  
 <400> 8  
  
 gcgtctagag tggaagagca ttccctcgca 30  
  
 30 SEQ ID NO:9  
  
 <210> 9  
 <211> 28  
 <212> DNA  
 <213> artificial: primer  
 <400> 9  
  
 35 gcgggtaccg atcgctgaa gatcgagt 28  
  
 40 SEQ ID NO:10  
  
 <210> 10  
 <211> 29  
 <212> DNA  
 <213> artificial: primer  
 <400> 10  
  
 45 gcgaagcttg ccggcgctgc agctctgtg 29  
  
 50