



US 20240052386A1

(19) **United States**

(12) **Patent Application Publication**
SUN et al.

(10) **Pub. No.: US 2024/0052386 A1**

(43) **Pub. Date: Feb. 15, 2024**

(54) **PROTEIN HAVING L-PROLINE EFFLUX FUNCTION, AND USE THEREOF**

Publication Classification

(71) Applicant: **TIANJIN INSTITUTE OF INDUSTRIAL BIOTECHNOLOGY, CHINESE ACADEMY OF SCIENCES, Tianjin (CN)**

(51) **Int. Cl.**
C12P 13/24 (2006.01)
C07K 14/34 (2006.01)
C12N 1/20 (2006.01)
C12N 9/12 (2006.01)
C12N 9/02 (2006.01)

(72) Inventors: **Jibin SUN, Tianjin (CN); Jiao LIU, Tianjin (CN); Ping ZHENG, Tianjin (CN); Moshi LIU, Tianjin (CN); Yu WANG, Tianjin (CN); Wenjuan ZHOU, Tianjin (CN); Guannan SUN, Tianjin (CN)**

(52) **U.S. Cl.**
CPC *C12P 13/24* (2013.01); *C07K 14/34* (2013.01); *C12N 1/205* (2021.05); *C12N 9/1217* (2013.01); *C12Y 207/02011* (2013.01); *C12N 9/0071* (2013.01); *C12Y 114/11002* (2013.01); *C12R 2001/15* (2021.05)

(21) Appl. No.: **18/260,861**

(22) PCT Filed: **Jan. 11, 2022**

(86) PCT No.: **PCT/CN2022/071346**

§ 371 (c)(1),

(2) Date: **Jul. 10, 2023**

(57) **ABSTRACT**

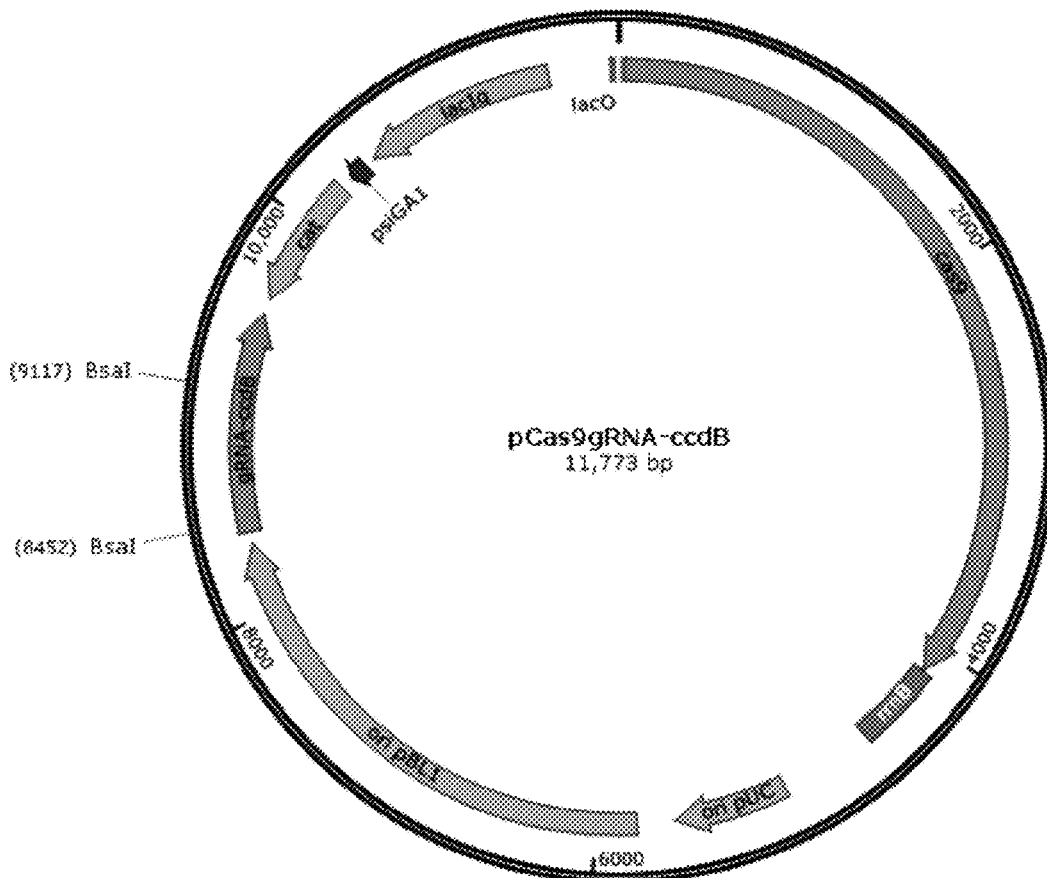
A protein having an L-proline efflux function and the use thereof are provided. A method for producing L-proline or hydroxyproline by means of using a protein ThrE is used for producing L-proline by means of enhancing the activity of a polypeptide, having an L-proline efflux function, in an L-proline-producing strain. Alternatively, the method is used for producing hydroxyproline by means of weakening the activity of a polypeptide, having an L-proline efflux function, in L-proline-producing host cells and enhancing the activity of a proline hydroxylase.

(30) **Foreign Application Priority Data**

Jan. 13, 2021 (CN) 202110045175.4

Specification includes a Sequence Listing.

Created with SnapGene™



Created with SnapGene™

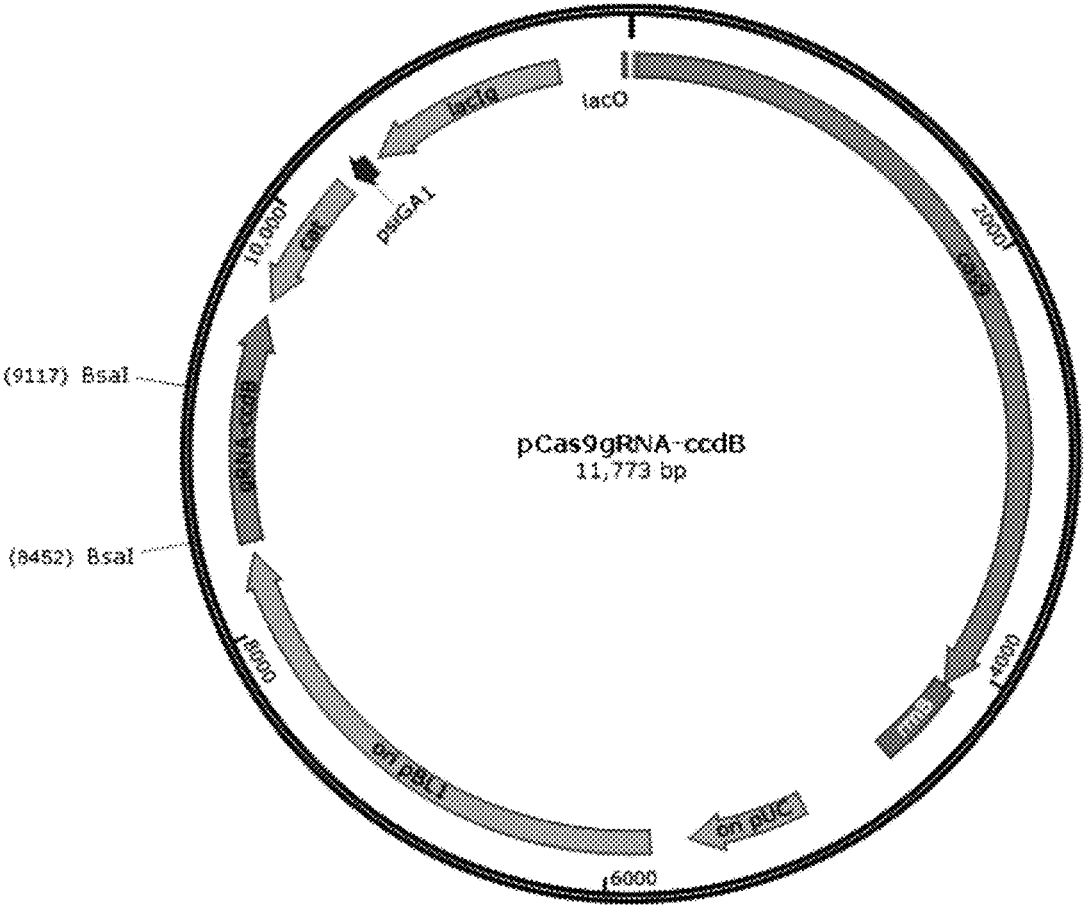


FIG. 1

Created with SnapGene™

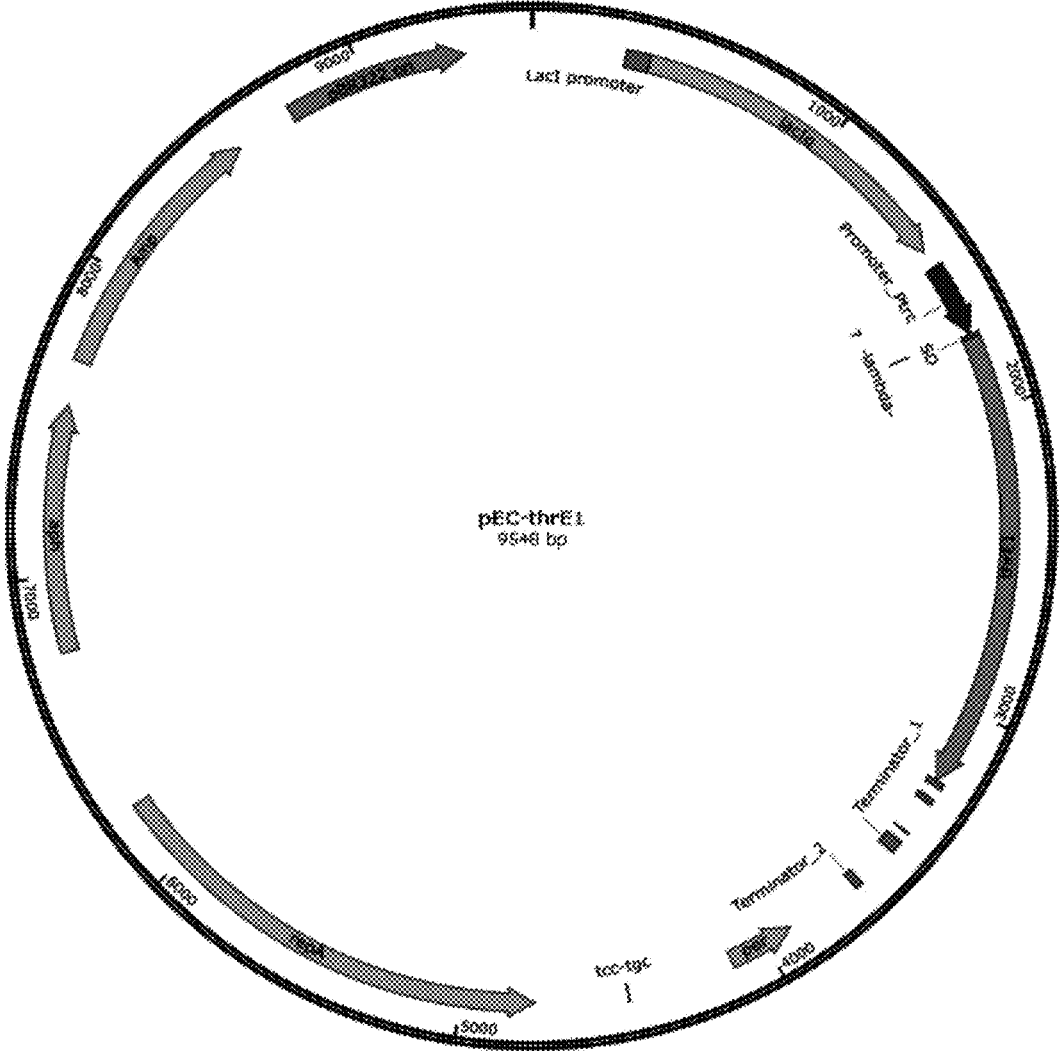


FIG. 2

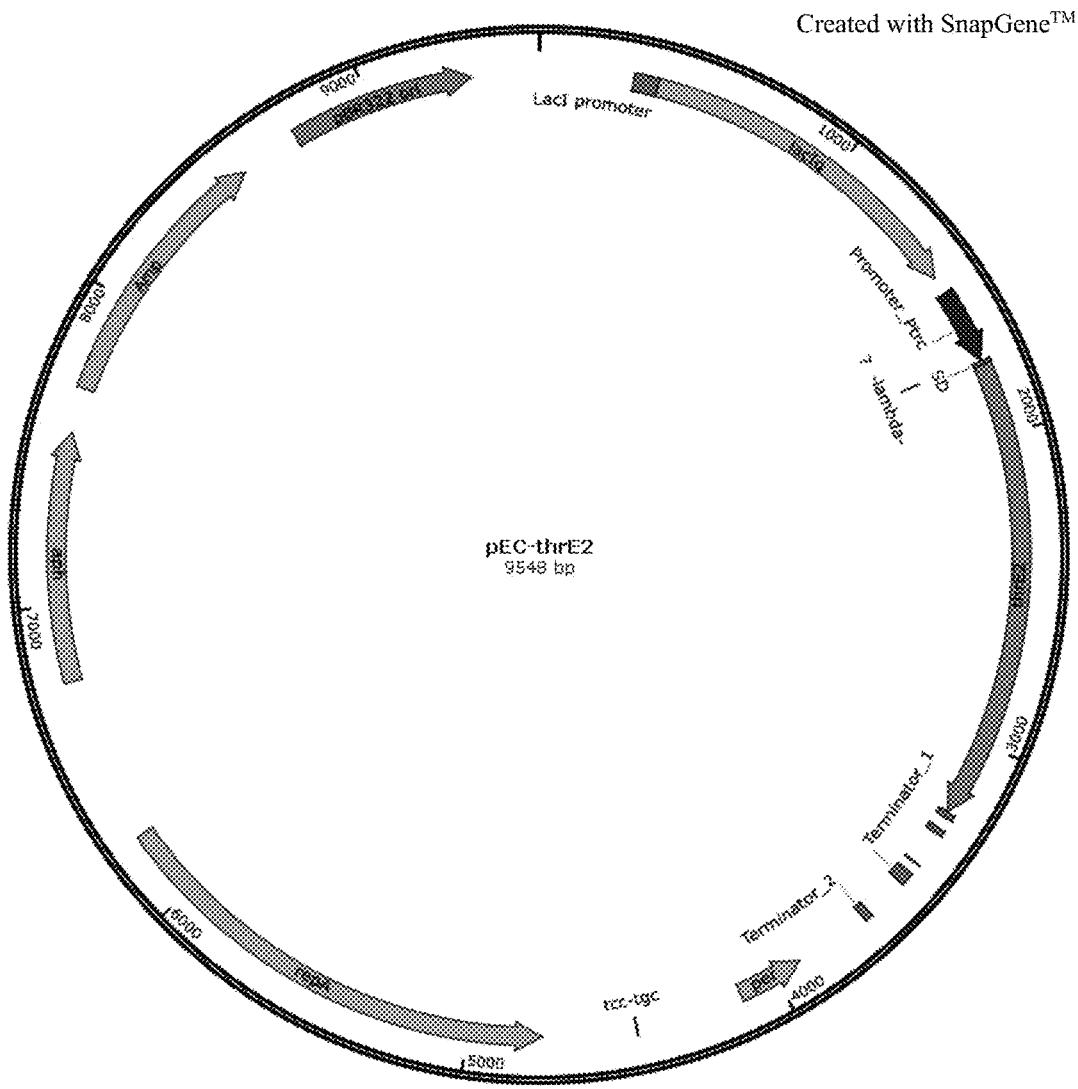


FIG. 3

**PROTEIN HAVING L-PROLINE EFFLUX
FUNCTION, AND USE THEREOF**

[0001] The present application claims priority to Chinese Patent Application No. 2021100451754 filed with

[0002] China National Intellectual Property Administration on Jan. 13, 2021, entitled "PROTEIN HAVING L-PROLINE EFFLUX FUNCTION, AND USE THEREOF", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0003] The present disclosure belongs to the field of molecular biology and bioengineering, and particularly relates to novel use of ThrE, a threonine efflux protein of *Corynebacterium glutamicum*, as an L-proline efflux protein, and a method for producing L-proline and a derivative thereof by using the protein ThrE.

BACKGROUND

[0004] L-proline is a naturally occurring non-essential amino acid in the human body and has a wide range of applications in the fields of clinics, biomaterials, industry, and the like. Methods for producing L-proline mainly include chemical methods and fermentation methods. The chemical extraction method has gradually lost its marketability due to the serious pollution and high cost, while the microbial fermentation method has become the most widely used method in the modern industry due to its advantages of low production cost, high productivity, high specificity, low environmental pollution, and the like. Currently, the commonly used industrial fermentation strains are *Bacillus* and *Escherichia*, commonly used *Escherichia* such as *Escherichia coli*, commonly used *Bacillus* such as *Corynebacterium glutamicum*, short bacillus such as *Brevibacterium flavum*, *Brevibacterium lactofermentus*, as well as some species of the genus *Arthrobacter* and some species of the genus *Microbacterium*. Among them, *Corynebacterium glutamicum* has physiological superiority and thus has become the most important strain used in industry for producing amino acids and other products. In *Corynebacterium*, L-proline is produced mainly by glutamic acid as a substrate catalyzed by γ -glutamyl kinase (glutamate-5-kinase, ProB), glutamate-semialdehyde dehydrogenase (ProA), and pyrroline-5-carboxylate reductase (ProC). In the prior art, L-proline is mainly produced by genetic modifications of ProA and ProB, the key enzymes in a synthetic pathway of L-proline. For example, a mutation at position 149 in the ProB protein derived from *Corynebacterium glutamicum* can relieve the feedback inhibition of L-proline and increase the yield of L-proline in engineered strains, as reported in CN101084312A. In the fermentation process, the timely transport of an amino acid synthesized in a microorganism to the outside of a cell through a transport pathway is also very important for promoting the biosynthesis of the amino acid. For example, in the production of lysine, the yield of lysine in an engineered strain can be remarkably increased by overexpression of a lysine efflux protein LysE (CN1283797C). However, the use of L-proline efflux protein to increase the yield of L-proline in the production of L-proline has not been reported. Therefore, there is an urgent need in the field to discover an effective L-proline efflux protein, so as to efficiently excrete L-proline produced in a cell to the outside of the cell and obtain a high-yield L-proline-producing engineered bacterium.

SUMMARY

[0005] In order to overcome the problems in the prior art, in the present disclosure, L-proline efflux proteins of *Corynebacterium glutamicum* were screened by a whole genome-scale membrane transporter inhibition library. Based on the screening in the present disclosure, it was found that the inhibition of thrE (Cg12622, which has been confirmed in the literature as the gene encoding the threonine efflux protein) gene resulted in a significant decrease in the specific productivity of L-proline. Overexpression of ThrE in a bacterium for producing L-proline, SZCgP1 (*Corynebacterium glutamicum* ATCC13869, in which a G149D mutation of ProB protein was introduced, with a codon mutation from GGT to GAT, see CN101084312A for details), resulting in a significantly higher extracellular proline concentration throughout the fermentation process and a significantly lower intracellular proline concentration throughout the fermentation process (especially during the rapid acidogenic phase), as compared to a control strain SZCgP1(pEC-control). The above results indicate that ThrE is an L-proline efflux protein.

[0006] An objective of the present disclosure is to provide the use of the proline efflux protein in producing L-proline or hydroxyproline.

[0007] Another objective of the present disclosure is to provide a strain producing L-proline or hydroxyproline.

[0008] Still, another objective of the present disclosure is to provide a method for producing L-proline or hydroxyproline.

[0009] Still, yet another objective of the present disclosure is to provide a method for constructing a strain producing L-proline or hydroxyproline.

[0010] In a first aspect of the present disclosure, provided is use of a polypeptide as a proline efflux protein in producing L-proline and hydroxyproline, the polypeptide being:

[0011] A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or

[0012] B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of an L-proline efflux function; or

[0013] C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function.

[0014] In another preferred embodiment, the polypeptide is a polypeptide derived from a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or 2, is formed by adding one or several, preferably 1-20, more preferably 1-15, more preferably 1-10, more preferably 1-3, and most preferably 1 amino acid residue at either or both ends of the amino acid sequence set forth in SEQ ID NO: 1 or 2, and has an L-proline efflux function.

[0015] In another preferred embodiment, the polypeptide has an amino acid sequence as set forth in SEQ ID NO: 1 or SEQ ID NO: 2.

[0016] In a second aspect of the present disclosure, provided is a strain for producing L-proline, wherein the strain expresses the following polypeptide:

[0017] A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or

- [0018] B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of an L-proline efflux function; or
- [0019] C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function.
- [0020] In another preferred embodiment, the polypeptide is a polypeptide derived from a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or 2, is formed by adding one or several, preferably 1-20, more preferably 1-15, more preferably 1-10, more preferably 1-3, and most preferably 1 amino acid residue at either or both ends of the amino acid sequence set forth in SEQ ID NO: 1 or 2, and has an L-proline efflux function.
- [0021] In another preferred embodiment, the polypeptide has an amino acid sequence as set forth in SEQ ID NO: 1 or SEQ ID NO: 2.
- [0022] In another preferred embodiment, the strain is a bacterium.
- [0023] In another preferred embodiment, the strain is selected from *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067.
- [0024] In another preferred embodiment, glutamate kinase in the bacterium is not subjected to feedback inhibition by L-proline or is subjected to attenuated feedback inhibition by L-proline.
- [0025] In another preferred embodiment, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the bacterium is enhanced.
- [0026] In a third aspect of the present disclosure, provided is a method for producing L-proline, the method comprising: culturing the strain described above to produce L-proline.
- [0027] In another preferred embodiment, the method further comprises a step of isolating L-proline from a fermentation broth.
- [0028] In a fourth aspect of the present disclosure, provided is a method for constructing a strain for producing L-proline, the method comprising enhancing the activity of the following polypeptide having an L-proline efflux function in the strain:
- [0029] A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or
- [0030] B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of an L-proline efflux function; or
- [0031] C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function.
- [0032] In another preferred embodiment, the polypeptide is a polypeptide derived from a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or 2, is formed by adding one or several, preferably 1-20, more preferably 1-15, more preferably 1-10, more preferably 1-3, and most preferably 1 amino acid residue at either or both ends of the amino acid sequence set forth in SEQ ID NO: 1 or 2, and has an L-proline efflux function.
- [0033] In another preferred embodiment, the polypeptide has an amino acid sequence as set forth in SEQ ID NO: 1 or SEQ ID NO: 2.
- [0034] In another preferred embodiment, the strain is a bacterium.
- [0035] In another preferred embodiment, the strain is selected from *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067.
- [0036] In another preferred embodiment, glutamate kinase in the bacterium is not subjected to feedback inhibition by L-proline or is subjected to attenuated feedback inhibition by L-proline.
- [0037] In another preferred embodiment, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the bacterium is enhanced.
- [0038] In a fifth aspect of the present disclosure, provided is a host cell in which the activity of a polypeptide having an L-proline efflux function is attenuated, wherein the polypeptide having the activity of the L-proline efflux function is:
- [0039] A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or
- [0040] B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of the L-proline efflux function; or
- [0041] C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function.
- [0042] In another preferred embodiment, the host cell is selected from *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067.
- [0043] In another preferred embodiment, glutamate kinase in the host cell is not subjected to feedback inhibition by L-proline or is subjected to attenuated feedback inhibition by L-proline.
- [0044] In another preferred embodiment, the activity of a proline hydroxylase in the host cell is enhanced.
- [0045] In another preferred embodiment, the proline hydroxylase is trans-proline-4-hydroxylase having a nucleotide sequence as set forth in SEQ ID NO: 5.
- [0046] In another preferred embodiment, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the bacterium is enhanced.
- [0047] In a sixth aspect of the present disclosure, provided is a method for producing hydroxyproline such as trans-4-hydroxy-L-proline and the like, the method comprising:

culturing the host cell described above to produce hydroxyproline such as trans-4-hydroxy-L-proline and the like.

[0048] In another preferred embodiment, the method further comprises a step of isolating hydroxyproline such as trans-4-hydroxy-L-proline and the like from a fermentation broth.

[0049] In a seventh aspect of the present disclosure, provided is a method for constructing a strain for producing hydroxyproline such as trans-4-hydroxy-L-proline and the like, wherein the method comprises:

[0050] (1) attenuating the activity of the following polypeptide having an L-proline efflux function in the strain, wherein the polypeptide having the activity of the L-proline efflux function is:

[0051] A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or

[0052] B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of the L-proline efflux function; or

[0053] C) a polypeptide with homology higher than 90%, 95%, 96%, 97%, 98%, or 99% with the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function;

[0054] (2) relieving or attenuating feedback inhibition by L-proline on glutamate kinase; and

[0055] (3) introducing a gene encoding a proline hydroxylase into the strain obtained in step (2).

[0056] In another preferred embodiment, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the bacterium is enhanced.

[0057] In another preferred embodiment, the proline hydroxylase is trans-proline-4-hydroxylase having a nucleotide sequence as set forth in SEQ ID NO: 5.

[0058] The beneficial effects of the present disclosure are as follows:

[0059] 1. In the present disclosure, after screening, ThrE, a threonine efflux protein, is found to be an L-proline efflux protein of *Corynebacterium glutamicum*. Production bacteria for L-proline, SZCgP1 and SZCgP3, are constructed based on strains ATCC13869 and ATCC13032 of *Corynebacterium glutamicum*, and overexpression of ThrE in the two strains can increase the yield of L-proline by 1.57 times and 2.56 times, respectively. Therefore, the present disclosure can be used in practice for producing L-proline by bacterial fermentation, is convenient to popularize and apply, and has considerable industrial application value.

[0060] 2. In the present disclosure, ThrE is also used in the production of hydroxyproline such as trans-4-hydroxy-L-proline and the like, and knockout of thrE gene remarkably reduces the yield of extracellular L-proline by-products and increases the ratio of hydroxyproline such as trans-4-hydroxy-L-proline and the like to L-proline, helping to obtain hydroxyproline such as trans-4-hydroxy-L-proline and the like with higher purity. The knockout of thrE gene is of great significance to the production of hydroxyproline products such as trans-4-hydroxy-L-proline and the like.

[0061] 3. The present disclosure provides a novel approach for constructing a hydroxyproline-producing strain by knocking out the L-proline efflux protein in the strain. This allows for an increase in the intracel-

lular concentration of L-proline, which serves as a substrate for hydroxylation by proline hydroxylase. As a result, the intracellular L-proline is converted into hydroxyproline, thereby offering valuable insights for the development of higher-yielding hydroxyproline-producing strains in the field of biotechnology.

BRIEF DESCRIPTION OF THE DRAWINGS

[0062] FIG. 1 shows a plasmid map of pCas9gRNA-ccdB;

[0063] FIG. 2 shows a plasmid map of pEC-thrE1; and

[0064] FIG. 3 shows a plasmid map of pEC-thrE2.

DETAILED DESCRIPTION

[0065] The technical solutions of the present disclosure will be further illustrated in detail with reference to the following specific examples. It will be appreciated that the following examples are merely exemplary illustrations and explanations of the present disclosure, and should not be construed as limiting the protection scope of the present disclosure. All techniques implemented based on the content of the present disclosure described above are included within the protection scope of the present disclosure.

[0066] Unless otherwise stated, the starting materials and reagents used in the following examples are all commercially available products or can be prepared using known methods. Where specific conditions are not indicated in experimental methods in the following examples, conventional conditions such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual* (New York: Cold Spring Harbor Laboratory Press, 1989) or conditions recommended by the manufacturer are followed.

Definitions and Description:

[0067] The term “polypeptide having an L-proline efflux function” used herein refers to a protein or a polypeptide that can excrete L-proline in a cell to the outside of the cell, which may be a protein set forth in SEQ ID NO: 1 or SEQ ID NO: 2, or a protein or a polypeptide that is obtained by mutating SEQ ID NO: 1 or SEQ ID NO: 2 and still has an L-proline efflux function.

[0068] The terms “protein”, “polypeptide” and “peptide” used herein are used interchangeably, have the meanings commonly understood by one of ordinary skill in the art, and refer to amino acid polymers of any length. The polymers may be linear or branched. They may comprise modified amino acids and may be interrupted by non-amino acids. The terms also encompass amino acid polymers that have been modified (e.g., disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other operation, such as conjugation with a labeling component).

[0069] The term “fusion protein” used herein refers to a protein obtained by fusing the mutant protein with a protein tag. The protein tag can be located at the N-terminus of the mutant protein or at the C-terminus of the mutant protein. The mutant protein and the protein tag may also have a spacer amino acid residue therebetween, and specifically may have fewer than 10 spacer amino acid residues therebetween.

[0070] The term “host cell” herein means any type of cell that is susceptible to transformation, transfection, transduction, and the like using the polypeptide having the activity of the L-proline efflux function of the present disclosure, a polynucleotide encoding the polypeptide, or an expression

vector. The term “recombinant microorganism” encompasses host cells which differ from parent cells after introduction of a transcription initiation element or a recombinant expression vector, specifically, the recombinant host cells are achieved by transformation. For example, suitable host cells for use in the present disclosure include, but are not limited to, *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067.

[0071] The term “transformation” herein has the meaning generally understood by those skilled in the art, i.e., the process of introducing exogenous DNA into a host. Methods for the transformation include any method for introducing a nucleic acid into a cell, including, but not limited to, electroporation, calcium phosphate precipitation, calcium chloride (CaCl₂) precipitation, microinjection, polyethylene glycol (PEG) method, DEAE-dextran method, cationic liposome method, and lithium acetate-DMSO method.

[0072] The culture of the host cell herein may be performed according to a conventional method in the art, including, but not limited to, well plate culture, shake-flask culture, batch culture, continuous culture, fed-batch culture, and the like, and various culture conditions such as temperature, time, pH of a medium, and the like may be appropriately adjusted according to actual circumstances.

[0073] The terms “comprise”, “have”, “include”, or “contain” used herein are intended to be inclusive or open-ended and do not exclude additional, unspecified elements or method steps.

[0074] The term “about” used herein means that a value includes the standard deviation of error for the device or method used to determine the value.

[0075] The term “or” used herein is defined as merely an alternative and “and/or”, but unless it is explicitly stated to be only an alternative or mutually exclusive between alternatives, the term “or” in the claims means “and/or”.

[0076] The selected/optional/preferred “numerical range” used herein includes both numerical endpoints at both ends of the range and all natural numbers covered between the numerical endpoints with respect to the aforementioned numerical endpoints.

[0077] The terms “wild-type” and “naturally occurring” used herein refer to an object that can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism, that can be isolated from a source in nature, and that has not been intentionally modified by humans in the laboratory, is naturally occurring.

[0078] The term “amino acid mutation” or “nucleotide mutation” used herein includes “substitution, repetition, deletion, or addition of one or more amino acids or nucleotides”. In the present disclosure, the term “mutation” refers to an alteration in nucleotide sequence or amino acid sequence. In a specific embodiment, the term “mutation” refers to “substitution”.

[0079] The term “native state” used herein refers to the activity of a polypeptide in a microorganism in an unmodified state, i.e., the activity in the native state.

[0080] The term “having the activity of an L-proline efflux function” used herein has the same or similar meaning as conventionally understood by those skilled in the art. It

means that the amino acid sequence of a fragment is a portion of the amino acid sequence of an intact protein or polypeptide, and has the same or similar function or activity as the intact protein or polypeptide. Specifically, in the present disclosure, it indicates any amino acid fragment obtained from the ThrE protein of the present disclosure and having an L-proline efflux function.

[0081] It is known to those skilled in the art that, for mutating a wild-type polypeptide in order to improve the activity, it is more important to find a site that can achieve the desired purpose. Therefore, based on the teachings of the present disclosure, a person skilled in the art will mutate the amino acid sequence set forth in SEQ ID NO: 1 or 2 and test the mutant for the relevant activity. In addition, it is not difficult for one of ordinary skill in the art to know that altering a few amino acid residues in certain regions of a polypeptide, e.g., in non-critical regions, does not substantially alter biological activity. For example, the sequence obtained by properly replacing some amino acids does not affect its activity (see Watson et al., *Molecular Biology of The Gene*, fourth edition, 1987, The Benjamin/Cummings Pub. Co. P224). Therefore, one of ordinary skill in the art would be able to implement such replacements and ensure that the resulting molecule retains the desired biological activity.

[0082] Therefore, it is apparent that further mutations of the L-proline efflux protein of the present disclosure result in further mutants that still have corresponding functions and activities. For example, it is well known to those skilled in the art that adding or removing several amino acid residues, e.g., preferably 1-20, more preferably 1-15, more preferably 1-10, more preferably 1-3, and most preferably 1 amino acid residue, to or from either end of a polypeptide does not affect the function of the resulting mutant. For example, for ease of purification, a skilled person usually makes the resulting protein have a 6xHis tag at either end, and such protein has the same function as a protein that does not have the 6xHis tag. Therefore, the present disclosure shall include conservative mutations obtained based on the present disclosure.

[0083] The term “conservative mutation” refers to a mutation that can normally maintain the function of a protein. A representative example of the conservative mutation is a conservative replacement.

[0084] As used in the present disclosure, the term “conservative replacement” relates to the replacement of an amino acid residue with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art and include those having basic side chains (e.g., lysine, arginine, and histidine), acidic side chains (e.g., aspartic acid and glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, and cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, and tryptophan), (3-branched chains (e.g., threonine, valine, and isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, and histidine).

[0085] As used in the present disclosure, the “conservative replacement” typically refers to the substitution of one type of amino acid at one or more sites in a protein. Such substitution may be conservative. Examples of replacements regarded as conservative replacements include replacement of Ala with Ser or Thr, replacement of Arg with Gln, His or Lys, replacement of Asn with Glu, Gln, Lys, His or Asp,

replacement of Asp with Asn, Glu or Gln, replacement of Cys with Ser or Ala, replacement of Gln with Asn, Glu, Lys, His, Asp or Arg, replacement of Glu with Gly, Asn, Gln, Lys or Asp, replacement of Gly with Pro, replacement of His with Asn, Lys, Gln, Arg or Tyr, replacement of Ile with Leu, Met, Val or Phe, replacement of Leu with Ile, Met, Val or Phe, replacement of Lys with Asn, Glu, Gln, His or Arg, replacement of Met with Ile, Leu, Val or Phe, replacement of Phe with Trp, Tyr, Met, Ile, or Leu, replacement of Ser with Thr or Ala, replacement of Thr with Ser or Ala, replacement of Trp with Phe or Tyr, replacement of Tyr with His, Phe or Trp, and replacement of Val with Met, Ile or Leu. In addition, the conservative mutations also include naturally occurring mutations caused by individual differences, differences in strain and species, and the like from which genes are derived.

[0086] The term “corresponding to” used herein has the meaning commonly understood by one of ordinary skill in the art. Specifically, the term “corresponding to” refers to a position in a sequence corresponding to a specified position in another sequence after alignment of the two sequences by homology or sequence identity. Thus, for example, in terms of “an amino acid residue corresponding to position 40 of the amino acid sequence set forth in SEQ ID NO: 1”, if a 6xHis tag is added to one end of the amino acid sequence set forth in SEQ ID NO: 1, then position 40 of the resulting mutant corresponding to the amino acid sequence set forth in SEQ ID NO: 1 may be position 46.

[0087] In a specific embodiment, the homology or sequence identity may be no less than 90%, preferably no less than 95%, and more preferably 96%, 97%, 98%, or 99% homology.

[0088] Methods for determining sequence homology or identity well known to one of ordinary skill in the art include, but are not limited to: *Computational Molecular Biology*, Ed. Lesk, A. M., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Ed. Smith, D. W., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part One, Ed. Griffin, A. M. and Griffin, H. G., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; *Sequence Analysis Primer*, Ed. Gribskov, M. and Devereux, J., M Stockton Press, New York, 1991; and Carillo, H. and Lipman, D., *SIAM J. Applied Math.*, 48:1073 (1988). A preferred method for determining identity is to obtain the greatest match between the sequences tested. Methods for determining identity are compiled in publicly available computer programs. Preferred computer program methods for determining identity between two sequences include, but are not limited to: GCG program package (Devereux, J. et al., 1984), BLASTP, BLASTN, and FASTA (Altschul, S. F. et al., 1990). The BLASTX program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S. et al., NCBI NLM NIH Bethesda, Md.20894; Altschul, S. et al., 1990). The well-known Smith Waterman algorithm can also be used to determine identity.

[0089] Unless defined otherwise, or clear from the background, all technical and scientific terms used in the present disclosure have the same meaning as commonly understood by one of ordinary skill in the art to which the present disclosure belongs.

EXAMPLE 1

Screening of L-Proline Efflux Proteins of *Corynebacterium Glutamicum* in *Corynebacterium Glutamicum* ATCC13869

[0090] The inventors screened L-proline efflux proteins of *Corynebacterium glutamicum* by a whole genome-scale membrane transporter inhibition library. Firstly, the cas9 gene of the pCas9 plasmid was mutated with D10A and H840A, and the Bsa I restriction enzyme site in the plasmid backbone was removed, resulting in the pdCas9 plasmid (LIU, Jiao, et al. Development of a CRISPR/Cas9 genome editing toolbox for *Corynebacterium glutamicum*. Microbial cell factories, 2017, 16.1: 205.). Subsequently, a gRNA-ccdB expression cassette was amplified from a pncas9 (D10A)-AID-gRNA-ccdB^{TS} plasmid (WANG, Yu, et al. Expanding targeting scope, editing window, and base transition capability of base editing in *Corynebacterium glutamicum*. Biotechnology and bioengineering, 2019, 116: 3016-3029) and cloned into the same position of pdCas9 to obtain pdCas9gRNA-ccdB as a CRISPRi plasmid that could be efficiently constructed and had a sequence as set forth in SEQ ID NO: 3. The inhibition efficiency of the system could reach 96%. Based on the above ingenious design, for the inhibition of the target gene, only a sgRNA with a 20 bp target DNA binding region was required to be designed, then two complementary primers of 24 bp were synthesized and annealed to form a double-stranded DNA fragment, and an inhibition plasmid of the target gene was constructed by Golden Gate cloning. The inventors also constructed a control plasmid pdCas9gRNA-control without the 20 bp DNA binding region. Whole genome-scale protein sequences of *Corynebacterium glutamicum* ATCC13869 were used in a TransportDB database for membrane transporter protein prediction to obtain 397 membrane transporters on the whole genome scale. Thereafter, inhibition plasmids of genes of the 397 membrane transporters were constructed based on the CRISPRi system, and the inhibition library plasmids and the control plasmid were separately introduced into a bacterium for producing L-proline, SZCgP1 (*Corynebacterium glutamicum* ATCC13869, in which a G149D mutation of proB gene was introduced, with a codon mutation from GGT to GAT, see CN101084312A for details), so as to obtain inhibition library strains. Finally, all of the strains were evaluated for the specific productivity of L-proline by fermentation on 96-well plates and genes with decreased specific productivity of L-proline after inhibition were potential L-proline efflux protein genes. Based on the above screening, the inventors found that the inhibition of thrE (Cg12622, which has been confirmed in the literature as an L-threonine efflux protein) gene resulted in a significant decrease in the specific productivity of L-proline. Primers for the construction of the CRISPRi plasmids of the thrE gene were thrE-F and thrE-R (Table 1). The control strain was SZCgP1(pdCas9gRNA-control), the inhibition plasmid of the thrE gene was pdCas9gRNA-thrE, and the inhibition strain of the thrE gene was SZCgP1(pdCas9gRNA-thrE).

TABLE 1

Primer	Nucleotide sequence	SEQ ID NO.
thrE-F	TTCACACTGCTC GAACCTGTACTT	6
thr E-R	AAACAGGTACAA GTTTCGAGCAGTG	7

[0091] To further confirm the effect of the thrE gene inhibition on L-proline production, evaluations were performed by fermentation using a 24-well deep-well plate. The seed culture medium contained the following ingredients (g/L): 5 g/L glucose, 1 g/L yeast powder, 3 g/L soybean peptone, 3 g/L urea, 0.5 g/L succinic acid, 1 g/L K₂HPO₄·3H₂O, 0.1 g/L MgSO₄·7H₂O, 0.01 mg/L biotin, 0.1 mg/L vitamin B1, and 20 g/L MOPS. The fermentation culture medium contained the following ingredients: 80 g/L glucose, 1 g/L yeast powder, 1 g/L soybean peptone, 1 g/L NaCl, 1 g/L ammonium sulfate, 6 g/L urea, 1 g/L K₂HPO₄·3H₂O, 0.45 g/L MgSO₄·7H₂O, 0.05 g/L FeSO₄·7H₂O, 0.4 mg/L biotin, 0.1 mg/L vitamin B1, and 40 g/L MOPS, and the initial pH was 7.2. Firstly, the strains were inoculated into a seed culture medium containing 15 µg/mL chloramphenicol and cultured for 8 h. The cultures were inoculated as seeds into a 24-well plate containing 15 µg/mL chloramphenicol and a 0.03 mM IPTG fermentation culture medium. The volume of the medium was 800 µL, and the inoculation amount was controlled with the initial OD₆₀₀ being 0.03 (as detected by a microplate reader). The seeds were cultured at 30° C. for 21 h, with the rotation speed of a plate shaker being 800 rpm. The experiment for each strain was performed in triplicate. After fermentation, OD₆₀₀ and the yield of L-proline were detected. The detection of L-proline was performed as follows: The fermentation solution was diluted with 3% (W/V) sulfosalicylic acid to a proper concentration. 1 mL of the dilution was taken, and 1 mL of an acid-ninhydrin complex (1.25 g of ninhydrin dissolved in 30 mL of glacial acetic acid and 20 mL of 6M H₃PO₄ by heating at 70° C.) and 1 mL of glacial acetic acid were added. The mixture was reacted in a boiling water bath at 100° C. for 45 min. After cooling, OD₅₂₀ was determined. A standard curve was plotted by adopting L-proline at concentrations of 0-100 mg/L, and the concentrations of the samples to be tested were calculated according to the standard curve. The results are shown in Table 2, indicating that the yield and specific productivity of L-proline of the strains decreased by 63% and 49%, respectively, after the inhibition of thrE gene expression.

TABLE 2

Effect of inhibition of thrE gene expression on yield and specific productivity of L-proline			
Strain	OD ₆₀₀	Yield of L-proline (g/L)	Specific productivity of L-proline (g/L/OD ₆₀₀)
SZCgP1(pdCas9gRNA-control)	13.60 ± 1.42	5.05 ± 0.50	0.37 ± 0.00
SZCgP1(pdCas9gRNA-thrE)	9.91 ± 0.64	1.87 ± 0.29	0.19 ± 0.03

EXAMPLE 2

Identification of L-Proline Efflux Proteins of *Corynebacterium Glutamicum* in *Corynebacterium Glutamicum* ATCC13869

[0092] (1) Construction of Knockout, Complementation and Overexpression Strains of thrE Gene in SZCgP1 Strains [0093] The inventors constructed knockout mutants based on CRISPR/Cas9 genome editing technology. Using a pCas9 plasmid (LIU, Jiao, et al. Development of a CRISPR/Cas9 genome editing toolbox for *Corynebacterium glutamicum*. *Microbial cell factories*, 2017, 16.1: 205.) as a template and cas9-1/cas9-2 as primers, a lacO mutation (a mutation from TGTGTGGAATTGTGAGCGGATAACAATTTACACA to TGTGTGGAATTGTGAGCGCTCACAAATTTACACA) and an RBS mutation (a mutation from AAAGGAGTTGAGA to AAAGGCACCCGAT) as manipulation elements were introduced on the primers, and a fragment containing cas9 was amplified. Thereafter, a plasmid backbone fragment containing a gRNA-ccdB expression cassette and a temperature-sensitive replication origin was amplified by using a pCas9(D10A)-AID-gRNA-ccdB^{TS} plasmid (WANG, Yu, et al. Expanding targeting scope, editing window, and base transition capability of base editing in *Corynebacterium glutamicum*. *Biotechnology and bioengineering*, 2019, 116: 3016-3029) as a template and gRNA-1/gRNA-2 as primers. The above 2 fragments were cloned and ligated by a one-step recombination kit from Vazyme to obtain a pCas9gRNA-ccdB plasmid that could be efficiently constructed and had a plasmid map as shown in FIG. 1 and a sequence as set forth in SEQ ID NO: 4. According to the reported genome sequence of *Corynebacterium glutamicum* ATCC13869, upstream and downstream homologous arms were amplified by using ATCC13869 genome as a template with thrE-1/ thrE-2 and thrE-3/ thrE-4 as primers, respectively. 3 plasmid backbone fragments were amplified by using a pCas9gRNA-ccdB plasmid as a template with cas9-3/cas9-4, cas9-5/cas9-6, and gRNA-3/gRNA-4 as primers, respectively, and a gRNA with 20bp target DNA binding region was also introduced. The above 5 fragments were cloned and ligated by a one-step recombination kit from Vazyme to obtain a pCas9gRNA-thrE plasmid. SZCgP1 strain was prepared to be competent, and about 2 µg of the pCas9gRNA-thrE plasmid was electrotransformed. 1 mL of a TSB culture medium pre-heated at 46° C. was added, and the heat shock was performed at 46° C. for 6 min. The mixture was incubated at 30° C. for 3 h, applied onto a TSB plate containing 15 µg/mL chloramphenicol and 0.05 mM IPTG, and cultured for 2 days. A correct mutant verified by PCR and sequencing was named as SZCgP2 strain.

[0094] To construct the overexpression plasmid, the thrE gene was cloned onto the pEC-ccdB plasmid, which is derived from the pEC-XK99E plasmid. The pEC-ccdB plasmid was digested using a Bsa I enzyme. According to the reported genome sequence of *Corynebacterium glutamicum* ATCC13869, a fragment of thrE1 gene was amplified with thrE-5/thrE-6 by using the genome of ATCC13869 as a template, and an amino acid sequence of a protein encoded by the thrE1 gene is as set forth in SEQ ID NO: 1. The two segments were cloned and ligated by a one-step recombination kit from Vazyme to obtain a pEC-thrE1 plasmid having a plasmid map as shown in FIG. 2. A control plasmid pEC-control, in which no gene was cloned, was also constructed. The pEC-thrE1 plasmid was transformed into the SZCgP2 strain to obtain a complementation strain SZCgP2 (pEC-thrE1). The pEC-thrE1 and pEC-control plasmids were transformed into SZCgP1 to obtain an overexpression strain SZCgP1(pEC-thrE1) and a control strain SZCgP1 (pEC-control), respectively.

TABLE 3

Primer	Nucleotide sequence	SEQ ID NO.
cas9-1	GCGCTCACAATTTCACACAGGAACAG AATTAATTAAGCTTAAAGGCACCCGAT ATGGATAAGAAATACTCAATAGGC	8
cas9-2	TCAGTCACCTCCTAGCTGACTCAAATC	9
gRNA-1	GTCAGCTAGGAGGTGACTGAAGCTTGG CTGTTTTGGCGGATG	10
gRNA-2	CTGTGTGAAATTGTGAGCGCTCACAAT TCC	11
Cas9-3	TCGAAGGGCACCAATAACTGC	12
Cas9-4	CTTTTACTTTACCAGCGTTTCTG	13
Cas9-5	AACGCTGGTGAAAGTAAAGATGC	14
Cas9-6	GAATGGAACGGATCAACGGTGAATTA CACTGTACTCTGTTGCGTC	15
gRNA-3	CCGTTTGATCCGTTCCATTCGTTTTAG AGCTAGAAATAGCAAG	16
gRNA-4	CAACCTGCCATCACGAGATTTTC	17
thrE-1	AATCTCGTGATGGCAGGTTGACGGGAG CCTAAGTGAATG	18
thrE-2	CCTTGGTAAATGCACGGTATGGGTCAG TGAGATCAATCGGG	19
thrE-3	CCATACCGTGCATTTACCAAGG	20
thrE-4	CAGTTATTGGTGCCTTCGACTTTGTG GTCTTCTTTGCGG	21
thrE-5	ACAGGCCAAAGGAGTTGAGAATGTTGA GTTTTGCGACCC	22
thrE-6	CCAAGCTTGCATGCCTGCAGTTACCTT TTATTACCGAATC	23

(2) Determination of Intracellular and Extracellular L-Proline Concentrations of Knockout, Complementation and Overexpression Strains of thrE Gene by Shake-Flask Fermentation

[0095] To further confirm whether the thrE gene is an L-proline efflux protein, intracellular and extracellular L-proline concentrations of knockout, complementation and overexpression strains were determined using shake-flask fermentation. The seed culture medium contained the following ingredients (g/L): 5 g/L glucose, 1 g/L yeast powder, 3 g/L soybean peptone, 3 g/L urea, 0.5 g/L succinic acid, 1 g/L $K_2HPO_4 \cdot 3H_2O$, 0.1 g/L $MgSO_4 \cdot 7H_2O$, 0.01 mg/L biotin, 0.1 mg/L vitamin B1, and 20 g/L MOPS. The fermentation culture medium contained the following ingredients: 80 g/L glucose, 1 g/L yeast powder, 1 g/L soybean peptone, 1 g/L NaCl, 1 g/L ammonium sulfate, 6 g/L urea, 1 g/L $K_2HPO_4 \cdot 3H_2O$, 0.45 g/L $MgSO_4 \cdot 7H_2O$, 0.05 g/L $FeSO_4 \cdot 7H_2O$, 0.4 mg/L biotin, 0.1 mg/L vitamin B1, and 40 g/L MOPS, and the initial pH was 7.2. Firstly, the strains were inoculated into a seed culture medium and cultured for 8 h. The cultures were inoculated as seeds into 250 mL triangular flasks containing a fermentation culture medium, the volume of the medium being 25 mL. 25 μ g/mL kanamycin and 0.05 mM IPTG were added as needed, and the inoculation amount was controlled with the initial OD_{600} being 0.1 (as determined by a spectrophotometer). The

mixture was cultured at 30° C. for 21 h, with the rotation speed of a shaker being 2200 rpm. The experiment for each strain was performed in triplicate. Samples were taken in the fermentation process to determine OD_{600} , intracellular L-proline concentration and extracellular L-proline concentration.

[0096] The intracellular and extracellular L-proline concentrations were determined as follows: 400 μ L of silicone oil was added into a 1.5 mL centrifuge tube in advance, and 400 μ L of the fermented bacterial solution was quickly pipetted and then slowly added into the tube. The mixture was centrifuged at 4° C. and 12000 rpm for 10 min, and the supernatant was obtained as an extracellular sample. The bacteria precipitate at the bottom of the centrifuge tube was cut off and placed in a 1.5 mL centrifuge tube. A proper amount of 3% sulfosalicylic acid was added according to the amount of the bacteria. The mixture was boiled at 100° C. for 20 min and centrifuged at 12000 rpm for 10 min, and the supernatant was obtained as an intracellular sample. The detection of L-proline was performed as described in Example 1, and the intracellular L-proline concentration was calculated based on the volume of *Corynebacterium glutamicum* reported in the literature.

[0097] The results of the determination of the intracellular and extracellular L-proline concentrations of each strain in the fermentation process are as shown in Tables 4 and 5. The extracellular L-proline concentration of the thrE gene-knockout strain SZCgP2 was significantly lower than that of the starting strain SZCgP1 throughout the fermentation process, and the intracellular L-proline concentration of the knockout strain was slightly higher than that of the starting strain SZCgP1 in the early stage (3-9 h) of the fermentation. The extracellular L-proline concentration of the complementation strain SZCgP2(pEC-thrE1) was significantly higher than that of the starting strain SZCgP1 throughout the fermentation process, and the intracellular L-proline concentration of the complementation strain was significantly lower than that of the starting strain during the rapid acidogenic phase (3-15 h). The extracellular L-proline concentration of the overexpression strain SZCgP1(pEC-thrE1) was significantly higher than that of the control strain SZCgP1(pEC-control) throughout the fermentation process, and the intracellular L-proline concentration of the overexpression strain was significantly lower than that of the control strain SZCgP1(pEC-control) throughout the fermentation process (especially during the rapid acidogenic phase). Based on the above results, it was indicated that ThrE is a proline efflux protein. The overexpression of ThrE in the SZCgP1 strain could increase the yield of L-proline, and the overexpression strain SZCgP1(pEC-thrE1) showed a 1.57-fold increase in the yield of extracellular L-proline at 21 h relative to the control bacterium SZCgP1(pEC-control) introduced with an empty plasmid, indicating that the overexpression of the L-proline efflux protein gene is of great significance in the modification of a strain producing L-proline.

TABLE 4

Intracellular L-proline concentrations of knockout, complementation and overexpression strains of thrE gene in <i>Corynebacterium glutamicum</i> ATCC13869					
Time	Intracellular L-proline concentration (mM)				
	SZCgP1	SZCgP2	SZCgP2 (pEC-thrE1)	SZCgP1 (pEC-thrE1)	SZCgP1 (pEC-control)
3 h	151 \pm 17	213 \pm 21	83 \pm 1	67 \pm 10	123 \pm 7
6 h	192 \pm 17	239 \pm 13	72 \pm 3	73 \pm 2	221 \pm 61

TABLE 4-continued

Intracellular L-proline concentrations of knockout, complementation and overexpression strains of thrE gene in <i>Corynebacterium glutamicum</i> ATCC13869					
Time	Intracellular L-proline concentration (mM)				
	SZCgP1	SZCgP2	SZCgP2 (pEC-thrE1)	SZCgP1 (pEC-thrE1)	SZCgP1 (pEC-control)
9 h	296 ± 18	315 ± 15	116 ± 19	106 ± 10	246 ± 37
12 h	287 ± 19	256 ± 14	112 ± 8	107 ± 5	228 ± 21
15 h	206 ± 37	235 ± 15	101 ± 2	92 ± 3	228 ± 11
18 h	103 ± 5	110 ± 6	105 ± 3	95 ± 2	159 ± 9
21 h	76 ± 6	70 ± 1	111 ± 4	105 ± 5	108 ± 9

TABLE 5

Extracellular L-proline concentrations of knockout, complementation and overexpression strains of thrE gene in <i>Corynebacterium glutamicum</i> ATCC13869					
Time	Extracellular L-proline concentration (g/L)				
	SZCgP1	SZCgP2	SZCgP2 (pEC-thrE1)	SZCgP1 (pEC-thrE1)	SZCgP1 (pEC-control)
3 h	0.05 ± 0.00	0.03 ± 0.00	0.08 ± 0.00	0.08 ± 0.00	0.04 ± 0.00
6 h	0.15 ± 0.00	0.03 ± 0.00	0.32 ± 0.03	0.30 ± 0.02	0.12 ± 0.01
9 h	0.65 ± 0.02	0.20 ± 0.04	1.40 ± 0.10	1.98 ± 0.26	0.47 ± 0.02
12 h	1.49 ± 0.37	0.47 ± 0.00	3.14 ± 0.35	3.09 ± 0.14	0.87 ± 0.09
15 h	5.41 ± 0.12	2.42 ± 0.05	11.61 ± 0.96	11.77 ± 1.14	4.40 ± 0.16
18 h	5.41 ± 1.09	3.40 ± 0.69	17.01 ± 0.96	16.40 ± 4.23	5.34 ± 1.48
21 h	5.56 ± 0.41	2.71 ± 0.29	17.87 ± 2.90	16.22 ± 4.31	6.32 ± 0.31

EXAMPLE 3

Use of L-Proline Efflux Protein in Production of L-Proline in *Corynebacterium Glutamicum* ATCC13032 Strain

[0098] Firstly, a strain producing L-proline, SZCgP3, was constructed based on a *Corynebacterium glutamicum* ATCC13032 strain, i.e., a G149K mutation of proB gene was introduced, with a codon mutation from GGT to AAG (see CN101084312A). The construction of a ThrE overexpression strain was similar to that of a *Corynebacterium glutamicum* 13869 strain. Overexpression of thrE was performed on a pEC-ccdB plasmid, a derivative plasmid of pEC-XK99E plasmid, and the pEC-ccdB plasmid was digested using a Bsa I enzyme. A fragment of thrE2 gene was amplified with thrE-5/thrE-6 by using the genome of *Corynebacterium glutamicum* ATCC13032 as a template, and an amino acid sequence of a protein encoded by the thrE2 gene is as set forth in SEQ ID NO: 2. The two segments were cloned and ligated by a one-step recombination kit from Vazyme to obtain a pEC-thrE2 plasmid having a plasmid map as shown in FIG. 3. The pEC-thrE2 and pEC-control plasmids were transformed into SZCgP3 to obtain an overexpression strain SZCgP3(pEC-thrE2) and a control strain SZCgP3(pEC-control), respectively. The yield of extracellular L-proline of the above strains was evaluated by shake-flask fermentation, and the fermentation was performed as described in Example 2. The results are shown in Table 6, indicating that the overexpression of ThrE in *Corynebacterium glutamicum* SZCgP3 could increase the yield of L-proline, and the overexpression strain SZCgP3(pEC-thrE2) showed a 2.56-fold increase in the yield of

L-proline at 21 h relative to the control bacterium SZCgP3 (pEC-control) introduced with an empty plasmid, further indicating that the overexpression of the L-proline efflux protein gene in *Corynebacterium glutamicum* is of great significance in the production of L-proline.

TABLE 6

Yield of extracellular L-proline of thrE gene overexpression strain in <i>Corynebacterium glutamicum</i> ATCC13032		
Time	Yield of extracellular L-proline (g/L)	
	SZCgP3(pEC-control)	SZCgP3(pEC-thrE2)
3 h	0.02 ± 0.00	0.03 ± 0.00
6 h	0.05 ± 0.00	0.17 ± 0.01
9 h	0.22 ± 0.02	0.80 ± 0.09

TABLE 6-continued

Yield of extracellular L-proline of thrE gene overexpression strain in <i>Corynebacterium glutamicum</i> ATCC13032		
Time	Yield of extracellular L-proline (g/L)	
	SZCgP3(pEC-control)	SZCgP3(pEC-thrE2)
12 h	0.59 ± 0.01	1.56 ± 0.07
15 h	1.00 ± 0.03	3.56 ± 0.22
18 h	3.85 ± 0.36	10.55 ± 0.36
21 h	3.81 ± 0.47	13.56 ± 1.26

EXAMPLE 4

Use of L-Proline Efflux Protein in Production of Hydroxyproline

[0099] Firstly, an expression plasmid pEC-mip4h was constructed for expressing trans-proline-4-hydroxylase, based on the pEC-XK99E plasmid, as described in Patent No. CN201710661945.1. The sequence of trans-proline-4-hydroxylase is provided as SEQ ID NO: 5. The pEC-mip4h plasmid and the pEC-XK99E control plasmid were transformed into an SZCgP1 strain to obtain an overexpression strain SZCgP1(pEC-m/p4h) and a control strain SZCgP1 (pEC-XK99E), respectively, and were transformed into an SZCgP2 strain to obtain an overexpression strain SZCgP2 (pEC-mip4h) and a control strain SZCgP2(pEC-XK99E), respectively. The seed culture medium contained the following ingredients (g/L): 5 g/L glucose, 1 g/L yeast powder,

3 g/L soybean peptone, 3 g/L urea, 0.5 g/L succinic acid, 1 g/L $K_2HPO_4 \cdot 3H_2O$, 0.1 g/L $MgSO_4 \cdot 7H_2O$, 0.01 mg/L biotin, 0.1 mg/L vitamin B1, and 20 g/L MOPS. The fermentation culture medium contained the following ingredients: 80 g/L glucose, 1 g/L yeast powder, 1 g/L soybean peptone, 1 g/L NaCl, 1 g/L ammonium sulfate, 6 g/L urea, 1 g/L $K_2HPO_4 \cdot 3H_2O$, 0.45 g/L $MgSO_4 \cdot 7H_2O$, 0.2 g/L $FeSO_4 \cdot 7H_2O$, 0.4 mg/L biotin, 0.1 mg/L vitamin B1, and 40 g/L MOPS, and the initial pH was 7.2. Firstly, the strains were inoculated into a seed culture medium and cultured for 8 h. The cultures were inoculated as seeds into a 24-well plate containing 800 μ L of a fermentation culture medium, and the inoculation amount was controlled with the initial OD_{600} being 0.1 (as detected by a microplate reader). The seeds were cultured at 30° C. for 18 h, with the rotation speed of a plate shaker being 800 rpm. The experiment for each strain was performed in triplicate.

[0100] After fermentation, OD_{600} and the yield of trans-4-hydroxy-L-proline and L-proline were detected. The detection of trans-4-hydroxy-L-proline was performed with reference to the China national standard GB/T 9695.23-2008, and the detection of L-proline was performed as described in Example 1. The results are shown in Table 7, indicating that the strains without expression of trans-proline-4-hydroxylase failed to produce trans-4-hydroxy-L-proline, and that the knockout of the L-proline efflux protein thrE gene could remarkably reduce the yield of extracellular L-proline by-products and increase the ratio of trans-4-hydroxy-L-proline to L-proline, helping to obtain trans-4-hydroxy-L-proline production with higher purity. It is expected that further enhancement of the activity of trans-proline-4-hydroxylase allows rapid transformation of intracellularly synthesized L-proline to trans-4-hydroxy-L-proline in the strain, and knockout of the L-proline efflux protein thrE gene also reduces the efflux of L-proline to the extracellular space, allowing the yield of trans-4-hydroxy-L-proline to be also increased. Therefore, the knockout of L-proline efflux protein thrE is of great significance to the production of trans-4-hydroxy-L-proline.

The pEC-Ubp4h plasmid and the pEC-XK99E control plasmid were transformed into an SZCgP1 strain to obtain an overexpression strain SZCgP1(pEC-Ubp4h) and a control strain SZCgP1(pEC-XK99E), respectively, and were transformed into an SZCgP2 strain to obtain an overexpression strain SZCgP2(pEC-Ubp4h) and a control strain SZCgP2(pEC-XK99E), respectively. The ingredients in the medium and the culture of the strains were referred to Example 4. The detection of trans-3-hydroxy-L-proline was performed with reference to Patent No. CN110804596A, and the detection of L-proline was performed as described in Example 1. The result shows that the strains without expression of trans-proline-3-hydroxylase failed to produce trans-3-hydroxy-L-proline, and that the knockout of the L-proline efflux protein thrE gene could remarkably reduce the yield of extracellular L-proline by-products and increase the ratio of trans-3-hydroxy-L-proline to L-proline, helping to obtain trans-3-hydroxy-L-proline production with higher purity. That is, the knockout of L-proline efflux protein thrE is of great significance to the production of trans-3-hydroxy-L-proline.

[0102] Hydroxyproline has four natural stereoisomers, namely trans-4-hydroxy-L-proline, cis-4-hydroxy-L-proline, trans-3-hydroxy-L-proline, and cis-3-hydroxy-L-proline, and it is known to those skilled in the art that the four different hydroxyprolines are obtained by directly catalyzing L-proline with corresponding trans-proline-4-hydroxylase, cis-proline-4-hydroxylase, trans-proline-3-hydroxylase, and cis-proline-3-hydroxylase, respectively. Documents in the prior art have also reported that different hydroxyprolines can be produced by overexpressing different proline hydroxylases based on strains producing L-proline (Zhang F, Liu H, Zhang T, Pijning T, Yu L, Zhang W, Liu W, Meng X. Biochemical and genetic characterization of fungal proline hydroxylase in echinocandin biosynthesis. *Appl Microbiol Biotechnol.* 2018 Sep;102(18):7877-7890. Chen K, Pang Y, Zhang B, Feng J, Xu S, Wang X, Ouyang P. Process optimization for enhancing production of cis-4-hydroxy-L-proline by engineered *Escherichia coli*. *Microb Cell Fact.*

TABLE 7

Effect of thrE gene knockout on the yield of extracellular trans-4-hydroxy-L-proline and L-proline				
Strain	OD_{600}	Yield of hydroxyproline (g/L)	Yield of proline (g/L)	Ratio of hydroxyproline to proline
SZCgP1(pEC-XK99E)	16.47 ± 0.94	—	4.25 ± 0.35	—
SZCgP1(pEC-mip4h)	13.16 ± 0.18	0.70 ± 0.01	2.94 ± 0.21	1:4.2
SZCgP2(pEC-XK99E)	15.52 ± 0.82	—	1.68 ± 0.08	—
SZCgP2(pEC-mip4h)	14.31 ± 0.22	0.70 ± 0.05	0.89 ± 0.05	1:1.3

EXAMPLE 5

Use of L-Proline Efflux Protein in Production of Hydroxyproline

[0101] To further verify the use of the L-proline efflux protein in the production of other hydroxyproline, in the present disclosure, a plasmid for expressing trans-proline-3-hydroxylase was first constructed, that is, an expression plasmid pEC-Ubp4h for expressing the trans-proline-3-hydroxylase Ubp4h as shown in Patent No. CN110804596A was artificially synthesized based on a pEC-XK99E plas-

mid. 2017 Nov. 22;16(1):210. Mori H, Shibasaki T, Yano K, Ozaki A. Purification and cloning of a proline 3-hydroxylase, a novel enzyme which hydroxylates free L-proline to cis-3-hydroxy-L-proline. *J Bacteriol.* 1997 September;179(18):5677-83). Therefore, for those skilled in the art, in Examples 4 and 5 of the present disclosure, it is demonstrated that the knockout of the L-proline efflux protein ThrE can increase intracellular L-proline concentration, and in the case that, based on this strain, the overexpression of trans-proline-4-hydroxylase can increase the yield of trans-4-hydroxy-L-proline, and the overexpression of trans-proline-3-hydroxylase can increase the yield of trans-3-hydroxy-L-

proline; it is apparent to those skilled in the art that the overexpression of different proline hydroxylases based on the knockout of the L-proline efflux protein ThrE can reduce the yield of extracellular L-proline as by-products and increase the corresponding ratio of hydroxyproline to L-proline, i.e., the knockout of the L-proline efflux protein is of great significance to the production of all of the above 4

hydroxyprolines. The embodiments of the present disclosure have been described above. However, the present disclosure is not limited to the embodiments described above. Any modification, equivalent, improvement, and the like made without departing from the spirit and principle of the present disclosure shall fall within the protection scope of the present disclosure.

 SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 23

<210> SEQ ID NO 1

<211> LENGTH: 489

<212> TYPE: PRT

<213> ORGANISM: *Corynebacterium glutamicum*

<400> SEQUENCE: 1

```

Met Leu Ser Phe Ala Thr Leu Arg Gly Arg Ile Ser Thr Val Asp Ala
1      5      10      15
Ala Lys Ala Ala Pro Pro Pro Ser Pro Leu Ala Pro Ile Asp Leu Thr
20     25     30
Asp His Ser Gln Val Ala Gly Val Met Asn Leu Ala Ala Arg Ile Gly
35     40     45
Asp Ile Leu Leu Ser Ser Gly Thr Ser Asn Ser Asp Thr Lys Val Gln
50     55     60
Val Arg Ala Val Thr Ser Ala Tyr Gly Leu Tyr Tyr Thr His Val Asp
65     70     75     80
Ile Thr Leu Asn Thr Ile Thr Ile Phe Thr Asn Ile Gly Val Glu Arg
85     90     95
Lys Met Pro Val Asn Val Phe His Val Val Gly Lys Leu Asp Thr Asn
100    105    110
Phe Ser Lys Leu Ser Glu Val Asp Arg Leu Ile Arg Ser Ile Gln Ala
115    120    125
Gly Ala Thr Pro Pro Glu Val Ala Glu Lys Ile Leu Asp Glu Leu Glu
130    135    140
Gln Ser Pro Ala Ser Tyr Gly Phe Pro Val Ala Leu Leu Gly Trp Ala
145    150    155    160
Met Met Gly Gly Ala Ile Ala Val Leu Leu Gly Gly Gly Trp Gln Val
165    170    175
Ser Leu Ile Ala Phe Ile Thr Ala Phe Thr Ile Ile Ser Thr Thr Ser
180    185    190
Phe Leu Gly Lys Lys Gly Leu Pro Thr Phe Phe Gln Asn Val Val Gly
195    200    205
Gly Phe Ile Ala Thr Leu Pro Ala Ser Ile Ala Tyr Ser Leu Ala Leu
210    215    220
His Phe Gly Leu Glu Ile Lys Pro Ser Gln Ile Ile Ala Ser Gly Ile
225    230    235    240
Val Val Leu Leu Ala Gly Leu Thr Leu Val Gln Ser Leu Gln Asp Gly
245    250    255
Ile Thr Gly Ala Pro Val Thr Ala Ser Ala Arg Phe Phe Glu Thr Leu
260    265    270
Leu Phe Thr Gly Gly Ile Val Ala Gly Val Gly Leu Gly Ile Gln Leu
275    280    285
Ser Glu Ile Leu His Val Met Leu Pro Ala Met Glu Ser Ala Ala Ala
290    295    300

```

-continued

Pro Asn Tyr Ser Ser Thr Phe Ala Arg Ile Ile Ala Gly Gly Val Thr
 305 310 315 320
 Ala Ala Ala Phe Ala Val Gly Cys Tyr Ala Glu Trp Ser Ser Val Ile
 325 330 335
 Ile Ala Gly Leu Thr Ala Leu Met Gly Ser Ala Phe Tyr Tyr Leu Phe
 340 345 350
 Val Val Tyr Leu Gly Pro Val Ser Ala Ala Ala Ile Ala Ala Thr Ala
 355 360 365
 Val Gly Phe Thr Gly Gly Leu Leu Ala Arg Arg Phe Leu Ile Pro Pro
 370 375 380
 Leu Ile Val Ala Ile Ala Gly Ile Thr Pro Met Leu Pro Gly Leu Ala
 385 390 395 400
 Ile Tyr Arg Gly Met Tyr Ala Thr Leu Asn Asp Gln Thr Leu Met Gly
 405 410 415
 Phe Thr Asn Ile Ala Val Ala Leu Ala Thr Ala Ser Ser Leu Ala Ala
 420 425 430
 Gly Val Val Leu Gly Glu Trp Ile Ala Arg Arg Leu Arg Arg Pro Pro
 435 440 445
 Arg Phe Asn Pro Tyr Arg Ala Phe Thr Lys Ala Asn Glu Phe Ser Phe
 450 455 460
 Gln Glu Glu Ala Glu Gln Asn Gln Arg Arg Gln Arg Lys Arg Pro Lys
 465 470 475 480
 Thr Asn Gln Arg Phe Gly Asn Lys Arg
 485

<210> SEQ ID NO 2
 <211> LENGTH: 489
 <212> TYPE: PRT
 <213> ORGANISM: Corynebacterium glutamicum

<400> SEQUENCE: 2

Met Leu Ser Phe Ala Thr Leu Arg Gly Arg Ile Ser Thr Val Asp Ala
 1 5 10 15
 Ala Lys Ala Ala Pro Pro Pro Ser Pro Leu Ala Pro Ile Asp Leu Thr
 20 25 30
 Asp His Ser Gln Val Ala Gly Val Met Asn Leu Ala Ala Arg Ile Gly
 35 40 45
 Asp Ile Leu Leu Ser Ser Gly Thr Ser Asn Ser Asp Thr Lys Val Gln
 50 55 60
 Val Arg Ala Val Thr Ser Ala Tyr Gly Leu Tyr Tyr Thr His Val Asp
 65 70 75 80
 Ile Thr Leu Asn Thr Ile Thr Ile Phe Thr Asn Ile Gly Val Glu Arg
 85 90 95
 Lys Met Pro Val Asn Val Phe His Val Val Gly Lys Leu Asp Thr Asn
 100 105 110
 Phe Ser Lys Leu Ser Glu Val Asp Arg Leu Ile Arg Ser Ile Gln Ala
 115 120 125
 Gly Ala Thr Pro Pro Glu Val Ala Glu Lys Ile Leu Asp Glu Leu Glu
 130 135 140
 Gln Ser Pro Ala Ser Tyr Gly Phe Pro Val Ala Leu Leu Gly Trp Ala
 145 150 155 160
 Met Met Gly Gly Ala Val Ala Val Leu Leu Gly Gly Gly Trp Gln Val

-continued

aggttctggg aaatacagac cgccacagta tcaaaaaaaaa tcttataggg gctcttttat	180
ttgacagtgg agagacagcg gaagcgactc gtctcaaacg gacagctcgt agaaggtata	240
cacgtcggaa gaatcgtatt tgttatctac aggagatfff ttcaaatgag atggcgaaa	300
tagatgatag tttctttcat cgacttgaag agtctttttt ggtggaagaa gacaagaagc	360
atgaacgtca tcctatffff ggaaatatag tagatgaagt tgcttatcat gagaaatatic	420
caactatcta tcactctcga aaaaaattgg tagattctac tgataaagcg gatttgcgct	480
taatctattt ggcccttagcg catatgatta agtttctgtg tcattttttg attgagggag	540
atftaaatcc tgataatagt gatgtggaca aactatttat ccagttggta caaacctaca	600
atcaattatt tgaagaaaac cctattaacg caagtggagt agatgctaaa gcgattcttt	660
ctgcacgatt gagtaaatca agacgattag aaaatctcat tgctcagctc cccggtgaga	720
agaaaaatgg cttatttggg aatctcattg ctttgtcatt gggtttgacc cctaatttta	780
aatcaaatftt tgatttggca gaagatgcta aattacagct ttcaaaagat acttacgatg	840
atgattftaga taatttattg gcgcaaatg gagatcaata tgctgatttg tttttggcag	900
ctaagaatftt atcagatgct atftttactftt cagataticct aagagftaaat actgaaataa	960
ctaaggctcc cctaticagct tcaatgatta aacgctacga tgaacaticat caagactftga	1020
ctctftftaaa agctfttagftt cgacaacaac ttccagaaaa gtataaagaa atctftftftg	1080
atcaatcaaaa aaacggatatic gcaggttata ttgatggggg agctagccaa gaagaatftft	1140
ataaatftftat caaaccaatftt ttagaaaaaa tggatggtac tgaggaatta ttggtgaaac	1200
taaatctgftga agattftgctg cgcaagcaac ggacctftftga caacggctct atftcccatc	1260
aaatftcactt gggftgagctg catgctattftt tgagaagaca agaagactftt tatccattftt	1320
taaaaagacaaa tctgtgagaag attgaaaaaa tctftgactftt tctgaattctct tattatgfttg	1380
gtccattggc gcgtggcaat agctgftftftg catggatgac tctggaagtct gaagaaacaa	1440
ttaccccatg gaatftftgaa gaagftftgctg ataaaggtgc ttcagctcaa tcaatftattg	1500
aacgcatgac aaactftftgat aaaaaatcttc caaatgaaaa agtactacca aaacatagftt	1560
tgctftftatga gtatftftacg gftftataacg aattgacaaa ggtcaaatat gfttactgaag	1620
gaatgcgaaa accagcattftt cftftcaggtg aacagaagaa agccattgftt gattftactct	1680
tcaaaaacaaa tctgaaaagta accgfttaagc aattaaaaaga agattatfttc aaaaaaatag	1740
aatgftftftga tagftgftgaa atfttcaggag ttgaagatag atfttaatgct tcaatftaggta	1800
cctaccatga tftftgctaaaa atfttataaag ataaagattftt tftftggataat gaagaaaatg	1860
aagataticftt agaggatatt gftfttaacat tgacctfttatt tgaagatagg gagatgattg	1920
aggaaaagact taaaaacatatic gctcacctct ttgatgataa ggtgatgaaa cagctftaac	1980
gtctgctgftta tactggtftg ggacgftftftg ctctgaaaaat gattaatggt atftagggata	2040
agcaatctgg caaaaacataa ttgattftftt tgaaatcaga tggftftftgct aatctgcaatft	2100
ttatgctgct gatccatgat gatagftftftga catftftaaaga agacattcaa aaagcacaag	2160
tgtctggaca aggcgatagftt ttacatgaac atattgcaaaa tfttagctggt agccctgctata	2220
ttaaaaaagg tattftftacag actgfttaaaag tftgftgatga atftggtcaaaa gtaatggggc	2280
ggcataagcc agaaaaatatic gfttattgaaa tggcacgftga aaatctagaca actcaaaaag	2340
gccagaaaaa tctctgagag cgtatgaaac gaatctgaaga aggtaticaaa gaatftaggaa	2400

-continued

gtcagattct	taaagagcat	cctggtgaaa	atactcaatt	gcaaaatgaa	aagctctatc	2460
tctattatct	ccaaaatgga	agagacatgt	atgtggacca	agaattagat	attaatcggt	2520
taagtgatta	tgatgtcgat	gccattgttc	cacaaagttt	ccttaaagac	gattcaatag	2580
acaataaggt	cttaacgcgt	tctgataaaa	atcgtggtaa	atcggataac	gttccaagtg	2640
aagaagtagt	caaaaagatg	aaaaactatt	ggagacaact	tctaaacgcc	aagttaatca	2700
ctcaacgtaa	gtttgataat	ttaacgaaag	ctgaacgtgg	aggtttgagt	gaacttgata	2760
aagctggttt	tatcaaacgc	caattgggtg	aaactcgcca	aatcactaag	catgtggcac	2820
aaattttggg	tagtgcgatg	aatactaaat	acgatgaaaa	tgataaaactt	attcgagagg	2880
ttaaagtgat	taccttaaaa	tctaaattag	tttctgactt	ccgaaaagat	ttccaattct	2940
ataaagtaog	tgagattaac	aattaccatc	atgcccatga	tgcgtatcta	aatgccgtcg	3000
ttggaactgc	tttgattaag	aaatatccaa	aacttgaatc	ggagtttgtc	tatggtgatt	3060
ataaagttta	tgatgttcgt	aaaatgattg	ctaagtctga	gcaagaaata	ggcaaagcaa	3120
ccgcaaaata	tttcttttac	tctaatatca	tgaacttctt	caaaacagaa	attacacttg	3180
caaatggaga	gattcgcgaa	cgccctctaa	tcgaaactaa	tggggaaact	ggagaaattg	3240
tctgggataa	aggggcgagat	tttgccacag	tgcgcaaagt	attgtccatg	ccccaagtca	3300
atattgtcaa	gaaaacagaa	gtacagacag	gcbgattctc	caaggagtca	attttaccaa	3360
aaagaaatc	ggacaagcct	attgctcgtg	aaaaagactg	ggatccaaaa	aaatattggtg	3420
gttttgatag	tccaacggta	gcttattcag	tccatgtggt	tgctaagggtg	gaaaaagggg	3480
aatcgaagaa	gttaaaatcc	gttaaagagt	tactagggat	cacaattatg	gaaagaagtt	3540
cctttgaaaa	aaatccgatt	gactttttag	aagctaaagg	atataaggaa	gttaaaaaag	3600
acttaatcat	taaaactaoc	aaatatagtc	tttttgagtt	agaaaacggg	cgtaaacgga	3660
tgctggctag	tgccggagaa	ttacaaaaag	gaaatgagct	ggctctgcca	agcaaatatg	3720
tgaatTTTTT	atattttagct	agtcattatg	aaaagttgaa	gggtagtcca	gaagataacg	3780
aacaaaaaca	attgtttgtg	gagcagcata	agcattattt	agatgagatt	attgagcaaa	3840
tcagtgaatt	ttctaagcgt	gttattttag	cagatgccaa	tttagataaa	gttcttagtg	3900
catataacaa	acatagagac	aaaccaatac	gtgaacaagc	agaaaatatt	attcatttat	3960
ttacgttgac	gaatccttga	gctcccgtg	cttttaaata	ttttgataca	acaattgatc	4020
gtaaacgata	tacgtctaca	aaagaagttt	tagatgccac	tcttatccat	caatccatca	4080
ctggtcttta	tgaaacacgc	attgatttga	gtcagctagg	agggtgactga	gcttggctgt	4140
tttggcggat	gagagaagat	tttcagcctg	atacagatta	aatcagaacg	cagaagcggg	4200
ctgataaaac	agaatttgcc	tggcggcagt	agcgcgggtg	tcccacotga	ccccatgccg	4260
aactcagaag	tgaaaccccg	tagcgcgat	ggtagtgtgg	ggctctgcca	tgcgagagta	4320
gggaactgcc	aggcatcaaa	taaaacgaaa	ggctcagtcg	aaagactggg	cctttcgttt	4380
tatctgttgt	ttgtcgggtg	acgctctcct	gagtaggaca	aatccgccgg	gagcggattt	4440
gaacgttgcg	aagcaacggc	ccggaggggtg	gcgggcagga	cgcccccat	aaactgccag	4500
gcatcaaatt	aagcagaagg	ccatcctgac	ggatggcctt	tttgogtttc	tacaaactct	4560
tttgtttatt	tttctaaata	cattcaata	tgtatccgct	catgagacaa	taaccctgat	4620
aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	4680

-continued

ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctgggta	4740
aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	4800
acagcggtaa	gatecctgag	agttttcgcc	cgaagaacg	ttttccaatg	atgagcactt	4860
ttgcttcttc	gctcactgac	tcgctgcgct	cggtcgttcg	gctgcggcga	gcggtatcag	4920
ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	ggataacgca	ggaagaaca	4980
tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgtttt	5040
tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	5100
gaaacccgac	aggactataa	agataccagg	cgtttcccc	tggaagetcc	ctcgtgcgct	5160
ctcctgttcc	gacctgccc	cttaccggat	acctgtccc	ctttctccct	tcgggaagcg	5220
tggcgctttc	tcaatgctca	cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	5280
agctgggctg	tgtgcacgaa	cccccgcttc	agcccgaccg	ctgcgcetta	tccggtaact	5340
atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	actggcagca	gccactggta	5400
acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	5460
actacggcta	cactagaagg	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	5520
tcggaaaaag	agttggtagc	tcttgatccg	gcaaaaaaac	caccgctggg	agcgggtggt	5580
tttttgttg	caagcagcag	attacgcgca	gaaaaaaagg	atctcaagaa	gaccccttga	5640
tctttcttac	ggggtctgac	gctcagtgga	acgaaaaactc	acgttaaggg	attttggtca	5700
tgagattatc	aaaaaggatc	ttcacctaga	tccttttggg	gtgggcgaag	aaactccagca	5760
tgagatcccc	gcgctggagg	atcatccagc	cattcggggg	cgttcactgg	ttccccttc	5820
tgatttctgg	catagaagaa	cccccgtgaa	ctgtgtggtt	ccggggggtg	ctgatTTTT	5880
cgagacttct	cgcgcaattc	cctagcttag	gtgaaaacac	catgaaaacac	tagggaaaca	5940
cccatgaaac	accatttagg	gcagtagggc	ggcttcttcg	tctagggctt	gcatttgggc	6000
ggtgatctgg	tcttttagcgt	gtgaaagtgt	gtcgtaggtg	gcgtgctcaa	tgcaactcgaa	6060
cgtcacgtca	tttaccgggt	cacgggtggc	aaagagaact	agtggttag	acattgtttt	6120
cctcgttgtc	ggtggtggtg	agcttttcta	gccgctcgg	aaacgcggcg	atcatgaact	6180
cttgagggtt	ttcacgcttc	tgcatgcctg	cgcgcttcat	gtcctcacgt	agtgcctaaag	6240
gaacgcgtgc	ggtgaccacg	acgggcttag	cctttgcctg	cgcttctagt	gcttcgatgg	6300
tggttgtgc	ctgcgcttgc	tgccctgta	gtgcctgttg	agcttctgt	agttgctgtt	6360
ctagctgtgc	cttggttgc	atgctttaa	actctagtag	ctttcctgcg	atatgtcatg	6420
cgcatgcgta	gcaaacattg	tctgcaact	cattcattat	gtgcagtgtc	cctgttacta	6480
gtcgtacata	ctcatattta	cctagtctgc	atgcagtgca	tgccatgca	gtcatgtcgt	6540
gctaattgtg	aaaacatgta	catgcagatt	gctgggggtg	cagggggcgg	agccaccctg	6600
tccatgcggg	gtgtggggct	tgccccgccc	gtacagacag	tgagcacccg	ggcacctagt	6660
cgcgataacc	ccccctaggt	atcggacacg	taacctccc	atgtcgatgc	aaatctttaa	6720
cattgagtac	gggtaagctg	gcacgcatag	ccaagctagg	cggccacca	acaccactaa	6780
aaattaatag	tccctagaca	agacaaacc	ccgtgcgagc	taccaactca	tatgcaagg	6840
ggccacataa	cccgaagggg	tttcaattga	caacctatgc	actagctaag	acaacgggca	6900
caacacccgc	acaaaactgc	actgcgcaac	cccgcacaac	atcgggtcta	ggtaacactg	6960

-continued

agtaacactg	aaatagaagt	gaacacctct	aaggaaccgc	aggccaatga	gggttctaag	7020
gtcactcgcg	ctagggcggtg	gcgtaggcaa	aacgtcatgt	acaagatcac	caatagtaag	7080
gctctggcgg	gggtccatag	gtggcgacgg	gacgaagctg	ttgcgggtgc	ctggctcgtct	7140
aacggtgctt	cgagtttga	gggtctgcaa	aactctcact	ctcgtggggg	gtcacctctg	7200
gctgaattgg	aagtcattggg	cgaacgccgc	attgagctgg	ctattgctac	taagaatcac	7260
ttggcgcgcg	gtggcgcgct	catgatgttt	gtgggcactg	ttcgacacaa	ccgctcacag	7320
tcatttgcgc	aggttgaagc	gggtattaag	actgcgtact	cttcgatggg	gaaaacatct	7380
cagtggaaga	aagaacgtgc	acggtaacggg	gtggagcaca	cctatagtga	ctatgaggtc	7440
acagactctt	gggcgaacgg	ttggcacttg	caccgcaaca	tgctgttgtt	cttggatcgt	7500
ccactgtctg	acgatgaact	caaggcgctt	gaggattcca	tgttttcccg	ctggctcgtct	7560
gggtgtggtta	aggccggtat	ggacgcgccca	ctgcgtgagc	acggggctca	acttgatcag	7620
gtgtctacct	ggggtggaga	cgctgcgaaa	atggcaacct	acctcgtcaa	gggcatgtct	7680
caggaactga	ctggctccgc	tactaaaacc	gcgtctaagg	ggctcgtcac	gccgtttcag	7740
atggttgata	tgttggcoga	tcaaacgcac	gccggcgagg	atatggacgc	tgttttggtg	7800
gctcgggtggc	gtgagtatga	gggtggttct	aaaaacctgc	gttcgtcctg	gtcacgtggg	7860
gctaagcgtg	ctttgggcat	tgattacata	gacgctgatg	tacgtcgtga	aatggaagaa	7920
gaactgtaca	agctcgcocg	tctggaagca	ccggaacggg	tcgaatcaac	ccgcgttgct	7980
gttgctttgg	tgaagccoga	tgattggaaa	ctgattcagt	ctgatttcgc	ggttaggcag	8040
tacgttctcg	attgcgtgga	taaggctaag	gacgtggccg	ctgcgcaacg	tgtcgttaat	8100
gaggtgctgg	caagtctggg	tgtggattcc	accccggtca	tgatcgttat	ggatgatgtg	8160
gacttggaoc	cggttctgcc	tactcatggg	gacgctacta	agcgtgatct	gaatgcggcg	8220
gtgttcgcgg	gtaatgagca	gactattctt	cgcacccact	aaaagcggca	taaaccctct	8280
tcgatatttt	gtgcgatgaa	tttatggcca	atgtcgcggg	ggcaaacctat	gatgggtctt	8340
gttgttggcg	tcccggaaaa	cgattccgaa	gcccacacct	tcatagaagg	cgcggttggg	8400
atthttctcc	acataagctg	gcaatgttgc	gacgcaacag	gtacagtgtg	attcatgaga	8460
ccacgcgtgg	atccggctta	ctaaaagcca	gataaacagta	tgcgtatttg	cgcgctgatt	8520
tttgcggtat	aagaatatat	actgatatgt	ataccogaag	tatgtcaaaa	agaggtatgc	8580
tatgaagcag	cgtattacag	tgacagttga	cagcgacagc	tatcagttgc	tcaaggcata	8640
tatgatgtca	atatctcccg	tctggttaagc	acaacctgc	agaatgaagc	ccgtcgtctg	8700
cgtgccgaac	gctggaaaagc	ggaaaatcag	gaagggatgg	ctgaggtcgc	ccgggttatt	8760
gaaatgaacg	gctcttttgc	tgacgagaa	aggggctggt	gaaatgcagt	ttaaggttta	8820
cacctataaa	agagagagcc	gttatcgtct	gtttgtggat	gtacagagtg	atattattga	8880
cacgcccggg	cgacggatgg	tgatccccct	ggccagtgc	cgctcgtctg	cagataaagt	8940
ctcccgtgaa	ctttaccocg	tggtgcata	cggggatgaa	agctggcgca	tgatgaccac	9000
cgatatggcc	agtgtgccgg	tatccgttat	cggggaagaa	gtggctgatc	tcagccaccg	9060
cgaaaatgac	atcaaaaacg	ccattaacct	gatgttctgg	ggaatataag	gtctcagttt	9120
tagagctaga	aatagcaagt	taaaataagg	ctagtcggtt	atcaacttga	aaaagtggca	9180
ccgagtcggt	gctttttttc	caggcatcaa	ataaaacgaa	aggctcagtc	gaaagactgg	9240

-continued

gccttctggt	ttatctgttg	ttgtcggtg	aacgctctcc	tgagtaggac	aatccgccc	9300
ggagcggatt	tgaacgttgc	gaagcaacgg	cccggagggt	ggcgggcagg	acgcccgcc	9360
taaaactgcc	ggcatcaaat	taagcagaag	gccatcctga	cgatggcct	ttttgcggtt	9420
ctacaaactc	ttttgttta	ttttctaaa	tacattcaaa	tatgtatccg	ctcatgaatt	9480
aattccgcta	gatgacgtgc	ggcttcgaaa	atctcgtgat	ggcaggttgg	gcgtcgttg	9540
gtcggtcatt	tcgaagggca	ccaataactg	ccttaaaaa	attacgccc	gcctgccac	9600
tcacgcagt	actgtttaa	ttcattaagc	attctgccga	catggaagcc	atcacagacg	9660
gcatgatgaa	cctgaatcgc	cagcggcatc	agcaccttgt	cgcttgcgt	ataatatttg	9720
cccatggtga	aaacgggggc	gaagaagttg	tccatattgg	ccacgtttaa	atcaaaactg	9780
gtgaaactca	cccagggtt	ggctgagacg	aaaaacatat	tctcaataaa	cccttaggg	9840
aaataggcca	ggtttccacc	gtaacacgcc	acatcttgcg	aatatatgtg	tagaaactgc	9900
cggaaatcgt	cgtggtatc	actccagacg	gatgaaaacg	tttcagtttg	ctcatggaaa	9960
acggtgtaac	aagggtgaac	actatcccat	atcaccagct	caccgtcttt	cattgccata	10020
cggaaactcgc	gatgagcatt	catcagggcg	gcaagaatgt	gaataaaggc	cggataaaac	10080
ttgtgcttat	ttttctttac	ggtctttaa	aaggccgtaa	tatccagctg	aacggctcgg	10140
ttataggtac	attgagcaac	tgactgaaat	gcctcaaaat	gttctttacg	atgccattgg	10200
gatatatcaa	cggtggtata	tccagtgatt	ttttctcca	ttttagcttc	cttagctcct	10260
gaaaatctcg	tcgaagctcg	gcggtttgt	cctactcaag	ctgatccgac	aaaatccaca	10320
cattatccca	ggtgtccgga	tcggtcaaat	acgctgccag	ctcatagacc	gtatccaaag	10380
catccggggc	tgatccccgc	cgccagggtg	gttttcttt	tcaccagtga	gacgggcaac	10440
agctgattgc	ccttcaccgc	ctggccctga	gagagttgca	gcaagcggtc	cacgtgggtt	10500
gcccagcag	gcaaaaatcc	tgtttgatgg	tggttaacgg	cgggatataa	catgagctgt	10560
cttcggatc	gtcgtatccc	actaccgaga	tatccgcacc	aacgcgcagc	ccggactcgg	10620
taatggcgcg	cattgcgccc	agcgcctct	gatcgttggc	aaccagcacc	gcagtgaggaa	10680
cgatgccctc	attcagcatt	tgcatgggtt	gttgaaaacc	ggacatggca	ctccagtcgc	10740
cttcccgttc	cgctatcggc	tgaatttgat	tgccagtgag	atatttatgc	cagccagcca	10800
gacgcagacg	cgccgagaca	gaacttaatg	ggcccgttaa	cagcgcgatt	tgctggtgac	10860
ccaatgcgac	cagatgctcc	acgccagtc	gcgtaccgtc	ttcatgggag	aaaataatac	10920
tggtgatggg	tgtctgttca	gagacatcaa	gaaataacgc	cggaacatta	gtgcaggcag	10980
cttccacagc	aatggcatcc	tggtcatcca	gcgatagtt	aatgatcagc	ccactgacgc	11040
gttgcgcgag	aagattgtgc	accgccgctt	tacaggcttc	gacgccgctt	cgttctacca	11100
tcgacaccac	cacgctggca	cccagttgat	cggcgcgaga	tttaatcgcc	gcgacaattt	11160
gcgacggcgc	gtgcagggcc	agactggagg	tggcaacgcc	aatcagcaac	gactgtttgc	11220
ccgccagttg	ttgtgccacg	cggttgggaa	tgtaattcag	ctccgccacc	gccgcttcca	11280
cttttcccgc	cgttttcgca	gaaacgtggc	tggcctggtt	caccacgcgg	gaaacggtct	11340
gataagagac	accggcatac	tctgcgacat	cgtataacgt	tactggtttc	acattcacca	11400
ccctgaattg	actctcttcc	ggcgctatc	atgccatacc	gcgaaaggtt	ttgcaccatt	11460
cgatgggtgc	aacgtaaatg	ccgcttcgcc	tccgcgcgcg	aattgcaagc	tgatccgggc	11520

-continued

```

ttatcgactg cacggtgcac caatgcttct ggcgtcaggc agccatcggga agctgtggta 11580
tggctgtgca ggtcgtaaat cactgcataa ttcgtgtcgc tcaaggcgca ctcccgttct 11640
ggataatggt ttttgcgccc acatcataac ggttctggca aatattctga aatgagctgt 11700
tgacaattaa tcatcggctc gtataatgtg tgggaattgtg agcggataac aatttcacac 11760
aggaaacaga att 11773

```

```

<210> SEQ ID NO 4
<211> LENGTH: 11773
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

```

```

<400> SEQUENCE: 4

```

```

aattaagctt aaaggcagcc gatatggata agaaatactc aataggctta gatatcggca 60
caaatagcgt cggatggggc gtgatcactg atgaatataa ggttccgtct aaaaagtcca 120
aggttctggg aaatacagac cgccacagta tcaaaaaaaaa tcttataggg gctcttttat 180
ttgacagtgg agagacagcg gaagcgactc gtctcaaacg gacagctcgt agaaggtata 240
cacgtcggaa gaatcgtatt tgttatctac aggagatttt tcaaatgag atggcgaaa 300
tagatgatag tttctttcat cgacttgaag agtctttttt ggtggaagaa gacaagaagc 360
atgaacgtca tcctattttt ggaatatag tagatgaagt tgcttatcat gagaaatatac 420
caactatcta tcatctcgga aaaaaattgg tagattctac tgataaagcg gatttgcgct 480
taatctattt ggccttagcg catatgatta agtttctgtg tcattttttg attgagggag 540
atntaaatcc tgataatagt gatgtggaca aactatttat ccagttggta caaacctaca 600
atcaattatt tgaagaaaac cctattaacg caagtggagt agatgctaaa gcgattcttt 660
ctgcacgatt gagtaaatca agacgattag aaaatctcat tgctcagctc cccggtgaga 720
agaaaaatgg cttatttggg aatctcattg ctttgtcatt gggtttgacc cctaatttta 780
aatcaaattt tgatttggca gaagatgcta aattacagct tcaaaaagat acttacgatg 840
atgatttaga taatttattg gcgcaaattg gagatcaata tgctgatttg ttttggcag 900
ctaagaattt atcagatgct attttacttt cagatatacct aagagtaaat actgaaataa 960
ctaaggctcc cctatcagct tcaatgatta aacgctacga tgaacatcat caagacttga 1020
ctcttttaaa agcttttagt cgacaacaac ttccagaaaa gtataaagaa atcttttttg 1080
atcaatcaaa aaacggatat gcaggttata ttgatggggg agctagccaa gaagaatttt 1140
ataaatttat caaaccaatt ttagaaaaaa tggatgggtac tgaggaatta ttggtgaaac 1200
taaatcgtga agatttgcgt cgcaagcaac ggacctttga caacggctct attccccatc 1260
aaattcactt gggtagctg catgctattt tgagaagaca agaagacttt tatccatttt 1320
taaaagacaa tcgtgagaag attgaaaaaa tcttgacttt tcgaattcct tattatggtg 1380
gtccattggc gcgtggcaat agtcgttttg catggatgac tcggaagtct gaagaacaaa 1440
ttaccccatg gaattttgaa gaagttgtcg ataaaggtgc ttcagctcaa tcatttattg 1500
aacgcatgac aaactttgat aaaaatcttc caaatgaaaa agtactacca aaacatagtt 1560
tgctttatga gtattttacg gtttataacg aattgacaaa ggtcaaatat gttactgaag 1620
gaatgcgaaa accagcattt ctttcagggtg aacagaagaa agccattggt gatttactct 1680

```

-continued

tcaaaacaaa	tcgaaaagta	accgttaagc	aattaaaga	agattatttc	aaaaaaatag	1740
aatgttttga	tagtggtgaa	atttcaggag	ttgaagatag	atttaatgct	tcattaggta	1800
cctaccatga	tttgctaaaa	attattaaag	ataaagattt	tttgataat	gaagaaaatg	1860
aagatatctt	agaggatatt	gttttaacat	tgaccttatt	tgaagatagg	gagatgattg	1920
aggaaagact	taaaacatat	gctcacctct	ttgatgataa	ggtgatgaaa	cagcttaaac	1980
gtcgcggtta	tactggttgg	ggcgttttgt	ctcgaaaatt	gattaatggt	attagggata	2040
agcaatctgg	caaaacaata	ttagattttt	tgaaatcaga	tggttttgcc	aatcgcaatt	2100
ttatgcagct	gatccatgat	gatagtttga	catttaaaga	agacattcaa	aaagcacaag	2160
tgtctggaca	aggcgatagt	ttacatgaac	atattgcaa	tttagctggt	agccctgcta	2220
ttaaaaaagg	tattttacag	actgtaaaag	ttggtgatga	attggtcaaa	gtaatggggc	2280
ggcataagcc	agaaaatata	gttattgaaa	tggcacgtga	aaatcagaca	actcaaaagg	2340
gccagaaaaa	ttcgcgagag	cgtatgaaac	gaatcgaaga	aggtatcaaa	gaattaggaa	2400
gtcagattct	taaagagcat	cctggtgaaa	atactcaatt	gcaaaatgaa	aagctctatc	2460
tctattatct	ccaaaatgga	agagacatgt	atgtggacca	agaattagat	attaatcggt	2520
taagtgatta	tgatgtcgat	cacattgttc	cacaaagttt	ccttaaagac	gattcaatag	2580
acaataaggt	cttaacgogt	tctgataaaa	atcgtggtaa	atcggataac	gttccaagtg	2640
aagaagtagt	caaaaagatg	aaaaactatt	ggagacaact	tctaaacgcc	aagttaatca	2700
ctcaacgtaa	gtttgataat	ttaacgaaag	ctgaacgtgg	aggtttgagt	gaacttgata	2760
aagctggttt	tatcaaacgc	caattgggtg	aaactcgcca	aatcactaag	catgtggcac	2820
aaattttggg	tagtcgcatg	aatactaaat	acgatgaaaa	tgataaaact	attcgagagg	2880
ttaaagtgat	taccttaaaa	tctaaattag	tttctgactt	ccgaaaagat	ttccaattct	2940
ataaagtagc	tgagattaac	aattaccatc	atgcccatga	tgcgtatcta	aatgccgtcg	3000
ttggaactgc	tttgattaag	aaatatocaa	aacttgaatc	ggagtttgtc	tatggtgatt	3060
ataaagttta	tgatgttcgt	aaaatgattg	ctaagtctga	gcaagaaata	ggcaaaagcaa	3120
ccgcaaaata	tttcttttac	tctaatatca	tgaacttctt	caaaacagaa	attacacttg	3180
caaatggaga	gattcgcaaa	cgccctctaa	tcgaaactaa	tggggaaact	ggagaaattg	3240
tctgggataa	aggcggagat	tttgccacag	tcgcgaaagt	attgtccatg	ccccaaagtca	3300
atattgtcaa	gaaaacagaa	gtacagacag	gctgattctc	caaggagtca	attttaccaa	3360
aaagaaatc	ggacaagcct	attgctcgta	aaaaagactg	ggatccaaaa	aaatattggtg	3420
gttttgatag	tccaacggta	gcttattcag	tcttagtggt	tgctaagggtg	gaaaaagggg	3480
aatcgaagaa	gttaaaatcc	gttaaagagt	tactagggat	cacaattatg	gaaagaagtt	3540
cctttgaaaa	aaatccgatt	gacttttttag	aagctaaagg	atataaggaa	gttaaaaaag	3600
acttaatcat	taaactacct	aaatatagtc	tttttgagtt	agaaaaagg	cgtaaacgga	3660
tgctggctag	tgccggagaa	ttacaaaaag	gaaatgagct	ggctctgcca	agcaaatatg	3720
tgaatttttt	atatttagct	agtcattatg	aaaagttgaa	gggtagtcca	gaagataacg	3780
aacaaaaaca	attgtttgtg	gagcagcata	agcattatct	agatgagatt	attgagcaaa	3840
tcagtgaatt	ttctaagcgt	gttatttttag	cagatgccaa	tttagataaa	gttcttagtg	3900
catatacaaa	acatagagac	aaaccaatac	gtgaacaagc	agaaaaatatt	attcatttat	3960

-continued

ttacgttgac gaatcttga gctcccctg cttttaaata tttgataca acaattgate	4020
gtaaacgata tacgtctaca aaagaagtt tagatgccac tcttatccat caatccatca	4080
ctggtcttta tgaaacacgc attgatttga gtcagctagg aggtgactga agcttgctg	4140
ttttggcggg tgagagaaga tttcagcct gatacagatt aaatcagaac gcagaagcgg	4200
tctgataaaa cagaatttgc ctggcggcag tagcgcggtg gtcccacctg accccatgcc	4260
gaactcagaa gtgaaacgcc gtagcggcga tggtagtggt gggctgccc atgagagagt	4320
agggaaactgc caggcatcaa ataaaacgaa aggctcagtc gaaagactgg gcctttcgtt	4380
ttatctgttg tttgtcggg aacgctctcc tgagtaggac aaatccgcc ggagcggatt	4440
tgaacgttgc gaagcaacgg ccggagggt ggcggcagg acgcccgcc taaactgcc	4500
ggcatcaaat taagcagaag gccatcctga cggatggcct ttttgcgtt ctacaaactc	4560
ttttgtttat ttttctaata acattcaaat atgtatccgc tcatgagaca ataacctga	4620
taaatgcttc aataatattg aaaaaggaag agtatgagta tcaacattt ccgtgtgcc	4680
cttattccct ttttgcggc attttgcct cctgtttttg ctcaaccaga aacgctggtg	4740
aaagtaaaag atgctgaaga tcagttgggt gcacgagtg gttacatcga actggatctc	4800
aacagcggta agatccttga gagtttctc cccgaagaac gttttccaat gatgagcact	4860
tttgcttcct cgctcactga ctgcctgcgc tcggtcgttc ggcctgcggcg agcggtatca	4920
gctcactcaa agcggttaat acggttatcc acagaatcag gggataacgc aggaaagaac	4980
atgtgagcaa aagccagca aaaggccagg aaccgtaaaa aggcgcggt gctggcgtt	5040
ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg	5100
cgaaaccga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc	5160
tctctgttc cgaccctgcc gcttacggga tacctgtccg cctttctccc ttcgggaagc	5220
gtggcgttt ctcaatgctc acgctgtagg tatctcagtt cgggtgtaggt cgttcgctcc	5280
aagctgggct gtgtgcacga acccccgtt cagcccagc gctgcgcctt atccgtaac	5340
tatcgtcttg agtccaaacc ggttaagacac gacttatcgc cactggcagc agccactggt	5400
aacaggatta gcagagcag gtatgtaggc ggtgctacag agttctttaa gtggtggcct	5460
aactacggct aactagaag gacagtatt ggtatctgcg ctctgctgaa gccagttacc	5520
ttcggaaaaa gagttgtag ctcttgatcc ggcaaaaaa ccaccgctgg tagcgggtgt	5580
ttttttgtt gcaagcagca gattacgcgc agaaaaaag gatctcaaga agatcctttg	5640
atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg gattttggtc	5700
atgagattat caaaaaggat cttcacctag atccttttgg ggtggcgaa gaactccagc	5760
atgagatccc cgcgctggag gatcatccag ccattcgggg tcgttcactg gttcccctt	5820
ctgatttctg gcataaga acccccgtg actgtgtggt tccgggggtt gctgattttt	5880
gcgagacttc tcgcgcaatt ccttagctta ggtgaaaaa ccatgaaaca ctagggaac	5940
acctatgaaa caccatttag ggcagtaggg cggctctctc gtctagggct tgcatttggg	6000
cggatgatctg gtcttttagc tgtgaaagtg tgtcgtagg ggcgtgctca atgcactcga	6060
acgtcacgctc atttaccggg tcacgggtgg caaagagaac tagtgggtta gacattgttt	6120
tccctggtgt cgggtggtgt gagcttttct agccgctcgg taaacgcggc gatcatgaa	6180
tcttggagggt tttcacggt ctgcatgcct gcgcgcttca tgcctcagc tagtgccaaa	6240

-continued

ggaacgcgtg	cggtgaccac	gacgggctta	gcctttgcct	gcgcttctag	tgettccgatg	6300
gtggcttggtg	cctgcgcttg	ctgcgcctgt	agtgcctggt	gagcttcttg	tagttgctgt	6360
tctagctgtg	ccttggttgc	catgctttaa	gactctagta	gctttcctgc	gatatgtcat	6420
gcgcatgcgt	agcaaacatt	gtcctgcaac	tcattcatta	tgtgcagtgc	tcctgttact	6480
agtcgtacat	actcatatth	acctagtctg	catgcagtgc	atgcacatgc	agtcgatgctg	6540
tgctaagtgtg	taaaacatgt	acatgcagat	tgctgggggt	gcagggggcg	gagccaccct	6600
gtccatgcgg	ggtgtggggc	ttgccccgcc	ggtacagaca	gtgagcaccg	gggcacctag	6660
tcgcggtatc	ccccctagg	tatcggacac	gtaaccctcc	catgtcgatg	caaatcttta	6720
acattgagta	cgggtaagct	ggcacgcata	gccaaagctag	gcgccacca	aacaccacta	6780
aaaattaata	gtccctagac	aagacaaacc	cccgtgcgag	ctaccaactc	atatgcacgg	6840
gggccacata	accgaaggg	gtttcaattg	acaaccatag	cactagctaa	gacaacgggc	6900
acaacaccgg	cacaaactcg	cactgcgcaa	ccccgcacaa	catcgggtct	aggtaaacct	6960
gagtaaacct	gaaatagaag	tgaacacctc	taaggaaccg	caggtcaatg	agggttctaa	7020
ggtcactcgc	gctaggcgt	ggcgtaggca	aaacgtcatg	tacaagatca	ccaatagtaa	7080
ggctctggcg	gggtgccata	ggtggcgcag	ggacgaagct	gttgcggtgt	cctggctcgc	7140
taacgggtgt	tcgcagtttg	agggtctgca	aaactctcac	tctcgtggg	ggtcatctct	7200
ggctgaattg	gaagtcatgg	gcgaacgcgg	cattgagctg	gctattgcta	ctaagaatca	7260
cttgccggcg	gggtggcgcg	tcgatgatgt	tgtagggcact	gttcgacaca	accgctcaca	7320
gtcatttgcg	caggttgaag	cggtatttaa	gactgcgtac	tcttcgatgg	tgaaaacatc	7380
tcagtggaag	aaagaacgtg	cacggtacgg	ggtggagcac	acctatagtg	actatgaggt	7440
cacagactct	tggggcaacg	ggtggcactt	gcaccgcaac	atgctgttgt	tcttgatcgc	7500
tccactgtct	gacgatgaac	tcaaggcgtt	tgaggattcc	atgttttccc	gctggtctgc	7560
tgggttggtt	aaggccggta	tggacgcgcc	actgcgtgag	cacggggtea	aacttgatca	7620
gggtgtctacc	tgggggtggag	acgctgcgaa	aatggcaacc	tacctcgcta	agggcatgct	7680
tcaggaactg	actggctccg	ctactaaaac	cgctctaaag	gggtcgtaca	cgccgtttca	7740
gatgttgat	atgttgcccg	atcaaaagca	cgccggcgag	gatatggacg	ctgttttggt	7800
ggctcgggtg	cgtgagtatg	aggttgggtc	taaaaacctg	cgttcgtcct	ggtcacgtgg	7860
ggctaagcgt	gctttgggca	ttgattacat	agacgctgat	gtacgctcgtg	aaatggaaga	7920
agaactgtac	aagctcgcgg	gtctggaagc	accggaaccg	gtcgaatcaa	cccgcgttgc	7980
tgttgctttg	gtgaagcccg	atgattggaa	actgattcag	tctgatttcg	cggttaggca	8040
gtacgttctc	gattgcgtgg	ataaggctaa	ggacgtggcc	gctgcgcaac	gtgtcgtcaa	8100
tgaggtgctg	gcaagtctgg	gtgtggatcc	caccccgctg	atgatcgtta	tgatgatgt	8160
ggacttgac	gcggttctgc	ctactcatgg	ggacgctact	aagcgtgatc	tgaatgcggc	8220
gggttccg	ggtaatgagc	agactattct	tcgcacccac	taaaagcggc	ataaaccccg	8280
ttcgatattt	tgtgcgatga	atthtatggtc	aatgtcgcgg	gggcaaacct	tgatgggtct	8340
tgttgttggc	gtcccggaaa	acgattccga	agcccaacct	ttcatagaag	gcggcgggtg	8400
aatthttctc	cacataagct	ggcaatggtg	cgacgcaaca	ggtacagtgt	aattcatgag	8460
accacgcgtg	gatccggctt	actaaaagcc	agataaacgt	atgcgtatth	gcgcgctgat	8520

-continued

ttttgctgta taagaatata tactgatatg tatacccgaa gtatgtcaaa aagaggtatg	8580
ctatgaagca gcgtattaca gtgacagttg acagcgacag ctatcagttg ctcaaggcat	8640
atatgatgtc aatatctocg gctcggtaag cacaacatg cagaatgaag cccgtcgtct	8700
gcgtgccgaa cgctggaaag cggaaaatca ggaagggatg gctgaggtcg cccggtttat	8760
tgaaatgaac ggctcttttg ctgacgagaa caggggctgg tgaaatgcag ttttaaggttt	8820
acacctataa aagagagagc cgttatcgtc tgtttgtgga tgtacagagt gatattattg	8880
acacgccccg gcgacggatg gtgatcccc tggccagtgc acgtctgctg tcagataaag	8940
tctcccgtga actttaccog gtgggtgcata tcggggatga aagctggcgc atgatgacca	9000
ccgatatggc cagtgtgccc gtatccgtta tcggggaaga agtggtgat ctcagccacc	9060
gcgaaaatga catcaaaaac gccattaacc tgatgttctg gggaaataa ggtctcagtt	9120
ttagagctag aaatagcaag ttaaaataag gctagtccgt tatcaacttg aaaaagtggc	9180
accgagtcgg tgcttttttt ccaggcatca aataaaaacg aaggctcagt cgaaagactg	9240
ggcctttcgt tttatctggt gtttgcggt gaacgctctc ctgagtagga caaatccgcc	9300
gggagcggat ttgaacgttg cgaagcaacg gcccgagggg tggcgggcag gacgcccc	9360
ataaactgcc aggcatacaa ttaagcagaa ggccatcctg acggatggcc tttttcgtt	9420
tctacaaact ctttttgttt atttttctaa atacattcaa atatgtatcc gctcatgaat	9480
taattccgct agatgacgtg cggcttcgaa aatctcgtga tggcaggttg ggcgtcgtt	9540
ggtcggctat ttcgaagggc accaataact gccttaaaaa aattacgccc cgccctgcca	9600
ctcatcgcag tactgttgta attcattaag cattctgccc acatggaagc catcacagac	9660
ggcatgatga acctgaatcg ccagcggcat cagcaccttg tcgcccgtcg tataaatatt	9720
gcccattggtg aaaacggggg cgaagaagtt gtccatattg gccacgttta aatcaaaact	9780
ggtgaaactc acccagggat tggctgagac gaaaaacata ttctcaataa acccttagg	9840
gaaataggcc aggttttccac cgtaacacgc cacatcttgc gaatatatgt gtgaaactg	9900
ccggaaatcg tcgtggtatt cactccagag cgatgaaaac gtttcagttt gctcatggaa	9960
aacggtgtaa caagggtgaa cactatocca tatcaccagc tcaccgtctt tcattgccat	10020
acggaaactc ggatgagcat tcatcaggcg ggcaagaatg tgaataaagg ccggataaaa	10080
cttgtgctta tttttcttta cgttctttaa aaaggccgta atatccagct gaacggctcg	10140
gttataggta cattgagcaa ctgactgaaa tgcctcaaaa tgttctttac gatgccattg	10200
ggatatatca accggtgat atccagtgat tttttctcc attttagctt ccttagctcc	10260
tgaaaatctc tcgaagctc ggccgatttg tctactcaa gctgatccga caaaatccac	10320
acattatccc aggtgtccgg atcggtaaaa tacgctgcca gctcatagac cgtatccaaa	10380
gcatccgggg ctgatcccc gcgccagggt ggtttttctt ttcaccagtg agacgggcaa	10440
cagctgattg cccttccacc cctggccctg agagagttgc agcaagcggg ccacgtggtt	10500
tgccccagca ggcgaaaatc ctgtttgatg gtggttaacg gcgggatata acatgagctg	10560
tcttcggtat cgtcgtatcc cactaccgag atatccgcac caacgcgcag cccggactcg	10620
gtaatggcgc gcattgcgcc cagcgcctc tgatcgttgg caaccagcat cgcagtggga	10680
acgatgccct cattcagcat ttgcatggtt tgttgaaaac cggacatggc actccagtcg	10740
ccttcccgtt ccgctatcgg ctgaatttga ttgcgagtga gatatttatg ccagccagcc	10800

-continued

```

agacgcagac ggcgagagac agaacttaat gggcccgccta acagcgcgat ttgctggtga 10860
cccaatgcga ccagatgctc cagcccagc cgcgtaccgt cttcatggga gaaaataata 10920
ctggtgatgg gtgtctggtc agagacatca agaaataacg ccggaacatt agtgcaggca 10980
gcttccacag caatggcadc ctggctatcc agcggatagt taatgatcag cccactgacg 11040
cgttgcgcga gaagattgtg caccgcccgt ttacaggctt cgacgcccgt tcggtctacc 11100
atcgacacca ccacgctggc acccagttga tcggcgcgag atttaatcgc cgcgacaatt 11160
tgcgacggcg cgtgcagggc cagactggag gtggcaacgc caatcagcaa cgactgtttg 11220
cccgcagatt gttgtgcac cggggtggga atgtaattca gctccgccat cgcgcttcc 11280
actttttccc gcgttttcgc agaaacgtgg ctggcctggt tcaccacgcg gaaaacggtc 11340
tgataagaga caccggcata ctctgcgaca tcgtataacg ttactggttt cacattcacc 11400
accctgaatt gactctcttc cgggctctat catgccatcc cgcgaaaggt tttgcacat 11460
tcgatggtgt caacgtaaat gccgcttcgc cttegcgcgc gaattgcaag ctgatccggg 11520
cttatcgact gcacggtgca ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt 11580
atggctgtgc aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc 11640
tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg aaatgagctg 11700
ttgacaatta atcatcggct cgtataatgt gtggaattgt gacgctcac aatttcacac 11760
aggaacaga att 11773

```

```

<210> SEQ ID NO 5
<211> LENGTH: 807
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

```

```

<400> SEQUENCE: 5
atggtgctgt acgagagtat cgacctgat cgcgaggacg ggttcctggt cgcggacccc 60
cttaccgaaa aggagacaga ccttctacgc gaacaggccg atcgtgagtt tctgcgggat 120
tctcccggcc ggatggtgga gaaggacggt ttcacggtgc gcggggtgca cgggtcgcac 180
gtggtgaaca gcacgttcgc cgggctggta cggcacccaa agatcgtcac accggccacg 240
cagctactgg gtggccccgt atatgtgcac cagttcaaga tcaatgcaa gaaggcgtg 300
accggcgatg tgtggccctg gcaccaggac tacatcttct ggaaccgggg tgacggaatg 360
cgtegccecg acgtggtgaa cgtcgcctgc ctgctggatg aggcgaccga tctgaacggt 420
ccactgctcg tcctgccccg ctcacacaag tgcggctctc tggaaagtggc gagggcgacg 480
accgtctgtg gcgacggggc gtggcggctc gacatctccg cagatctgga ctacgccatc 540
gacgcgccgc tgctgtccaa gctcaccgag actacaaaag tgaccgccat caaagggccc 600
cccggctega tcctctctct cgatcccctg ctctgtacac cctcgggctg gaacatggct 660
ccctacgatc ggcgcatgat cctcgtcacc tacaaccggg tggacaatcc actcggagag 720
gtgcccagcc ctgcaccgga cttcctcgcc gcccgggaca acacgccggt caggccgctc 780
gaccccgcg cggaccttct cgcctaa 807

```

```

<210> SEQ ID NO 6
<211> LENGTH: 24
<212> TYPE: DNA

```

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 6

ttcacactgc tcgaacttgt acct 24

<210> SEQ ID NO 7
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 7

aaacaggtac aagttcgagc agtg 24

<210> SEQ ID NO 8
<211> LENGTH: 78
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 8

gcgctcacia tttcacacag gaaacagaat taattaagct taaaggcacc cgatatggat 60

aagaaatact caataggc 78

<210> SEQ ID NO 9
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 9

tcagtcacct cctagctgac tcaaatc 27

<210> SEQ ID NO 10
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 10

gtcagctagg aggtgactga agcttggctg ttttggcgga tg 42

<210> SEQ ID NO 11
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 11

ctgtgtgaaa ttgtgagcgc tcacaattcc 30

<210> SEQ ID NO 12
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 12

tcgaaggcca ccaataactg c 21

<210> SEQ ID NO 13

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 13

cttttacttt caccagcgtt tctg 24

<210> SEQ ID NO 14

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 14

aacgctggtg aaagtaaaag atgc 24

<210> SEQ ID NO 15

<211> LENGTH: 45

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 15

gaatggaacg gatcaaacgg tgaattacac tgtacctggt gcgtc 45

<210> SEQ ID NO 16

<211> LENGTH: 43

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 16

ccgtttgatc cgttcattc gttttagagc tagaaatagc aag 43

<210> SEQ ID NO 17

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 17

caacctgcc aacagagatt ttc 23

<210> SEQ ID NO 18

<211> LENGTH: 40

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 18

-continued

aatctcgtga tggcaggtg acgggagcct aagtgaaatg 40

<210> SEQ ID NO 19
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 19

ccttggtaaa tgcacggtat gggtcagtga gatcaatcgg g 41

<210> SEQ ID NO 20
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 20

ccataccgtg catttaccaa gg 22

<210> SEQ ID NO 21
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 21

cagttattgg tgcacctcga ctttgggtc ttctttgagg 40

<210> SEQ ID NO 22
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 22

acaggccaaa ggagttgaga atgttgagtt ttgagacc 39

<210> SEQ ID NO 23
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthesized

<400> SEQUENCE: 23

ccaagcttgc atgctgcag ttacctttaa ttaccgaatc 40

1. Use of a polypeptide in producing L-proline or hydroxyproline, wherein the polypeptide is:

- A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or
- B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of an L-proline efflux function; or
- C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function;

preferably, the polypeptide has an amino acid sequence as set forth in SEQ ID NO: 1 or SEQ ID NO: 2.

2. A strain for producing L-proline, wherein the strain expresses the following polypeptide:

- A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or
- B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of an L-proline efflux function; or
- C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function;

preferably, the polypeptide has an amino acid sequence as set forth in SEQ ID NO: 1 or SEQ ID NO: 2;

preferably, the strain is a bacterium;

preferably, the strain is selected from *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC 13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067;

more preferably, glutamate kinase in the bacterium is not subjected to feedback inhibition by L-proline or is attenuated feedback inhibition by L-proline;

even more preferably, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the bacterium is enhanced.

3. A method for producing L-proline, wherein the method comprises: culturing the strain of claim 2 to produce L-proline;

preferably, the method further comprises a step of isolating L-proline from a fermentation broth.

4. A method for constructing the strain for producing L-proline of claim 2, wherein the method comprises enhancing the activity of the following polypeptide in the strain:

- A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or
- B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of an L-proline efflux function; or
- C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function;

preferably, the polypeptide has an amino acid sequence as set forth in SEQ ID NO: 1 or SEQ ID NO: 2;

preferably, the strain is a bacterium;

preferably, the strain is selected from *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067;

more preferably, glutamate kinase in the bacterium is not subjected to feedback inhibition by L-proline or is subjected to an attenuated feedback inhibition by L-proline;

even more preferably, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the bacterium is enhanced.

5. A host cell for producing hydroxyproline, wherein the activity of a polypeptide having an L-proline efflux function in the host cell is attenuated, wherein the polypeptide having the activity of the L-proline efflux function is:

- A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2; or
- B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of the L-proline efflux function; or
- C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function;

preferably, the host cell is selected from *Escherichia* and *Corynebacterium*, preferably *Escherichia coli* and *Corynebacterium glutamicum*, more preferably *Corynebacterium glutamicum*, and more particularly *Corynebacterium glutamicum* ATCC 13032, *Corynebacterium glutamicum* ATCC13869, *Corynebacterium glutamicum* B253, and *Corynebacterium glutamicum* ATCC 14067;

more preferably, glutamate kinase in the host cell is not subjected to feedback inhibition by L-proline or is subjected to an attenuated feedback inhibition by L-proline;

even more preferably, the activity of a proline hydroxylase in the host cell is enhanced;

even more preferably, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the host cell is enhanced.

6. The host cell as claimed in claim 5, wherein the proline hydroxylase is trans-proline-4-hydroxylase having a nucleotide sequence as set forth in SEQ ID NO: 5.

7. A method for producing hydroxyproline, wherein the method comprises: culturing the host cell of claim 5 to produce hydroxyproline;

preferably, the method further comprises a step of isolating hydroxyproline from a fermentation broth.

8. A method for constructing the host cell for producing hydroxyproline of claim 5, wherein the method comprises:

(1) attenuating the activity of a polypeptide having an L-proline efflux function in the cell, wherein the polypeptide having the activity of the L-proline efflux function is:

A) a polypeptide having an amino acid sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 2 and having the activity of the L-proline efflux function; or

B) a polypeptide having some amino acids added or deleted at either or both ends of the polypeptide of A) but still having the activity of the L-proline efflux function; or

C) a polypeptide having higher than 90%, 95%, 96%, 97%, 98%, or 99% homology to the polypeptide of A) and derived from *Corynebacterium* with the activity of an L-proline efflux function;

(2) relieving or attenuating feedback inhibition by L-proline on glutamate kinase; and

(3) introducing a gene encoding a proline hydroxylase into the cell obtained in step (2);

preferably, the activity of glutamate kinase and/or glutamate-semialdehyde dehydrogenase and/or pyrroline-5-carboxylate reductase in the host cell is enhanced.

9. The method for constructing the host cell for producing hydroxyproline as claimed in claim 8, wherein the proline hydroxylase is trans-proline-4-hydroxylase having a nucleotide sequence as set forth in SEQ ID NO: 5.

* * * * *