

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 October 2007 (25.10.2007)

PCT

(10) International Publication Number  
**WO 2007/119575 A1**

(51) International Patent Classification:  
*C12N 9/88* (2006.01) *C12P 13/06* (2006.01)

(21) International Application Number:  
PCT/JP2007/056756

(22) International Filing Date: 22 March 2007 (22.03.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
2006109216 24 March 2006 (24.03.2006) RU

(71) Applicant (for all designated States except US): **AJINOMOTO CO., INC.** [JP/JP]; 15-1, Kyobashi 1-chome, Chuo-ku, Tokyo, 1048315 (JP).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SMIRNOV, Sergey, Vasilievich** [RU/RU]; bldg. 2, #94, 29, Yartsevskaya pr., Moscow, 121552 (RU). **SAMSONOVA, Natalia, Nikolaevna** [RU/RU]; bldg. 1, #47, 31, Sevastopolskiy pr., Moscow, 117186 (RU). **KODERA, Tomohiro** [JP/JP]; c/o AJINOMOTO CO., INC., 1-1, Suzuki-cho, Kawasaki-ku, Kawasaki-shi, Kanagawa, 2108681 (JP). **MATSUI, Kazuhiko** [JP/JP]; c/o AJINOMOTO CO., INC., 1-1, Suzuki-cho, Kawasaki-ku, Kawasaki-shi, Kanagawa, 2108681 (JP).

(74) Agents: **KAWAGUCHI, Yoshiyuki** et al.; Acropolis 21 Building 6th floor, 4-10, Higashi Nihonbashi 3-chome, Chuo-ku, Tokyo, 1030004 (JP).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL ALDOLASE AND PRODUCTION PROCESS OF 4-HYDROXY-L-ISOLEUCINE

(57) Abstract: A novel aldolase is described. 4-hydroxy-3-methyl-2-keto-pentanoic acid, which is useful as an intermediate in the synthesis of 4-hydroxy-L-isoleucine, may be synthesized from acetaldehyde and  $\alpha$ -ketobutyric acid using a novel aldolase, which is derived from the genus *Arthrobacter*.

WO 2007/119575 A1

## DESCRIPTION

Novel aldolase and production process of 4-hydroxy-L-isoleucine

5

### TECHNICAL FIELD

The present invention relates to the microbiological industry, and specifically to a novel aldolase and methods for manufacturing 4-hydroxy-L-isoleucine or a salt thereof.

### BACKGROUND ART

- 10 4-hydroxy-L-isoleucine is an amino acid which can be extracted and purified from fenugreek seeds (*Trigonella foenum-graecum* L. *leguminosae*). 4-hydroxy-L-isoleucine displays an insulintropic activity of great interest because its stimulating effect is clearly dependent on the plasma glucose concentration in the medium, as demonstrated both in isolated perfused rat pancreas and human pancreatic islets (Sauvaire, Y. et al, Diabetes, 15 47: 206–210, (1998)). Such dependency on glucose has not been confirmed with sulfonylureas (Drucker, D. J., Diabetes 47: 159–169, (1998)), which are the only insulintropic drug currently used to treat type II diabetes [or non-insulin-dependent diabetes (NIDD) mellitus (NIDDM)]. As a consequence, hypoglycemia remains a common undesirable side effect of sulfonylurea treatment (Jackson, J., and Bessler, R. 20 Drugs, 22: 211–245; 295–320, (1981); Jennings, A. et al. Diabetes Care, 12: 203–208, (1989)). Methods for improving glucose tolerance (Am. J. Physiol. Endocrinol., Vol. 287, E463-E471, 2004) are also known. This glucometabolism enhancement activity, and its potential application to pharmaceuticals and health foods, has been reported (Japanese Patent Application Laid-Open No. Hei 6-157302).
- 25 Due to its particular insulintropic action, 4-hydroxy-L-isoleucine, which is only found in plants, might be considered as a novel secretagogue for the treatment of type II diabetes. This is because Type II diabetes is characterized by defective insulin secretion associated with various degrees of insulin resistance (Broca, C. et al, Am. J. Physiol. 277 (Endocrinol. Metab. 40): E617–E623, (1999)).
- 30 A method of oxidizing iron, ascorbic acid, 2-oxyglutaric acid, and oxygen-dependent isoleucine by utilizing dioxygenase activity in fenugreek extract has been reported as a method for manufacturing 4-hydroxy-L-isoleucine (Phytochemistry, Vol.

44, No. 4, pp. 563-566, 1997). However, this method is unsatisfactory as a method of manufacturing 4-hydroxy-L-isoleucine because the activity of the enzyme is inhibited by the substrate at isoleucine concentrations of 20 mM and above, the enzyme has not been identified, the enzyme is derived from plant extracts and cannot be readily obtained in large quantities, and the enzyme is unstable.

An efficient eight-step synthesis of optically pure (2S,3R,4S)-4-hydroxyisoleucine with 39% overall yield has been disclosed. The key steps of this synthesis involve the biotransformation of ethyl 2-methylacetoacetate to ethyl (2S,3S)-2-methyl-3-hydroxy-butanoate with *Geotrichum candidum* and an asymmetric Strecker synthesis (Wang, Q. et al, Eur. J. Org. Chem., 834-839 (2002)).

A short six-step chemoenzymatic synthesis of (2S,3R,4S)-4-hydroxyisoleucine with total control of stereochemistry, the last step being the enzymatic resolution by hydrolysis of a N-phenylacetyl lactone derivative using the commercially available penicillin acylase G immobilized on Eupergit C(E-PAC), has also been disclosed (Rolland-Fulcrand, V. et al, J. Org. Chem., 873-877 (2004)).

But currently, there have been no reports of producing 4-hydroxy-L-isoleucine by enzymatic transamination of 4-hydroxy-3-methyl-2-keto-pentanoic acid or by any other enzymatic conversion from any other starting materials.

## DISCLOSURE OF THE INVENTION

As a result of extensive research conducted in consideration of the aforementioned problems, the inventors of the present invention have found an aldolase that may be preferably used in the synthesis of the desired precursor of 4-hydroxy-L-isoleucine (free form and a salt form thereof, and may be referred to as "4HIL"; hereinafter the same). This precursor is 4-hydroxy-3-methyl-2-keto-pentanoic acid (free form and a salt form thereof, and may be referred to as "HMKP"; hereinafter the same), and is present in certain microbial species. It has also been found that 4-hydroxy-L-isoleucine is produced by combining the aldolase enzyme with an aminotransferase and/or dehydrogenase, thereby leading to completion of the present invention.

Objects of the present invention include providing a novel aldolase and DNA encoding the aldolase, and a method for producing 4-hydroxy-L-isoleucine using the aldolase. The above objects were achieved by finding the novel aldolase of the present

invention.

Therefore, it is an object of the present invention to provide a DNA selected from the group consisting of:

- (a) A DNA comprising the nucleotide sequence of SEQ ID No: 1;
- 5 (b) A DNA that hybridizes under stringent conditions with a DNA having a nucleotide sequence complementary to the nucleotide sequence of SEQ ID No: 1 and encodes a protein having aldolase activity;
- (c) A DNA that encodes a protein comprising the amino acid sequence of SEQ ID No: 2;
- 10 (d) A DNA that encodes a protein having the amino acid sequence of SEQ ID NO: 2, except that it contains a substitution, deletion, insertion, addition, or inversion of one or several amino acid residues and wherein the protein has aldolase activity; and
- (e) A DNA that encodes a protein having an amino acid sequence that is at least 70% homologous to the amino acid sequence of SEQ ID NO: 2 and wherein the protein has
- 15 aldolase activity.

It is a further object of the present invention to provide a recombinant DNA comprising the DNA as described above and a vector DNA.

It a further object of the present invention to provide an isolated cell transformed with the recombinant DNA as described above.

- 20 It is still a further object of the present invention to provide a process for producing a protein having aldolase activity comprising cultivating the cell as described above in a medium, and accumulating a protein having aldolase activity in the medium, cells, or both.

Another object of the invention is to provide a protein selected from the group consisting of:

- 25 (f) a protein comprising the amino acid sequence of SEQ ID No: 2;
- (g) a protein comprising the amino acid sequence of SEQ ID No: 2, except that it contains a substitution, deletion, insertion, addition, or inversion of one or several amino acid residues and wherein the protein has aldolase activity; and
- (h) a protein that is at least 70% homologous to the amino acid sequence of SEQ ID
- 30 No: 2 and has aldolase activity.

It is a further object of the invention to provide a protein comprising:

- (A) an activity that catalyze production of 4-hydroxy-3-methyl-2-keto-pentanoic acid

from acetaldehyde and  $\alpha$ -ketobutyric acid;

(B) an activity that is dependent on bivalent cations selected from the group consisting of  $\text{Zn}^{2+}$ ,  $\text{Mn}^{2+}$  and  $\text{Mg}^{2+}$ ;

(C) a molecular weight of about 186kDa as measured by gel filtration, and

5 (D) a molecular weight per subunit of 27kDa as measured by SDS-PAGE.

It is another object of the invention to provide a method for manufacturing 4-hydroxy-L-isoleucine or a salt thereof, comprising:

(1) reacting acetaldehyde and  $\alpha$ -ketobutyric acid in an aqueous solvent in the presence of at least one aldolase selected from the group consisting of:

10 (a) a protein comprising the amino acid sequence of SEQ ID No: 2,

(b) a protein comprising the amino acid sequence of SEQ ID No: 2, except that it contains a substitution, deletion, insertion, addition or inversion of one or several amino acid residues 2 and wherein the protein has aldolase activity, and

(c) a protein that is at least 70% homologous to the amino acid sequence of SEQ ID  
15 No: 2 and has aldolase activity;

(2) converting the produced 4-hydroxy-3-methyl-2-keto-pentanoic acid to 4-hydroxy-L-isoleucine; and

(3) isolating the produced 4-hydroxy-L-isoleucine.

It is an object of the invention to provide the method as described above, wherein  
20 the 4-hydroxy-3-methyl-2-keto-pentanoic acid is converted to 4-hydroxy-L-isoleucine by an aminotransferase and/or a dehydrogenase, wherein the aminotransferase and/or dehydrogenase have a transamination activity in the presence of an amino group donor.

It is an object of the present invention to provide a method for manufacturing 4-hydroxy-L-isoleucine or a salt thereof, comprising:

25 (1) reacting acetaldehyde and  $\alpha$ -ketobutyric acid in an aqueous solvent in the presence of a bacterium containing an aldolase selected from the group consisting of:

(a) a protein comprising the amino acid sequence of SEQ ID No: 2,

(b) a protein comprising the amino acid sequence of SEQ ID No: 2, except that it contains a substitution, deletion, insertion, addition or inversion of one or several amino  
30 acid residues and wherein the protein has aldolase activity, and

(c) a protein that is at least 70% homologous to the amino acid sequence of SEQ ID No: 2 and has aldolase activity,

and wherein the aqueous solvent also contains an aminotransferase and/or a dehydrogenase which catalyzes conversion of 4-hydroxy-3-methyl-2-keto-pentanoic acid to 4-hydroxy-L-isoleucine in the presence of an amino group donor; and

(2) isolating the produced 4-hydroxy-L-isoleucine.

5 It is an object of the invention to provide the method as described above, wherein the aminotransferase is a branched-chain amino acid aminotransferase.

It is an object of the invention to provide the method as described above, wherein the branched-chain amino acid aminotransferase is derived from a bacterium selected from the group consisting of *Escherichia* and *Bacillus*.

10 It is an object of the invention to provide the method as described above, wherein the amino group donor is selected from the group of L-amino acids.

It is an object of the invention to provide the method as described above, wherein the branched-chain amino acid aminotransferase is derived from a bacterium selected from the group consisting of *Escherichia* and *Bacillus*.

15 It is an object of the invention to provide the method as described above, wherein the bacterium has been modified to enhance at least one of the activities of the aldolase and the branched-chain amino acid aminotransferase.

It is an object of the invention to provide the method as described above, wherein the activities of the aldolase and the branched-chain amino acid aminotransferase are  
20 enhanced by increasing the expression of the aldolase and/or the branched-chain amino acid aminotransferase.

It is an object of the invention to provide the method as described above, wherein the expression of the aldolase and/or the branched-chain amino acid aminotransferase is increased by modifying an expression control sequence of the gene encoding the aldolase  
25 and/or branched-chain amino acid aminotransferase, or by increasing the copy number of the gene encoding the aldolase and/or branched-chain amino acid aminotransferase.

It is an object of the invention to provide the method as described above, wherein the bacterium belongs to the genus *Escherichia*, *Pseudomonas*, *Corynebacterium*, *Arthrobacter*, *Aspergillus* or *Bacillus*.

30 It is an object of the invention to provide the method as described above, wherein the bacterium belongs to *Escherichia coli*, *Arthrobacter simplex*, *Corynebacterium glutamicum*, *Arthrobacter globiformis*, *Arthrobacter sulfureus*, *Arthrobacter viscosus*, or

*Bacillus subtilis*.

It is an object of the invention to provide the method as described above, wherein the bacterium is a bacterial culture, cells, or treated cells.

It is an object of the invention to provide the method as described above, wherein  
5 the 4-hydroxy-L-isoleucine is selected from the group consisting of  
(2S,3S,4S)-4-hydroxyisoleucine, (2S,3R,4R)-4-hydroxyisoleucine,  
(2S,3S,4R)-4-hydroxyisoleucine, (2S,3R,4S)-4-hydroxyisoleucine, and combinations  
thereof.

10

### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a flowchart of the process for producing aldolase.

Fig. 2 (photograph) shows purification of HMKP-aldolase from *Arthrobacter simplex* AKU626 (NBRC 12069) (hereinafter abbreviated as asiHPAL). Lanes: 1, 4 –  
protein molecular weight marker (Fermentas, Lithuania). 2 – Crude cell lysate (45 µg  
15 applied), 3 – asiHPAL preparation (0.5 µg applied).

Fig. 3 shows determination of the asiHPAL oligomeric structure. A –  
Determination of the Mw of the asiHPAL monomer using a calibrated SDS – PAGE gel.  
PageRuler™ Protein Ladder (Fermentas, Lithuania) was used as protein markers.  
Experimental data (black circle) were fitted (black line) by linear regression analysis using  
Sigma Plot 8 software. B - Determination of native Mw of the asiHPAL using SEC on the  
20 Superdex™ 200 HR 10/30A (Pharmacia) columns calibrated by Molecular Weight Protein  
Markers (Sigma). Experimental data (black circle) were fitted (black line) by linear  
regression analysis using Sigma Plot 8 software.

Fig. 4 shows structures of primers for PCR-amplification and sequencing of  
25 *asiHPAL* gene. A) – N-terminal sequence of HMKP aldolase from *Arthrobacter simplex*  
(NBRC 12069). B) – degenerative primer asiN10; I = deoxyinosine phosphate. C) –  
flanking primer SVS88.

Fig. 5 (photograph) shows amplification of a 0.9 kb DNA fragment of *Arthrobacter simplex* chromosome containing the gene encoding HMKP-aldolase. Electrophoresis of  
30 amplified DNA fragment. Tracks: 1 – DNA size marker; 3 – aliquot of reaction mixture B; 2  
– resulting 0.9 kb DNA fragment digested by *NcoI* restrictase.

Fig. 6 shows the strategy for sequencing the amplified DNA fragment.

Fig. 7 shows the alignment of asiHPAL and HHDE aldolase from *Brevibacterium linens* BL2.

### DESCRIPTION OF THE PREFERRED EMBODIMENTS

5 In the present specification, the term "4-hydroxy-L-isoleucine" or "4HIL" refers to a single compound or a diastereomer mixture of (2S,3S,4S)-4-hydroxyisoleucine, (2S,3R,4R)-4-hydroxyisoleucine, (2S,3S,4R)-4-hydroxy-isoleucine, and/or (2S,3R,4S)-4-hydroxyisoleucine, or any combination thereof.

The term "bacterium" as employed in the present specification includes an  
10 enzyme-producing bacteria, a mutant and a genetic recombinant thereof in which the targeted enzymatic activity exists or has been enhanced, and the like.

The following provides a detailed explanation of [I] aldolase, [II] a process for producing 4-hydroxy-L-isoleucine using the aldolase of the present invention with reference to the accompanying drawings.

15 [I] Aldolase

According to research by the inventors of the present invention, bacterial strains were confirmed to contain aldolase having the ability to form 4-hydroxy-3-methyl-2-keto-pentanoic acid in the genus *Arthrobacter*.

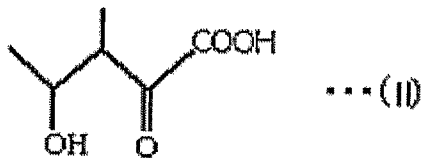
The inventors of the present invention found that *Arthrobacter simplex* possesses  
20 aldolase activity which can catalyze a reaction in which 4-hydroxy-3-methyl-2-keto-pentanoic acid is generated from acetaldehyde and  $\alpha$ -ketobutyric acid (both the free form and a salt form thereof; hereinafter the same). The inventors of the present invention purified and isolated the novel aldolase from cultivated microbial cells.

25 Furthermore, the inventors of the present invention determined the amino acid sequence of the aldolase by purifying the native aldolase from *Arthrobacter simplex* AKU626 (NBRC12069) strain (hereinafter abbreviated as asiHPAL). Furthermore, the inventors of the present invention also synthesized a DNA molecule of about 30 base pairs deduced from the amino acid sequence of the aldolase, and isolated and obtained the  
30 entire length of DNA that encodes asiHPAL by PCR using this DNA molecule.

The DNA encoding the asiHPAL is shown in SEQ ID No: 1. Furthermore, the amino acid sequence of asiHPAL encoded by the nucleotide sequence of SEQ ID No: 1 is



shown in SEQ ID No: 2. SEQ ID No: 2 is the amino acid sequence of asiHPAL encoded by the nucleotide sequence of SEQ ID No: 1. AsiHPAL of SEQ ID No: 2 possesses the aldolase activity, and catalyzes the reaction in which 4-hydroxy-3-methyl-2-keto-pentanoic acid (HMKP) shown in the following formula (II) is synthesized from one molecule of  
 5    acetaldehyde and one molecule of  $\alpha$ -ketobutyric acid.



Next, a detailed explanation is provided of (1) the DNA encoding aldolase, (2) the properties of aldolase and (3) the process for producing aldolase, in that order.

(1)     The DNA Encoding Aldolase

10        The aldolase gene of the present invention having the nucleotide sequence of SEQ ID No: 1 was isolated from the chromosomal DNA of *Arthrobacter simplex* strain AKU626 (NBRC12069) as described in Example section. The amino acid sequence of SEQ ID No: 2 demonstrates homology of 55% with known HHDE-aldolase (2, 4-dihydroxyhept-2-ene-1,7-dioic acid aldolase) (Submitted (08-APR-2005) National Center  
 15    for Biotechnology Information, NIH, Bethesda, MD 20894, USA) derived from *Brevibacterium linens* BL2 (Fig. 7).

Homology was calculated by using the "Vector NTI" genetic analytical software (informax®, Invitrogen Life Science Software).

20        The following provides an explanation of the method for obtaining the DNA encoding aldolase from aldolase-producing bacteria.

First, a partial amino acid sequence of the purified aldolase is determined. At this time, a partial amino acid sequence (SEQ ID No: 3) of the aldolase derived from *Arthrobacter simplex* strain AKU626 (NBRC12069) was determined by using the automated degradation method of Edman (Edman, P., Acta Chem. Scand. 4, 227 (1950))  
 25    using the 491cLC Protein Sequencer (Applied Biosystems, USA)

The nucleotide sequence of a DNA that encodes this amino acid sequence is then able to be deduced based on the amino acid sequence. Universal codons are employed to deduce the nucleotide sequence of the DNA.

A DNA molecule of about 30 base pairs was then synthesized based on the

deduced nucleotide sequence. The method used to synthesize the DNA molecule is disclosed in Tetrahedron Letters, 22, 1859 (1981). Furthermore, the DNA molecule may also be synthesized using the Synthesizer made by Applied Biosystems. The DNA molecule may be used as a probe to isolate the entire length of DNA that encodes aldolase from the chromosomal DNA library of a microorganism that produces aldolase. Alternatively, the DNA may also be used as a primer to amplify the DNA that encodes aldolase by PCR. However, since the DNA amplified by PCR may not contain the entire length of the aldolase DNA, the entire length of the DNA is isolated from a chromosomal DNA library of a microorganism that produces aldolase using the DNA amplified by PCR as a probe.

The procedure for PCR is described in publications such as White, T.J. et al., Trends Genet. 5, 185 (1989). The method for isolating a chromosomal DNA, as well as the method for isolating a desired DNA molecule from a gene library using a DNA molecule as a probe, are described in publications such as Molecular Cloning, 3rd edition, Cold Spring Harbor Laboratory Press (2001).

A method for determining the nucleotide sequence of an isolated DNA that encodes aldolase is described in A Practical Guide to Molecular Cloning, John Wiley & Sons, Inc. (1985). Furthermore, the nucleotide sequence may be determined by using the DNA Sequencer made by Applied Biosystems. A DNA encoding aldolase derived from *Arthrobacter simplex* strain AKU626 (NBRC12069) is shown in SEQ ID No: 1.

The DNA that encodes aldolase which catalyzes the reaction in which HMKP is formed from acetaldehyde and  $\alpha$ -ketobutyric acid is not only the DNA shown in SEQ ID No: 1. This is because there may be differences in nucleotide sequences from each species and strain among *Arthrobacter* species.

Therefore, the DNA of the present invention not only includes the isolated DNA encoding aldolase, but also includes a DNA encoding aldolase in which mutations have been artificially induced as long as the DNA encodes aldolase which is still able to catalyze the above-mentioned reaction. Methods for artificially inducing mutations include commonly used methods such as introducing site-specific mutations, as described in Method. in Enzymol., 154 (1987).

A DNA that hybridizes under stringent conditions with a DNA which is complementary to the nucleotide sequence of SEQ ID No: 1, and encodes a protein having

- aldolase activity is also included in the DNA of the present invention. As used herein, the "stringent conditions" refer to those conditions under which a specific hybrid is formed and a non-specific hybrid is not formed. Although it is difficult to numerically express these conditions explicitly, by way of example, mention is made of those conditions under which
- 5 DNA molecules having higher homology e.g. preferably 70% or more, more preferably 80% or more, still more preferably 90% or more, and particularly preferably 95% or more homology, hybridize with each other, while DNA molecules having lower homology do not hybridize with each other, or those conditions under which hybridization occurs under usual washing conditions in Southern hybridization, that is, at a salt concentration corresponding
- 10 to 0.1×SSC and 0.1% SDS at 37°C, preferably 0.1×SSC and 0.1% SDS at 60°C, and more preferably 0.1×SSC and 0.1% SDS at 65°C. The length of the probe may be suitably selected, depending on the hybridization conditions, and usually varies from 100 bp to 1 kbp. Furthermore, "aldolase activity" means the synthesis of HMKP from acetaldehyde and  $\alpha$ -ketobutyric acid. However, when a nucleotide sequence hybridizes
- 15 under stringent conditions with a nucleotide sequence which is complementary to the nucleotide sequence of SEQ ID No: 1, the nucleotide sequence encodes an aldolase protein which preferably retains aldolase activity of 10% or more, preferably 30% or more, more preferably 50% or more, and still more preferably 70% or more, under conditions of 37°C and pH 8.
- 20 Furthermore, a DNA encoding a protein which is substantially identical to the aldolase encoded by the DNA of SEQ ID No: 1 is also included in the DNA of the present invention. Namely, the following DNAs are also included in the DNA of the present invention:
- (a) a DNA having the nucleotide sequence of SEQ ID No: 1;
- 25 (b) a DNA that hybridizes under stringent conditions with a DNA having a nucleotide sequence complementary to the nucleotide sequence of SEQ ID No: 1, and encodes a protein having aldolase activity;
- (c) a DNA that encodes the protein having the amino acid sequence of SEQ ID No: 2;
- 30 (d) a DNA that encodes the protein having the amino acid sequence of SEQ ID NO: 2, except having substitutions, deletions, insertions, additions or inversions of one or several amino acid residues, wherein the protein maintains aldolase activity; and

(e) a DNA that encodes a protein having an amino acid sequence that is at least 70% homologous, preferably at least 80% homologous, more preferably at least 90% homologous and still more preferably at least 95% homologous to the amino acid sequence of SEQ ID NO:2, and having the aldolase activity.

5 Here, "one or several" refers to a range over which the 3-dimensional structure of a protein or aldolase activity is not significantly impaired, and more specifically, a range of 1 to 78, preferably 1 to 52, more preferably 1 to 26, and still more preferably 1 to 13, especially preferably 1 to 5.

The substitution, deletion, insertion, addition or inversion of one or several amino  
10 acid residues should be conservative mutation(s) so that the activity can be maintained. The representative conservative mutation is a conservative substitution. Examples of conservative substitutions include substitution of Ser or Thr for Ala, substitution of Gln, His or Lys for Arg, substitution of Glu, Gln, Lys, His or Asp for Asn, substitution of Asn, Glu or Gln for Asp, substitution of Ser or Ala for Cys, substitution of Asn, Glu, Lys, His, Asp or Arg  
15 for Gln, substitution of Asn, Gln, Lys or Asp for Glu, substitution of Pro for Gly, substitution of Asn, Lys, Gln, Arg or Tyr for His, substitution of Leu, Met, Val or Phe for Ile, substitution of Ile, Met, Val or Phe for Leu, substitution of Asn, Glu, Gln, His or Arg for Lys, substitution of Ile, Leu, Val or Phe for Met, substitution of Trp, Tyr, Met, Ile or Leu for Phe, substitution of Thr or Ala for Ser, substitution of Ser or Ala for Thr, substitution of Phe or Tyr for Trp,  
20 substitution of His, Phe or Trp for Tyr, and substitution of Met, Ile or Leu for Val.

Furthermore, "aldolase activity" refers to the synthesis of HMKP from acetaldehyde and  $\alpha$ -ketobutyric acid as described above. However, when the amino acid sequence contains a substitution, deletion, insertion, addition or inversion of one or several amino acid residues in the amino acid sequence of SEQ ID No: 2, this protein preferably  
25 retains aldolase activity of 10% or more, preferably 30% or more, more preferably 50% or more, and still more preferably 70% or more, of protein having the amino acid sequence of SEQ ID No: 2 under conditions of 37°C and pH 8.

## (2) Properties of Aldolase

Next, an explanation is provided of the properties of purified aldolase derived from  
30 *Arthrobacter simplex* strain AKU626 (NBRC12069) (asiHPAL).

The asiHPAL of the present invention has the amino acid sequence of SEQ ID No: 2 as was clearly determined by the previously described gene isolation and analysis.

However, the present invention also includes a protein having an amino acid sequence that contains a substitution, deletion, insertion, addition or inversion of one or several amino acid residues in the amino acid sequence of SEQ ID No: 2, and which also has aldolase activity.

Namely, the aldolase of the present invention includes the following proteins:

- 5 (f) a protein having the amino acid sequence of SEQ ID No: 2;
- (g) a protein having an amino acid sequence that contains a substitution, deletion, insertion, addition or inversion of one or a several amino acid residues in the amino acid sequence of SEQ ID No: 2, and having aldolase activity; and
- (h) a protein that is at least 70% homologous, preferably at least 80% homologous,  
10 more preferably at least 90% homologous and still more preferably at least 95% homologous to the amino acid sequence of SEQ ID NO:2, and having aldolase activity.

Here, the definitions of "several" and "aldolase activity" are the same as defined in section (1), DNA Encoding Aldolase.

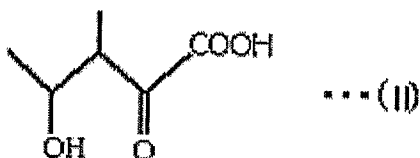
The aldolase of the present invention catalyzes the synthesis of  
15 4-hydroxy-3-methyl-2-keto-pentanoic acid (HMKP) by aldol condensation reaction from acetaldehyde and  $\alpha$ -ketobutyric acid.

The aldolase activity may be measured by determination of the HMKP formation from acetaldehyde and  $\alpha$ -ketobutyric acid or the 4HIL formation after subsequent conversion from HMKP to 4HIL by using high-performance liquid chromatography (HPLC).

20 The aldolase of the present invention is able to catalyze the synthesis of HMKP by aldol condensation reaction from acetaldehyde and  $\alpha$ -ketobutyric acid. Two microbial enzymes capable of catalyzing the aldol condensation reaction using 2 molecules of  $\alpha$ -keto acid (or substituted  $\alpha$ -keto acid) as a substrate have been reported thus far. They are MhpE aldolase derived from *E. coli* (Appl. Environ. Microbiol., Vol. 64, No. 10, 4093-4094,  
25 1998) and the aldolase reported by Sugiyama et al. (WO2004-018672). However, the quantities of HMKP produced were reported to be consistently less than 1  $\mu$ M, which is not a sufficient amount for use in a further transamination process to produce 4-hydroxy-L-isoleucine. Therefore, there have been no findings or reports describing the production of HMKP with industrial applicability, and it is completely unknown as to whether  
30 it is possible to synthesize HMKP using this enzyme. Namely, the aldolase of the present invention differs from aldolases that have been reported thus far in that it is able to catalyze the synthesis of HMKP by aldol condensation of acetaldehyde and  $\alpha$ -ketobutyric acid.

Next, the following is a description of the enzymatic properties for purified asiHPAL.

asiHPAL catalyzes the formation of HMKP, which is represented by the following general formula (II):



from acetaldehyde and  $\alpha$ -ketobutyric acid. Thus, the process of producing HMKP from acetaldehyde and  $\alpha$ -ketobutyric acid using the asiHPAL of the present invention is also described herein.

Furthermore, the activity of asiHPAL strictly depends on  $\text{Zn}^{2+}$ ,  $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$  ions, and is completely blocked in the presence of EDTA. So, asiHPAL may belong to the Type II aldolases (see Table 3).

Since the molecular weight of asiHPAL is about 186 kDa as measured by gel filtration and about 27 kDa as measured by SDS-PAGE, asiHPAL might have a hexameric structure.

Therefore, the present invention also includes a protein as defined by following characteristics:

- (A) has an activity of catalyzing production of 4-hydroxy-3-methyl-2-keto-pentanoic acid from acetaldehyde and  $\alpha$ -ketobutyric acid;
- (B) the activity is dependent on one or more bivalent cations, including  $\text{Zn}^{2+}$ ,  $\text{Mn}^{2+}$  and  $\text{Mg}^{2+}$ ; and
- (C) the molecular weight as measured by gel filtration is about 186 kDa, and the molecular weight per subunit as measured by SDS-PAGE is about 27 kDa.

### (3) Process for Producing Aldolase

Next, the process of producing the aldolase of the present invention is described. There are two ways to produce the aldolase of the present invention: (i) cultivating an aldolase-producing microorganism to produce aldolase, and (ii) preparing a transformant by recombinant DNA technology and cultivating the transformant to produce aldolase.

#### (i) Process for producing Aldolase by Microbial Cultivation

Microorganisms which are sources of native aldolase include microorganisms belonging to the genus *Arthrobacter*.

Any microorganisms belonging to the genus *Arthrobacter* may be used in the present invention provided they are microorganisms that produce aldolase which catalyzes the synthesis of a precursor keto acid (HMKP) from acetaldehyde and  $\alpha$ -ketobutyric acid. Preferably, these microorganisms include *Arthrobacter simplex* (AKU626, NBRC12069), *Arthrobacter globiformis* (AKU625 strain, NBRC12140 strain), *Arthrobacter sulfureus* (AKU635 strain, NBRC12678 strain) and *Arthrobacter viscosus* (AKU636 strain, NBRC13497 strain). Among these, *Arthrobacter simplex* (AKU626, NBRC12069) is particularly preferable. The particulars on the deposits of these microorganisms are indicated below.

Of the above bacteria, those strains the designation of which begins with AKU may be obtained from the Laboratory of Fermentation Physiology and Applied Microbiology, Division of Applied Life Sciences, Faculty of Agriculture, Graduate School, Kyoto University.

Those strains the designation of which begins with NBRC (formerly IFO) can be obtained from the National Institute of Technology and Evaluation (2-5-8 Kazusakamatari, Kisarazu-shi, Chiba, Postal Code 292-0818). The registration numbers of individual strains are recorded in the NBRC catalog:  
(<http://www.nbrc.nite.go.jp/NBRC2/NBRCDispSearchServlet?lang=en>).

Although the microorganism may be cultivated either as liquid cultivation and/or solid cultivation, an industrially advantageous method is the deep-aerated stir cultivation. Carbon sources, nitrogen sources, inorganic salts, and other trace nutrient elements commonly used in microbial cultivating may be used in the media. Any nutrient source may be used so long it is usable by the chosen microbial strain.

Culturing is conducted under aerobic conditions by shake culturing, deep ventilation stir culturing, or the like. The temperature may be within a range in which the microorganisms grow and aldolase is produced. Thus, although the conditions are not strict, the temperature is normally 10 to 50°C and preferably 15 to 42°C. The cultivating time varies according to other cultivating conditions. For example, the microorganisms may be cultivated until the greatest amount of aldolase is produced, and this is normally about 5 hours to 7 days, and preferably about 10 hours to 96 hours.

Following cultivation, the microbial cells are recovered by centrifugation (e.g.,

10,000xg for 10 minutes). Since the majority of the aldolase is present in the cells, the aldolase is solubilized by disrupting or lysing the microbial cells. Ultrasonic disruption, French press disruption, or glass bead disruption may be used to disrupt the microbial cells. When lysing the cells, egg white lysozyme, peptidase treatment, or a suitable combination  
5 thereof may be used.

When aldolase is purified from an aldolase-producing microorganism, although an enzyme solubilizing solution is used, if undisrupted or unlysed residue or debris remains, re-centrifuging the solubilization solution and removing any residue or debris that precipitates is advantageous to purification.

10 All commonly used methods for purifying typical enzymes may be employed to purify the aldolase, examples of which include ammonium sulfate precipitation, gel filtration chromatography, ion exchange chromatography, hydrophobic interaction chromatography and hydroxyapatite chromatography. As a result, an aldolase-containing fraction with higher specific activity may be obtained.

15 (ii) Production Process Using Recombinant DNA Technology

Next, the process for producing aldolase using a recombinant DNA technology is described. There are numerous known examples of producing useful proteins such as enzymes and physiologically active substances using recombinant DNA technology. This technology enables mass production of useful proteins which are present only in trace  
20 amounts in nature.

Fig. 1 is a flowchart of the process for producing the aldolase of the present invention.

First, a DNA is prepared that encodes the aldolase of the present invention (Step S1).

25 Next, the DNA is ligated with a vector DNA to produce a recombinant DNA (Step S2), and cells are transformed with this recombinant DNA to produce a transformant (Step S3). Then, the transformant is cultivated in a medium, and the aldolase is produced, and accumulates in the medium, cells, or both (Step S4).

Subsequently, the process proceeds to Step S5 where the aldolase is recovered  
30 and purified.

Large amounts of HMKP may be produced by using the purified aldolase produced at Step S5, or the medium and/or the cells in which containing the aldolase (Step



S4), in an aldol condensation reaction (Step S6).

The vector DNA ligated with the aldolase gene allows for expression of the aldolase of the present invention.

Aldolase genes which are ligated into the vector DNA include the previously  
5 described DNA as in [I].

When using recombinant DNA technology for large scale production of proteins, cells such as bacterial cells, *Actinomyces* cells, yeast cells, mold cells, plant cells, and animal cells may be used as host cells. Examples of bacterial cells for which host-vector systems have been developed include *Escherichia* species, *Pseudomonas* species,  
10 *Arthrobacter* species, *Corynebacterium* species, *Aspergillus* species, and *Bacillus* species. *Escherichia coli* or *Corynebacterium glutamicum* are preferably used in the present invention. This is because there is much known regarding mass production of proteins using *Escherichia coli* or *Corynebacterium glutamicum*. The following describes the process for producing aldolase using a transformed *E. coli*.

15 A promoter which is typically used for heterogeneous protein production in *E. coli* may be used to express the DNA encoding aldolase. Powerful promoters include the T7 promoter, trp promoter, lac promoter, tac promoter, and PL promoter.

To produce aldolase as a fused protein, a gene that encodes another protein, preferably a hydrophilic peptide, is ligated either upstream or downstream of the aldolase  
20 gene. The gene that encodes the other protein may function to increase the amount of the fused protein which is produced, and/or enhance the solubility of the fused protein after the denaturation and regeneration steps. The T7 gene 10,  $\beta$ -galactosidase gene, dehydrofolate reductase gene, interferon- $\gamma$  gene, interleukin-2 gene, polyhistidine gene, glutathione S-transferase gene, and prochymosin gene are examples of the gene which  
25 can be fused to the aldolase gene.

When ligating to the aldolase gene, the codon reading frames must match. The genes may either be ligated in a suitable restriction enzyme site or using a synthetic DNA of an appropriate sequence.

In order to increase the amount produced, it is preferable to couple a transcription  
30 terminating sequence, such as a terminator, downstream from the fused protein gene. The T7 terminator, fd phage terminator, T4 terminator, tetracycline resistance gene terminator, and *E. coli* trpA gene terminator may be used.

Multi-copy vectors are preferable to introduce a gene that encodes aldolase or a fused protein of aldolase into *E. coli*. Plasmids with a replication starting point derived from Col E1 such as pUC plasmids, pBR322 plasmids, or their derivatives may be used. A "derivative" here refers to a plasmid that has a base alteration such as a substitution, deletion, insertion, addition, or inversion. These alteration may include those caused by mutagenic treatment using a mutagen or UV, spontaneous mutations, or random mutations.

It is preferable that the vector has a marker such as an ampicillin resistance gene so that the transformant may be easily selected. Examples of such plasmids include commercially available expression vectors having a powerful promoter (such as pUC (Takara), pPROK (Clontech), and pKK233-2 (Clontech)).

The recombinant DNA is obtained by ligating the chosen promoter, the gene encoding aldolase or the fused protein of aldolase and another protein, and the terminator in that order, with a vector DNA.

When *E. coli* is transformed with the recombinant DNA and then cultivated, aldolase or the fused protein of aldolase with another protein is expressed and produced. A strain that is typically used for expression of heterogeneous genes may be used as the transformed host, and *E. coli* strain JM109(DE3) and *E. coli* strain JM109 are particularly preferable. The transformation method and method for selecting the transformant are described in, for example, Molecular Cloning, 3rd edition, Cold Spring Harbor Laboratory Press (2001).

When expressing the fused protein, the aldolase may be cut out using a restriction protease, such as blood coagulation factor Xa or kallikrein. These proteases recognize sequences which are not present in aldolase.

A medium which is typically used for cultivating *E. coli* may be used as the production medium. Examples include M9-casamino acid medium and LB medium. Furthermore, conditions for cultivating and inducing production may be appropriately selected depending the type of the marker and promoter which is used, as well as the type of host microorganism.

The following method may be used to recover the aldolase or fused protein. If the aldolase or the fused protein is solubilized within the microbial cells, then the crude enzyme solution may be used after disrupting or lysing the recovered cells. Furthermore,

the aldolase or the fused protein may also be purified by precipitation, filtration, column chromatography, or other common techniques, as necessary. An antibody of the aldolase or its fused protein may also be used for purification.

When a protein inclusion body forms, it may be solubilized with a denaturant.

- 5 Although it may be solubilized with microbial cell protein, in consideration of the subsequent purification procedure, it is preferable to take out the inclusion body and then solubilize it. A method known in the prior art may be used to recover the inclusion body from the microbial cell. For example, the inclusion body may be recovered by disrupting the microbial cell, followed by centrifugal separation. Examples of denaturants that solubilize
- 10 protein inclusion bodies include guanidine hydrochloride (e.g., 6 M, pH 5-8) and urea (e.g., 8 M).

- The protein inclusion body may be regenerated as an active protein by removing these denaturants by a treatment such as dialysis. Dialysis solutions such as Tris-HCl buffer or phosphate buffer may be used for dialysis, and the concentration may be from 20
- 15 mM to 0.5 M, and the pH may be from pH 5 to pH 8.

The protein concentration during the regeneration step is preferably maintained at about 500 µg/ml or less. In order to prevent the regenerated aldolase from crosslinking to itself, the dialysis temperature is preferably 5°C or lower. Furthermore, the activity may be restored by removing the denaturant by dilution and ultrafiltration, in addition to dialysis.

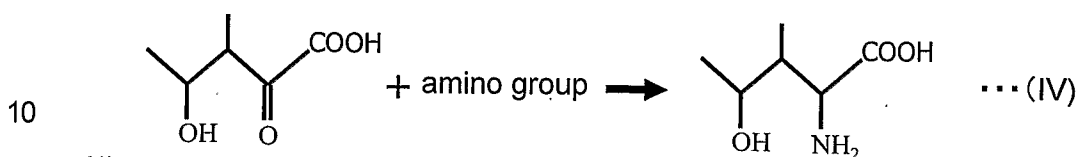
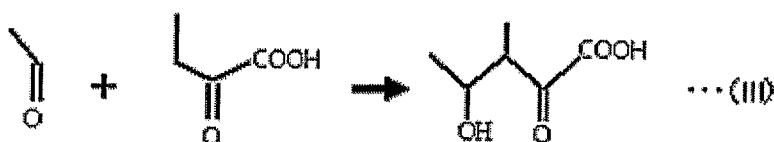
- 20 When the aldolase gene is derived from bacteria belonging to the genus *Arthrobacter*, the aldolase may be expressed and produced using the *Arthrobacter* species bacteria as a host. Examples of such host cells and recombinant expression systems used therein include a recombinant expression method in *Arthrobacter* sp. (Shaw P. C. et al., J Gen Microbiol. 134 (1988) p.903-911), a recombinant expression method in
- 25 *Arthrobacter nicotinovorans* (Sandu C. et al., Appl Environ Microbiol. 71(2005) p8920-8924), and a recombinant expression method in *Arthrobacter* sp. (Morikawa, M. et al., Appl Microbiol Biotechnol., 42 (1994), p.300-303). Also, expression systems developed for Coryneform bacteria are also applicable in *Arthrobacter* species (Sandu C. et al.). However, the present invention is not limited to the *Arthrobacter* species bacteria
- 30 expression system for aldolase as described herein.

#### [II] Method for producing 4-hydroxy-L-isoleucine

The method for producing 4-hydroxy-L-isoleucine (4HIL) of the present invention

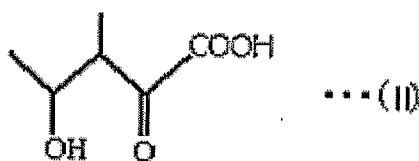
is a two step reaction: (1) enzymatic aldol condensation reaction to produce 4-hydroxy-3-methyl-2-keto-pentanoic acid (HMKP) from acetaldehyde and  $\alpha$ -ketobutyric acid (reaction formula (III) below), and (2) conversion from HMKP to 4HIL, that is, enzymatic transamination producing 4HIL from HMKP (reaction formula (IV) below).

- 5 These steps may be continuously performed in this order or almost simultaneously performed in one container. Each step is described below.



(1) Producing 4-hydroxy-3-methyl-2-keto-pentanoic acid (HMKP)

HMKP (II) is produced by reacting acetaldehyde and  $\alpha$ -ketobutyric acid :



15 wherein the reaction is performed in the presence of an aldolase that catalyzes the reaction.

There are no particular limitations on the aldolase that is used to catalyze the reaction, and any protein may be used as long as the protein is capable of catalyzing the synthesis of HMKP represented by the general formula (II) by aldol condensation of acetaldehyde and  $\alpha$ -ketobutyric acid.

20 A preferable example of such an aldolase is the aldolase described in section [1]. In the method of the present invention, aldolase may be used in any form, such as in a bacterium (including a culture, bacterial cells, or treated cells), a purified enzyme, or a crude enzyme, so long as it is able to catalyze the production of HMKP. When a bacterium is used as the source of the aldolase, both a bacterium which naturally produces the aldolase, 25 such as microorganisms belonging to the genus *Arthrobacter*, and a recombinant microorganism which has been transformed with a recombinant DNA as described in section [1] are encompassed.

For example, when producing HMKP using an aldolase-producing bacterium or bacterial cells that have been transformed with a recombinant DNA, the substrate may be added directly to the culture media during cultivation, or the substrate may be bacterial cells or washed bacterial cells that have been separated from the culture. Furthermore, 5 bacterial cells that have been disrupted or lysed may be used directly, or the aldolase may be recovered from the treated bacterial cells and used as a crude enzyme solution, or a purified enzyme may be used. Namely, as long as aldolase activity is present, the aldolase-activity containing fraction may be used in the process for producing a 4HIL of the present invention.

10 The aldol condensation reaction is performed in a reaction solution containing acetaldehyde,  $\alpha$ -ketobutyric acid, and aldolase or aldolase-containing composition, adjusted to a suitable temperature of 20 to 50°C, and allowing to stand undisturbed, shaking or stirring for 30 minutes to 5 days while maintaining at pH 5 to 12.

The speed of the reaction may also be increased by adding a bivalent cation such 15 as  $Mg^{2+}$ ,  $Mn^{2+}$ ,  $Fe^{2+}$ ,  $Zn^{2+}$  or  $Co^{2+}$  to the reaction mixture.  $Zn^{2+}$  may be used preferably in terms of cost and so forth.

When adding these bivalent cations to the reaction solution, although any salt may be used provided it does not hinder the reaction,  $ZnCl_2$ ,  $ZnSO_4$ ,  $MgSO_4$ , and so forth may be used preferably. The concentrations of these bivalent cations may be suitably 20 determined by simple preliminary studies. These bivalent cations may be added within the range of 0.01 mM to 10 mM, preferably 0.1 mM to 5 mM.

The HMKP may be either separated or purified according to known techniques, or not, particularly when produced by a recombinant microorganism which expresses both aldolase and aminotransferase. Separation and purification methods include contacting 25 the HMKP with an ion exchange resin to adsorb basic amino acids, followed by elution and crystallization. Another purification method includes when the elution product is discolored and filtration with activated charcoal is necessary, then crystallization can be performed to obtain HMKP.

The process for producing HMKP of the present invention makes it possible to 30 form a precursor keto acid (HMKP) of 4-hydroxy-L-isoleucine (4HIL) from acetaldehyde and  $\alpha$ -ketobutyric acid. Since HMKP may be used to derive 4-hydroxy-L-isoleucine by amination at position 2, it is useful as an intermediate in 4-hydroxy-L-isoleucine synthesis.

The process for producing 4-hydroxy-L-isoleucine from HMKP is described next.

(2) Converting 4-hydroxy-3-methyl-2-keto-pentanoic acid to 4-hydroxy-L-isoleucine

Once HMKP is produced as described above (1), HMKP is converted to 4HIL.

Methods for converting HMKP to 4HIL are not particularly limited, and either

5 chemical or enzymatic methods may be used. As for enzymatic conversion, enzymatic transamination is preferable. In this step, "transamination" means the reaction whereby an amino group is transferred from a donor compound, for example, L-glutamic acid or L-glutamate, to an acceptor compound with a keto-group, for example,  
4-hydroxy-3-methyl-2-keto-pentanoic acid or the like. Although the amino group donor is  
10 not particularly limited as long as it has an amino group, L-amino acids such as L-glutamic acid and branched-chain amino acids are preferably used.

In the present invention, "enzymatic transamination" means a transamination reaction which is carried out by an aminotransferase (transaminase) or a dehydrogenase enzyme. Particularly a bacterial aminotransferase (transaminase) is preferred. These  
15 enzymes may be used in any form, such as in a bacterium (including a culture, bacterial cells, or treated cells), a purified enzyme, or a crude enzyme, so long as it incorporates the above aminotransferase and/or dehydrogenase. To reduce the cost for producing 4HIL by simplifying the process, adding the substrate directly to the culture solution is most preferable.

20 The aminotransferase and/or dehydrogenase used in this step is not particularly limited as long as HMKP can be converted to 4HIL. A branched-chain amino acid aminotransferase (BCAT) is preferable. For example, the aminotransferase encoded by the *ilvE* gene, the aromatic aminotransferase encoded by *tyrB* gene, the aspartate aminotransferase encoded by *aspC* gene, the valine-pyruvate aminotransferase encoded  
25 by *avtA* gene in *E. coli*, the aminotransferase encoded by *ywaA* gene in *Bacillus subtilis*, and the like are exemplified

An exemplary protein sequence of BCAT is COG0115 (SEQ ID NO: 13). Also, the protein belonging to BCAT is shown as EC 2.6.1.42.

Practically all branched-chain amino acid aminotransferases exhibit broad  
30 substrate specificity, so most of them can use HMKP as a substrate for the amino-group transfer.

The branched-chain amino acid aminotransferase catalyzes the transfer of an

amino-group from L-glutamic acid to different  $\alpha$ -keto acids, such as  $\alpha$ -keto-isovaleric acid, 2-keto-3-methylvaleric acid, and 2-keto-4-methylpentanoic acid, resulting in formation of L-valine, L-isoleucine, and L-leucine, respectively.

5 The branched-chain amino acid aminotransferases from a great majority of microorganisms are known, and the nucleotide sequences of genes encoding these aminotransferases have also been disclosed.

The *ilvE* gene encodes the IlvE protein, which is a branched-chain amino acid aminotransferase from *Escherichia coli*, (synonyms include B3770, IlvE, branched-chain amino acid: 2-oxoglutaric acid aminotransferase, BCAT, transaminase B, leucine  
10 transaminase, valine transaminase, and isoleucine transaminase). This branched-chain amino acid aminotransferase may be abbreviated as ecoBCAT. The *ilvE* gene is located between the *ilvM* and *ilvD* genes on the chromosome of *E. coli* strain K-12. The nucleotide sequence of the *ilvE* gene is known (nucleotide positions: 3950507 to 3951436; GenBank accession no. NC\_000913.2; gi:49175990) (SEQ ID NO: 14). The nucleotide sequence  
15 of the *ilvE* gene and the amino acid sequence of the IlvE protein encoded by the *ilvE* gene are shown in SEQ ID NO: 14 and SEQ ID NO: 15, respectively.

The *ywaA* gene encodes the branched-chain amino acid aminotransferase from *Bacillus subtilis*. This branched-chain amino acid aminotransferase may be abbreviated as bsuBCAT. The *ywaA* gene is located between the *dltE* and *licH* genes on the  
20 chromosome of *B. subtilis* strain 168. The nucleotide sequence of the *ywaA* gene is known (nucleotide positions: 3956412 to 3957503; GenBank accession no. NC\_000964.2; gi:50812173) (SEQ ID NO: 16). The nucleotide sequence of the *ywaA* gene and the amino acid sequence of the YwaA protein encoded by the *ywaA* gene are shown in SEQ ID NO: 16 and SEQ ID NO: 17, respectively.

25 Other genes encoding branched-chain amino acid aminotransferases from other microorganisms can be identified by homology to known genes encoding branched-chain amino acid aminotransferases, followed by evaluation of the activity of the proteins encoded by the genes.

Homology between two amino acid sequences can be determined using  
30 well-known methods, for example, the computer program BLAST 2.0, which calculates three parameters: score, identity, and similarity.

Therefore, the *ilvE* gene from *E. coli* and *ywaA* gene from *B. subtilis* can be

obtained by PCR (polymerase chain reaction; refer to White, T.J. et al., Trends Genet., 5, 185 (1989)) utilizing primers based on the known nucleotide sequence of the gene. Genes encoding branched-chain amino acid aminotransferases from other microorganisms can be obtained in a similar manner.

5           Since there may be some differences in DNA sequences between bacterial strains, the above-described genes encoding aldolase or BCAT are not limited to the nucleotide sequences shown in SEQ ID NOS: 1, 14 and 16 but may also include nucleotide sequences similar to those shown in SEQ ID NOS: 1, 14 and 16. Therefore, the protein variants encoded by the above-described genes may have a homology of not less than  
10   70%, preferably not less than 80%, more preferably not less than 90%, and most preferably not less than 95%, with respect to the entire amino acid sequences shown in SEQ ID NOS. 2, 15 and 17, as long as the proteins are able to catalyze the reactions.

Moreover, the above-described genes may be represented by variants which can hybridize under stringent conditions with the nucleotide sequences shown in SEQ ID NOS:  
15   1, 14 and 16 or with probes prepared based on these nucleotide sequences, provided that they encode functional proteins. "Stringent conditions" used here is the same as aforementioned section, [1] (1) A DNA Encoding Aldolase.

The treated bacterial cells employed in the present invention may be in the form of dried bacterial mass, freeze-dried bacterial mass, products treated with surfactants or  
20   organic solvents, enzyme-treated products, ultrasound-treated products, mechanically ground products, solvent-treated products, protein fractions of bacterial mass, immobilized products of bacterial mass, and processed bacterial mass.

Production of 4HIL from HMKP by transamination may be continuously performed after the aldol reaction, subsequent separation and/or purification of produced  
25   HMKP; or it may be simultaneously performed in the same reaction solution by allowing the aldolase and the aminotransferase to coexist. When the reactions are performed in the same reaction solution (one pot reaction), the aldolase and the aminotransferase may be co-expressed in a bacterium. Or these enzymes may be prepared separately as described above and added into the reaction solution. A bacterium (host cell) that  
30   co-expresses the DNA encoding aldolase and the DNA encoding aminotransferase may be prepared by co-transfection with the expression vectors each containing the aldolase and aminotransferase DNAs, or transformation with the expression vector which contains



both the aldolase and aminotransferase DNA in a form capable of expression in the host cell. Furthermore, increasing the expression of the genes encoding aldolase and the branched-chain amino acid aminotransferase is the preferable method for enhancing the activities of the aldolase and the aminotransferase.

5           The phrase "increasing the expression of the gene" means that the expression of the gene is higher than that of a non-modified strain, for example, a wild-type strain. Examples of such modifications include increasing the copy number of expressed gene(s) per cell, increasing the expression level of the gene(s), and so forth. The copy number of an expressed gene can be measured, for example, by restricting the chromosomal DNA  
10 followed by Southern blotting using a probe based on the gene sequence, fluorescence in situ hybridization (FISH), and the like. The level of gene expression can be measured by various known methods including Northern blotting, quantitative RT-PCR, and the like. The amount of the expressed protein can be measured by known methods including SDS-PAGE followed by immunoblotting assay (Western blotting analysis), and the like.

15           "Transformation of a bacterium with DNA encoding a protein" means introducing the DNA into a bacterium, for example, by conventional methods. Transformation of this DNA will result in an increase in expression of the gene encoding the protein of present invention, and will enhance the activity of the protein in the bacterial cells. Methods of transformation include any known methods that have hitherto been reported. For  
20 example, treating recipient cells with calcium chloride so as to increase permeability of the cells to DNA has been reported for *Escherichia coli* K-12 (Mandel, M. and Higa, A., J. Mol. Biol., 53, 159 (1970)) and may be used.

          Methods of enhancing gene expression include increasing the gene copy number. Introducing a gene into a vector that is able to function in the bacterium of the present  
25 invention increases the copy number of the gene. For such purposes, multi-copy vectors can be preferably used. The multi-copy vector is exemplified by pBR322, pMW119, pUC19, pET22b, or the like.

          Enhancing gene expression may also be achieved by introducing multiple copies of the gene into the bacterial chromosome by, for example, homologous recombination,  
30 Mu integration, or the like. For example, one act of Mu integration allows for up to 3 copies of the gene to be introduced into the bacterial chromosome.

          Increasing the copy number of the gene can also be achieved by introducing

multiple copies of the gene into the chromosomal DNA of the bacterium. In order to introduce multiple copies of the gene into a bacterial chromosome, homologous recombination is carried out using a sequence which is present in multiple copies as targets in the chromosomal DNA. Sequences having multiple copies in the chromosomal DNA include, but are not limited to repetitive DNA, or inverted repeats present at the end of a transposable element. Also, as disclosed in US patent No. 5,595,889, it is possible to incorporate the gene into a transposon, and transfer it to introduce multiple copies of the gene into the chromosomal DNA.

Enhancing gene expression may also be achieved by placing the DNA of the present invention under the control of a potent promoter. For example, the Ptac promoter, the lac promoter, the trp promoter, the trc promoter, the PR, or the PL promoter of lambda phage are all known to be potent promoters. The use of a potent promoter can be combined with multiplication of gene copies.

Alternatively, the effect of a promoter can be enhanced by, for example, introducing a mutation into the promoter to increase the transcription level of a gene located downstream of the promoter. Furthermore, it is known that substitution of several nucleotides in the spacer region between the ribosome binding site (RBS) and the start codon, especially the sequences immediately upstream of the start codon, profoundly affect the mRNA translatability. For example, a 20-fold range in the expression levels was found, depending on the nature of the three nucleotides preceding the start codon (Gold et al., *Annu. Rev. Microbiol.*, 35, 365-403, 1981; Hui et al., *EMBO J.*, 3, 623-629, 1984). Previously, it was shown that the *rhtA23* mutation is an A-for-G substitution at the -1 position relative to the ATG start codon (ABSTRACTS of 17th International Congress of Biochemistry and Molecular Biology in conjunction with 1997 Annual Meeting of the American Society for Biochemistry and Molecular Biology, San Francisco, California August 24-29, 1997, abstract No. 457).

Moreover, it is also possible to introduce a nucleotide substitution into the promoter region of the gene on the bacterial chromosome, which results in stronger promoter function. The alteration of the expression control sequence can be performed, for example, in the same manner as the gene substitution using a temperature-sensitive plasmid, as disclosed in International Patent Publication WO 00/18935 and Japanese Patent Application Laid-Open No. 1-215280.

Methods for the preparation of plasmid DNA include, but are not limited to digestion and ligation of DNA, transformation, selection of an oligonucleotide as a primer, and the like, or other methods well known to one skilled in the art. These methods are described, for instance, in "Molecular Cloning A Laboratory Manual, Third Edition", Cold  
5 Spring Harbor Laboratory Press (2001).

### Examples

The present invention will be explained in further detail with reference to the examples shown below, however, the invention is not limited thereto.

#### 10 Example 1. Purification of asiHPAL.

The purification protocol includes the following procedures.

*Step1:* 1 ml of overnight bacterial culture (grown for 12 h at 34 °C) of *Arthrobacter simplex* AKU626 was used to inoculate 5 liters of LB-broth [8X(375 ml in 1 L flask)]. Cells were cultivated at an optimal temperature for about 24 hours. Then, cells were harvested  
15 by centrifugation (16000 x g) at 4 °C, and re-suspended in 30 ml of buffer A [50 mM KH<sub>2</sub>PO<sub>4</sub> (pH 7.4 adjusted by KOH)] supplemented by 1mM PMSF (phenylmethylsulfonyl fluoride).

*Step2:* Cells were disrupted by 3-5 passages through a French pressure cell (max P = 2,5 Psi) followed by centrifugation to remove debris. Protein preparation was passed  
20 through Sephadex G-15 column (2.6 X 28 cm) equilibrated with buffer A.

*Step3:* Anion-exchange chromatography (AEC 1) was carried out using ÄKTAbasic100 system supplemented with 50 ml DEAE (fast flow) column (d = 1.6 cm). 40 - 50 ml of the protein preparation obtained from *Step2* were applied to a column equilibrated with buffer A. The elution was performed at a flow rate of 2.5 ml/min with a liner  
25 gradient of 0-0.5 M NaCl in buffer A (10 CV (column volumes)). Each 10ml-fraction was collected. Active fractions were pooled and desalted as described in "step2" item.

*Step4:* Anion-exchange FPLC (AEC 2) was carried out using the ÄKTAbasic100 system, and supplemented with an 1.6 ml "Source15Q" column (Amersham Pharmacia Biotech). The protein preparation obtained from *Step3* was applied to column equilibrated  
30 with buffer A. The elution was carried out at flow rate 1 ml/min by liner 0-0.5 M NaCl in buffer A (40 CV). 2 ml fractions were collected. Active fractions were pooled (Tables 1, 2).

*Step5:* Hydrophobic interaction chromatography (HIC) was carried out using the

ÄKTAbasic100 system, and supplemented with an 1 ml "Resource PHE" column (Amersham Pharmacia Biotech). The protein concentration in the preparation obtained from *Step 5* was adjusted to 0.8 mg/ml and then ammonium sulphate was added up to final concentration 1.5 M. The protein solution was applied to a column equilibrated with buffer **A**, and supplemented with 1.5 M ammonium sulfate. Elution was performed at flow rate of 1 ml/min with a liner gradient from 1.5 M to 0 M ammonium sulfate in buffer **A** (30 CV). 1 ml fractions were collected. Active fractions were pooled (Tables 1, 2).

**Table 1. Purification of the asiHPAL.**

Step	Volume (ml)	Protein conc. (mg/ml)	Total protein (mg)	Specific <sup>1)</sup> activity (nmoles/mg/min)	Total Activity <sup>2)</sup> (nmoles/min)	Yield % <sup>3)</sup>	Purification rate <sup>4)</sup>
Lysate	120	12.6	1512	0.6	903	100.00	1
AEC1	20	0.7	14	30	401	45	48
AEC2	4	1	4	100	402	45	167
HIC	1	0.1	0.1	1419	142	16	2365
SEC	1	0.04	0.04	1490	60	7	2483

<sup>1)</sup> Specific activity was determined by monitoring via HPLC of the time-dependent 4HIL formation in the bsuBCAT/asiHPAL bi-enzymatic reaction with the following composition [100mM L-glutamate (pH 8 adjusted by pH 8.0), 100mM  $\alpha$ -ketobutyrate, 100 mM acetaldehyde ], 1mM ZnCl<sub>2</sub>, and 0.5  $\mu$ g purified bsuBCAT) and an aliquot of the active fraction of asiHPAL was separated. All reactions were carried out at 37 °C.

<sup>2)</sup> Calculated as (Total protein) x (Specific activity).

<sup>3)</sup> Calculated as 100% x (Total activity/Total activity in crude lysate).

<sup>4)</sup> Calculated as (Specific activity/Specific activity in crude cell lysate).

**Table 2. Chromatographic elution parameters.** <sup>a)</sup>

Chromatographic stages.			
AEC 1 NaCl, M	AEC 2 NaCl, M	HIC (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , M	SEC V <sub>e</sub> / V <sub>0</sub>
0.35 – 0.37	0.31 – 0.35	0.19 – 0.1	1.5

<sup>a)</sup> Data are presented as a salts concentration interval within HPAL activity eluted

*Step6:* Size exclusion chromatography (**SEC**) was performed using the ÄKTAbasic100 system, and supplemented with a Superdex™ 200 HR 10/30A (Amersham Pharmacia Biotech) column. The protein preparation obtained from *Step5* was applied to a column equilibrated with buffer **A**, and supplemented with 100 mM NaCl. Isocratic elution

occurred at a flow rate of 0.5 ml/min. 1 ml fractions were collected. Active fractions were pooled (Tables 1, 2; Fig. 1).

The specific activity of asiHPAL was determined by HPLC monitoring of the time-dependent 4HIL formation in the *bsuYwaA*/asiHPAL bi-enzymatic reaction with  
 5 following composition [100mM L-glutamate (pH 8 adjusted by pH 8.0), 100mM  $\alpha$ -ketobutyrate, 100 mM acetaldehyde], 1mM  $ZnCl_2$ , and 0.5  $\mu$ g purified His-tag-*bsuBCAT* protein (branched-chain amino acid aminotransferase from *Bacillus subtilis*) and an aliquot of the elution fraction. All reactions were carried out at 37 °C. Cloning the branched-chain amino acid aminotransferase from *Bacillus subtilis* (*bsuBCAT*) and HPLC measuring the  
 10 4HIL formation are described in the Examples 5 and 6, respectively.

The molecular weight of the asiHPAL monomer was determined using SDS-PAGE (Fig. 2, A). Its average value was estimated to be 27 kDa. The molecular weight of the native asiHPAL was determined using analytical SEC on the Superdex™ 200 HR 10/30A (Pharmacia) columns calibrated by Molecular Weight Protein Markers (Sigma) (Fig.  
 15 2, B). Its average value was estimated to be 186 kDa. So, asiHPAL is assumed to be the hexamer.

The dependence of HMKP-aldolase on metal ions was investigated. It was established that asiHPAL activity strictly depends on the  $Zn^{2+}$ ,  $Mg^{2+}$  and  $Mn^{2+}$  ions and is completely blocked in the presence of EDTA. So, it was assumed that asiHPAL is a Type II  
 20 aldolases (Table 3).

**Table 3. asiHPAL activity depending on the metal  $^{2+}$  ions.**

Protein	Cofactor (2mM concentration)			
	EDTA	$Zn^{2+}$	$Mg^{2+}$	$Mn^{2+}$
asiHPAL	ND. <sup>a)</sup>	1 <sup>b)</sup>	0.5	0.5

<sup>a)</sup> non detected

<sup>b)</sup> Activities were measured as well as described in the Table 7 footnote. The value of the asiHPAL activity measured in the presence of the  $Zn^{2+}$  ion is taken as 1.

## **Example 2. Determination of asiHPAL N-terminal sequence.**

### **2.1 Western blotting of the asiHPAL.**

asiHPAL was immobilized on the Sequi-Blot PVDF membrane (Bio-Rad) using  
 25 trans-Blot SD sell (Bio-Rad). Optimized blotting conditions are as follows: Dun carbonate

transfer buffer: 10 mM NaCHO<sub>3</sub>, 3 mM Na<sub>2</sub>CO<sub>3</sub> without methanol, six pieces extra thick filter paper/membrane sandwich, starting current 5.5 mA/cm<sup>2</sup>, transfer time – 1 hour).

## 2.2 Sequencing of N-terminal end of asiHPAL.

Determination of asiHPAL N-terminal sequence was done using the 491cLC Protein Sequencer (Applied Biosystems, USA). 26 cycles were performed. As the result, the following amino acid sequence was determined:  
NH<sub>2</sub>-Pro-Phe-Pro-Val-Glu-Leu-Pro-Asp-Asn-Phe-Ala-Lys-Arg-Val-Thr-Asp-Ser-Asp-Ser-Ala-Gln-Val-Gly-Leu-Phe-Ile..-COOH (SEQ ID NO: 3).

An alignment of this N-terminal sequence with all known proteins (BLAST search) revealed a single protein with a similar N-terminal sequence. It is the HHDE-aldolase (2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase) from *Brevibacterium linens* BL2 (HHDE\_BLI) (Fig. 3). So, this is a Type II aldolase and its native substrate, 2,4-dihydroxyhept-2-ene-1,7-dioic acid, is structurally similar to HMKP (indeed, both have an hydroxyl group at C<sub>4</sub> position, carbonyl group at C<sub>2</sub> position, and carboxyl group at C<sub>1</sub> position). Moreover, the Mw of the HHDE\_BLI subunit is 27 kDa which is consistent with the experimentally obtained subunit Mw of asiHPAL. Also *Brevibacterium linens* is closely related to *Arthrobacter simplex*.

Thus, it could be assumed that asiHPAL purified from *Arthrobacter simplex* is a homologue of HHDE aldolase from *Brevibacterium linens* BL2.

### Example 3. Cloning of *asiHPAL* gene encoding HMKP-aldolase from *Arthrobacter simplex* (NBRC 12069).

To amplify the DNA fragment from the chromosome of *Arthrobacter simplex* which contains the *asiHPAL* gene, PCR was performed using the degenerative primer asiN10 (SEQ ID NO: 4), which was designed based on the N-terminal amino acid sequence of asiHPAL determined previously (Figure 4). Primer asiN10 contains a degenerative part on the 3'-end thereof and has a sequence of 21 nucleotides, which are not complementary to the *asiHPAL* gene. This is necessary for further amplification of the PCR product. PCR was performed in a volume of 40μl with 0.1 μg of *Arthrobacter simplex* chromosomal DNA and primer asiN10 (20 pmoles). PCR conditions were as follows: 95 °C – 10 sec; 53 °C – 20 sec; 72 °C – 40 sec; 50 cycles.

Surprisingly, as a result of the PCR, a single PCR product was obtained (Figure 5,

track 3). Primer asiN10 appeared to be suitable for use as a forward and reverse primer simultaneously. Furthermore, this was proved by a second round of PCR using the primer SVS88 (SEQ ID NO: 5) (10 pmoles), part of the primer asiN10, which contains the non-complementary 21 nucleotides and 2 µl of reaction mixtures from the first PCR as a  
5 template. PCR conditions were as follows: 95 °C – 10 sec; 65 °C – 20 sec; 72 °C – 40 sec; 25 cycles. A strong single PCR product was observed.

Restriction analysis of the PCR product revealed a unique *Nco*I site which divides the DNA fragment into two fragments: large – 0.6 kb and small – 0.3 kb (Figure 5, track 2). Both fragments were sequenced separately using the SVS88 primer. 471 nucleotides of  
10 the 0,6 kb fragment and 262 nucleotides of the 0.3 kb fragment were determined.

Then, primers SVS\_108 (SEQ ID NO: 6) and SVS\_109 (SEQ ID NO: 7) were used to complete the sequencing of the entire DNA fragment. The strategy for sequencing the amplified DNA fragment is depicted in Figure 6. Analysis of the nucleotide sequence revealed that the N-terminal 26 amino acid residues coded by asiHPAL ORF are identical  
15 to the N-terminal sequence of asiHPAL determined in Example 2.

An alignment of the ORF with the amino acid sequence of HHDE aldolase from *Brevibacterium linens* BL2 revealed high homology between the enzymes (Fig. 7).

#### Example 4. Production of 4-hydroxy-L-isoleucine

20 To test the specific activity of recombinant asiHPAL aldolase coded by the asiHPAL gene from *Arthrobacter simplex* during 4-hydroxy-L-isoleucine production, the asiHPAL gene can be amplified using primers 1 (SEQ ID NO: 8) and 2 (SEQ ID NO: 9), and chromosomal DNA of *Arthrobacter simplex* as a template. Primer 1 contains an *Hind*III restriction site at the 5'-end thereof, primer 2 contains a *Bam*HI restriction site at the 5'-end  
25 thereof. The resulting PCR fragment can be digested with *Hind*III and *Bam*HI restrictases and ligated into plasmid pMW119, which had been previously treated with the same restrictases. Thus, plasmid pMW119- asiHPAL was obtained. Then, *E. coli* strain BW25113 can be transformed with plasmid pMW119- asiHPAL and the resulting strain BW25113/pMW-asiHPAL can be cultivated in the LB medium containing IPTG (1 mmol).

30 Formation of 4-hydroxy-L-isoleucine can be tested in the *bsuYwaA*/asiHPAL bi-enzymatic reaction with the following composition [100mM L-glutamate (pH 8 adjusted by pH 8.0), 100mM α-ketobutyrate, 100 mM acetaldehyde], 1mM ZnCl<sub>2</sub>, 0.5 µg purified

His-tag-bsuBCAT protein (branched-chain amino acid aminotransferase from *Bacillus subtilis*) and an aliquot of BW25113/pMW-asiHPAL cell lysate. All reactions can be carried out at 37 °C. Cloning the branched-chain amino acid aminotransferase from *Bacillus subtilis* (bsuBCAT) and HPLC measuring the 4HIL formation are described in the

5 Examples 5 and 6, respectively.

Example 5. Cloning and efficient expression of the BCAT from *Bacillus subtilis*.

BCAT from *Bacillus subtilis* was cloned and expressed using the pET expression system (Novagen, Madison, WI, USA) as his<sub>6</sub>-tag derivatives.

10 To construct the pET-HT-IlvE-BSU plasmid, the *ywaA* gene from *Bacillus subtilis* encoding BCAT aminotransferase (Berger, B.J et al, *J Bacteriol.*, 185(8), 2418-31 (2003)) was amplified by PCR using the chromosomal DNA of *B. subtilis* strain 168 as a template and primers P3 (SEQ ID NO: 10) and P4 (SEQ ID NO: 11) as “upstream” and

15 “downstream” primers, respectively. Primer P5 contains the *NcoI* restriction site and six codons coding for histidine at the 5'-end thereof, primer P6 contains the *NotI* restriction site at the 5'-end thereof. The resulting PCR fragment was digested with *NcoI* restrictase and ligated with plasmid pET-15(b+) which had been previously digested with the same restrictase. Then, the linear ligated DNA fragment was used as a template for

20 PCR-amplification, and oligonucleotides T7 (Novagen, SEQ ID NO: 12) and P4 (SEQ ID NO: 11) were used as a primers. The resulting PCR fragment containing the *ywaA* gene under the control of T7 promoter of the plasmid pET-15(b+) was digested by *XbaI* and *NotI* restrictases and ligated into the pET-22(b+) vector which had been previously treated with the same restrictases. Thus, plasmid pET-HT-IlvE-BSU was obtained. This plasmid was used to transform *E. coli* and His-tag-bsuBCAT protein was purified from the transformant.

25

Example 6. HPLC measurement of 4-hydroxy-L-isoleucine

HPLC analysis: High pressure chromatograph (Waters, USA) with spectrofluorometer 1100 series (Agilent, USA) was used. The chosen detection wave range: excitation wavelength at 250 nm, range of emission wavelengths were 320-560 nm.

30 The separation by accq-tag method was performed in a column Nova-Pak<sup>TM</sup> C18 150 × 3,9 mm, 4μm (Waters, USA) at +400 °C. Injection volume of the sample was 5μl. The formation of amino acid derivatives and their separation was performed according to



Waters manufacturer's recommendation (Liu, H. et al, J. Chromatogr. A, 828, 383-395 (1998); Waters accq-tag chemistry package. Instruction manual. Millipore Corporation, pp.1-9 (1993)). To obtain amino acid derivatives with 6-aminoquinolil-N-hydroxysuccinimidyl carbamate, the kit Accq-Fluor™ (Waters, USA) was used. The analysis by accq-tag method was performed using concentrated Accq-tag Eluent A (Waters, USA). All solutions were prepared using Milli-Q water, standard solutions were stored at + 4 °C.

#### INDUSTRIAL APPLICABILITY

10        A novel aldolase is described that catalyzes an aldol condensation reaction reaction of acetaldehyde and  $\alpha$ -ketobutyric acid, and may be preferably used to synthesize 4-hydroxy-3-methyl-2-keto-pentanoic acid (HMKP). The aldolase is also useful as an intermediate in 4-hydroxy-L-isoleucine synthesis.

## CLAIMS

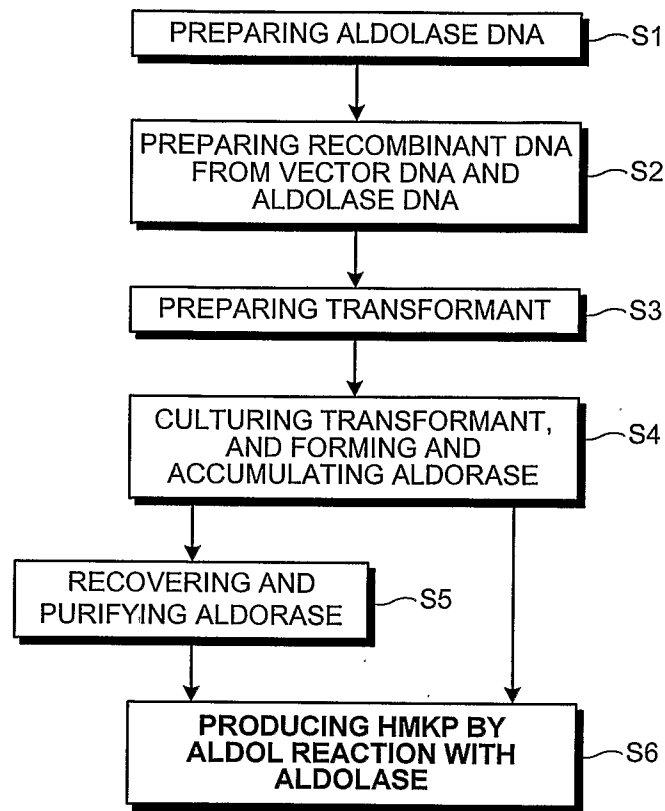
1. A DNA selected from the group consisting of:
  - (a) A DNA comprising the nucleotide sequence of SEQ ID No: 1;
  - 5 (b) A DNA that hybridizes under stringent conditions with a DNA having a nucleotide sequence complementary to the nucleotide sequence of SEQ ID No: 1, and encodes a protein having aldolase activity;
  - (c) A DNA that encodes a protein comprising the amino acid sequence of SEQ ID No: 2;
  - 10 (d) A DNA that encodes a protein having the amino acid sequence of SEQ ID NO: 2, except that it contains a substitution, deletion, insertion, addition, or inversion of one or several amino acid residues and wherein the protein has aldolase activity; and
  - (e) A DNA that encodes a protein having an amino acid sequence that is at least 70% homologous to the amino acid sequence of SEQ ID NO: 2 and wherein the protein has  
15 aldolase activity.
2. A recombinant DNA comprising the DNA according to claim 1 and a vector DNA.
3. An isolated cell transformed with the recombinant DNA according to claim 2.
4. A process for producing a protein having aldolase activity comprising cultivating  
20 the cell according to claim 3 in a medium, and accumulating a protein having aldolase activity in the medium, cells, or both.
5. A protein selected from the group consisting of:
  - (a) a protein comprising the amino acid sequence of SEQ ID No: 2;
  - (b) a protein comprising the amino acid sequence of SEQ ID NO: 2, except that it  
25 contains a substitution, deletion, insertion, addition, or inversion of one or several amino acid residues and wherein the protein has aldolase activity; and
  - (c) a protein that is at least 70% homologous to the amino acid sequence of SEQ ID No: 2 and has aldolase activity.
6. A protein comprising
  - (A) an activity that catalyzes production of 4-hydroxy-3-methyl-2-keto-pentanoic acid  
30 from acetaldehyde and  $\alpha$ -ketobutyric acid;
  - (B) an activity that is dependent on bivalent cation selected from the group consisting of  $\text{Zn}^{2+}$ ,  $\text{Mn}^{2+}$ , and  $\text{Mg}^{2+}$ ;

- (C) a molecular weight of about 186kDa as measured by gel filtration, and  
(D) a molecular weight of about 27kDa per subunit as measured by SDS-PAGE.
7. A method for manufacturing 4-hydroxy-L-isoleucine or a salt thereof, comprising:
- (1) reacting acetaldehyde and  $\alpha$ -ketobutyric acid in an aqueous solvent in the  
5 presence of at least one aldolase selected from the group consisting of:
- (a) a protein comprising the amino acid sequence of SEQ ID No: 2,  
(b) a protein comprising the amino acid sequence of SEQ ID NO: 2, except that it  
contains a substitution, deletion, insertion, addition, or inversion of one or several amino  
acid residues and wherein the protein has aldolase activity, and  
10 (c) a protein that is at least 70% homologous to the amino acid sequence of SEQ ID  
No: 2 and has aldolase activity;
- (2) converting the produced 4-hydroxy-3-methyl-2-keto-pentanoic acid to  
4-hydroxy-L-isoleucine; and  
(3) isolating the produced 4-hydroxy-L-isoleucine.
- 15 8. The method according to claim 7, wherein the  
4-hydroxy-3-methyl-2-keto-pentanoic acid is converted to 4-hydroxy-L-isoleucine by an  
aminotransferase and/or a dehydrogenase, wherein the aminotransferase and/or  
dehydrogenase have transamination activity in the presence of an amino group donor.
9. A method for manufacturing 4-hydroxy-L-isoleucine or a salt thereof, comprising  
20 the steps of:
- (1) reacting acetaldehyde and  $\alpha$ -ketobutyric acid in an aqueous solvent in the  
presence of a bacterium containing an aldolase selected from the group consisting of:
- (a) a protein comprising the amino acid sequence of SEQ ID No: 2,  
(b) a protein comprising an amino acid sequence of SEQ ID NO: 2, except that it  
25 contains a substitution, deletion, insertion, addition, or inversion of one or several amino  
acid residues 2 and wherein the protein has aldolase activity, and  
(c) a protein that is at least 70% homologous to the amino acid sequence of SEQ ID  
No: 2 and has aldolase activity,  
and wherein the aqueous solvent also contains an aminotransferase and/or a  
30 dehydrogenase which catalyzes conversion of 4-hydroxy-3-methyl-2-keto-pentanoic  
acid to 4-hydroxy-L-isoleucine in the presence of an amino group donor; and  
(2) isolating 4-hydroxy-L-isoleucine.

10. The method according to any of claims 8 and 9, wherein the aminotransferase is a branched-chain amino acid aminotransferase.
11. The method according to claim 10, wherein the branched-chain amino acid aminotransferase is derived from a bacterium selected from the group consisting of  
5 *Escherichia* and *Bacillus*.
12. The method according to any of claims 8 and 9, wherein the amino group donor is selected from the group of L-amino acids.
13. The method according to claim 12, wherein the branched-chain amino acid aminotransferase is derived from a bacterium selected from the group consisting of  
10 *Escherichia* and *Bacillus*.
14. The method according to claim 9, wherein the bacterium has been modified to enhance at least one of the activities of the aldolase and the branched-chain amino acid aminotransferase.
15. The method according to claim 14, wherein the activities of the aldolase and the  
15 branched-chain amino acid aminotransferase are enhanced by increasing the expression of the aldolase and/or the branched-chain amino acid aminotransferase.
16. The method according to claim 15, wherein the expression of the aldolase and/or the branched-chain amino acid aminotransferase is increased by modifying an expression control sequence of the gene encoding the aldolase and/or the gene encoding  
20 branched-chain amino acid aminotransferase, or by increasing the copy number of the gene encoding the aldolase and/or the gene encoding branched-chain amino acid aminotransferase.
17. The method according to any of claims 9 to 16, wherein the bacterium belongs to the genus *Escherichia*, *Pseudomonas*, *Corynebacterium*, *Arthrobacter*, *Aspergillus* or  
25 *Bacillus*.
18. The method according to any of claims 9 to 16, wherein the bacterium belongs to *Escherichia coli*, *Arthrobacter simplex*, *Corynebacterium glutamicum*, *Arthrobacter globiformis*, *Arthrobacter sulfureus*, *Arthrobacter viscosus*, or *Bacillus subtilis*.
19. The method according to any of claims 9 to 18, wherein the bacterium is a  
30 bacterial culture, cells, or treated cells.
20. The method according to any of claims 7 to 19, wherein the 4-hydroxy-L-isoleucine is selected from the group consisting of

(2S,3S,4S)-4-hydroxyisoleucine, (2S,3R,4R)-4-hydroxyisoleucine,  
(2S,3S,4R)-4-hydroxyisoleucine, (2S,3R,4S)-4-hydroxyisoleucine, and combinations  
thereof.

FIG.1



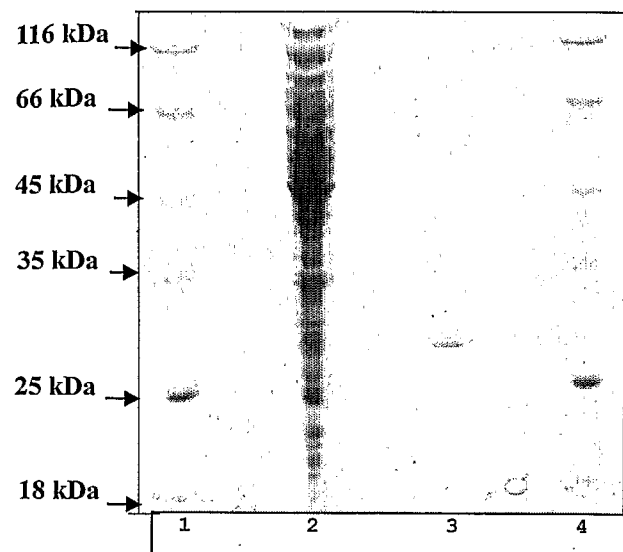


FIG. 2

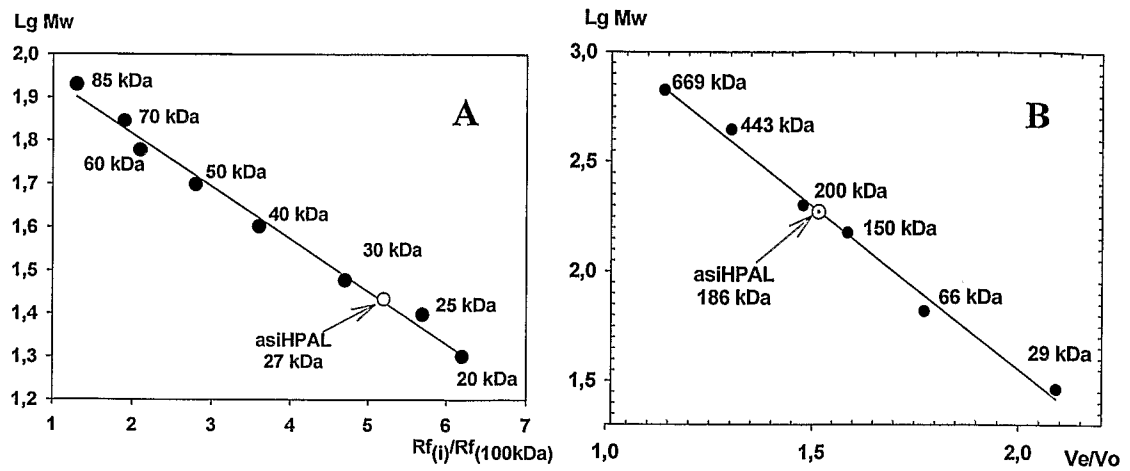


FIG. 3



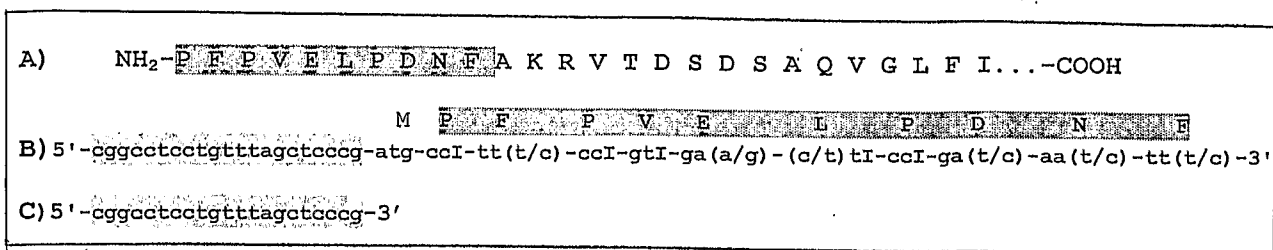


FIG. 4

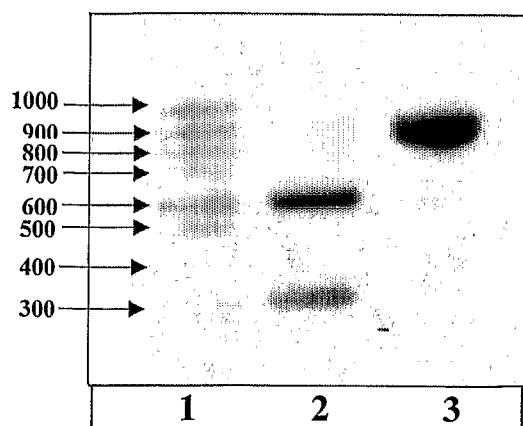


FIG. 5

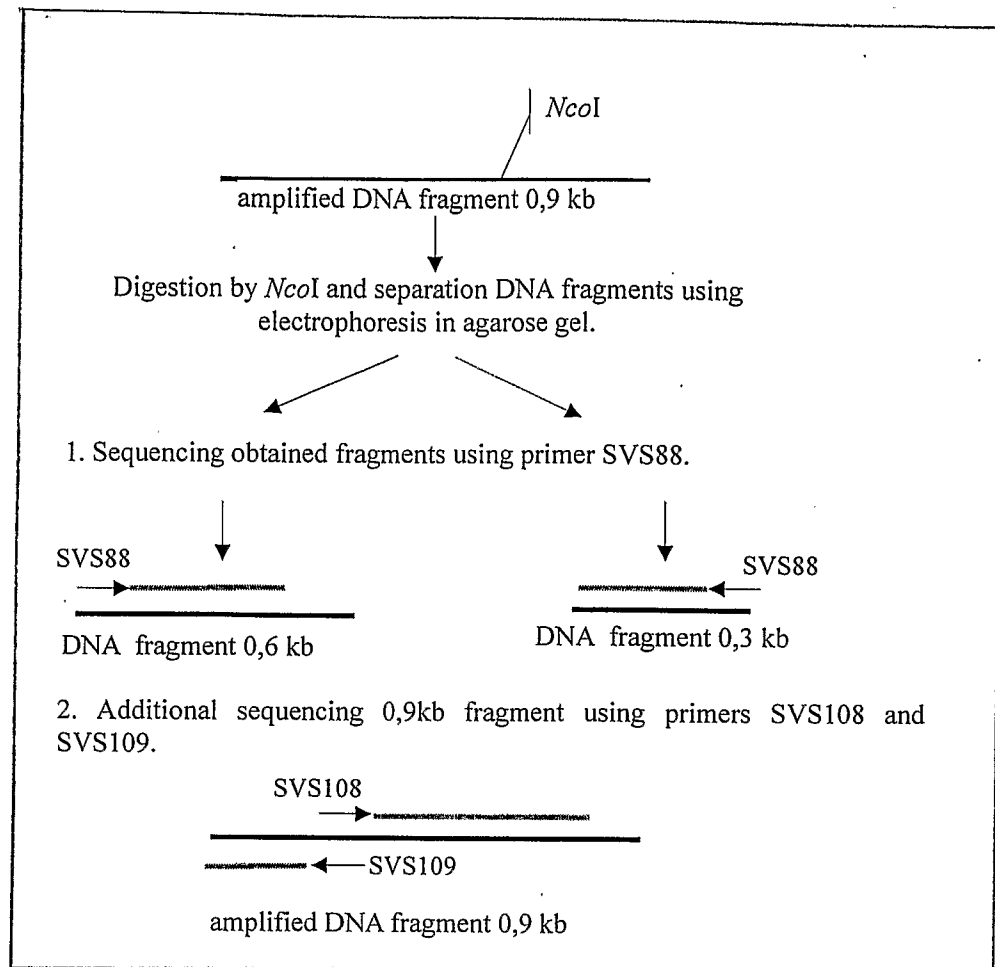


FIG. 6

		1		70
HHDE_BLI	(1)	MPFQVELPQTFTQTVAKLGAGEHLAGMWVCSGSPVAAEIAAASGMQWVLIDAEHSPIDLQITTSLLQAMN		
asiHPAL	(1)	MPFPVELEPDNFAKRVTDSDS--AQVGLFISSGSETNAEIVASAGFDWLLIDAEHSPYGLETVTSLLRTVA		
		71		140
HHDE_BLI	(71)	GYPATPVVRVPVNDQVLIKQYLDLGAQNLLVPMVDTPADAEAAVRSVYYPARGVRGVGSALARASRWNAV		
asiHPAL	(69)	AYPATPVVRVPVNDTVLIKQYLDLGAQNLMVPMVHNAEQAEKAVAAAMHYPPRGVRGIGAALARSSRFNGV		
		141		210
HHDE_BLI	(141)	PNYLARAEDFVSLTIQIESATAVDNAAETAAVDGVDVAVFVGPSDLAASMGLLGQQTHPDVTDVAVLRTFDA		
asiHPAL	(139)	DDYLNKASETVSLTVQVESAEAVENAAEVAAVDGVDAIFIGPSDLAASMGLLGQQQHPAVLAAVDTTTKA		
		211		262
HHDE_BLI	(211)	VKAAGKLVGVNAFDPDQARKYLDAGASFVLVGADVGLMMNGARAWAKTIVQD stop		
asiHPAL	(180)	VRDAGKLVGINAFNLAQAQAYIXAGASFVCGADVQQLASATRALVEKEKG- stop		

FIG. 7

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/JP2007/056756

**A. CLASSIFICATION OF SUBJECT MATTER**  
INV. C12N9/88 C12P13/06

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, Sequence Search, WPI Data, FSTA, EMBASE, MEDLINE, BIOSIS, CHEM ABS Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE UniProt [Online] 24 January 2006 (2006-01-24), "2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase (EC &lt;A HREF="http://srs.ebi.ac.uk/srsbin/cgi-bin/ wgetz?[enzyme-ECNumber:4.1.2.*]+-e"&gt;4.1.2. -&lt;/A&gt;)."  XP002442263 retrieved from EBI accession no. UNIPROT:Q2T4H6 Database accession no. Q2T4H6 see sequence</p> <p style="text-align: center;">----- -/--</p>	1-5

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

12 July 2007

Date of mailing of the international search report

31/07/2007

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Bassias, Ioannis

## INTERNATIONAL SEARCH REPORT

International application No

PCT/JP2007/056756

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE UniProt [Online]  23 November 2004 (2004-11-23), "Putative  2,4-dihydroxyhept-2-ene-1,7-dioic acid  aldolase."  XP002442264  retrieved from EBI accession no.  UNIPROT:Q5YW20  Database accession no. Q5YW20  see sequence</p>	1-5
P,X	<p>-----</p> <p>DATABASE EMBL [Online]  16 December 2006 (2006-12-16), "Synthetic  construct HKP(HMKP)-aldolase gene,  complete cds."  XP002442265  retrieved from EBI accession no.  EMBL:EF117323  Database accession no. EF117323  see sequence</p>	1-20
T	<p>-&amp; SMIRNOV S V ET AL: "A novel strategy  for enzymatic synthesis of  4-hydroxyisoleucine: identification of an  enzyme possessing HMKP  (4-hydroxy-3-methyl-2-keto-pentanoate)  aldolase activity" FEMS MICROBIOL LETT,  vol. 273, no. 1, August 2007 (2007-08),  pages 70-77, XP002442257</p>	
P,X	<p>-----</p> <p>WO 2006/093322 A (AJINOMOTO KK [JP];  KODERA TOMOHIRO [JP]; MATSUI KAZUHIKO  [JP]; NIO NOR)  8 September 2006 (2006-09-08)  the whole document, in particular p.36-40  and claims</p>	1-20
P,X	<p>-----</p> <p>DATABASE UniProt [Online]  12 December 2006 (2006-12-12),  "2-dehydro-3-deoxyglucarate aldolase (EC  &lt;A  HREF="http://srs.ebi.ac.uk/srsbin/cgi-bin/  wgetz?[enzyme-ECNumber:4.1.2.20]+-e"&gt;4.1.2  .20&lt;/A&gt;)." XP002442266  retrieved from EBI accession no.  UNIPROT:AOKOTO  Database accession no. AOKOTO  see sequence</p> <p>-----</p> <p style="text-align: center;">-/--</p>	1-5

## INTERNATIONAL SEARCH REPORT

International application No

PCT/JP2007/056756

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>DATABASE UniProt [Online] 6 February 2007 (2007-02-06), "Putative 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase (HpaI)." XP002442267 retrieved from EBI accession no. UNIPROT:A1RBH6 Database accession no. A1RBH6 see sequence</p> <p>-----</p>	1-5
A	<p>WO 97/32577 A (GESTION JOUVENET SOC CIV DE [FR]; SAUVAIRE YVES [FR]; RIBES GERARD [FR] 12 September 1997 (1997-09-12)</p> <p>-----</p>	
A	<p>BROCA C ET AL: "4-HYDROXYISOLEUCINE: EFFECTS OF SYNTHETIC AND NATURAL ANALOGUES ON INSULIN SECRETION" EUROPEAN JOURNAL OF PHARMACOLOGY, AMSTERDAM, NL, vol. 390, no. 3, March 2000 (2000-03), pages 339-345, XP000908995 ISSN: 0014-2999</p> <p>-----</p>	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/JP2007/056756

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 2006093322	A	08-09-2006	NONE	
WO 9732577	A	12-09-1997	AU FR	22-09-1997 12-09-1997
			2031997 A 2745718 A1	