J. Biochem. 126, 624-631 (1999)

Steroid Monooxygenase of *Rhodococcus rhodochrous*: Sequencing of the Genomic DNA, and Hyperexpression, Purification, and Characterization of the Recombinant Enzyme[1]

Shingo Morii,* Shizue Sawamoto,* Yuhji Yamauchi,* Masahiko Miyamoto,* Masafumi Iwami,† and Eiji Itagaki*,[2]

*Departments of *Chemistry and †Biology, Faculty of Science, Kanazawa University, Kakuma-machi, Kanazawa 920-1192*

Received June 15, 1999; accepted July 13, 1999

Steroid monooxygenase of *Rhodococcus rhodochrous* is a Baeyer-Villigerase catalyzing the insertion of an oxygen atom between the C_{17}- and C_{20}-carbons of progesterone to produce testosterone acetate. The 5.1-kbp-long *Bam*HI DNA fragment containing the steroid monooxygenase gene, *smo*, was cloned from the chromosomal DNA and sequenced. The *smo* gene is 1,650 nucleotides long, starts with a TTG codon, and ends with a TGA codon. The deduced amino acid sequence indicates that the enzyme protein consist of 549 amino acid residues with a molecular mass of 60,133. Thus, the molecular mass of the holoenzyme is 60,919. The amino acid sequence is highly homologous (41.2% identity) to that of cyclohexanone monooxygenase of *Acinetobacter* sp. In the upstream of the *smo* gene, the genes of heat shock proteins, *dnaK*, *grpE*, and *dnaJ*, located on the complementary strand, and the DNA-inserts of pSMO and pD1, which contains the *ksdD* gene, were joined at the *Bam*HI site of the *dnaJ* gene. The *smo* gene was modified at the initiation codon to ATG and ligated with an expression vector to construct a plasmid, pSMO-EX, and introduced into *Escherichia coli* cells. The transformed cells hyperexpressed the steroid monooxygenase as an active and soluble protein at more than 40 times the level in *R. rhodochrous* cells. Purification of the recombinant monooxygenase from the *E. coli* cells by simplified procedures yielded about 2.3 mg of enzyme protein/g wet cells. The purified recombinant steroid monooxygenase exhibited indistinguishable molecular and catalytic properties from those of the *R. rhodochrous* enzyme.

Key words: biological Baeyer-Villiger reaction, hyperexpression, recombinant enzyme, sequence, steroid monooxygenase.

Steroid monooxygenase is a flavoenzyme, Baeyer-Villigerase, catalyzing a Baeyer-Villiger oxidation of a C_{21}-20-ketosteroid to the acetate ester of a C_{19}-17α-hydroxysteroid (*1-4*). It exhibits a unique monooxygenase reaction greatly different from the well-known reactions of many hydroxylases. In the previous reports, we described the isolation and characterization of steroid monooxygenases from a fungus, *Cylindrocarpon radicicola* (*1, 2*), and from a bacterium, *Rhodococcus rhodochrous* (*5*). These two enzymes have different properties; in particular, the fungal enzyme is a bifunctional one catalyzing the conversion of a 20-ketosteroid to a steroid ester, oxidative esterification, and the ring expansion of a 17-ketosteroid to a D-homo-17α-oxasteroid, oxidative lactonization. The latter type of reaction was reported to be the reaction catalyzed by cyclohexanone monooxygenases from *Acinetobacter* NCIB 9871 and *Pseudomonas* NCIB 9872 (*6-8*). This enzyme is the only Baeyer-Villigerase of which the gene has been cloned and sequenced to deduce the amino acid sequence (*9*). Characterization of the differences between Baeyer-Villigerase and hydroxylase is an interesting and important subject as to understanding the biological monooxygenation.

In the present study, we have cloned and sequenced the bacterial steroid monooxygenase gene, *smo*, of *R. rhodochrous*. It became evident that the steroid monooxygenase and cyclohexanone monooxygenase are highly homologous enzymes in amino acid sequence, suggesting the presence of a characteristic structure in their protein molecules. During the study we also preliminarily elucidated the gene organization around the *smo* gene of this organism, and clarify the relation of the *smo* and *ksdD* genes of 3-ketosteroid-Δ^1-dehydrogenase on the chromosomal DNA (*10*). To confirm that the isolated clone actually encoded the monooxygense, we designed an expression plasmid carrying the enzyme gene and hyperexpressed it as the active monooxygenase in *E. coli* cells. The recombinant enzyme was purified to homogeneity from the *E. coli* cells on a large scale, and

[1] The sequence data reported in this paper have been deposited in DDBJ, EMBL, and GenBank under accession No. AB010439.

[2] To whom correspondence should be addressed. Fax: +81-762-64-5742, E-mail: itagaki@cacheibm.s.kanazawa-u.ac.jp

Abbreviations: bp, base pair(s); Da, Dalton; IPTG, isopropyl-β-D-thiogalactopyranoside; *ksd*, 3-ketosteroid-Δ^1-dehydrogenase gene; ORF, open reading frame; PCR, polymerase chain reaction; SD, Shine-Dalgarno; SDS-PAGE, polyacrylamide electrophoresis in the presence of dodecyl sulfate; *smo*, steroid monooxygenase gene.

characterized in terms of its molecular and catalytic properties.

MATERIALS AND METHODS

Bacterial Strains and Plasmids—*R. rhodochrous* IFO 3338 was used as a source of DNA and protein. *E. coli* JM109 was purchased from Toyobo, and *E. coli* BL21-(DE3) and plasmid pET28a(+) were from Novagen. The pCR™II vector and *E. coli* INVαF′, and pUC18 were from Invitrogen and Pharmacia, respectively.

DNA Manipulations and Sequencing—Steroid monooxygenase was purified from *R. rhodochrous* IFO 3338 (*5*). The amino acid sequences of the NH_2-terminal and tryptic peptides of steroid monooxygenase were determined with an automated gas phase amino acid sequencer. Total DNA of *R. rhodochrous* was prepared as described previously (*11*) and used as a template for polymerase chain reaction (PCR). The DNA primers used for PCR were: SMO-N, 5′-ATGAACGGNCAGCAYCCNCG-3′, which was designed based on the amino terminal sequence, MNGQHPR; and SMO-2, 5′-TCSGCSACRTGYTCSAGGTA-3′, which was designed based on the amino acid sequence, YLEHVAD, from a tryptic peptide. PCR was carried out with 0.1 nmol of each primer, 1 μg of total DNA, 200 nmol of each dNTP, and 2.5 U of *Taq*-DNA polymerase in PCR buffer (Takara Shuzo) in a total volume of 100 μl at 94°C for 30 s for denaturation, 60°C for 60 s for annealing, and 72°C for 60 s for extension. The reaction was continued for 30 cycles. DNA fragments of about 300 bp were obtained. The fragments were purified, ligated with the pCR™II-vector, and then introduced into *E. coli* INVαF′ (Invitrogen). Transformants were selected and one plasmid was isolated. The nucleotide sequence of the inserted DNA was determined with a cycle sequencing kit using a model A.L.F. DNA sequencer (Pharmacia). The 329-bp DNA fragment contained the desired base sequences for the N-terminal and tryptic peptide amino acid sequences of the steroid monooxygenase. It was labeled with [α-^{32}P]dCTP with a *redi*prime™ DNA labeling system (Amersham) and then hybridized with the genomic DNA, digested previously with one of six endonucleases, of *R. rhodochrous*. The *Bam*HI DNA fragment of about 5.1-kbp that gave a positive signal was purified, ligated with the pUC18 plasmid digested previously with *Bam*HI, and then introduced into *E. coli* JM109. Plasmid screening was carried out by colony hybridization with the ^{32}P-labeled probe, and a plasmid, named pSMO (7.8 kbp), containing the desired DNA fragment was isolated. The inserted DNA in pSMO was restricted with *Pst*I and then subcloned into the pUC18 plasmid to obtain the pSMO-1, pSMO-2, and pSMO-3 plasmids. These contained 0.7, 1.0, and 3.4-kbp inserted DNA, respectively. pSMO was subcloned into a series using a deletion kit (Takara Shuzo) and sequenced with a Thermo sequenase fluorescent labeled primer cycle sequence kit with 7-deaza-dGTP (Amersham). Both strands were sequenced over the entire length using a series of synthetic primers.

Computer analyses of the obtained sequence data were carried out with DNASIS-MAC (Hitachi Software).

Construction of the Expression Plasmid—Primers SMO-NCO and SMO-BAM were designed from the nucleotide sequence shown in Fig. 1 and synthesized: SOM-NCO, 5′-CATG**CCATGG**ACGGCCAGCATC-3′, based on the base sequence from position 56 to 77, in which the first ten nucleotide bases were replaced with the underlined nucleotides to make a *Noc*I-restriction site (bold letters), and the initiation codon, TTG, was changed to ATG. Through the replacement, the second amino acid, Asn (ACC), was changed to Asp (GAC). SMO-BAM, 5′-AT**GGATCC**TCAC-CCACCCGGAG-3′, based on the complementary sequence from position 1823 to 1844, in which a *Bam*HI-restriction site (bold letters) is present. PCR with the primers and pSMO-3 as a template was carried out for 10 cycles comprising 94°C for 60 s for denaturation, 68°C for 60 s for annealing, and 72°C for 60 s for extension, and the 1.6-kbp DNA fragment was obtained. The product was then digested with *Nco*I and *Bam*HI, and the obtained fragment was ligated with the pET28a(+) vector digested previously with *Nco*I and *Bam*HI to construct pSMO-EX. The plasmid was introduced into *E. coli* JM109, and then transformants were selected. The inserted plasmid was examined by means of the colony PCR method using primers SMO-NCO and SMO-BAM, and the sizes of the PCR products were determined by agarose-gel electrophoresis. The insert plasmid was then examined by sequencing to confirm the whole nucleotide sequence of the *smo* gene.

Hyperexpression—*E. coli* BL21(DE3) cells were transformed with pSMO-EX and then cultured in 50 ml of LB medium containing 0.002% kanamycin for 15 h at 37°C. The cells were transferred to 1 liter of the new medium containing the same concentration of kanamycin and 0.02% riboflavin, and then cultured at 30°C. At 2 h after the inoculation, IPTG was added to a final concentration of 0.1 mM and the cultures were continued further for 6 h at 30°C. Then the cells were harvested, washed with cold saline and stored at −80°C.

Purification—Purification was carried out at 4°C by the modified methods used for the purification of *R. rhodochrous* steroid monooxygenase (*5*). The standard buffer used for the purification was 30 mM Tris-HCl buffer, pH 7.4, containing 100 μM EDTA, 1 μM FAD, and 50 μM dithiothreitol. Cells, 57 g wet weight, were suspended in 120 ml of the standard buffer and disrupted by sonication in an ice bath. A crude cell extract was obtained by centrifugation at 10,000 × *g* for 20 min and applied on a DEAE-cellulose column (3 × 35 cm) equilibrated with the buffer. The column was developed with a 1-liter linear gradient of 0–0.6 M KCl in the standard buffer at the flow rate of 0.45 ml/min. The active fractions were collected, combined, and concentrated. After reducing the KCl concentration with Tris buffer, the enzyme preparation was applied on a second DEAE-cellulose column (1.5 × 25 cm) equilibrated with the buffer and eluted with a 400-ml linear gradient of 0–0.5 M KCl in the standard buffer at the flow rate of 0.2 ml/min. The active fractions were collected and concentrated. The enzyme preparation was subjected to a preparative gel electrophoresis in a polyacrylamide gel (8.5 × 5 × 0.2 cm) at 4°C. The gel containing the yellow-colored enzyme was cut out, and the enzyme was eluted from the gel and concentrated. The purified enzyme was stored at −80°C.

N-Terminal Sequence Determination—The amino terminal sequence of the recombinant steroid monooxygenase was determined with an Applied Biosystems model 476A protein sequencer, and phenylthiohydantoin derivatives were detected *via* an on-line system.

Spectrophotometric Experiments—The spectrophotometric properties of the recombinant enzyme were examined as described in the previous paper (*5*).

Gel Electrophoresis, Enzyme Assay, and Protein Determination—Gel electrophoresis, enzyme activity assaying and protein determination were carried out by the methods described (*5*).

Materials—All restriction enzymes were obtained from Boehringer Mannheim. *Taq*-DNA polymerase and a Deletion kit for Kilo-sequencing were purchased from Takara Shuzo. The Autocycle™ Sequencing kit was from Pharmacia, and [α-^{32}P]dCTP and the *redi*prime™ DNA labeling system were from Amersham. All other chemicals used were of the highest purity available.

RESULTS

Cloning and Sequencing of the Steroid Monooxygenase Gene—A 5.1-kbp *Bam*H1 DNA fragment of pSMO was digested with *Pst*I, and then subcloned to yield pSMO-1, pSMO-2, and pSMO-3. The steroid monooxygenase gene was localized in the 3.4-kbp *Pst*I-*Bam*HI DNA fragment of pSMO-3. pSMO and its subclones were sequenced, and the middle part of the determined nucleotide sequence, a 1,844-bp region encompassing the *smo* gene of steroid monooxygenase, is shown in Fig. 1. The ORF of 1,650 nucleotides long encoding the 549 amino acid residues of the oxygenase was identified and designated as *smo*. The ORF is in the reversed direction, and begins with TTG as the start codon and ends with TGA as the stop codon (Fig. 1). The deduced amino acid sequences of the N-terminus of steroid monooxygenase and of the tryptic peptides used for the design of primers are in perfect agreement with those determined by Edman degradation of the purified *R. rhodochrous* enzyme. The calculated molecular mass is 60,133 Da, which is the size of the apoenzyme, and that of the holoenzyme bound with an FAD molecule is 60,919 Da, which is good agreement with the molecular mass determined by SDS-gel electrophoresis (*5*). Comparison of the predicted amino acid sequence of *R. rhodochrous* steroid monooxygenase with those of the monooxygenases from other microorganisms revealed 41.2% identity and 55.2% similarity with the cyclohexanone monooxygenase from *Acinetobacter* sp. NCIB 98871, a Baeyer-Villiger enzyme catalyzing the oxidation of cyclohexanone to ϵ-caprolactone (*9*) (Fig. 2). The sequence shows lower similarities (less than 10% identity) with those of bacterial monooxygenases, salicylate hydroxylase from *Pseudomonas putida* S1 (*12*) and *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescence* (*13*). Sequence comparison showing the conserved amino acids in the enzymes (Fig. 2) revealed consensus sequences [D-X_2-V-X-G-X-G-X_2-G-L-Y-A] and [V-X-A-F-X] of a nucleotide-binding site between positions 23 and 51. These sequences correspond to a putative site that binds the ADP moiety of FAD (*14*). The second concensus sequences, [R-V-G-V-I-G-T-G-S-X-G-X-Q-X-I-X-X-X] and [L-X-V-F-Q], of the second nucleotide-binding site are between positions 193 and 221. These sequences also correspond to a putative site that binds the ADP moiety of NADPH (*15*), a hydrogen donor of the enzyme. Eight highly conserved sequence regions can also be seen in Fig. 2, positions 23-73, 172-207, 217-235, 270-290, 340-362, 385-399, 421-442, and 491-515, these exhibit from 68 to 79% similarity. The *smo* gene is preceded by a typical ribosome-binding site, GGAAGG, at positions 48-53 (Fig. 1), and a palindrome sequence is located just downstream (position 1733) from the termination codon.

Upstream Sequence of the smo Gene—An attempt was made to reveal the genes located upstream of the *smo* gene. The nucleotide sequences of the inserts of pSMO1, pSMO2, and pSMO3 suggested the presence of three ORFs on the complementary strand of the *smo* gene. Even though the ORFs were sequenced in part, the obtained sequence data for the three ORFs were found to exhibit high similarity (over 70%) with those of three heat shock protein genes, *dnaK*, *grpE*, and *dnaJ*, of *Mycobacterium tuberculosis*, respectively, through a computer aided BLAST search of the GeneBank data base (accession number X58406) (*16*). The heat shock protein genes are located on the complementary DNA strand of the upstream of the *smo* gene and are arrayed in the same order of *M. tuberculosis*, as summarized in the gene map of pSMO in Fig. 3. In the previous report, we described the presence of the *dnaJ* heat shock gene in the upstream of the *ksdD* gene (*10*). The nucleotide sequence determined from the *Bam*HI-site of pSMO-1 showed high similarity, 84.6%, to the N-terminus of *dnaJ* of *M. tuberculosis*, in which the initiation codon, GTG, SD-nucleotide sequence, GAGAGGAGGAGA, and *Bam*HI site are conserved in both organisms. Conjugation of the nucleotide sequences of *DnaJ* from pD1 and from pSMO at the *Bam*HI site revealed the presence of an ORF of 1,056 bp, of which the deduced sequence of 352 amino acid residues indicated 70.1% identity and 80.9% similarity to the DnaJ protein of *M. tuberculosis* (data not shown). This was confirmed by means of a PCR experiment on the genomic DNA with primers, pSMO-DnaJ, 5′-ACGCAACG-CAGGACGAGATC-3′, and pD1-DnaJ, 5′-GAAGTCGAG-TGTCGTCTCGG-3′, yielding a DNA fragment of 400 bp, as expected from the nucleotide sequence of *DnaJ*, and two fragments of 270 and 130 bp from the *Bam*HI digest of the 400-bp fragment. The results indicated definitely that the inserted DNA fragments in pSMO and pD1 are joined at the *Bam*HI site to form the *dnaJ* gene of *R. rhodochrous*, as summarized in the gene map in Fig. 3.

Expression of the smo Gene of R. rhodochrous in E. coli—We designed and constructed an efficient heterologous hyperexpression system, plasmid pSMO-EX, from pSMO-3. It was introduced into *E. coli* BL21(DE3) cells to express the *smo* gene. The gene product exhibiting the expected molecular mass (60 kDa) on SDS-PAGE was hyperexpressed upon induction of the expression system with IPTG. The protein was produced at a lower level without the IPTG treatment. Expression of the active enzyme protein was also confirmed by assaying of the production of testosterone acetate from progesterone in the presence of NADPH, and of the progesterone-dependent oxidation of NADPH with a cell-extract of hyperexpressed *E. coli* cells. The time course kinetics of IPTG-induction of the enzyme by the *E. coli* cells carrying pSMO-EX showed that the enzyme activity was detected in the cells at 2 h after the addition of the inducer and reached the maximum level at 4 h.

Purification and Characterization of the Recombinant Steroid Monooxygenase—The recombinant enzyme was purified from *E. coli* cells harboring pSMO-EX by a simple

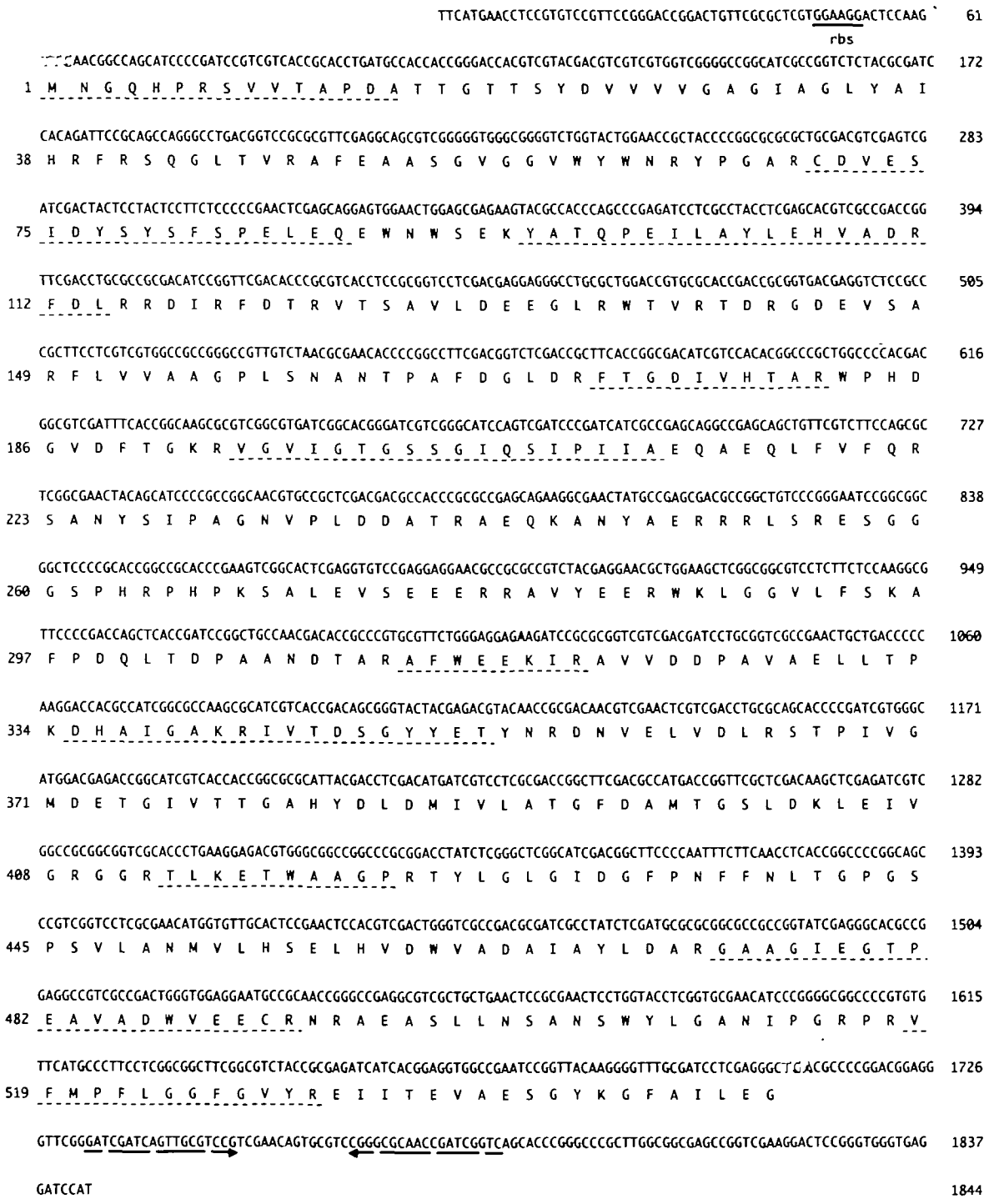

Fig. 1. **Sequence of the 1,844-bp nucleotides of the *Bam*H1 insert in plasmid pSMO and the deduced amino acid sequence of steroid monooxygenase from *Rhodococcus rhodochrous*.** The numbering at the left refers to the amino acid positions and that at the right to the nucleotide positions. The dotted-underlined amino acid sequences indicate the sequences determined by Edman degradation of the enzyme protein or peptides. The relevant initiation and termination codons are indicated by outlined letters. The putative ribosome-binding sequence (rbs) is indicated by a bold underline (–) upstream of the initiation codon, TTG. An inverted repeat (possible terminator) sequence is indicated by inverted bold arrows.

method with a high yield, as shown in Table I. The purification protocol involved repeated DEAE-cellulose ion-exchange chromatographies with linear gradient elution, followed by a preparative gel electrophoresis. The degree of purification of the fractions obtained at each step was analyzed by SDS-PAGE (Fig. 4), and the activities of the samples were measured by NADPH-oxidation assaying (Table I). The steroid monooxygenase was efficiently obtained with a high degree of purity in a yield of 131 mg protein/57 g wet cells. The purified enzyme exhibited a

high specific activity of 947 nmol/min·mg protein. The mass of the recombinant enzyme protein was confirmed to be 60 kDa by SDS-PAGE, on which the enzyme migrated the same distance as the *R. rhodochrous* enzyme (Fig. 4, lane 1). The N-terminal sequence of the recombinant enzyme was confirmed to be MetAspGlyGlnHis, in which the second residue replaced Asn. The recombinant monooxygenase migrated a slightly greater distance than the native enzyme on native gel electrophoresis due to the increase in the negative charge of Asp as the second amino acid residue of the N-terminus.

The spectroscopic visible absorption characteristics of the recombinant steroid monooxygenase were recorded in different redox states (Fig. 5). Absorption maxima were observed at 278, 376, and 438 nm, and shoulders around 360 and 465 nm, with an A_{278}/A_{438} ratio of 10. This is typical of a flavoprotein showing a large hypsochromic shift and the same as in the cases of the *R. rhodochrous* (*5*) and *C. radicicolla* enzymes (*1*). The recombinant enzyme also exhibited perturbation spectra on binding of progesterone with a large increase in the absorption intensities at 487 nm and a shoulder at 455 nm, and decreases in the intensities at 412 and 396 nm. Upon reduction with NADPH in the presence of progesterone under anaerobic conditions shown in Fig. 6, a dramatic and stoichiometric decrease occurred in the absorbance in the 320–550 nm range due to the reduction of FAD, as seen for the *Cylindrocarpon* enzyme

```
SMO   MNGQHPRSVVTAPDATTGTTSYDVVVVGAGIAGLYAIHRFRSQ-GLTVRAFEAASGVGGV   59
CHMO                MSQKMDFDAIVIGGGFGGLYAVKKLRDELELKVQAFDKATDVAGT   45
SMO   WYWNRYPGARCDVESIDYSYSFSPELEQEWNWSEKYATQPEILAYLEHVADRFDLRRDIR  119
CHMO  WYWNRYPGALTDTETHLYCYSWDKELLQSLQIKKKYVQGPDVRKYLQQVAEKHDLKKSYQ  105
SMO   FDTRVTSAVLDEEGLRWTVRTDRGDEVSARFLVVAAGPLSNANTPAFDGLDRFTGDIVHT  179
CHMO  FNTAVQSAHYNEADALWEVTTEYGDKYTARFLITALGLLSAPNLPNIKGINQFKGELHHT  165
SMO   ARWPHDGVDFTGKRVGVIGTGSSGIQSIPIIAEQAEQLFVFQRSANYSIPAGNVPLDDAT  239
CHMO  SRWP-DDVSFEGKRVGVIGTGSTGVQVITAVATLAKHLTVFQRSAQYSVPIGNDPLSEED  224
SMO   RAEQKANYAERRRLSRESGGGSPHRPHPKSALEVSEEERRAVYEERWKLGGVL-FS-KAF  297
CHMO  VKAIKDNYDKSLGWCMNSALAFALNESTVPAMSVSAEERKAVFEKAWQTGGGFRFMFETF  284
SMO   PDQLTDPAANDTARAFWEEKIRAVVDDPAVAELLTPKDHAIGAKRIVTDSGYYETYNRDN  357
CHMO  GDIATNMEANIEAQNFIKGKIAEIVKDPAIAQKLMPQD--LTAKRPLCDSGYYNTFNRDN  342
SMO   VELVDLRSTPIVGMDETGIVTTGAHY-DLDMIVLATGFDAMTGSLDKLEIVGRGGRTLKE  416
CHMO  VRLEDVKANPIVEITENGVKLENGDFVELDLMICATGFDAVDGNYVRMDIQGKNGLAMKD  402
SMO   TWAAGPRTYLGLGIDGFPNFFNLTGPGSPSVLANMVLHSELHVDWVADAIAYLDARGAAG  476
CHMO  YWKEGPSSYMGVTVNNYPNMFMVLGPNGPP--TNLPPSIESQVEWISDTIQYTVENNVES  460
SMO   IEGTPEAVADWVEECRNRAEASLLNSANSWYLGANIPGRPRVFMPFLGGFGVYREIITEV  536
CHMO  IEATKEAEEQWTQTCANIAEMTLFPKAQSWIFGANIPGKKNTVYFYLGGLKEYRTCASNC  520
SMO   AESGYKGFAILEG                                               549
CHMO  KNHAYEGFDIQLQRSDIKQPANA                                      543
```

Fig. 2. **Alignment of protein sequences of *Rhodococcus rhodochrous* steroid monooxygenase and *Acinetobacter* cyclohexanone monooxygenase.** ":" indicates an identical amino acid and "." a conserved substitution. A dashed line, "-", indicates a gap in the sequence. SMO, steroid monooxygenase; and CHMO, cyclohexanone monooxygenase. The adenine-binding motifs for FAD and NADPH are indicated by the underlined sequences of positions 23–37 and 47–51, and positions 193–211 and 217–221 of SMO, respectively. The underlined sequences of CHMO indicate the same motifs as in SMO.

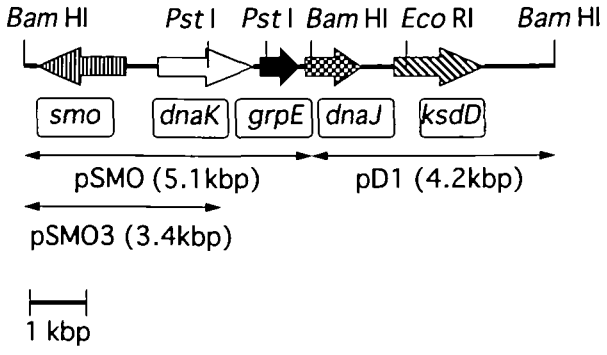

Fig. 3. **Schematic presentation of the gene organization of the *Bam*HI-DNA fragment of *R. rhodochrous*.** The arrows represent cloned and subcloned DNA fragments from the chromosomal DNA from *R. rhodochrous*. The pSMO3 insert DNA, and the 3′-part of the pSMO insert were sequenced. The pD1 insert was sequenced in the previous study (*10*). The ORFs, *dnaK*, *grpE*, and *dnaJ*, were found in this order.

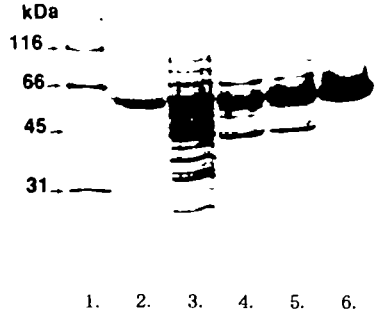

Fig. 4. **SDS-PAGE of the recombinant steroid monooxygenase from *E. coli* containing pSMO-EX at various stages of the purification.** Lane 1, molecular weight markers; lane 2, native enzyme from *R. rhodochrous* (60 kDa); lane 3, crude extract; lane 4, eluent of 1st DEAE-cellulose chromatography; lane 5, eluent of 2nd DEAE-cellulose chromatography; and lane 6, eluent of preparative gel electrophoresis. A sample of 10-μ protein was loaded on each lane.

TABLE 1. **Summary of purification of the recombinant steroid monooxygenase.**

Step	Total volume (ml)	Total protein (mg)	Total activity (U*)	Specific activity (U/mg)	Yield (%)	Purification factor
			($\times 10^3$)			
Crude extract	440	3,406	506	149	100	1
1st DEAE-cellulose	180	738	342	463	67	3.1
2nd DEAE-cellulose	55	315	215	682	43	4.6
Preparative Gel Electrophoresis	16.6	131	125	947	25	6.4

*1 U = 1 nmol testosterone acetate produced/min. The purification was carried out with 57 g wet cells.

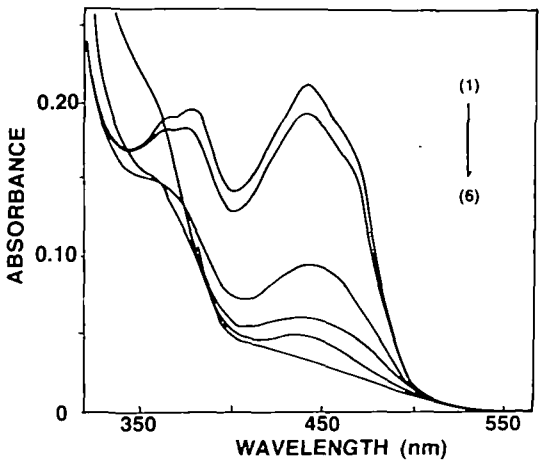

Fig. 5. **Spectrophotometric analysis of the recombinant steroid monooxygenase.** Visible absorption spectra of the oxidized and NADPH-reduced recombinant enzyme were monitored at 25°C in the 340-600 nm range. The sample consisted of 18.9 μM purified recombinant steroid monooxygenase in 30 mM Tris-HCl buffer, pH 8.4, containing 0.1 mM EDTA and 50 μM progesterone in a volume of 2-ml. The reduced form was prepared by addition of NADPH under anaerobic conditions in a Thunberg cuvette. (1) Oxidized form; (2) 4.3 μM; (3) 12.5 μM; (4) 29 μM; and (5) 44 μM NADPH added. (6) Dithionite-reduced form.

(*1*). Complete reduction of the SMO-enzyme, however, could not be attained even with the addition of a twofold excess amount of NADPH. These spectrophotometric properties indicate that the FAD molecule in the recombinant enzyme is oriented specifically in the same manner as for the *Rhodococcus* flavoprotein.

The purified recombinant steroid monooxygenase catalyzed the conversion of progesterone to testosterone acetate with the concomitant oxidation of NADPH, as judged on HPLC and TLC, and spectral analysis, with apparent K_m and V_{max} values of 55 μM and 850 mol/min·mol at pH 8.3 and 37°C for progesterone, and an apparent K_m value of 0.44 μM for NADPH, respectively. These were near the values of the wild type monooxygenase under the same assay conditions. The recombinant enzyme exhibited the same specificity pattern for substrate steroids as the native enzyme (*5*).

DISCUSSION

The present study disclosed the nucleotide sequence of the *smo* gene of 1,650 nucleotide length and the sequence of 549 amino acid residues of *R. rhodochrous* steroid monooxygenase was deduced. A nucleotide study of the flanking region of the *smo* gene revealed the orientation of genes of the steroid metabolizing enzymes, steroid monooxygenase and 3-ketosteroid-Δ^1-dehydrogenase (*10*), on the chromosomal DNA. The monooxygenase gene showed a high content of G+C of 68.7% in the coding region, and a high preferential usage (92.1%) of C- and G-ending codons. These values for *ksdD* of this organism were 68 and 87.25%, respectively (*10*).

It is noteworthy that the *smo* gene is encompassing on the complementary strand of chromosomal DNA, on which the *ksdD* gene is encoded, and TTG is the initiation codon which is used at a far lower frequency. This is supported by the findings that the amino-terminal sequence determined from the purified steroid monooxygenase shows good agreement with the sequence deduced from the base sequence of the *smo* gene, and the SD sequence is located 10 bp upstream from TTG, which is used for methionine as the initiation codon. The codon usages were reported for the genes of a regulatory factor protein (*nhlD*) of the low-molecular weight nitrile hydratase from *R. rhodochrous* J1 (*17*) and other proteins (*18-28*). Induced expression of the steroid monooxygenase was lower than that of 3-ketosteroid-Δ^1-dehydrogenase in *R. rhodochrous* cells, suggesting that the initiation codon, TTG, functions at a lower efficiency on expression of the *smo* gene (*5*).

The amino acid sequence of steroid monooxygenase of *R. rhodochrous* is highly homologus to that of cyclohexanone monooxygenase of *Acinetobacter* sp., a Baeyer-Villigerase flavoprotein catalyzing the oxygenative ring expansion of cyclohexanone (*9*), but less so to those of flavoprotein hydroxylases of *P. fluorescence* and *P. putida* S1 (*12, 13*). The N-terminus sequence (positions 1-18) of steroid monooxygenase from *C. radicicolla*, containing the GXGXXG motif of adenylate binding, exhibits 67% similarity to that of *R. rhodochrous* (*5*). The bacterial steroid monooxygenase only catalyzes Baeyer-Villiger oxidation of 21-ketosteroid, and the enzyme from a fungus, *C. radicicolla*, catalyses bifunctionally the oxidation of 21-ketosteroid, oxidative ester formation, and the 17-ketosteroid, oxidative ring-expansion reaction. The *Acinetobacter* enzyme acts on cyclohexanone and various other compounds to expand the ring structure like the fungal enzyme (*6-8*). The substrates of steroid monooxygenase and cyclohexanone monooxygenase are aliphatic or alicyclic ketones with different structures, but the enzymes are highly homologous in amino acid sequence; this suggests the presence of a characteristic protein structure common to Baeyer-Villiger oxygenases. The sequences at positions 23-36 and positions 48-51 of steroid monooxygenase are found in both enzymes and were confirmed to be the binding motif for adenine of FAD in flavoprotein. The sequence is preceded and followed by many hydrophobic amino acid residues, yielding a $\beta\alpha\beta$-structure. The conserved motif and FAD fingerprint in the flavoprotein hydroxylases are not observed in the sequences of steroid monooxygenase and cyclohexanone monooxygenase (*29*). The second consensus sequence at positions 193-214 might be the binding motif for adenine of NADPH, that was found in adrenodoxin reductase and others (*15*). In six other regions of highly conserved sequences of the *R. rhodochrous* enzyme seen in Fig. 3, the 6-8th regions might constitute the substrate binding domain of the enzyme. This will become clearer when the amino acid sequence of fungal steroid monooxygenase is determined and compared with those of the two enzymes.

On construction of the expression vector of the steroid monooxygenase, the initiation codon was changed from a less-frequently used codon, TTG, to a generally used codon, ATG, to give good expression efficiency in the host cells of *E. coli*, and the fourth base adenine was replaced with guanine to make a *Nco*I site (CCATGG), which became the ligation site of the 5′-terminal of the gene as to plasmid pET28a(+). The base change resulted in the replacement of the second amino acid, Asn, with Asp. Other amino acids

were not changed in the construction. The designed expression system for the steroid monooxygenase could efficiently express and produce a large amount of the enzyme on induction with IPTG, like 3-ketosteroid-Δ^1-dehydrogenase (*10*). The product could be extracted from the *E. coli* cells by brief sonication as a soluble flavoenzyme protein. The specific activity of steroid monooxygenase in the crude extract of the *E. coli* cells is 40 times greater than that in the case of *R. rhodochrous* (*5*). The monooxygenase could be purified by a simplified method to homogeneity with a high yield. The recombinant enzyme has the same molecular mass as the native enzyme, but its mobility on native gel electrophoresis is slightly different from that of the native enzyme due to that the amino acid residue was changed from Asn to Asp at position 2. This replacement does not affect the catalytic properties of the recombinant enzyme. The purified enzyme exhibits the same spectrophotometric properties as the native type enzyme. The reaction kinetic parameters of the enzyme are nearly the same as those of the native type enzyme (*5*).

The steroid monooxygenase gene is located about 4.6 kbp away from the 3-ketosteroid-Δ^1-dehydrogenase gene and is encoded on the complementary strand of DNA in the reverse direction. Three heat shock protein genes are present between them and in the upstream of the dehydrogenase genes (Fig. 3). Therefore, the two steroid metabolizing enzymes do not construct an operon of genes for steroid metabolizing enzyme proteins and control genes for the expression could not be found in the upstream region of each enzyme gene. The meaning of this orientation is not clear and this is a problem to be resolved for understanding regulation of the expression of the enzymes.

In conclusion, this is the first report on determination of the nucleotide sequence and amino acid sequence of a bacterial Baeyer-Villiger oxygenase, steroid monooxygenase. Comparison of Baeyer-Villigerases revealed the high homology in the amino acid sequence, suggesting the presence of a specified protein structure in the Baeyer-Villiger enzymes. The constructed expression plasmid of the steroid monooxygenase efficiently expressed the enzyme. The hyperexpressed enzyme was indistinguishable in molecular and catalytic properties from the native monooxygenase. Thus, it was confirmed that the DNA fragment cloned from total DNA of *R. rhodochrous* encompasses the steroid monooxygenase gene, *smo*. This hyperexpression system will be useful for studying the structure and function of the amino acid residues in the active site for catalysis of the unique oxygenase reaction.

REFERENCES

1. Itagaki, E. (1986) Studies on steroid monooxygenase from *Cylindrocarpon radicicola* ATCC 11011. Purification and characterization. *J. Biochem.* **99**, 815-824
2. Itagaki, E. (1986) Studies on steroid monooxygenase from *Cylindrocarpon radicicola* ATCC11011. Oxygenative lactonization of androstenedione to testololactone. *J. Biochem.* **99**, 825-832
3. Katagiri, M. and Itagaki, E. (1991) A steroid ketone monooxygenase from *Cylindrocarpon radicicola* in *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.) Vol. II, pp. 101-108, CRC Press, Boca Raton
4. Itagaki, E. and Katagiri, M. (1984) Steroid monooxygenase from *Cylindrocarpon radicicola*. An FAD-containing Baeyer-Villiger type oxygenase in *Flavins and Flavoproteins* (Bray, R.C., Engel, P.C., and Mayhew, S.G. eds.) pp. 639-642, Walter de Gruyter, Berlin
5. Miyamoto, M., Matsumoto, J., Iwaya, T., and Itagaki, E. (1995) Bacterial steroid monooxygenase catalyzing the Baeyer-Villiger oxidation of C_{21}-ketosteroids from *Rhodococcus rhodochrous*: the isolation and characterization. *Biochim. Biophys. Acta* **1251**, 115-124
6. Donoghue, N.A., Norris, D.B., and Trudgill, P.W. (1976) The purification and properties of cyclohexanone oxygenase from *Nocardia globerula* CL1 and *Acinetobacter* NCIB 9871. *Eur. J. Biochem.* **63**, 175-192
7. Griffin, M. and Trudgill, P.W. (1976) Purification and properties of cyclohexanone oxygenase from *Pseudomonas* NCIB 9872. *Eur. J. Biochem.* **63**, 199-209
8. Ryerson, C.C., Ballou, D.P., and Walsh, C.T. (1982) Mechanistic studies on cyclohexanone oxygenase. *Biochemistry* **21**, 2644-2655
9. Chen, Y.-C.J., People, O.P., and Walsh, C.T. (1988) *Acinetobacter* cyclohexanone monooxygenase: Gene cloning and sequence determination. *J. Bacteriol.* **170**, 781-789
10. Morii, S., Fujii, C., Miyoshi, T., Iwami, M., and Itagaki, E. (1998) 3-Ketosteroid-Δ^1-dehydrogenase of *Rhodococcus rhodochrous*: Sequencing of the genomic DNA and hyperexpression, purification, and characterization of the recombinant enzyme. *J. Biochem.* **124**, 1026-1032
11. Weeks, D.P., Beerman, N., and Griffith, O.M. (1986) A small scale five-hour procedure for isolating multiple samples of CsCl-purified DNA: application to isolations from mammalian, insect, higher plant, algal, yeast, and bacterial sources. *Anal. Biochem.* **152**, 376-385
12. Suzuki, K., Mizuguchi, M., Ohnishi, K., and Itagaki, E. (1996) Structure of chromosomal DNA coding for *Pseudomonas putida* S-1 salicylate hydroxylase. *Biochim. Biophys. Acta* **1275**, 154-156
13. Hofsteenge, J., Weijer, W.J., Jekel, P.A., and Beintema, J.J. (1983) *p*-Hydroxybenzoate hydroxylase from *Pseudomonas fluorescence*. I. Completion of the elucidation of the primary structure. *Eur. J. Biochem.* **133**, 91-108
14. Wierenga, R.K., Terpstra, P., and Hol, W.G.J. (1986) Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**, 101-107
15. Hanukoglu, I. and Gutfinger, T. (1989) cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites in oxidoreductases. *Eur. J. Biochem.* **180**, 479-484
16. Lathigra, R.B., Young, D.B., Sweetser, D., and Young, R.A. (1988) A gene from *Mycobacterium tuberculosis* which is homologous to the DnaJ heat shock protein of *E. coli*. *Nucleic Acids Res.* **16**, 1636
17. Komeda, H., Kobayashi, M., and Shimizu, S. (1996) A novel gene cluster including the *Rhodococcus rhodochrous* J1 *nhlBA* genes encoding a low molecular mass nitrile hydratase (L-NHase) induced by its reaction product. *J. Biol. Chem.* **271**, 15796-15802
18. Raden, B. and Rutberg, L. (1984) Nucleotide sequence of the temperate *Bacillus subtilis* bacteriophage SPO2 DNA polymerase gene L. *J. Virol.* **52**, 9-15
19. Harwood, C.R., Williams, D.M., and Lovett, P.S. (1983) Nucleotide sequence of a *Bacillus pumilus* gene specifying chloramphenical acetyltransferase. *Gene* **24**, 163-169
20. Golderer, G., Dlaska, M., Grobner, P., and Piendl, W. (1995) TTG serves as an initiation codon for the ribosomal protein MvaS7 from the archaeon *Methanococcus vannielii*. *J. Bacteriol.* **177**, 5994-5996
21. Chopra, A.K., Peterson, J.W., Chary, P., and Prasad, R. (1994) Molecular characterization of an enterotoxin from *Salmonella typhimurium*. *Microb. Pathog.* **16**, 85-98
22. Nakao, M., Nakayama, T., Kakudo, A., Inohara, M., Harada, M., Omura, F., and Shibano, Y. (1994) Structure and expression of a gene coding for thermostable α-glucosidase with a broad substrate specificity from *Bacillus* sp. SAM1606. *Eur. J. Biochem.* **220**, 293-300
23. Black, P.N., DiRusso, C.C., Metzger, A.K., and Heimert, T.L.

(1992) Cloning, sequencing, and expression of the fadD gene of *Escherichia coli* encoding acyl coenzyme A synthetase. *J. Biol. Chem.* **267**, 25513-25520
24. Kitazono, A., Yoshimoto, T., and Tsuru, D. (1992) Cloning, sequencing, and high expression of the proline iminopeptidase gene from *Bacillus coagulans*. *J. Bacteriol.* **174**, 7919-7925
25. Nishiyama, M., Horinouchi, S., and Beppu, T. (1991) Characterization of an operon encoding succinyl-CoA synthetase and malate dehydrogenase from *Thermus flavus* AT-62 and its expression in *Escherichia coli*. *Mol. Gen. Genet.* **226**, 1-9
26. Sin, K., Nakamura, A., Kobayashi, K., Masaki, H., and Uozumi, T. (1991) Cloning and sequencing of a cyclodextrin glucanotransferase gene from *Bacillus ohbensis* and its expression in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **35**, 600-605
27. Suzuki, T., Kazama, S., Hirai, A., Akihama, T., and Kadowaki, K. (1991) The rice mitochondrial nad3 gene has an extended reading frame at its 5′ end: nucleotide sequence analysis of rice *trnS*, *nad3*, and *rps12* genes. *Curr. Genet.* **20**, 331-337
28. Mauch, L., Bichler, V., and Brandsch, R. (1990) Functional analysis of the 5′ regulatory region and the UUG translation initiation codon of the *Arthrobacter oxidans* 6-hydroxy-D-nicotine oxidase gene. *Mol. Gen. Genet.* **221**, 427-434
29. Eppink, M.H.M., Schreuder, H.A., and Van Berkel, W.J.H. (1997) Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. *Protein Sci.* **6**, 2454-2458