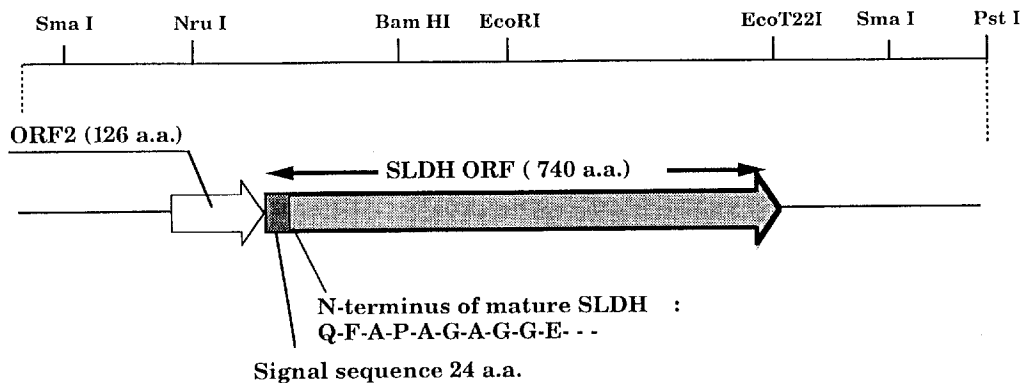




(72) HOSHINO, Tatsuo, JP
(72) OJIMA, Setsuko, JP
(72) SHINJOH, Masako, JP
(72) TOMIYAMA, Noribumi, JP
(72) MIYAZAKI, Taro, JP
(71) F. HOFFMANN-LA ROCHE AG, CH
(51) Int.Cl.⁶ C12N 15/53, C12N 1/21, C12N 9/04, C12P 19/02
(30) 1997/08/21 (97114432.4) EP
(54) **GENE DE LA D-SORBITOL-DESHYDROGENASE**
(54) **D-SORBITOL DEHYDROGENASE GENE**



(57) ADN renfermant une séquence de nucléotides codant une protéine, définie comme en A ou en B et possédant une activité de sorbitol-déshydrogénase : A) protéine renfermant la séquence d'acides aminés de la position 1 à 716 de la séquence décrite dans SEQ ID NO: 2; B) protéine dérivée de la protéine A par substitution, délétion, insertion ou addition d'un ou de plusieurs acides aminés dans la séquence d'acides aminés définie en A; vecteurs renfermant ce type d'ADN, hôtes transformés par ces vecteurs; méthode pour préparer ces protéines et leur utilisation.

(57) A DNA comprising a nucleotide sequence which encodes a protein as defined by (a) or (b) and having sorbitol dehydrogenase activity: (a) a protein having the amino acid sequence from the position 1 to 716 of the sequence described in SEQ ID NO: 2, or (b) a protein derived from the protein of (a) by substitution, deletion, insertion or addition of one or more amino acids in the amino acid sequence defined in (a), vectors, comprising such DNAs, hosts transformed by such vectors, a process for the preparation of such proteins and their use.



Abstract

A DNA comprising a nucleotide sequence which encodes a protein as defined by (a) or (b) and having sorbitol dehydrogenase activity:

5 (a) a protein having the amino acid sequence from the position 1 to 716 of the sequence described in SEQ ID NO: 2, or

(b) a protein derived from the protein of (a) by substitution, deletion, insertion or addition of one or more amino acids in the amino acid sequence defined in (a), vectors, comprising such DNAs, hosts transformed by such
10 vectors, a process for the preparation of such proteins and their use.

F.HOFFMANN-LA ROCHE AG, CH-4070 Basle/Switzerland

Ref. 13'738

D-Sorbitol Dehydrogenase Gene

The present invention relates to a novel DNA coding for sorbitol dehydrogenase of a microorganism belonging to acetic acid bacteria including
5 the genus *Gluconobacter* and the genus *Acetobacter*, an expression vector containing the said DNA, and recombinant organisms containing the said expression vector. Furthermore, the present invention relates to a process for producing recombinant D-sorbitol dehydrogenase and a process for producing L-sorbose by utilizing the said recombinant enzymes or recombinant
10 organisms containing the said expression vector.

L-Sorbose is an important intermediate in the actual industrial process of vitamin C production, which is mainly practiced by Reichstein method (Helvetica chimica Acta, 17: 311, 1934). In the process, the only microbial conversion is the L-sorbose production from D-sorbitol by *Gluconobacter* or
15 *Acetobacter* strains. The conversion is considered to be carried out by NAD/NADP independent D-sorbitol dehydrogenase (SLDH). L-Sorbose is also a well known substrate in the art for microbiologically producing 2-keto-L-gulonic acid, which is a useful intermediate in the production of vitamin C.

It is known that there are NAD/NADP-independent D-sorbitol
20 dehydrogenase which catalyzes the oxidation of D-sorbitol to L-sorbose. The said D-sorbitol dehydrogenase was isolated and characterized from *Gluconobacter suboxydans* var α IFO 3254 (E. Shinagawa et al., Agric Biol. Chem., 46: 135-141, 1982), and found to consist of three subunits with the molecular weight of 63 kDa, 51 kDa, and 17 kDa; the largest subunit is
25 dehydrogenase containing FAD as a cofactor, the second one is cytochrome c AB/vs / 26.6.98

- 2 -

and the smallest one is a protein with unknown function; and shows its optimal pH at 4.5. Such SLDH was also purified and characterized from *G. suboxydans* ATCC 621 (KCTC 2111) (E-S Choi et al., FEMS Microbiol. Lett. 125:45-50, 1995) and found to consist of three subunits with the molecular weight of 75 kDa, 50 kDa, and 14 kDa; the large subunit is dehydrogenase containing pyrroloquinoline quinone (PQQ) as a cofactor, the second one is cytochrome *c* and the small one is a protein with unknown function. The inventors also purified and characterized the NAD/NADP-independent SLDH from *G. suboxydans* IFO 3255 (T. Hoshino et al., EP 728840); the SLDH consists of one kind of subunit with the molecular weight of 79.0 +/- 0.5 kDa and shows its optimal pH at 6 to 7 and shows dehydrogenase activity on mannitol and glycerol as well as on D-sorbitol.

Although several SLDHs have been purified, their genes have not been cloned yet. It is useful to clone the SLDH gene for efficient production of the SLDH enzyme and for constructing recombinant organism having enhanced SLDH activity to improve the production yield of L-sorbose. It is also useful to introduce the said SLDH gene into desired organisms, for example, *Gluconobacter* converting L-sorbose to 2-keto-L-gulonic acid for constructing recombinant microorganisms which directly produce 2-keto-L-gulonic acid from D-sorbitol.

The present invention provides a nucleotide sequence (gene) coding for D-sorbitol dehydrogenase (SLDH) originating from a microorganism belonging to acetic acid bacteria including the genus *Gluconobacter* and the genus *Acetobacter*; a DNA molecule comprising said nucleotide sequence as well as a combination of the said DNA with a DNA comprising a nucleotide sequence of a protein functional in activating the said SLDH in vivo, expression vectors carrying the DNA comprising SLDH nucleotide sequence or the said combination of the DNAs; recombinant organisms carrying the expression vectors; a process for producing the recombinant SLDH; and a process for producing L-sorbose by utilizing the recombinant SLDH or the recombinant organisms.

The present invention is also directed to functional derivatives of the present case. Such functional derivatives are defined on the basis of the amino acid sequences of the present invention by addition, insertion, deletion and/or substitution of one or more amino acid residues of such sequences wherein such derivatives still have SLDH activity measured by an assay known in the art or specifically described herein. Such functional derivatives

- 3 -

can be made either by chemical peptide synthesis known in the art or by recombinant means on the basis of the DNA sequences as disclosed herein by methods known in the state of the art and disclosed e.g. by Sambrook et al. (Molecular Cloning, Cold Spring Harbour Laboratory Press, New York, USA, second edition 1989). Amino acid exchanges in proteins and peptides which do not generally alter the activity of such molecules are known in the state of the art and are described, for example, by H. Neurath and R. L. Hill in *The Proteins* (Academic Press, New York, 1979, see especially Figure 6, page 14). The most commonly occurring exchanges are: Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, Asp/Gly as well as these in reverse.

Furthermore the present invention is directed to DNA sequences encoding the polypeptides with SLDH activity and polypeptides functional in activating the said SLDH in vivo as disclosed e.g. in the sequence list as SEQ ID NO:2 and NO:3 as well as their complementary strands, or those which include these sequences, DNA sequences which hybridize under standard conditions with such sequences or fragments thereof and DNA sequences, which because of the degeneration of the genetic code, do not hybridize under standard conditions with such sequences but which code for polypeptides having exactly the same amino acid sequence.

Standard conditions for hybridization mean in this context the conditions which are generally used by a man skilled in the art to detect specific hybridization signals and which are described, e. g. by Molecular Cloning, Cold Spring Harbour Laboratory Press, New York, USA, second edition 1989, or preferably so called stringent hybridization and stringent washing conditions a man skilled in the art is familiar with and which are described, e. g. in Sambrook et al. (s. a.).

In the following a brief description of the drawings is given.

Figure 1 illustrates the partial amino acid sequences of SLDH polypeptide and oligonucleotides. The amino acid sequences in boldface in the figure were used for synthesizing two oligonucleotide sequences (S7 and S6R) for PCR. Arrow shows direction of DNA synthesis. All primers were degenerate DNA mixtures having bias for *Gluconobacter* codon usage.

- 4 -

Figure 2 illustrates a restriction map of the SLDH gene cloned in the present invention and the genetic structure of the DNA region encoding SLDH and ORF2.

Figure 3 illustrates the nucleotide sequence encoding SLDH and ORF2 with upstream and downstream sequences and illustrates deduced amino acid sequences of SLDH and ORF2. Figure 3 also illustrates putative ribosome-binding sites (SD sequences) of SLDH and ORF2 genes and the possible transcription terminator sequences.

Figure 4 illustrates the steps for constructing plasmids pTNB114, pTNB115, and pTNB116 for the expression of SLDH and/or ORF2 genes in *E. coli* under the control of lac promoter.

Figure 5 illustrates the steps for constructing plasmids pTNB110 which carries ORF2 and SLDH genes under the control of common pA promoter (described below in Example 6) and pTNB143 which carries ORF2 gene under the control of pA located adjacent to the gene in the upstream and SLDH gene under the control of another pA promoter located adjacent to the SLDH gene in its upstream.

The inventors have isolated SLDH gene together with a gene functionable in developing SLDH activity in vivo by DNA recombinant techniques and determined the nucleotide sequences.

The present invention provides DNA sequences encoding *Gluconobacter* SLDH and a gene functionable in developing SLDH activity in vivo herein after referred to as ORF2 gene, expression vectors containing the said DNA for SLDH and host cells carrying the said expression vectors.

Briefly, the SLDH and/or ORF2 gene(s), the DNA molecule containing the said gene(s), the recombinant expression vector and the recombinant organism utilized in the present invention can be obtained by the following steps:

(1) Isolating chromosomal DNA from the microorganisms which can provide SLDH activity that converts D-sorbitol to L-sorbose and constructing the gene library with the chromosomal DNA in an appropriate host cell, e. g. *E. coli*.

- 5 -

- (2) Cloning SLDH and/or ORF2 gene(s) from a chromosomal DNA by colony-, plaque-, or Southern-hybridization, PCR (polymerase chain reaction) cloning, Western-blot analysis and the like.
- (3) Determining the nucleotide sequence of the SLDH and/or ORF2 gene(s) obtained as above by conventional methods to select DNA molecule containing said SLDH and/or ORF2 gene(s) and constructing the recombinant expression vector on which SLDH and/or ORF2 gene(s) can express efficiently.
- (4) Constructing recombinant organisms carrying SLDH and/or ORF2 gene(s) by an appropriate method for introducing DNA into host cell, e. g. transformation, transduction, transconjugation and/or electroporation.

The materials and the techniques used in the above aspect of the present invention are exemplified in details as follows:

A total chromosomal DNA can be purified by a procedure well known in the art. The aimed gene can be cloned in either plasmid or phage vectors from a total chromosomal DNA typically by either of the following illustrative methods:

(i) The partial amino acid sequences are determined from the purified proteins or peptide fragments thereof. Such whole protein or peptide fragments can be prepared by the isolation of such a whole protein or by peptidase-treatment from the gel after SDS-polyacrylamide gel electrophoresis. Thus obtained protein or fragments thereof are applied to protein sequencer such as Applied Biosystems automatic gas-phase sequencer 470A. The amino acid sequences can be utilized to design and prepare oligonucleotide probes and/or primers with DNA synthesizer such as Applied Biosystems automatic DNA sequencer 381A. The said probes can be used for isolating clones carrying the objective gene from a gene library of the strain carrying the objective gene with the aid of Southern-, colony- or plaque-hybridization.

(ii) Alternatively, for the purpose of selecting clones expressing aimed protein from the gene library, immunological methods with antibody prepared against the aimed protein can be applied.

- 6 -

(iii) The DNA fragment of the aimed gene can be amplified from the total chromosomal DNA by PCR method with a set of primers, i.e. two oligonucleotides synthesized according to the amino acid sequences determined as above. Then a clone carrying the aimed-whole gene can be isolated from the gene library constructed, e.g. in *E. coli* by Southern-, colony-, or plaque-hybridization with the PCR product obtained above as the probe.

(iv) A further alternative way of the cloning is screening of the clone complementing SLDH-deficient strain constructed by conventional mutation with chemical mutagenesis or by recombinant DNA techniques e.g. with transposon Tn5 to disrupt aimed gene.

DNA sequences which can be made by the polymerase chain reaction by using primers designed on the basis of the DNA sequences disclosed therein by methods known in the art are also an object of the present invention. It is understood that the DNA sequences of the present invention can also be made synthetically as described, e.g. in EP 747 483.

Above mentioned antibody can be prepared with purified SLDH protein or its peptide fragment as an antigen by such method described in Methods in Enzymology, vol. 73, p 46, 1981.

Once a clone carrying the desired gene is obtained, the nucleotide sequence of the aimed gene can be determined by a well known method such as dideoxy chain termination method with M13 phage (Sanger F. et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, 1977).

The desired gene expressing the D-sorbitol dehydrogenase activity of the present invention is illustrated in Fig. 2 and Fig. 3. This specific gene encodes the SLDH enzyme having 716 amino acid residues together with a signal peptide of 24 amino acid residues. The inventors found an open reading frame just upstream of the above SLDH structure gene and designated it as ORF2. This ORF2 gene encodes a protein having 126 amino acid residues, and was suggested to be functional in providing the desired enzymatic activity to the recombinant SLDH of the present invention, in particular when the said SLDH is expressed in a host cell of a different genus from acetic acid bacteria.

To express the desired gene/nucleotide sequence isolated from acetic acid bacteria including genus *Gluconobacter* and genus *Acetobacter* efficiently, various promoters can be used; for example, the original promoter of the gene, promoters of antibiotic resistance genes such as kanamycin resistant gene of

Tn5 (D. E. Berg, and C. M. Berg. 1983. Bio/Technology 1:417-435), ampicillin resistant gene of pBR322, and β -galactosidase of *E. coli* (lac), trp-, tac-, trc-promoter, promoters of lambda phage and any promoters which can be functional in a host organism. For this purpose, the host organism can be selected from microorganisms including bacteria such as *Escherichia coli*,
5 *Pseudomonas putida*, *Acetobacter xylinum*, *Acetobacter pasteurianus*, *Acetobacter aceti*, *Acetobacter hansenii*, and *Gluconobacter albidus*, *Gluconobacter capsulatus*, *Gluconobacter cerinus*, *Gluconobacter dioxyceticus*, *Gluconobacter gluconicus*, *Gluconobacter industrius*,
10 *Gluconobacter melanogenus*, *Gluconobacter nonoxygluconicus*, *Gluconobacter oxydans*, e.g. *Gluconobacter oxydans* DSM 4025, which had been deposited as DSM 4025 on March 17, 1987 under the conditions of the Budapest Treaty at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, BRD, *Gluconobacter oxydans* subsp. *sphaericus*, *Gluconobacter roseus*, *Gluconobacter rubiginosus*, *Gluconobacter suboxydans*, mammalian
15 cells and plant cells.

For expression, other regulatory elements, such as a Shine-Dalgarno (SD) sequence (for example, AGGAGG etc. including natural and synthetic sequences operable in the host cell) and a transcriptional terminator (inverted
20 repeat structure including any natural and synthetic sequence operable in the host cell) which are operable in the host cell into which the coding sequence will be introduced can be used with the above described promoters.

For the expression of membrane-bound polypeptides, like the SLDH protein of the present invention, a signal peptide, which contains usually 15 to
25 50 amino acid residues and is totally hydrophobic, is preferably associated. A DNA encoding a signal peptide can be selected from any natural and synthetic sequence operable in the desired host cell.

A wide variety of host/cloning vector combinations may be employed in cloning the double-stranded DNA. The cloning vector is generally a plasmid or
30 phage which contains a replication origin, regulatory elements, a cloning site including a multi-cloning site and selection markers such as antibiotic resistance genes including resistance genes for ampicillin, tetracycline, kanamycin, streptomycin, gentamicin, spectinomycin etc.

Preferred vectors for the expression of the gene of the present invention in *E. coli* is selected from any vectors usually used in *E. coli*, such as pBR322 or its derivatives including pUC18 and pBluescript II (Stratagene Cloning Systems, CA, USA), pACYC177 and pACYC184 (J. Bacteriol., 134:1141-1156, 5 1978) and their derivatives, and a vector derived from a broad host range plasmid such as RK2 (C. M. Thomas, Plasmid 5: 10, 1981) and RSF1010 (P. Guerry et al., J. Bacteriol. 117: 619-630, 1974). A preferred vector for the expression of the nucleotide sequence of the present invention in bacteria including *Gluconobacter*, *Acetobacter* and *P. putida* is selected from any 10 vectors which can replicate in *Gluconobacter*, *Acetobacter* and/or *P. putida*, as well as in a preferred cloning organism such as *E. coli*. The preferred vector is a broad-host-range vector such as a cosmid vector like pVK100 (V. C. Knauf et al., Plasmid 8: 45-54, 1982) and its derivatives and RSF1010. Copy number and stability of the vector should be carefully considered for stable and 15 efficient expression of the cloned gene and also for efficient cultivation of the host cell carrying the cloned gene. DNA molecules containing transposable elements such as Tn5 can be also used as a vector to introduce the desired gene into the preferred host, especially on a chromosome. DNA molecules containing any DNAs isolated from the preferred host together with the gene 20 of the present invention is also useful to introduce this gene into the preferred host, especially on a chromosome. Such DNA molecules can be transferred to the preferred host by applying any of a conventional method, e.g. transformation, transduction, transconjugation or electroporation, which are well known to those skilled in the art, considering the nature of the host and 25 the DNA molecule.

Useful hosts may include microorganisms, mammalian cells, and plant cells and the like. As a preferable microorganism, there may be mentioned bacteria such as *E. coli*, *P. putida*, *A. xylinum*, *A. pasteurianus*, *A. aceti*, *A. hansenii*, *Gluconobacter albidus*, *Gluconobacter capsulatus*, *Gluconobacter* 30 *cerinus*, *Gluconobacter dioxyaceticus*, *Gluconobacter gluconicus*, *Gluconobacter industrius*, *Gluconobacter melanogenus*, *Gluconobacter nonoxygluconicus*, *Gluconobacter oxydans*, *Gluconobacter oxydans subsp. sphaericus*, *Gluconobacter roseus*, *Gluconobacter rubiginosus*, *Gluconobacter suboxydans*, and any bacteria which are capable of expressing recombinant 35 SLDH and/or ORF2 gene(s). Functional equivalents, subcultures, mutants and variants of said microorganism can be also used in the present invention. A preferred strain is *E. coli* K12 or its derivatives, *P. putida*, *Gluconobacter*, or *Acetobacter* strains.

The SLDH and/or ORF2 gene(s)/nucleotide sequences provided in this invention are ligated into a suitable vector containing a regulatory region such as a promoter, a ribosomal binding site and a transcriptional terminator operable in the host cell described above with a well-known methods in the art to produce an expression vector. When the SLDH and ORF2 genes are cloned in combination, the two genes can be cloned either in tandem or separately on the same plasmid and also on the chromosomal DNA. Figures 4 and 5 exemplifies the form of the combined cloning of SLDH and ORF2 genes on the plasmid. One may clone also one gene on a plasmid and the other one on chromosomal DNA.

To construct a recombinant microorganism carrying a recombinant expression vector, various gene transfer methods including transformation, transduction, conjugal mating (Chapters 14 and 15, Methods for general and molecular bacteriology, Philipp Gerhardt et al. ed., American Society for Microbiology, 1994), and electroporation can be used. The method for constructing a recombinant organism may be selected from the methods well-known in the field of molecular biology. Usual transformation systems can be used for *E. coli*, *Pseudomonas*, *Gluconobacter* and *Acetobacter*. A transduction system can also be used for *E. coli*. Conjugal mating system can be widely used in Gram-positive and Gram-negative bacteria including *E. coli*, *P. putida* and *Gluconobacter*. A preferred conjugal mating is disclosed in WO89/06688. The conjugation can occur in liquid medium or on a solid surface. The preferred recipient for SLDH and/or ORF2 production is selected from *E. coli*, *P. putida*, *Gluconobacter* and *Acetobacter*. To the recipient for conjugal mating, a selective marker is usually added; for example, resistance against nalidixic acid or rifampicin is usually selected. Natural resistance can be also used; e.g. resistance against polymyxin B is useful for many *Gluconobacters*.

The present invention provides recombinant SLDH. One can increase the production yield of the SLDH enzyme by introducing the gene of SLDH provided by the present invention into organisms including acetic acid bacteria including the genus *Gluconobacter* and the genus *Acetobacter*. One can also produce active SLDH in microorganisms beside *Gluconobacter* by using the SLDH gene of the present invention in combination with the ORF2 gene of the present invention. The recombinant SLDH can be immobilized on a solid carrier for solid phase enzyme reaction? The present invention also provides recombinant organisms. One can produce L-sorbose from D-sorbitol

with the recombinant organisms. One can also produce L-sorbose from D-sorbitol even in a host organism beside acetic acid bacteria by introducing the SLDH gene in combination with the ORF2 gene of the present invention.

Examples

5 **Example 1. Determination of amino acid sequences of SLDH**

(1) N-terminal amino acid sequence analyses of lysylendopeptidase-treated SLDH polypeptides

Partial amino acid sequences of peptides No.1, 3 and 8 prepared from SLDH protein were determined (Fig. 1; SEQ ID NO:4 to 6). Purified SLDH of
10 *G. suboxydans* IFO 3255 (T. Hoshino et al., EP 728840) was digested with lysylendopeptidase; the reaction mixture (25 ml) contained 1.2 mM of lysylendopeptidase and 3 nmol of the SLDH protein in 100 mM Tris-HCl buffer, pH9.0, and the reaction was carried out at 37°C for 15 hours. The resulting peptide fragments were separated by HPLC (column : protein C-4,
15 VYDAC, California, USA) with acetonitrile/isopropanol gradient in 0.1% TFA at a flow rate of 1.0 ml/min. Elution of the peptides was monitored by UV absorbance at 214 nm, and the peak fractions were collected manually and subjected to N-terminal amino acid sequence analysis with an amino acid sequencer (Applied Biosystems model 470A, The Perkin Elmer Corp., Conn.,
20 USA).

Example 2. Cloning of partial SLDH gene by PCR

PCR was conducted with chromosomal DNA of *G. suboxydans* IFO 3255 and degenerate oligonucleotide DNA primers, s7 and s6R whose sequences are shown in Fig. 1. The PCR amplification was carried out with thermostable
25 polymerase Amplitaq (Perkin-Elmer, Roche Molecular Systems Inc., NJ, USA), using a thermal cycler (ZYMOREACTOR II AB-1820, ATTO, Tokyo, Japan). The following reaction mixture (50 µl) was used for PCR: 200 µM of dNTPs, 10 ~ 20 pmol of each primer (54 ~ 288 degeneracy), 6 ng of chromosomal DNA of *G. suboxydans* IFO 3255 and 0.5 unit of the DNA
30 polymerase in the buffer provided from the supplier. The reaction consisted of 30 cycles of 1) denaturation step at 94°C for 1 min; 2) annealing step at 48°C for 1 min; 3) synthesis step at 72°C for 2 min. Consequently, 1.6 kb DNA was amplified and cloned in *E. coli* vector, pUC57/T (MBI Fermentas, Vilnius, Lithuania), which has 3' -ddT-tailed ends for direct ligation of an amplified
35 DNA fragment to obtain a recombinant plasmid pMT20. The cloned DNA was

subjected to nucleotide sequencing by the method of dideoxy-chain termination (F. Sanger et al, Proc. Natl. Acad. Sci. USA, 74:5463-5467, 1977); the 1.6 kb fragment encoded the peptides No. 3 and No. 8.

Example 3. Complete cloning of the SLDH gene

5 (1) Construction of gene library of *G. suboxydans* IFO 3255

The chromosomal DNA of *G. suboxydans* IFO 3255 was prepared from the cells grown on MB agar plate for 2 days. The chromosomal DNA (160 µg) was partially digested with 20 units of *Eco* RI in 500 µl of reaction mixture. Portions of the sample were withdrawn at 5, 10, 15, 30, and 60 minutes and
10 the degree of the digestion was detected by agarose gel electrophoresis. Former four portions which contained partially-digested DNA fragments were combined and subjected to preparative gel electrophoresis (agarose: 0.6%). Fragments of 15 - 35 kb were cut out and electroeluted from the gel. The eluate was filtered and precipitated with sodium acetate and ethanol at -80°C.
15 The DNA fragments were collected by centrifugation and suspended in 200 µl of 10 mM Tris-HCl, pH8.0, buffer containing 1 mM EDTA.

In parallel, 1.8 µg of a cosmid vector pVK100 was completely digested with *Eco* RI and treated with calf intestine alkaline phosphatase. The linearized and dephosphorylated pVK100 was ligated with 15 - 35 kb *Eco* RI
20 fragments of the chromosomal DNA of *G. suboxydans* IFO 3255 (5 µg) with the ligation kit (Takara Shuzo, Kyoto, Japan) in 20 separate tubes under the condition recommended by the supplier to obtain highly polymerized DNA. The ligated DNA was then used for in vitro packaging according to the method described by the supplier (Amersham Japan). The resulting phage particles
25 were used to infect *E. coli* ED8767, a host for the genomic library. Consequently, 4,271 colonies were obtained and all of the colonies tested (20 colonies) possessed the insert DNAs with the average size of about 25 kb.

(2) Colony hybridization to obtain complete SLDH gene

The cosmid library described above was screened to isolate the clone
30 carrying complete SLDH gene by colony hybridization method with ³²P-labeled 1.6 kb DNA of pMT20. One clone was isolated and designated pSLII, which carry about 25 kb insert in pVK100 vector. From the pSLII, 6.2 kb *Pst* I fragment was isolated and cloned into pUC18 to obtain plasmid pUSLIIP. The 6.2 kb *Pst* I fragment containing ORF2 and SLDH genes was isolated
35 from pUSLIIP and subcloned into pCRII vector (Invitrogen Corporation, CA,

- 12 -

USA) to produce pCRSLIIP. From the resulting plasmid, the DNA fragment containing the 6.2 kb *Pst* I fragment was isolated as *Hind* III-*Xho* I fragment and cloned between *Hind* III and *Xho* I sites of pVK100 to obtain pVKSLIIP.

(3) Expression of SLDH gene in *E. coli*

5 To confirm whether 6.2 kb *Pst* I-fragment encodes the aimed SLDH, cells of *E. coli* JM109 carrying pUSLIIP was subjected to Western-blot analysis with anti-SLDH antibody as described above. Immuno-positive proteins with a molecular weight of about 80 kDa were observed in the transformant, indicating that the *Pst* I-fragment encodes the polypeptides with the molecular
10 weight of the intact SLDH (79 kDa +/- 0.5 kDa).

(4) Construction of SLDH-deficient *Gluconobacter*, strain 26A11, as the test strain for SLDH-activity complementation

Transposon Tn5 mutagenesis was performed with *G. melanogenus* IFO 3293 as the parent. Tn5, a transposable element coding for *Kmr*, causes null
15 mutations at random on DNA of its host organism and widely used as a mutagen in Gram-negative bacteria. The IFO 3293 was selected as the parent in the following reasons: (i) it produced L-sorbose from D-sorbitol, (ii) it showed immuno-positive polypeptide of about 80 kDa in Western-blot analysis with the antibody prepared against SLDH purified from *G. suboxydans* IFO
20 3255 and (iii) its frequency for generating Tn5 mutants was much higher than that of *G. suboxydans* IFO 3255.

G. melanogenus IFO 3293 was cultivated in a test tube containing 5 ml of the MB medium containing 25 g/l of mannitol, 5 g/l of yeast extract (Difco Laboratories,), 3 g/l of Bactopepton (Difco) at 30°C overnight. *E. coli* HB101
25 (pRK2013) [D. H. Figurski, Proc. Natl. Acad. Sci. USA 76: 1648 - 1652, 1979] and *E. coli* HB101 (pSUP2021) [R. Simon, et al., BIO/TECHNOL. 1:784-791, 1983] were cultivated in test tubes containing 5 ml of LB medium with 50 µg/ml of kanamycin at 37°C overnight. The cells was separately collected by centrifugation and suspended in the half volume of MB medium. The each cell
30 suspension was mixed in the ratio of 1:1:1 and the mixture was placed on the nitrocellulose filter on the surface of MB agar plate. The plate was incubated at 27°C overnight and the resulting cells on the filter was scraped, suspended in the appropriate volume of MB medium and spread on the selection plate (MPK plate), MB containing 10 µg/ml of polymyxin B and 50 µg/ml of
35 kanamycin. The MPK plate was incubated at 27°C for 3 to 4 days.

The resulting 3,436 Tn5 mutants were subjected to the immuno-dot blot screening with the anti-SLDH antibody. The cells of each strain were independently suspended in 50 µl Laemmli buffer consisting of 62.5 mM Tris-HCl (pH6.5), 10% glycerol, 2% SDS and 5% β-mercaptoethanol in 96-well microtiter plate and incubated at 60°C for 2 hours. The Cell Free Extracts (CFEs) were stamped on the nitrocellulose filter and the immuno-positive samples in the CFEs were screened with AP conjugate substrate kit (Bio-RAD Laboratories, Richmond, Calif., USA). As a result, only one strain 26A11 without positive signal in the immuno-dot blot screening was obtained.

SLDH-deficiency of the strain 26A11 was confirmed by Western-blot analysis; it expressed at most 1/500 amount of SLDH compared with its parent strain. 26A11 was not a complete SLDH-deficient strain but the strain with SLDH gene repressed by Tn5 insertion; the insertion site was found to be close to the C-terminus by determining the nucleotide sequence around Tn5-insertion point. Next, a resting cell reaction was conducted to examine the whole SLDH activity in 26A11 and the wild *Gluconobacter* strains. In the potassium phosphate buffer 100 mM (pH7.0) containing 2% D-sorbitol, 26A11 slightly converted D-sorbitol to L-sorbose, whereas the wild strains IFO 3293 and IFO 3255 completely did it in 39.5 hr at 30°C.

(5) Expression of SLDH gene in 26A11

To confirm the SLDH activity of the SLDH clones obtained, complementation test was conducted. Plasmids pSLII and pVKSLIIP were introduced into 26A11 by a conjugal mating. The transconjugant carrying pSLII or pVKSLIIP restored the activity of SLDH in a mini-resting cell reaction and showed immuno-reactive polypeptide of about 80 kDa in Western-blot analysis.

(6) Nucleotide sequencing of the SLDH gene

Plasmid pUSLIIP was used for nucleotide sequencing of SLDH and ORF2 genes. Determined nucleotide sequence (SEQ ID NO: 1; 3,481 bp) revealed that ORF of SLDH gene (2,223 bp, nucleotide No. 572 to 2794 in SEQ ID NO: 1) encoded the polypeptide of 740 amino acid residues (SEQ ID NO: 2), in which there were three amino acid sequences (Peptides No. 1, 3, and 8 shown in SEQ ID NO: 4 to 6) determined from the purified SLDH polypeptide. In addition to the SLDH ORF, one more ORF, ORF2, was found just upstream of

- 14 -

SLDH ORF as illustrated in Figs. 2 and 3. The ORF of ORF2 (381 bp, nucleotide No. 192 to 572 in SEQ ID NO: 1) encoded the polypeptide of 126 amino acid residues (SEQ ID NO. 3).

The 4th amino acid sequence of Peptide No.1 was determined as Glu by the amino acid sequencer, but it was Ala according to the DNA sequence. The 11th amino acid sequence of Peptide No.3 was determined as Gln by the amino acid sequencer, but it was Pro according to the DNA sequence. A signal peptide-like region (SEQ ID NO: 8) is possibly included in the deduced amino acid sequence: it contains (i) many hydrophobic residues, (ii) a positively-charged residues near N-terminus, and (iii) Ala-Xaa-Ala site as a cleaved signal. The actual signal sequence was determined as described in Example 3 (7). The putative ribosome-binding site (Shine-Dalgarno, SD, sequence) for SLDH gene was located at 8 bp upstream of the initiation codon (AGAGGAG at nucleotide No. 558 - 564 of Seq ID NO: 1). The putative SD sequence for ORF2 gene was located at 10 bp upstream of the initiation codon (GGGAGG at nucleotide No. 177 to 182 of Seq ID NO: 1). There were some inverted repeat sequences immediately downstream the SLDH structure gene (nucleotide sequences of No. 2803 - 2833 and of No. 2838 - 2892) as illustrated in Fig. 3; they may function as transcription termination loops for SLDH gene. For ORF2 gene, the inverted repeat sequence was found at No. 684 - 704.

Homology search for SLDH and ORF2 was performed with the programs of mpBlast (NCBI, Bethesda, Md. USA) and Motifs in GCG (Genetics Computer Group, University Research Park, WI, USA). SLDH polypeptide had the sequence commonly conserved in quinoprotein at the region near C-terminus (within amino acid residue No. 632 to 692 of SEQ ID No. 2) and the other sequence identified as a quinoprotein motifs (Prosite No. PS00363: [DN]W.{3}G[RK].{6}[FY]S.{4}[LIVM]N.{2}NV.{2}L[RK]; amino acid residue No. 79 to 107 of SEQ ID No. 2).

ORF2 showed homology with the N-terminal region of the membrane-bound PQQ-dependent D-glucose dehydrogenase (GDH) of *G. oxydans*, *E. coli*, *Acinetobacter calcoaceticus*, which is known as a membrane-spanning region to bind the GDH to the membrane. Identities of ORF2 to the N-terminal region of the GDHs of *G. oxydans*, *E. coli*, *A. calcoaceticus* were 30%, 32%, and 37%, respectively. The ORF2 protein may function as an anchoring protein to make the SLDH membrane-bound type.

Example 4. Determination of N-terminal and C-terminal sequences of mature SLDH polypeptide

Direct sequencing of the N-terminus gave no results, indicating that the
5 N-terminus is blocked. Then, SLDH polypeptide was treated with the
endoproteinase Lys-C (Wako, Osaka Japan) in 0.1 M Tris-HCl at 37°C for 20
hours with a substrate-to-enzyme ratio of 20:1 (w/w). Total digest was
analyzed by reversed phase HPLC (RP300, 1 mm x 25 cm, Applied
Biosystems, Foster City, CA) and each peak was subjected to mass
10 spectrometer (TSQ700 triple quadrupole instrument, Finnigan-MAT, San
Jose, California) for determining molecular weight. One of the digest
described as SEQ ID No. 7 was assigned as the N-terminal sequence by the
mass spectrometric analysis and amino acid composition analysis together
with the amino acid composition predicted from determined nucleotide
15 sequence shown as SEQ ID NO: 1. Further analysis with collisional induced
dissociation (CID) was carried out to confirm the identity of the peptide with
N-terminal sequence. The N-terminus was determined to be pyroglutamyl
residue.

Since the N-terminus of SLDH was determined to be Gln-Phe-Ala-Pro-
20 Ala-Gly-Ala-Gly-Gly-Glu-Pro-Ser-Ser-Ser-Val-Pro-Gly-Pro-Gly-Asn-Ala-Ser-
Glu-Pro-Thr-Glu-Asn-Ser-Pro-Lys as shown in SEQ ID NO. 7, the signal
sequence was confirmed to be 24 amino acid residue long with the sequence of
Met-Arg-Arg-Pro-Tyr-Leu-Leu-Ala-Thr-Ala-Ala-Gly-Leu-Ala-Leu-Ala-Cys-Ser-
Pro-Leu-Ile-Ala-His-Ala as listed as SEQ ID NO: 8. The C-terminal sequence
25 was also determined by using the peptide recovered from V8 protease digest to
be Pro-Asp-Ala-Ile-Lys-Gln (SEQ ID NO: 9).

Example 5. Expression of the SLDH and/or ORF2 gene(s) in *E. coli*

From the pCRSLIIP described in Example 3 (2), plasmids carrying SLDH
30 gene with or without ORF2 gene under the lac promoter control were
constructed as illustrated in Fig. 4. The resulting three plasmids are
pTNB114 carrying SLDH and ORF2 genes, pTNB115 carrying SLDH gene
and ORF2 gene truncated at its N-terminus containing ribosome binding site
and start codon (ATG) and pTNB116 carrying mostly-truncated ORF2 gene
35 and intact SLDH gene. These three plasmids were introduced into *E. coli* by

- 16 -

conventional transformation. The production of the SLDH polypeptide was detected by Western-blot analysis with cell free extracts of the resulting transformants and the SLDH activity was assayed with resting cells. The resting cell reaction was carried out in the reaction mixture consisting of 0.3% NaCl, 1% CaCO₃, 4% D-sorbitol, and 1 mM PMS with or without 1 µg/ml of PQQ at room temperature for 17 hours. The SLDH activity to produce L-sorbose was analyzed by TLC assay with Silica gel 60 F254, 0.25 mm, Merck with the developing solvent consisting of *n*-propanol-H₂O-1% H₃PO₄-HCOOH (400:100:10:1) and spray reagent of naphthoresorcinol. Consequently, SLDH polypeptide was detected in all transformants, even without ORF2 gene expression, in Western-blot analysis, but SLDH activity to produce L-sorbose was detected only in transformant carrying pTNB114 containing intact ORF2 gene under the resting cell reaction condition in the presence of PQQ.

15 **Example 6. Expression of the SLDH and/or ORF2 gene(s) in *E. coli***

Figure 5 illustrates construction steps of pTNB110 and pTNB143. Plasmid pTNB110 carrying ORF2 and SLDH genes under control of the promoter of Enzyme A gene (pA) of DSM4025 (T. Hoshino et al., European Patent Application No. 9611500.8) was constructed by inserting *Hind* III-*Kpn* I fragment containing pA, *Kpn* I-*Xho* I fragment from pTNB114 between *Hind* III-*Xho* I of pUC18. Plasmid pTNB143 carrying ORF2 and SLDH genes as independent expression units, ORF2 gene with pA and SLDH gene with pA, was constructed by inserting 0.9 kb *Sal* I fragment from pTNB141 and 3.2 kb *Hind* III-*Xho* I fragment from pTNB135 into pUC57/pCRII hybrid vector (*Sca* I-*Hind* III fragment with plac from pUC57 and *Xho* I to *Sca* I fragment without plac from pCRII) as shown in Fig. 5. The plasmids, pTNB110 and pTNB143, were introduced into *E. coli* by conventional transformation. The resulting transformants were subjected to Western-blot analysis and resting cell reaction as described in Example 5. Consequently, both of the transformants carrying pTNB110 and pTNB143 produced SLDH protein and showed SLDH activity to produce L-sorbose from D-sorbitol in the presence of PQQ.

Example 7. Expression of the SLDH gene in *G. oxydans* DSM 4025

The plasmid pTNB136 for the expression of SLDH gene in *G. oxydans* DSM 4025 having strong L-sorbose and L-sorbosone dehydrogenase activities together with weak D-sorbitol dehydrogenase activity to produce L-sorbose, was constructed by inserting *Hind* III-*Xho* I fragment from pTNB135 (Fig. 5) between *Hind* III and *Xho* I sites of pVK100. The plasmid pTNB136 and its vector pVK100 were introduced by conjugal mating into strain GOBΔK, which is a mutant of *G. oxydans* DSM 4025 whose gene of Enzyme B (EP 832 974) having D-sorbitol dehydrogenase to produce L-sorbose is deleted by replacing two *Eco* RI fragments containing Enzyme B gene with kanamycin resistant gene cassette (1.28 kb *Eco* RI fragment of pUC4K; Pharmacia Uppsala, Sweden). The gene disruption was conducted by the recombinant DNA techniques well-known in the art. The resulting transconjugant, GOBΔK carrying pTNB136 or pVK100, produced 5 g/L or below 2 g/l of 2KGA in 10% D-sorbitol, respectively, by flask fermentation conducted at 30°C for 4 days (medium: 10% D-sorbitol, 1.6% urea, 0.05% glycerol, 0.25% MgSO₄·7H₂O, 3% corn steep liquor, 6.25% baker's yeast wet cells and 1.5% CaCO₃). Western-blot analysis with anti-SLDH antibody revealed that the transconjugants carrying pTNB136 expressed the immuno-reactive SLDH polypeptides.

- 18 -

SEQUENCE LISTING

- (1) GENERAL INFORMATION
- (i) APPLICANT NAME: F. HOFFMANN-LA ROCHE AG
STREET: Grezacherstrasse 124
CITY: Basle
COUNTRY: Switzerland
POSTAL CODE: CH-4002
TELEPHONE: 061 - 688 25 11
FAX: 061 - 688 13 95
TELEX: 962292/965542 hlr c
 - (ii) TITLE OF INVENTION: D-Sorbitol dehydrogenase gene
 - (iii) NUMBER OF SEQUENCES: 9
 - (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC Compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: MS word ver 5.1a
 - (v) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: CA 2,239,802
 - (B) FILING DATE: August 5, 1998
 - (C) CLASSIFICATION
- (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3481 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) ORIGINAL SOURCE:
 - ORGANISM: Gluconobacter suboxydans
 - STRAIN: IFO 3255
 - (iv) FEATURE:
 - FEATURE KEY: CDS
 - POSITION: 192..572
 - SEQUENCING METHOD: E
 - FEATURE KEY: CDS
 - POSITION: 572..2794
 - SEQUENCING METHOD: E
- ACAAATCATA CTGGCGGCGC TGTAGTGACA ATTCCGGCGG GTTAAAGAGA ATATTTTTTT 60
- GGTGACAGGC CACAACAAAT TTTTGTACC TCAAACACAG TTTTGTAGA GCATTGAAA 120
- ACGAAGTCCG ATGGACCTGA ACTGAATATG GATTTACCGT CCGGAGGATT CAGTTTGGGA 180
- GGCATTCCGGT TATGCCAAAT CTTCAAGGTA ATAGGACTCT GACGGAGTGG CTGACGCTGC 240
- TTCTCGGGGT CATCGTCCTT CTTGTGGGCC TGTTCCTCGT CATTGGGGGT GCTGACCTCG 300
- CGATGCTGGG CGGCTCTACC TACTATGTTT TCTGTGGCAT CCTCCTGGTT GCTAGCGGCG 360

- 19 -

TATTCATGCT CATGGGCCGC ACGCTTGGTG CCTTCCTGTA TCTGGGTGCC CTGGCCTACA 420
CGTGGGTCTG GTCCTTCTGG GAAGTCGGTT TCAGCCCCAT CGATCTTCTG CCCC GCGCTT 480
TCGGCCCCGAC CATCCTTGGC ATTCTCGTTG CCCTGACCAT TCCGGTCCTG CGCCGCATGG 540
AAAGCCGTCG TACTCTCAGA GGAGCCGTCT GATGCGCCGG CTTACCTTC TAGCAACAGC 600
CGCAGGACTC GCCCTTGCCT GTTCGCCGCT CATCGCTCAT GCACAGTTTG CTCCCGCAGG 660
GGCTGGCGGC GAACCTTCCT CGTCAGTTCC TGGGCCAGGA AATGCGAGCG AGCCCACCGA 720
AAACTCTCCG AAAAGTCAGA GCTACTTCGC AGGACCGTCG CCCTATGCCC CGCAGGCTCC 780
TGGCGTAAAC GCAGCCAACC TGCCGGACAT TGAGTCAATC GATCCCTCGC AGGTCCC GGC 840
CATGGCTCCG CAGCAGAGTG CCAATCCGGC ACGTGGAGAC TGGGTTGCTT ACGGACGTGA 900
CGATCATCAG ACGCGATACT CTCCGCTTTC GGAAATCACG CCTGAGAACG CAAGCAAGCT 960
CAAGGTCGCT TTCGTCTACC ACACGGGGAG TTATCCGCGT CCGGGACAGG TGAACAAATG 1020
GGCCGCCGAA ACCACGCCGA TCAAGGTTGG TGACGGTCTC TACACATGTT CCGCCATGAA 1080
CGACATCATC AAGCTGGATC CGGCTACGGG TAAGCAGATC TGGCGTCGGA ACGTGGATGT 1140
CAAATACCAC TCCATTCCCT ATACCGCTGC CTGTAAGGGT GTGACGTATT TCACGTCCTC 1200
CGTGGTGCCG GAAGGCCAGC CCTGCCACAA TCGCCTTATC GAAGGCACGC TGGATATGCG 1260
TCTGATTGCG GTTGACGCGG AGACAGGGGA TTTCTGCCCT AATTTCGGTC ATGGTGGTCA 1320
GGTCAACCTG ATGCAGGGTC TGGGTGAGTC TGTTCGGGC TTCGTCTCCA TGACGGCACC 1380
TCCACCGGTC ATCAACGGCG TCGTGGTTGT AAACCACGAA GTGCTCGACG GTCAGCGCCG 1440
CTGGGCTCCG TCCGGTGTGA TCCGTGGTTA CGATGCTGAA AGTGGCAAAT TCGTATGGGC 1500
CTGGGACGTC AACAAATCCG GACGATCACA GCCAGCCTAC CGGGTAACCG TCATTACAGC 1560
CGTGGAACGC CGAATTCCTG GGCTACCTGA CAGGCGACAA CGAGGAGGGT CTCGTTTACG 1620
TCCCGACAGG AACTCTGCTG CTGACTATTA CAGCGCCCTG CGTAGTGATG CTGAAAACAA 1680
GGTGTCTCC GCTGTTGTCG CCATTGACGT CAAGACGGGT TCTCCGCGCT GGGTCTTCCA 1740
GACGGCTCAT AAGGACGTCT GGGATTATGA CATCGGTTCA CAGGCGACCC TGATGGATAT 1800
GCCTGGCCCC GATGGCCAGA CGGTTCTGTC TCTCATCATG CCGACCAAGC GTGGCCAGAC 1860
GTTCTGTGCTT GACCGTCGTA CCGGCAAGCC AATTCTGCCG GTTGAAGAAC GCCCAGCTCC 1920
GTCCCCTGGT GTTATTCCGG GTGACCCGCG TTCTCCGACG CAGCCATGGT CCGTCGGGAT 1980
GCCGGCCCTT CGCGTGCCGG ATCTGAAAGA GACAGACATG TGGGGTATGT CCCCATCGA 2040

- 20 -

TCAGCTCTTC TGCCGTATCA AGTTCCGCCG TGCGAACATAT GTGGGTGAGT TCACACCACC 2100
 GAGCGTTGAC AAGCCGTGGA TTGAATATCC GGGCTATAAC GGTGGCAGTG ACTGGGGCTC 2160
 CATGTCCTAT GATCCGCAGT CCGGCATCCT GATTGCGAAC TGGAACATCA CACCGATGTA 2220
 CGACCAGCTC GTAACCCGCA AGAAGGCAGA CTCCCTCGGC CTGATGCCGA TCGATGACCC 2280
 CAACTTCAAG CCAGGTGGCG GTGGTGCCGA AGGTAACGGC GCCATGGACG GAACGCCTTA 2340
 CGGTATCGTC GTGACACCGT TCTGGGATCA GTACACGGGC ATGATGTGCA ACCGTCCGCC 2400
 CTACGGTATG ATCACAGCCA TCGACATGAA GCACGGCCAG AAGGTTCTGT GGCAGCATCC 2460
 GCTCGGAACG GCTCGCGCCA ACGGTCCATG GGGTCTGCCA ACAGGTCTGC CATGGGAAAT 2520
 CGGCACTCCG AACAAATGGTG GTTCGGTTGT GACCGGTGGC GGTCTGATCT TCATCGGTGC 2580
 GGCAACGGAT AACCAGATCC GCGCGATTGA TGAACACACT GGCAAGGTTG TCTGGAGCGC 2640
 AGTCCTCCCC GCGGGCGGTC AGGCCAATCC GATGACGTAT GAAGCCAATG GTCACCAGTA 2700
 CGTTGCCATC ATGGCTGGCG GTCATCACTT CATGATGACG CCAGTGTCTG ACCAGCTTGT 2760
 GGTTTACGCA CTGCCGGATG CCATCAAGCA GTAATTAAGT CCTGTGGCGG ATGTGTCATG 2820
 CATATCCGCC AACTCCATC GTCAGAAGGA GACTTTCGTG CTAGCCATGC AGGGAAGTCT 2880
 CCTTTTGACG TTTTGGCTC TTTCCAGCGA GCGGGCAGTC TGAAACGGGG CTTCGTCTGG 2940
 CTCGTACTTT CAGAATGGCT CGTCGCACCC TCATGACTGC CCACTCCCC GTTATCTTGC 3000
 AGGTTCTGCC AGCCCTCAGC ACGGGCGGCC TGGAGCGGGG AGCTATTGAA ATTGCGGCTG 3060
 CCATCACACA GGCTGGTGGC AAGGCCATTG TCGCTTCGAA GACGGGTCTT CTTCCTTGTG 3120
 AACTCCGCCA CGTCGGAGCA GTGCATGTGC CGCTGGATCT CAAATCGAAA TCGCCGTTTT 3180
 CTGTTCCGGC CCGTGCCCGT GAACTCCAGA AACTGATCCG GGAGCAGCAG GTTGATCTGG 3240
 TTCACGCCCG GTCCCGTATT CCGGCATGGG CCGCCTGGCT CGCCTGCCGC CGCGAGAACA 3300
 TTCCTTTTGT GACAACGTGG CATGGCGTCC ACGAGGCTGG CTGGTGGGGC AAGAAATTCT 3360
 ACAATTCCGGT GCTGGCCCCG GGTGCAAGGG TCATCGCAAT TTCGCACTAC ATTTCCGGGC 3420
 GTCTTTCAGG GCAGTACGGC GTTCAGGCAG ATCGTCTTCG AACCATTCG CGTGGTGCCG 3480

A

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 740 residues

- 21 -

(B) TYPE: amino acid
 (C) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (iii) ORIGINAL SOURCE:
 ORGANISM: Gluconobacter suboxydans
 STRAIN: IFO 3255
 (iv) FEATURE:
 FEATURE KEY: sig peptide
 POSITION: -24..-1
 SEQUENCING METHOD: E
 FEATURE KEY: mat peptide
 POSITION: 1..716
 SEQUENCING METHOD: E

Met Arg Arg Pro Tyr Leu Leu Ala Thr Ala Ala Gly Leu Ala Leu Ala
 -24 -20 -15 -10
 Cys Ser Pro Leu Ile Ala His Ala Gln Phe Ala Pro Ala Gly Ala Gly
 -5 1 5
 Gly Glu Pro Ser Ser Ser Val Pro Gly Pro Gly Asn Ala Ser Glu Pro
 10 15 20
 Thr Glu Asn Ser Pro Lys Ser Gln Ser Tyr Phe Ala Gly Pro Ser Pro
 25 30 35 40
 Tyr Ala Pro Gln Ala Pro Gly Val Asn Ala Ala Asn Leu Pro Asp Ile
 45 50 55
 Glu Ser Ile Asp Pro Ser Gln Val Pro Ala Met Ala Pro Gln Gln Ser
 60 65 70
 Ala Asn Pro Ala Arg Gly Asp Trp Val Ala Tyr Gly Arg Asp Asp His
 75 80 85
 Gln Thr Arg Tyr Ser Pro Leu Ser Glu Ile Thr Pro Glu Asn Ala Ser
 90 95 100
 Lys Leu Lys Val Ala Phe Val Tyr His Thr Gly Ser Tyr Pro Arg Pro
 105 110 115 120
 Gly Gln Val Asn Lys Trp Ala Ala Glu Thr Thr Pro Ile Lys Val Gly
 125 130 135
 Asp Gly Leu Tyr Thr Cys Ser Ala Met Asn Asp Ile Ile Lys Leu Asp
 140 145 150
 Pro Ala Thr Gly Lys Gln Ile Trp Arg Arg Asn Val Asp Val Lys Tyr
 155 160 165
 His Ser Ile Pro Tyr Thr Ala Ala Cys Lys Gly Val Thr Tyr Phe Thr
 170 175 180
 Ser Ser Val Val Pro Glu Gly Gln Pro Cys His Asn Arg Leu Ile Glu
 185 190 195 200
 Gly Thr Leu Asp Met Arg Leu Ile Ala Val Asp Ala Glu Thr Gly Asp
 205 210 215
 Phe Cys Pro Asn Phe Gly His Gly Gly Gln Val Asn Leu Met Gln Gly
 220 225 230
 Leu Gly Glu Ser Val Pro Gly Phe Val Ser Met Thr Ala Pro Pro Pro
 235 240 245
 Val Ile Asn Gly Val Val Val Val Asn His Glu Val Leu Asp Gly Gln
 250 255 260
 Arg Arg Trp Ala Pro Ser Gly Val Ile Arg Gly Tyr Asp Ala Glu Ser
 265 270 275 280
 Gly Lys Phe Val Trp Ala Trp Asp Val Asn Asn Ser Gly Arg Ser Gln
 285 290 295

- 22 -

Pro Ala Tyr Arg Val Thr Val Ile Thr Ala Val Glu Arg Arg Ile Pro
 300 305 310
 Gly Leu Pro Asp Arg Arg Gln Arg Gly Gly Ser Arg Leu Arg Pro Asp
 315 320 325
 Arg Asn Ser Ala Ala Asp Tyr Tyr Ser Ala Leu Arg Ser Asp Ala Glu
 330 335 340
 Asn Lys Val Ser Ser Ala Val Val Ala Ile Asp Val Lys Thr Gly Ser
 345 350 355 360
 Pro Arg Trp Val Phe Gln Thr Ala His Lys Asp Val Trp Asp Tyr Asp
 365 370 375
 Ile Gly Ser Gln Ala Thr Leu Met Asp Met Pro Gly Pro Asp Gly Gln
 380 385 390
 Thr Val Pro Ala Leu Ile Met Pro Thr Lys Arg Gly Gln Thr Phe Val
 395 400 405
 Leu Asp Arg Arg Thr Gly Lys Pro Ile Leu Pro Val Glu Glu Arg Pro
 410 415 420
 Ala Pro Ser Pro Gly Val Ile Pro Gly Asp Pro Arg Ser Pro Thr Gln
 425 430 435 440
 Pro Trp Ser Val Gly Met Pro Ala Leu Arg Val Pro Asp Leu Lys Glu
 445 450 455
 Thr Asp Met Trp Gly Met Ser Pro Ile Asp Gln Leu Phe Cys Arg Ile
 460 465 470
 Lys Phe Arg Arg Ala Asn Tyr Val Gly Glu Phe Thr Pro Pro Ser Val
 475 480 485
 Asp Lys Pro Trp Ile Glu Tyr Pro Gly Tyr Asn Gly Gly Ser Asp Trp
 490 495 500
 Gly Ser Met Ser Tyr Asp Pro Gln Ser Gly Ile Leu Ile Ala Asn Trp
 505 510 515 520
 Asn Ile Thr Pro Met Tyr Asp Gln Leu Val Thr Arg Lys Lys Ala Asp
 525 530 535
 Ser Leu Gly Leu Met Pro Ile Asp Asp Pro Asn Phe Lys Pro Gly Gly
 540 545 550
 Gly Gly Ala Glu Gly Asn Gly Ala Met Asp Gly Thr Pro Tyr Gly Ile
 555 560 565
 Val Val Thr Pro Phe Trp Asp Gln Tyr Thr Gly Met Met Cys Asn Arg
 570 575 580
 Pro Pro Tyr Gly Met Ile Thr Ala Ile Asp Met Lys His Gly Gln Lys
 585 590 595 600
 Val Leu Trp Gln His Pro Leu Gly Thr Ala Arg Ala Asn Gly Pro Trp
 605 610 615
 Gly Leu Pro Thr Gly Leu Pro Trp Glu Ile Gly Thr Pro Asn Asn Gly
 620 625 630
 Gly Ser Val Val Thr Gly Gly Gly Leu Ile Phe Ile Gly Ala Ala Thr
 635 640 645
 Asp Asn Gln Ile Arg Ala Ile Asp Glu His Thr Gly Lys Val Val Trp
 650 655 660
 Ser Ala Val Leu Pro Gly Gly Gly Gln Ala Asn Pro Met Thr Tyr Glu
 665 670 675 680
 Ala Asn Gly His Gln Tyr Val Ala Ile Met Ala Gly Gly His His Phe
 685 690 695
 Met Met Thr Pro Val Ser Asp Gln Leu Val Val Tyr Ala Leu Pro Asp
 700 705 710
 Ala Ile Lys Gln
 715

- 23 -

- (4) INFORMATION FOR SEQ ID NO:3:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 126 residues
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (iii) FRAGMENT TYPE: internal fragment
 - (iv) ORIGINAL SOURCE:
 - ORGANISM: Gluconobacter suboxydans
 - STRAIN: IFO 3255
 - (v) FEATURE:
 - FEATURE KEY: mat peptide
 - POSITION: 1..126
 - SEQUENCING METHOD: E

```

Met Pro Asn Leu Gln Gly Asn Arg Thr Leu Thr Glu Trp Leu Thr Leu
 1           5           10           15
Leu Leu Gly Val Ile Val Leu Leu Val Gly Leu Phe Phe Val Ile Gly
      20           25           30
Gly Ala Asp Leu Ala Met Leu Gly Gly Ser Thr Tyr Tyr Val Leu Cys
      35           40           45
Gly Ile Leu Leu Val Ala Ser Gly Val Phe Met Leu Met Gly Arg Thr
      50           55           60
Leu Gly Ala Phe Leu Tyr Leu Gly Ala Leu Ala Tyr Thr Trp Val Trp
      65           70           75           80
Ser Phe Trp Glu Val Gly Phe Ser Pro Ile Asp Leu Leu Pro Arg Ala
      85           90           95
Phe Gly Pro Thr Ile Leu Gly Ile Leu Val Ala Leu Thr Ile Pro Val
      100           105           110
Leu Arg Arg Met Glu Ser Arg Arg Thr Leu Arg Gly Ala Val
      115           120           125

```

- (5) INFORMATION FOR SEQ ID NO:4:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 residues
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (iii) FRAGMENT TYPE: internal fragment
 - (iv) ORIGINAL SOURCE:
 - ORGANISM: Gluconobacter suboxydans
 - STRAIN: IFO 3255
 - (v) FEATURE:
 - SEQUENCING METHOD: E

```

Lys Trp Ala Glu Glu Thr Xaa Pro
 1           5

```

- (6) INFORMATION FOR SEQ ID NO:5:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 residues

- 24 -

- (B) TYPE: amino acid
- (C) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) FRAGMENT TYPE: internal fragment
- (iv) ORIGINAL SOURCE:
ORGANISM: Gluconobacter suboxydans
STRAIN: IFO 3255
- (v) FEATURE:
SEQUENCING METHOD: E

Lys Ser Gln Ser Tyr Phe Ala Gly Pro Ser Gln Tyr Ala Pro Gln Ala
 1 5 10 15
 Pro Gly Val Asn Ala Xaa Asn Leu
 20

- (7) INFORMATION FOR SEQ ID NO:6:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 residues
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (iii) FRAGMENT TYPE: internal fragment
 - (iv) ORIGINAL SOURCE:
ORGANISM: Gluconobacter suboxydans
STRAIN: IFO 3255
 - (v) FEATURE:
SEQUENCING METHOD: E

Lys Val Leu Trp Gln His Pro Leu Gly Thr Ala Arg Xaa Asn Gly Pro
 1 5 10 15

- (8) INFORMATION FOR SEQ ID NO:7:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 residues
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (iii) FRAGMENT TYPE: N-terminal fragment
 - (iv) ORIGINAL SOURCE:
ORGANISM: Gluconobacter suboxydans
STRAIN: IFO 3255
 - (v) FEATURE:
SEQUENCING METHOD: E

Gln Phe Ala Pro Ala Gly Ala Gly Gly Glu Pro Ser Ser Ser Val Pro
 1 5 10 15
 Gly Pro Gly Asn Ala Ser Glu Pro Thr Glu Asn Ser Pro Lys
 20 25 30

- (9) INFORMATION FOR SEQ ID NO:8:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 residues
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide

- 25 -

- (iii) ORIGINAL SOURCE:
 ORGANISM: Gluconobacter suboxydans
 STRAIN: IFO 3255
- (iv) FEATURE:
 FEATURE KEY: sig peptide
 POSITION: 1..24
 SEQUENCING METHOD: E

```

Met Arg Arg Pro Tyr Leu Leu Ala Thr Ala Ala Gly Leu Ala Leu Ala
 1           5                   10                   15
Cys Ser Pro Leu Ile Ala His Ala
          20
  
```

(10) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 residues
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) FRAGMENT TYPE: C-terminal fragment
- (iv) ORIGINAL SOURCE:
 ORGANISM: Gluconobacter suboxydans
 STRAIN: IFO 3255
- (v) FEATURE:
 SEQUENCING METHOD: E

```

Pro Asp Ala Ile Lys Gln
 1           5
  
```

- 26 -

Claims:

1. A DNA comprising a nucleotide sequence which encodes a protein as defined by (a) or (b) and having sorbitol dehydrogenase activity:
 - 5 (a) a protein having the amino acid sequence from the position 1 to 716 of the sequence described in SEQ ID NO: 2, or
 - (b) a protein derived from the protein of (a) by substitution, deletion, insertion or addition of one or more amino acids in the amino acid sequence defined in (a).
- 10 2. A DNA according to claim 1, which encodes sorbitol dehydrogenase of a microorganism belonging to acetic acid bacteria.
3. A DNA according to claim 2, wherein the said microorganism belongs to the genus *Gluconobacter* or the genus *Acetobacter*.
4. A DNA according to any one of claims 1- 4, which is selected from the
15 group consisting of:
 - (e) a DNA comprising a nucleotide sequence from the position 644 to 2791 or from the position 572 to 2791 of the sequence represented by SEQ ID NO: 1, and
 - (f) a DNA which is capable of hybridizing to the DNA defined in (e) and
20 encodes the protein having the function of the protein defined in (a) of claim 1.
5. A combination of DNAs comprising;
 - (1) a DNA as defined in any one of claims 1 - 4; and
 - (2) a DNA encoding ORF2 having a nucleotide sequence from the position
192 to 569 of the sequence represented by SEQ ID NO: 1 and/or a DNA
25 encoding an ORF2 derivative with substitution, deletion, insertion or addition of one or more amino acids of the sequence defined in SEQ ID NO: 3 and with the function equivalent to the one of ORF2.
6. A combination of DNAs according to claim 5 comprising;

- 27 -

- (1) a DNA having a nucleotide sequence from the position 644 to 2791 or from the position 572 to 2791 of the sequence represented by SEQ ID NO: 1; or a DNA which is capable of hybridizing to this DNA which encodes the protein having the function of the protein defined claim 1; and
- 5 (2) a DNA encoding ORF2 having a nucleotide sequence from the position 192 to 569 of the sequence represented by SEQ ID NO: 1 or a DNA encoding an ORF2 derivative with substitution, deletion, insertion or addition of one or more amino acids of the sequence defined in SEQ ID NO: 3 with the function equivalent to one of ORF2.
- 10 7. An expression vector comprising a DNA as claimed in any one of claims 1 - 6.
8. A combination of DNAs according to claims 5 or 6, wherein the respective DNAs are separately carried on different expression vectors or on the same expression vector.
- 15 9. A combination of DNAs according to claims 5 or 6, which are in a tandem form as described in SEQ ID NO: 1.
10. The expression vector of claim 7, which is functional in a microorganism selected from those belonging to the genus *Gluconobacter*, *Acetobacter* or *E. coli*.
- 20 11. A recombinant organism which has been transformed by an expression vector as claimed in any one of claims 7 to 10.
12. A recombinant organism having the DNA as claimed in any one of claims 1 - 4 or the combination of the DNAs as defined in any one of claims 5, 6, 8 or 9 on a chromosomal DNA of the host organism.
- 25 13. The recombinant organism as claimed in claim 11 or 12, wherein the host organism is a microorganism selected from those belonging to the genus *Gluconobacter*, *Acetobacter* or *E. coli*.
14. A process for producing recombinant D-sorbitol dehydrogenase which comprises cultivating the recombinant organism as claimed in any one of
- 30 claims 11 - 13 in an appropriate medium and recovering the said recombinant D-sorbitol dehydrogenase from the culture.

- 28 -

15. A recombinant D-sorbitol dehydrogenase produced by the expression of the DNA as claimed in any one of claims 1 - 4 or a combination of DNAs as claimed in any one of claims 5, 6, 8 or 9.

16. A recombinant D-sorbitol dehydrogenase produced by the process claimed
5 in claim 14.

17. A recombinant D-sorbitol dehydrogenase according to claim 15 or 16, which is immobilized on a solid carrier for solid phase enzymatic reaction.

18. A process for producing L-sorbose which comprises converting D-sorbitol into L-sorbose with the aid of the recombinant D-sorbitol dehydrogenase
10 claimed in claim 16 or 17..

19. A process for producing L-sorbose which comprises converting D-sorbitol into L-sorbose by fermentation of the recombinant organism claimed in any one of claims 11 - 13 in an appropriate medium.

Peptide No.1 K W A E E T X P

s7 : 20mer

Peptide No.3 **K S Q S Y F A G P S Q Y A P Q A P G V N A X N L**

5' CAG TCG TAT TTC GCA GGA CC 3'
 AGC C G C
 T C T
 T

s6R : 20mer

Peptide No.8 **K V L W Q H P L G T A R X N G P**

3' GAG ACC GTC GTG GGG GAG CC 5'
 C A C C
 A A A

The amino acid sequence in boldface were used for synthesizing two oligonucleotide sequences (S7 and s6R) for PCR. Arrows show direction of DNA synthesis. The primers were degenerate DNA mixtures having bias for *Gluconobacter* codon usage.

Fig. 1

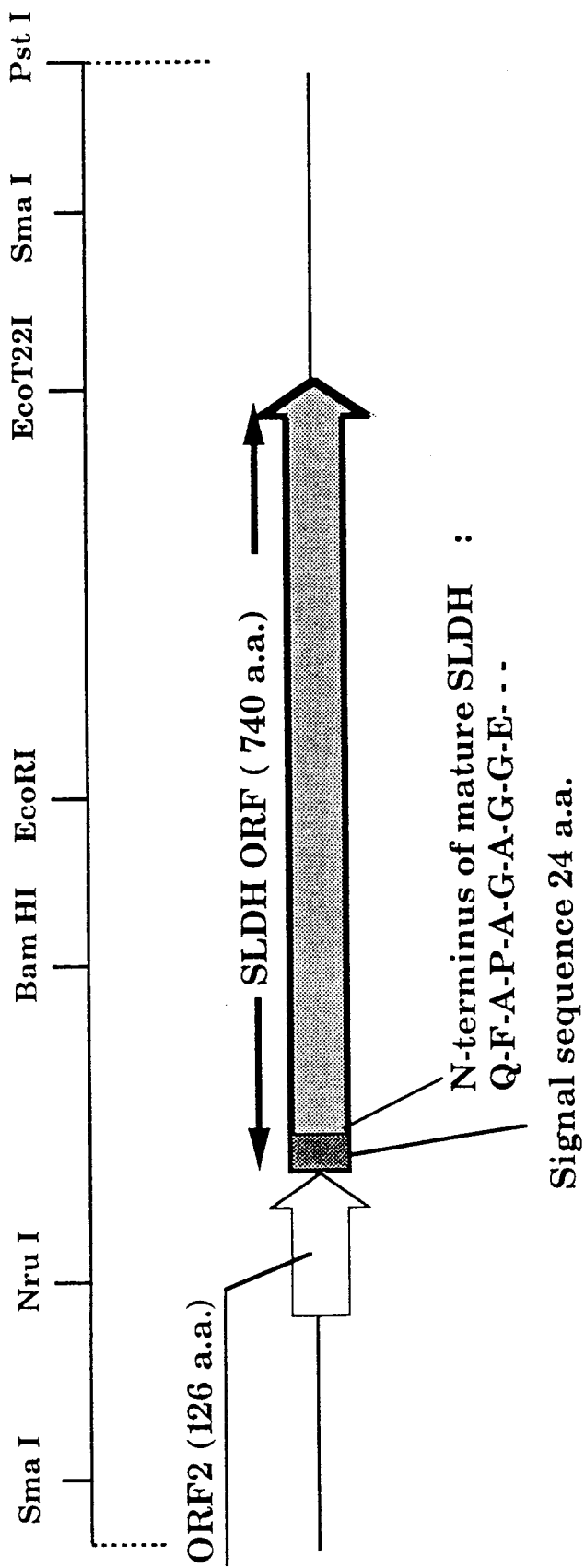


Fig. 2

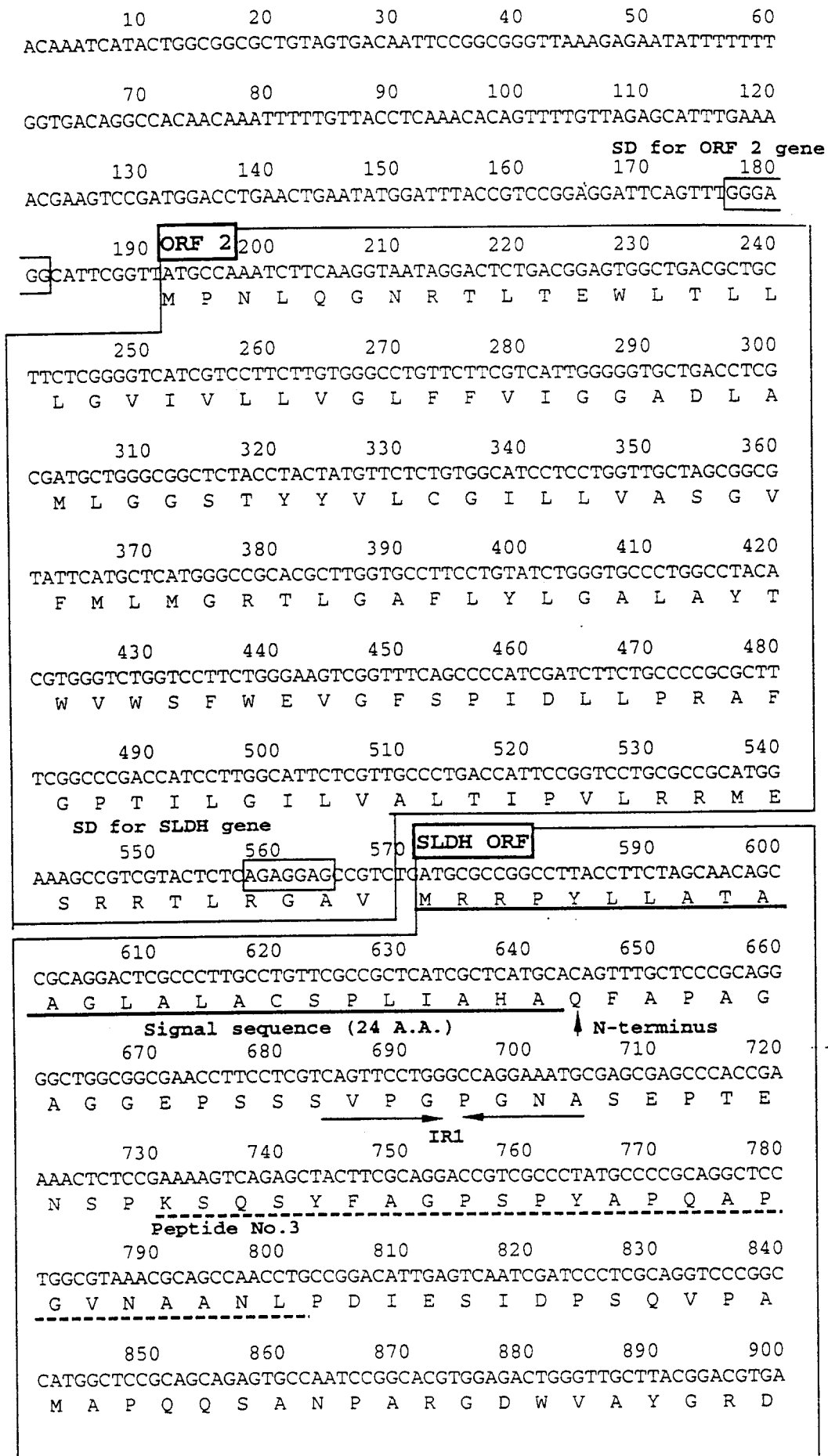


Fig. 3a

910	920	930	940	950	960
CGATCATCAGACGCGATACTCTCCGCTTTTCGGAAATCACGCCTGAGAACGCAAGCAAGCT					
D H Q T R Y S P L S E I T P E N A S K L					
970	980	990	1000	1010	1020
CAAGGTCGCTTTTCGTCTACCACACGGGGAGTTATCCGCGTCCGGGACAGGTGAACAAATG					
K V A F V Y H T G S Y P R P G Q V N K W					
				Peptide No.1	
1030	1040	1050	1060	1070	1080
GGCCGCCGAAACCACGCCGATCAAGGTTGGTGACGGTCTCTACACATGTTCCGCCATGAA					
A A E T T P I K V G D G L Y T C S A M N					

1090	1100	1110	1120	1130	1140
CGACATCATCAAGCTGGATCCGGCTACGGGTAAGCAGATCTGGCGTCGGAACGTGGATGT					
D I I K L D P A T G K Q I W R R N V D V					
1150	1160	1170	1180	1190	1200
CAAATACCACTCCATTCCCTATAACCGCTGCCTGTAAGGGTGTGACGTATTTACGTCCTC					
K Y H S I P Y T A A C K G V T Y F T S S					
1210	1220	1230	1240	1250	1260
CGTGGTGCCGGAAGGCCAGCCCTGCCACAATCGCCTTATCGAAGGCACGCTGGATATGCC					
V V P E G Q P C H N R L I E G T L D M R					
1270	1280	1290	1300	1310	1320
TCTGATTGCGGTTGACGCGGAGACAGGGGATTTCTGCCCTAATTTTCGGTCATGGTGGTCA					
L I A V D A E T G D F C P N F G H G G Q					
1330	1340	1350	1360	1370	1380
GGTCAACCTGATGCAGGGTCTGGGTGAGTCTGTTCCGGGCTTCGTCTCCATGACGGCACC					
V N L M Q G L G E S V P G F V S M T A P					
1390	1400	1410	1420	1430	1440
TCCACCGTCAACGGCGTCTGGTGTAAACCACGAAGTGCTCGACGGTCAGCGCCG					
P P V I N G V V V V N H E V L D G Q R R					
1450	1460	1470	1480	1490	1500
CTGGGCTCCGTCGGTGTGATCCGTTACGATGCTGAAAGTGGCAAATTCGTATGGGC					
W A P S G V I R G Y D A E S G K F V W A					
1510	1520	1530	1540	1550	1560
CTGGGACGTCAACAATTCGGACGATCACAGCCAGCCTACCGGGTAACCGTCATTACAGC					
W D V N N S G R S Q P A Y R V T V I T A					
1570	1580	1590	1600	1610	1620
CGTGAACGCCGAATTCCTGGGCTACCTGACAGGCGACAACGAGGAGGGTCTCGTTTACG					
V E R R I P G L P D R R Q R G G S R L R					
1630	1640	1650	1660	1670	1680
TCCCGACAGGAACCTCTGCTGCTGACTATTACAGCGCCCTGCGTAGTGATGCTGAAAACAA					
P D R N S A A D Y Y S A L R S D A E N K					
1690	1700	1710	1720	1730	1740
GGTGTCTCCGCTGTTGTGCGCCATTGACGTCAAGACGGGTTCTCCGCGCTGGGTCTTCCA					
V S S A V V A I D V K T G S P R W V F Q					
1750	1760	1770	1780	1790	1800
GACGGCTCATAAGGACGTCTGGGATTATGACATCGGTTACAGGCGACCCTGATGGATAT					
T A H K D V W D Y D I G S Q A T L M D M					

Fig. 3b

1810 1820 1830 1840 1850 1860
 GCCTGGCCCCGGATGGCCAGACGGTTCCTGCTCTCATCATGCCGACCAAGCGTGGCCAGAC
 P G P D G Q T V P A L I M P T K R G Q T

1870 1880 1890 1900 1910 1920
 GTTCGTGCTTGACCGTCGTACCGGCAAGCCAATTCTGCCGGTTGAAGAACGCCAGCTCC
 F V L D R R T G K P I L P V E E R P A P

1930 1940 1950 1960 1970 1980
 GTCCCCTGGTGTATTCCGGGTGACCCGCGTTCTCCGACGCAGCCATGGTCCGTCGGGAT
 S P G V I P G D P R S P T Q P W S V G M

1990 2000 2010 2020 2030 2040
 GCCGGCCCTTCGCGTGCCGGATCTGAAAGAGACAGACATGTGGGGTATGTCCCCCATCGA
 P A L R V P D L K E T D M W G M S P I D

2050 2060 2070 2080 2090 2100
 TCAGCTCTTCTGCCGTATCAAGTTCGCCGTGCGAACTATGTGGGTGAGTTCACACCACC
 Q L F C R I K F R R A N Y V G E F T P P

2110 2120 2130 2140 2150 2160
 GAGCGTTGACAAGCCGTGGATTGAATATCCGGGCTATAACGGTGGCAGTGA CTGGGGCTC
 S V D K P W I E Y P G Y N G G S D W G S

2170 2180 2190 2200 2210 2220
 CATGTCCTATGATCCGCAGTCCGGCATCCTGATTGCGAACTGGAACATCACACCGATGTA
 M S Y D P Q S G I L I A N W N I T P M Y

2230 2240 2250 2260 2270 2280
 CGACCAGCTCGTAACCCGCAAGAAGGCAGACTCCCTCGGCCTGATGCCGATCGATGACCC
 D Q L V T R K K A D S L G L M P I D D P

2290 2300 2310 2320 2330 2340
 CAACTTCAAGCCAGGTGGCGGTGGTGGCCGAAGGTAACGGCGCCATGGACGGAACGCCTTA
 N F K P G G G G A E G N G A M D G T P Y

2350 2360 2370 2380 2390 2400
 CGGTATCGTCTGACACCGTTCCTGGGATCAGTACACGGGCATGATGTGCAACCGTCCGCC
 G I V V T P F W D Q Y T G M M C N R P P

2410 2420 2430 2440 2450 2460
 CTACGGTATGATCACAGCCATCGACATGAAGCACGGCCAGAAGTTCTGTGGCAGCATCC
 Y G M I T A I D M K H G Q K V L W Q H P

Peptide No. 8

2470 2480 2490 2500 2510 2520
 GCTCGGAACGGCTCGCGCCAACGGTCCATGGGGTCTGCCAACAGGTCTGCCATGGGAAAT
 L G T A R A N G P W G L P T G L P W E I

2530 2540 2550 2560 2570 2580
 CGGCACTCCGAACAATGGTGGTTCGGTTGTGACCGGTGGCGGTCTGATCTTCATCGGTGC
 G T P N N G G S V V T G G G L I F I G A

2590 2600 2610 2620 2630 2640
 GGCAACGGATAACCAGATCCGCGGATTGATGAACACACTGGCAAGGTTGTCTGGAGCGC
 A T D N Q I R A I D E H T G K V V W S A

2650 2660 2670 2680 2690 2700
 AGTCCTCCCCGGCGGGTCCAGGCCAATCCGATGACGTATGAAGCCAATGGTCACCGAGTA
 V L P G G G Q A N P M T Y E A N G H Q Y

Fig. 3c

```

2710      2720      2730      2740      2750      2760
CGTTGCCATCATGGCTGGCGGTCATCACTTCATGATGACGCCAGTGTCTGACCAGCTTGT
V A I M A G G H H F M M T P V S D Q L V

2770      2780      2790      2800      2810      2820
GGTTTACGCACTGCCGGATGCCATCAAGCAGTAATTAAGTCCTGTGGCGGATGTGTCATG
V Y A L P D A I K Q *
                                     IR2

2830      2840      2850      2860      2870      2880
CATATCCGCCCACTCCATCGTCAGAAGGAGACTTTCGTGCTAGCCATGCAGGGAAGTCT
←----- IR2 -----→

2890      2900      2910      2920      2930      2940
CCTTTTGACGTTTTTTGGCTCTTTCCAGCGAGCGGGCAGTCTGAAACGGGGCTTCGTCTGG
IR2'

2950      2960      2970      2980      2990      3000
CTCGTACTTTCAGAATGGCTCGTCGCACCCTCATGACTGCCCACTCCCCCGTTATCTTGC

3010      3020      3030      3040      3050      3060
AGGTTCTGCCAGCCCTCAGCACGGGCGGCTGGAGCGGGGAGCTATTGAAATTGCGGCTG

3070      3080      3090      3100      3110      3120
CCATCACACAGGCTGGTGGCAAGGCCATTGTCGCTTTCGAAGACGGGTCTCTTTTGTGC

3130      3140      3150      3160      3170      3180
AACTCCGCCACGTCGGAGCAGTGCATGTGCCGCTGGATCTCAAATCGAAATCGCCGTTTT

3190      3200      3210      3220      3230      3240
CTGTTGGGCGCCGTGCCCGTGAAGTCCAGAACTGATCCGGGAGCAGCAGGTTGATCTGG

3250      3260      3270      3280      3290      3300
TTCACGCCCCGGTCCCGTATTCCGGCATGGGCCCGCCTGGCTCGCCTGCCCGCGAGAACA

3310      3320      3330      3340      3350      3360
TTCCTTTTCGTGACAACGTGGCATGGCGTCCACGAGGCTGGCTGGTGGGGCAAGAAATTCT

3370      3380      3390      3400      3410      3420
ACAATTCGGTGTGTCGGCCCGGGTGCAAGGGTCATCGCAATTTTCGCACTACATTTCCGGGC

3430      3440      3450      3460      3470      3480
GTCTTTCAGGGCAGTACGGCGTTCAGGCAGATCGTCTTCGAACCATTCCGCGTGGTGCCG

```

3481

A

→ : IR1 and IR2 (or 2') are inverted repeats as possible transcription terminators for ORF2 and SLDH II genes, respectively.

Fig. 3d

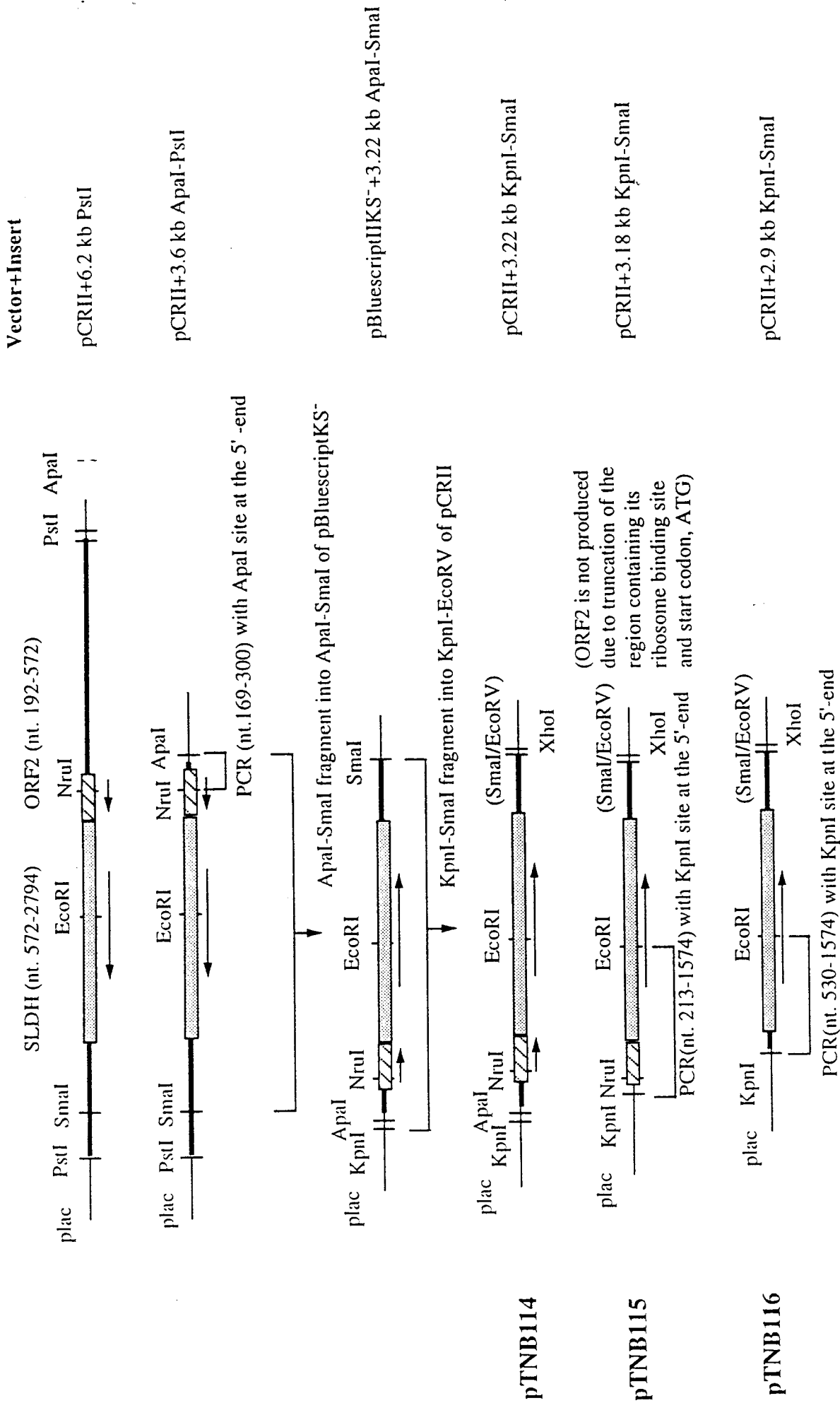


Fig. 4

Vector + Insert

pUC18 + 3.5 kb HindIII-XhoI

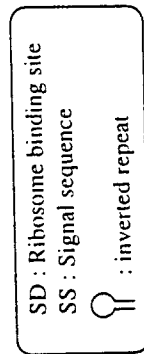
*1 Promoter of Enzyme A gene (pA) of DSM4025 was cloned from DNA fragment carrying the said gene e.g. chromosomal DNA of DSM 4025 itself or pSSA202 (T. Hoshino et al., EP application No. 9611500. 8) into HindIII-BamHI of pUC18 to produce pUCAP by using PCR with a set of primers:
 5'-CCAAAGCTTATCGCCGGCTTTGTCA T-3'
 5'-CCGGATCCGGTTGGTGCAGCGGTCA-3'

pUC18 + 3.2 kbHindIII-XhoI

*2 From promoter to signal sequence of Enzyme A gene of DSM4025 was subcloned from pSSA202 by using PCR with a set of primers:
 5'-CCAAAGCTTATCGCCGGCTTTGTCA T-3'
 5'-GCTAGCAAAGGGGGTTCGGG-3'

pUC57 + 0.9 kb Sall

*3 PCR was carried out against pTNB110 with a set of primers:
 5'-GTCGACATCGCCGGCTTTGTCA-3'
 5'-GCCGGGCACTAGACGGGCTCC-3'



pUC57/pCRII + 0.9 kb Sall and 3.2 kb HindIII-XhoI

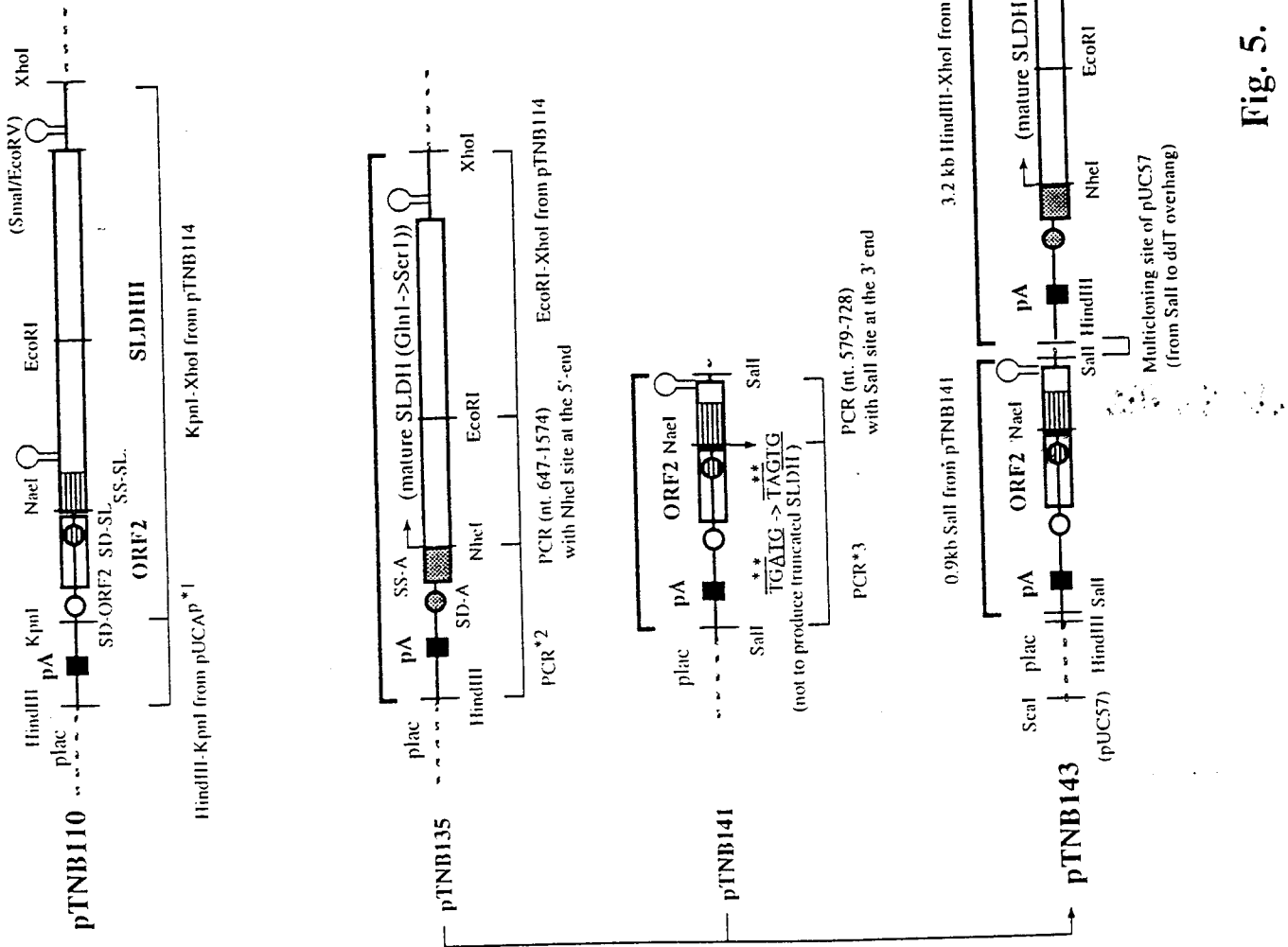


Fig. 5.

