

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(10) **DK/EP 3269809 T3**

(12) **Oversættelse af
europæisk patentskrift**

- (51) Int.Cl.: **C 12 N 15/09 (2006.01)** **C 12 N 1/15 (2006.01)** **C 12 N 1/19 (2006.01)**
C 12 N 1/21 (2006.01) **C 12 N 5/10 (2006.01)** **C 12 N 9/00 (2006.01)**
C 12 P 19/34 (2006.01) **C 12 P 21/00 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2022-09-12**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2022-07-27**
- (86) Europæisk ansøgning nr.: **16764874.0**
- (86) Europæisk indleveringsdag: **2016-03-11**
- (87) Den europæiske ansøgnings publiceringsdag: **2018-01-17**
- (86) International ansøgning nr.: **JP2016057707**
- (87) Internationalt publikationsnr.: **WO2016148044**
- (30) Prioritet: **2015-03-13 JP 2015051202**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Chugai Seiyaku Kabushiki Kaisha, 5-1, Ukima 5-chome, Kita-ku, Tokyo 115-8543, Japan**
- (72) Opfinder: **OHTA, Atsushi, c/o Chugai Seiyaku Kabushiki Kaisha, 200, Kajiwara, Kamakura-shi, Kanagawa 247-8530, Japan**
YAMAGISHI, Yusuke, c/o Chugai Seiyaku Kabushiki Kaisha, 200, Kajiwara, Kamakura-shi, Kanagawa 247-8530, Japan
MATSUO, Atsushi, c/o CHUGAI SEIYAKU KABUSHIKI KAISHA, 135, Komakado 1-chome, Gotemba-shi, Shizuoka 412-8513, Japan
- (74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Strandvejen 70, 2900 Hellerup, Danmark**
- (54) Benævnelse: **MODIFICERET AMINOACYL-tRNA-SYNTETASE OG ANVENDELSE DERAFF**
- (56) Fremdragne publikationer:
EP-A1- 1 964 916
JP-A- 2009 528 824
KRISTOPHER JOSEPHSON ET AL: "Ribosomal Synthesis of Unnatural Peptides", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 127, no. 33, 1 August 2005 (2005-08-01), pages 11727-11735, XP055236675, US ISSN: 0002-7863, DOI: 10.1021/ja0515809
SHIMIZU YOSHIHIRO ET AL: "Cell-free translation reconstituted with purified components", NATURE BIOTECHNOLOGY, GALE GROUP INC, US, vol. 19, no. 8, 1 August 2001 (2001-08-01), pages 751-755, XP002677144, ISSN: 1087-0156, DOI: 10.1038/90802
LIU D R ET AL: "Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NATIONAL ACADEMY OF SCIENCES, US, vol. 94, 1 September 1997 (1997-09-01), pages 10092-10097, XP002954766, ISSN: 0027-8424, DOI: 10.1073/PNAS.94.19.10092

Fortsættes ...

- YANAGISAWA, T. ET AL.: 'Multistep engineering of Pyrrolysyl-tRNA Synthetase to genetically encode N epsilon-(o-Azidobenzyloxycarbonyl) lysine for site-specific protein modification' *CHEM. BIOL.* vol. 15, no. 11, 2008, pages 1187 - 1197, XP025678151
- KOBAYASHI, T. ET AL.: 'Recognition of non-alpha-amino substrates by Pyrrolysyl-tRNA Synthetase' *J. MOL. BIOL.* vol. 385, no. 5, 2009, pages 1352 - 1360, XP026034771
- HARTMAN, M.C.T. ET AL.: 'An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides' *PLOS ONE* vol. 2, no. 10, 2007, pages 1 - 15, XP008128032
- HARTMAN, M.C.T. ET AL.: 'Enzymatic aminoacylation of tRNA with unnatural amino acids' *PROC. NATL. ACAD. SCI. USA* vol. 103, no. 12, 2006, pages 4356 - 4361, XP055311246
- ITOH, Y. ET AL.: 'Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon *Pyrococcus horikoshii*' *RNA BIOLOGY* vol. 5, no. 3, 2008, pages 169 - 177, XP055311254
- MERMERSHTAIN, I. ET AL.: 'Idiosyncrasy and identity in the prokaryotic phe-system: Crystal structure of *E. Coli* phenylalanyl-tRNA synthetase complexed with phenylalanine and AMP' *PROTEIN SCI.* vol. 20, 2011, pages 160 - 167, XP055311262
- DOUBLIE, S. ET AL.: 'Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase' *STRUCTURE* vol. 3, no. 1, 1995, pages 17 - 31, XP004587824
- SEVER, S. ET AL.: 'Escherichia coli tryptophanyl-tRNA synthetase mutants selected for tryptophan auxotrophy implicate the dimer interface in optimizing amino acid binding' *BIOCHEMISTRY* vol. 35, no. 1, 1996, pages 32 - 40, XP055311271
- CUSACK, S. ET AL.: 'The 2 Angstrom crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue' *EMBO J.* vol. 19, no. 10, 2000, pages 2351 - 2361, XP002459775
- LEE, K.W. ET AL.: 'Molecular modeling study of the editing active site of Escherichia coli leucyl-tRNA synthetase: Two amino acid binding sites in the editing domain' *PROTEINS* vol. 54, no. 4, 2004, pages 693 - 704, XP008131999
- FUKUNAGA, R. ET AL.: 'Structural basis for non-cognate amino acid discrimination by the valyl-tRNA synthetase editing domain' *J. BIOL. CHEM.* vol. 280, no. 33, 2005, pages 29937 - 29945, XP055311274
- WU, N. ET AL.: 'A genetically encoded photocaged amino acid' *J. AM. CHEM. SOC.* vol. 126, no. 44, 2004, pages 14306 - 14307, XP002459777
- SANKARANARAYANAN, R. ET AL.: 'The structure of threonyl-tRNA synthetase-tRNAThr Complex enlightens its repressor activity and reveals an essential zinc ion in the active site' *CELL* vol. 97, no. 3, 1999, pages 371 - 381, XP055311278
- HOUNTONDJI, C. ET AL.: 'Crucial role of conserved lysine 277 in the fidelity of tRNA aminoacylation by Escherichia coli valyl-tRNA synthetase' *BIOCHEMISTRY* vol. 41, no. 50, 2002, pages 14856 - 14865, XP055311280
- ZHAI, Y. ET AL.: 'Two conserved threonines collaborate in the Escherichia coli leucyl-tRNA synthetase amino acid editing mechanism' *BIOCHEMISTRY* vol. 44, no. 47, 2005, pages 15437 - 15443, XP055311285
- CHEN, J.-F. ET AL.: 'Effect of alanine-293 replacement on the activity, ATP binding, and editing of Escherichia coli leucyl-tRNA synthetase' *BIOCHEMISTRY* vol. 40, no. 5, 2001, pages 1144 - 1149, XP055311287
- KAWAKAMI, T. ET AL.: 'In vitro selection of multiple libraries created by genetic code reprogramming to discover macrocyclic peptides that antagonize VEGFR2 activity in living cells' *ACS CHEM. BIOL.* vol. 8, no. 6, 2013, pages 1205 - 1214, XP055311288
- TERASAKA, N. ET AL.: 'Recent developments of engineered translational machineries for the incorporation of non-canonical amino acids into polypeptides' *INT. J. MOL. SCI.* vol. 16, no. 3, 20 March 2015, pages 6513 - 6531, XP055311291
- PERONA, J.J. ET AL.: 'Structural diversity and protein engineering of the aminoacyl-tRNA synthetase' *BIOCHEMISTRY* vol. 51, no. 44, 2012, pages 8705 - 8729, XP055311293

DESCRIPTION

Technical Field

[0001] Disclosed are aminoacyl-tRNA synthetases (ARSs) that aminoacylate tRNA with the corresponding N-methyl amino acid more efficiently than natural aminoacyl-tRNA synthetases, and use thereof. More specifically, the disclosed aminoacyl-tRNA synthetases have modified amino acid sequences, which are able to aminoacylate a tRNA with one of six N-methyl-substituted amino acids corresponding to the tRNA, namely, N-methyl-phenylalanine, N-methyl-valine, N-methyl-serine, N-methyl-threonine, N-methyl-tryptophan, and N-methyl-leucine, more efficiently than natural ARSs, and uses thereof. The aminoacyl-tRNA synthetases with modified amino acid sequences can be used to produce peptides that selectively and regioselectively contain N-methyl amino acids with high efficiency.

Background Art

[0002] Generally in organisms on earth, information stored in DNA (= an information-storing substance) defines, via RNA (= an information-transmitting substance), the structures of proteins (= functional substances) and functions resulting from the structures. Polypeptides and proteins are composed of 20 types of amino acids. The information stored in DNA, which is composed of four kinds of nucleotides, is transcribed into RNA and then translated into amino acids, which make polypeptides and proteins.

[0003] During translation, tRNA plays the role of an adaptor that matches a stretch of three nucleotides to one amino acid, and aminoacyl-tRNA synthetase (aminoacyl-tRNA synthetase; ARS) is involved in the attachment of tRNA to its amino acid.

[0004] ARSs are enzymes that specifically attach an amino acid to its corresponding tRNA. There are 20 types of ARSs, each corresponding to each of the 20 types of naturally-occurring amino acids with a few exceptions, in every biological species. Of the 20 kinds of proteinaceous amino acids, ARS precisely acylates tRNA having an anticodon corresponding to an arbitrary codon, using the specific amino acid assigned to the codon. In other words, a tRNA synthetase corresponding to a certain amino acid can distinguish a tRNA corresponding to that amino acid from tRNAs corresponding to other amino acids, and does not attach other amino acids.

[0005] In translation of an mRNA into a polypeptide chain, a tRNA bound to its corresponding amino acid (aminoacyl-tRNA) pairs with the appropriate codon on the mRNA starting from the initiation codon and form hydrogen bonds with the mRNA on ribosomes. This is followed by peptidyl transfer to the amino acid on the adjacent aminoacyl-tRNA bound corresponding to the next codon. The first tRNA that released the amino acid by the transfer is liberated from the

mRNA and can reattach its corresponding amino acid catalyzed by ARS. Translation is terminated upon reaching the mRNA's stop codon and upon arrival of a protein called the termination factor, which releases the polypeptide chain from the ribosome. Proteins within the living body are produced via such processes. The proteins then exert important physiological functions in the living body.

[0006] Meanwhile, in the field of pharmaceuticals of which bioactivities are exerted similarly in living bodies, the possibility of creating new pharmaceuticals using compounds having a molecular weight from 500 to 2000, called "middle molecules", is being anticipated in the field of drug discovery where it was considered difficult to create drugs with conventional low molecular weight compounds having a molecular weight less than 500. A representative example is cyclosporine A, a naturally-derived middle molecule drug, which is a peptide consisting of 11 residues that is produced by microorganisms, inhibits the intracellular target cyclophilin, and can be orally administered.

[0007] The Cyclosporine A peptide is characterized by including the non-natural amino acids "N-methyl amino acids" as components. Triggered by this, multiple studies have recently reported that the incorporation of N-methyl amino acids into peptides can increase drug-likeness of the peptides, and this is being further applied to drug discovery (Non Patent Literature 1, 2, 3). Particularly, it has come to be known that the incorporation of N-methyl amino acids leads to decrease in hydrogen bond-donating hydrogens, acquisition of protease resistance, and fixed conformation, and contributes to membrane permeability and metabolic stability (Non Patent Literature 1, 4, Patent Literature 1).

[0008] This paves way to the conception of drug discovery methods that select pharmaceutical candidate substances from a library of diverse peptides containing multiple N-methyl amino acids. In terms of diversity and ease of screening, much anticipated are mRNA display libraries of N-methyl amino acid-containing peptides, and such, which use a cell-free translation system (Non Patent Literature 9, 10, Patent Literature 1). First, a display library is constructed, which is a collection of molecules forming one-for-one pairs between an enormous variety of RNA or DNA molecules (genotype) or the like and peptides encoded by these molecules (phenotype). The display library is then allowed to bind to a target protein or the like, followed by washing to remove non-bound molecules and selecting a molecule in the library, which is contained scarcely, in extremely small amounts. The selected RNA (DNA) molecule can be then sequenced to easily obtain the sequence information of the bound peptide. Particularly, an mRNA display library and a ribosome display library utilizing a cell-free translation system can be used to easily analyze 10^{12-14} diverse types of molecules (Non Patent Literature 11). Recently, a method that can prepare peptides containing non-proteinaceous amino acids using a reconstituted cell-free translation system has been also developed, enabling combination with display techniques to construct an N-methyl amino acid-containing peptide display library (Non Patent Literature 10).

[0009] Some methods for preparing N-methyl amino acid-containing peptides by translation of mRNAs are known so far. These methods are performed by separately preparing beforehand

"N-methyl aminoacyl-tRNAs" and adding these to a translation system.

[0010] First, in the pdCpA method developed by Hecht *et al.* (Non Patent Literature 5), an N-methyl aminoacyl-tRNA is prepared beforehand by ligating pdCpA (5'-phospho-2'-deoxyribocytidylriboadenosine) acylated with a chemically-synthesized non-natural N-methyl amino acid to a tRNA lacking 3'-terminal CA obtained by transcription using T4 RNA ligase. This method has been used to introduce amino acids including N-methylalanine and N-methylphenylalanine (Non Patent Literature 6). However, when the present inventors attempted to introduce multiple N-methyl amino acids by preparing several complexes between non-natural N-methyl amino acids and tRNAs using the pdCpA method and adding the complexes to a cell-free translation system, translational efficiency decreased. Particularly, N-methylvaline could not be introduced (unpublished data).

[0011] As a different method, Suga *et al.* reported a method for aminoacylating tRNAs with N-methyl amino acids activated in advance by esterification using an artificial RNA catalyst (Flexizyme) (Patent Literature 2), and successfully introduced multiple types of N-methyl amino acids by translation. This method can be applied to various side chain structures, but aminoacylation efficiency of amino acids with nonaromatic side chains is in many cases 40 to 60% which is by no means high, and particularly, N-methylvaline introduction has not been confirmed as with the pdCpA method (Non Patent Literature 3).

[0012] Moreover, in the method of Szostak *et al.*, aminoacyl-tRNAs are obtained by using tRNAs extracted from *Escherichia coli* and wild-type ARSs, and then preparing N-methyl aminoacyl-tRNAs via a chemical N-methylation reaction consisting of three steps. However, side chains that are efficiently translated in this method are limited to only the three side chains valine, leucine, and threonine. This method also requires cumbersome operations and further results in contamination of trace amounts of natural amino acids (the starting material) resulting from incomplete progression of the N-methylation reaction. Such difficulties in controlling the reaction directly affect the purity of products (Non Patent Literature 2).

[0013] Moreover, since these techniques add N-methyl aminoacyl-tRNAs prepared outside a translation system to a translation reaction solution, N-methyl aminoacyl-tRNAs are not reproduced in the translation system and thus are only consumed in the translation process. This requires addition of large amounts of N-methyl aminoacyl-tRNAs, but this addition itself of large amounts of tRNAs contributes to the reduction of peptide yield (Non Patent Literature 7). Further, instability of aminoacyl-tRNAs in the translation solution becomes problematic. Aminoacyl-tRNAs have been shown to be hydrolyzed under physiological conditions at pH 7.5 due to the presence of ester bonds between amino acids and tRNAs (Non Patent Literature 12). Half-life of aminoacyl-tRNAs depends on amino acid side chains, the shortest being 30 minutes. Hydrolysis of aminoacyl-tRNA is suppressed when it forms a complex with an aminoacyl-tRNA-elongation factor Tu (EF-Tu), but aminoacyl-tRNAs in excess of the concentration of EF-Tu present in the translation system are hydrolyzed. That is to say, as the translation reaction proceeds, deacylation of aminoacyl-tRNAs added at the start of translation proceeds, and at the end, aminoacyl-tRNAs having N-methyl amino acids are exhausted. In

fact, Szostak *et al.* perceived this depletion as a problem and added N-methyl aminoacyl-tRNAs twice, at the start of translation and during the reaction, when synthesizing a polypeptide containing multiple N-methyl amino acids (Non Patent Literature 2).

[0014] The above-mentioned problems regarding introduction of N-methyl amino acids can be solved if there are ARSs for N-methyl amino acids having functions similar to natural ARSs for natural amino acids, but ARSs have an ability to precisely recognize their substrates and thus have limitations. As an exception, it was reported that natural HisRS and PheRS could be used to introduce N-methylhistidine and N-methylphenylalanine into peptides in a cell-free translation system (Non Patent Literature 8, 13). Also, Hartman *et al.* confirmed that natural ARSs were used to aminoacylate tRNAs with the six N-methyl amino acids N-methylvaline, N-methylleucine, N-methyl aspartic acid, N-methylhistidine, N-methyllysine, and N-methyltryptophan (Non Patent Literature 14). However, even though a subsequent report using a cell-free translation system and natural ARSs that analyzed translational synthesis of peptides containing N-methylhistidine and N-methyl aspartic acid by mass spectrum reported a certain level of yield, in the case of peptides containing N-methylvaline, N-methylleucine, N-methyllysine, and N-methyltryptophan, it was shown that the efficiency of translational synthesis (ribosomal synthesis) is very low (Non Patent Literature 8). These reports reveal that aminoacylation with N-methyl amino acids and translational introduction of N-methyl amino acids into peptides using natural ARSs has been confirmed substantially in only the three cases of N-methylphenylalanine, N-methylhistidine, and N-methyl aspartic acid.

[0015] There are some prior art in which natural ARSs have been modified to give them the function to catalyze the attachment of non-natural amino acids to tRNAs (Patent Literature 3, 4, 5). Even though these are modified ARSs that catalyze the attachment of non-natural amino acids to tRNAs, and substrates of these modified ARSs are amino acids mainly having side chain derivatives of phenylalanine and tyrosine, modified ARSs having N-methyl amino acids as substrates, and modified ARSs that can introduce multiple N-methyl amino acid residues into peptides have not been known.

Citation List

Patent Literature

[0016]

Patent Literature 1: WO2013/100132

Patent Literature 2: WO2007/066627

Patent Literature 3: WO2003/014354

Patent Literature 4: WO2007/103307

Patent Literature 5: WO2002/085923

Non Patent Literature

[0017]

Non Patent Literature 1: R. S. Lokey et al., Nat. Chem. Biol. 2011, 7(11), 810-817.

Non Patent Literature 2: J. W. Szostak et al., J. Am. Chem. Soc. 2008, 130, 6131-6136.

Non Patent Literature 3: T. Kawakami et al., Chemistry & Biology, 2008, Vol. 15, 32-42.

Non Patent Literature 4: H. Kessler et al., J. Am. Chem. Soc. 2012, 134, 12125-12133

Non Patent Literature 5: S.M. Hecht et al., J. Biol. Chem. 253 (1978) 4517-4520.

Non Patent Literature 6: Z. Tan et al., J. Am. Chem. Soc. 2004, 126, 12752-12753.

Non Patent Literature 7: A. O. Subtelný et al., Angew Chem Int Ed 2011 50 3164.

Non Patent Literature 8: M. C. T. Hartman et al., PLoS one, 2007, 10, e972.

Non Patent Literature 9: S. W. Millward et al., J. Am. Chem. Soc., 2005, 127, 14142-14143

Non Patent Literature 10: Y. Yamagishi et al., Chem. Biol., 18, 1562-1570, 2011

Non Patent Literature 11: H. R. Hoogenboom, Nature Biotechnol. 23, 1105-1116, 2005

Non Patent Literature 12: J. R. Peacock et al., RNA, 20, 758-64, 2014

Non Patent Literature 13: T. Kawakami, ACS Chem Biol., 8, 1205-1214, 2013

Non Patent Literature 14: M. C. T. Hartman et al., Proc Natl Acad Sci USA., 103, 4356-4361, 2006

Summary

[Problems to be Solved]

[0018] An objective of the present disclosure is to provide modified ARSs that have been modified to increase reactivity, and which use N-methyl amino acids as substrates. More

specifically, an objective of the present disclosure is to provide a novel modified ARS that catalyzes the acylation reaction in which tRNAs attach non-natural N-methyl amino acids, particularly N-methyl-phenylalanine, N-methyl-valine, N-methyl-serine, N-methyl-threonine, N-methyltryptophan, and N-methylleucine, without using large amounts of tRNAs in order to efficiently produce peptides containing multiple N-methyl amino acids, and uses thereof.

[Means for Solving the Problems]

[0019] The invention relates to the embodiments as defined in the claims. In particular the invention relates to a modified valyl-tRNA synthetase (ValRS) which incorporates an N-methyl valine more efficiently than the ValRS having the amino acid sequence SEQ ID NO:24, wherein said modified ValRS is selected from the following (a) and (b):

1. (a) a ValRS modified at a position(s) corresponding to asparagine at position 43 and/or threonine at position 45 and/or threonine at position 279 of ValRS from *Escherichia coli* having the amino acid sequence SEQ ID NO:24, and
2. (b) a ValRS having (i) glycine or alanine at a position corresponding to asparagine at position 43 and/or (ii) serine at a position corresponding to threonine at position 45 and/or (iii) glycine or alanine at a position corresponding to threonine at position 279 of ValRS from *Escherichia coli* having the amino acid sequence SEQ ID NO:24,

wherein the modified ValRS of (a) and (b) comprises an amino acid sequence having at least 90% identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 3 to 5 and 182 and 183.

[0020] The invention also relates to a polynucleotide encoding the modified ValRS of the invention.

[0021] The invention further relates to a vector comprising a polynucleotide encoding the modified ValRS of the invention.

[0022] In addition, the invention relates to a host cell comprising a polynucleotide encoding the modified ValRS of the invention or a vector comprising a polynucleotide encoding the modified ValRS of the invention.

[0023] The invention additionally relates to a method producing the modified ValRS of the invention comprising the step of culturing the host cell as described directly above.

[0024] The invention further relates to a method for producing a tRNA acylated with an N-methylvaline, comprising the step of contacting the N-methylvaline with a tRNA in the presence of the modified ValRS of the invention.

[0025] The invention still further relates to a method for producing a polypeptide comprising an

N-methylvaline, comprising the step of performing translation in the presence of the modified ValRS of the invention and the N-methylvaline.

[0026] In order to obtain ARSs having increased reactivity with N-methyl amino acids, the present inventors obtained multiple ARS genes that employ different amino acids as substrates and introduced mutations into the genes to construct mutated ARS genes encoding ARSs with altered amino acid sequences. These modified ARSs were expressed and collected, and were incubated with tRNAs in the presence of unmodified amino acids or N-methyl amino acids to determine aminoacylation reaction. As a result of estimating the conformation formed in the N-methyl amino acid-ARS interaction, and after much trial and error, the present inventors successfully obtained modified ARSs, including multiple ARSs such as phenylalanyl-tRNA synthetase (PheRS), seryl-tRNA synthetase (SerRS), valyl-tRNA synthetase (ValRS), threonyl-tRNA synthetase (ThrRS), leucyl-tRNA synthetase (LeuRS), and tryptophanyl-tRNA synthetase (TrpRS), with increased activity in the aminoacylation reaction with N-methyl amino acids, compared to wild-type ARSs.

[0027] For example, 0.1 μM wild-type PheRS hardly incorporated N-methylphenylalanine into ribosomally-synthesized peptides even when 1 mM N-methylphenylalanine was added. On the other hand, 0.1 μM modified PheRS clearly incorporated N-methylphenylalanine into peptides even when 0.25 mM N-methylphenylalanine was added (Example 1). The amount of phenylalanine and N-methylphenylalanine translationally introduced (ribosomally introduced) into peptides were measured by mass spectroscopy using MALDI-TOF MS. The ratio of the peak values (the peak intensity of the peptide containing N-methylphenylalanine / the peak intensity of the peptide containing phenylalanine) at the time of 0.25 mM N-methylphenylalanine addition was 0.8 when using wild-type PheRS, whereas it dramatically increased to 12.4 when using the modified PheRS α subunit (Example 1). When sequences containing two consecutive and three consecutive phenylalanines were allowed to translate, peptides containing two consecutive and three consecutive N-methylphenylalanines were confirmed to be synthesized respectively, and the efficiency was significantly higher than that using the pdCpA method. Also, ValRS was used to perform translational synthesis in the presence of 5 mM N-methylvaline, and the peptide products were analyzed with mass spectroscopy. A peptide incorporated with unmodified valine was detected as the main product when using wild-type ValRS, whereas a peptide incorporated with N-methylvaline was observed as the main product when using modified ValRS (Example 2). Furthermore, the selectivity to N-methylvaline was successfully increased by introducing mutations into the editing domain of ValRS and decreasing aminoacylation activity with unmodified Val (Example 7). When sequences containing two consecutive and three consecutive valines were allowed to translate, peptides containing two consecutive and three consecutive N-methylvalines were confirmed to be synthesized respectively (Example 2). For SerRS, an N-methylserine-incorporated translation product was detected as the main product when modified SerRS was used under conditions in which an unmodified serine-incorporated translation product is detected as the main product when using wild-type SerRS (Example 3). Even for ThrRS, an N-methyl Thr-incorporated translation product was observed when modified ThrRS was used under conditions in which an unmodified Thr-incorporated translation product is detected when

using wild-type ThrRS; whereas it was demonstrated that translation products incorporated with unmodified Thr were hardly observed and the peptide with N-methyl Thr introduced was synthesized with higher purity compared to that achieved with wild-type ThrRS (Example 4). For TrpRS, unmodified Trp was detected as a main product in using wild-type TrpRS, whereas the translation product with N-methyl Trp incorporated was observed as a main product using the modified TrpRS (Example 5). For LeuRS, translation products containing N-methyl Leu were not observed in using wild-type LeuRS, whereas it is revealed that translation products with N-methyl Leu incorporated were produced as much as translation products containing unmodified Leu when modified LeuRSs were used (Example 6). Thus, the modified ARSs as disclosed herein can be used to introduce N-methyl amino acids into peptides more efficiently than wild-type ARS.

[0028] The present disclosure provides ARSs having reactivity with N-methyl amino acids. Specifically, the present disclosure provides an aminoacyl-tRNA synthetase (aminoacyl-tRNA synthetase; ARS) that has an altered amino acid sequence and is able to incorporate any N-methyl amino acid, particularly the six N-methyl-substituted amino acids of N-methylphenylalanine, N-methyl-valine, N-methyl-serine, N-methyl-threonine, N-methyltryptophan, and N-methylleucine more efficiently than natural ARSs and use thereof. The ARS with an altered amino acid sequence as described herein can be used to produce peptides selectively and regioselectively containing any N-methyl amino acid from among these N-methyl amino acids with high efficiency.

[0029] Further disclosed is a method for producing polypeptides containing non-natural amino acids using a modified ARS as described herein. More specifically, disclosed is a method for producing polypeptides containing N-methylphenylalanine, N-methylvaline, N-methylserine, N-methylthreonine, N-methyltryptophan, and N-methylleucine using ARSs for phenylalanine, valine, serine, threonine, tryptophan, and leucine, respectively, with altered amino acid sequences.

[Effects of the Disclosure]

[0030] The modified ARSs as disclosed herein can be used to efficiently attach N-methylphenylalanine, N-methylvaline, N-methylthreonine, N-methylserine, N-methyltryptophan, and N-methylleucine to tRNAs corresponding to natural phenylalanine, valine, threonine, serine, tryptophan, and leucine, respectively, without complicated reactions.

[0031] Methods using modified ARSs as disclosed herein require no stoichiometric amount of tRNA, can synthesize peptides into which multiple N-methyl amino acids are introduced with a high translational efficiency, and is useful for generating a highly diverse peptide library.

[0032] In order to investigate the effect on translational efficiency provided by the ARS's characteristic feature of "providing a continuous supply of aminoacyl-tRNA during the translation reaction", the present inventors compared the introduction efficiencies of N-

methylphenylalanine in two methods: a method using the modified PheRS05 (SEQ ID NO: 2) obtained by the disclosed application and the pdCpA method in which aminoacyl-tRNA is not expected to be regenerated during the translation reaction. As a result, the translational efficiency of the method using the modified ARS was higher, and particularly, when two consecutive and three consecutive N-methylphenylalanines were introduced, the target peptide was synthesized approximately 4 to 8 times more (unpublished data). Thus, the present disclosure enables more efficient production of N-methyl amino acid-containing polypeptides, which were conventionally hard to produce.

Brief Description of Drawings

[0033]

[Figure 1] Figure 1 shows evaluation of aminoacylation activities of modified ARSs. The bands of the synthetic peptide acylated with N-methylphenylalanine using PheRS04 and PheRS05 were detected more strongly than when using wild-type PheRS (lane 8, 9 vs. 12, 13).

[Figure 2] Figure 2 shows the confirmation by electrophoresis of peptides ribosomally synthesized using mutant PheRS. The bands of the N-methylphenylalanine-containing peptide synthesized using the modified PheRS were detected more strongly than that using wild-type PheRS (lanes 2, 3 vs. 5, 6).

[Figure 3-1] Figure 3-1 shows detection by mass spectroscopy of peptides ribosomally synthesized using the modified PheRS. Mass spectrum of the peptide ribosomally synthesized using (a) 0.1 μ M wt PheRS and 0.25 mM Phe, (b) 0.1 μ M wt PheRS and 0.25 mM MePhe, or (c) 0.1 μ M wt PheRS and 1 mM MePhe.

[Figure 3-2] Figure 3-2 shows detection by mass spectroscopy of peptides ribosomally synthesized using the modified PheRS. Mass spectrum of the peptide ribosomally synthesized using (d) 0.1 μ M PheRS05 and 0.25 mM Phe, (e) 0.1 μ M PheRS05 and 0.25 mM MePhe, or (f) 0.1 μ M PheRS05 and 1 mM MePhe.

[Figure 4] Figure 4 shows aminoacylation reaction performed using modified ValRSs. tRNA acylated with N-methylvaline using the mutant 13 (ValRS13) was observed more than tRNA acylated using wild-type ValRS (lane 10).

[Figure 5] Figure 5 shows the results of mass spectrometry analysis of peptides translated using the modified ValRSs. Mass spectrum of the peptides translated using (a) wild-type ValRS, (b) ValRS13, or (c) ValRS04. The peptide containing MeVal was observed as a main product when ValRS13 was used.

[Figure 6] Figure 6 shows aminoacylation reaction performed using ValRS13-11. tRNA acylated with N-methylvaline using the mutant 13-11 (ValRS13-11) was observed more than tRNA acylated using wild-type ValRS or the mutant 13 (ValRS13).

[Figure 7] Figure 7 shows comparison of activities between ValRS13 and ValRS13-11. Mass spectra of the N-methyl peptide-containing peptides translated using ValRS13 ((a), (c), (e)) and ValRS13-11 ((b), (d), (f)). It can be seen that the N-methylvaline-containing target peptide translated using ValRS13-11 had higher purity.

[Figure 8] Figure 8 shows aminoacylation reaction performed using modified SerRSs. tRNA acylated with N-methylserine was observed when mutants 03, 35, and 37 were used (lanes 3, 25, 27).

[Figure 9] Figure 9 shows the results of mass spectrometry analysis of peptides ribosomally synthesized using the modified SerRSs. Mass spectrum of the N-methylserine-containing peptide translated using (a) wild-type SerRS, (b) SerRS03, (c) SerRS05, (d) SerRS35, or (e) SerRS37. The MeSer-containing target peptide synthesized using each modified SerRS had higher purity compared to that synthesized using wild-type SerRS.

[Figure 10] Figure 10 shows the results of mass spectrometry analysis of peptides ribosomally synthesized using modified ThrRSs. Mass spectrum of the N-methylthreonine-containing peptide translated using (a) wild-type ThrRS, (b) ThrRS03, or (c) ThrRS14. The MeThr-containing target peptide synthesized using each modified ThrRS had higher purity compared to that synthesized using wild-type ThrRS.

[Figure 11] Figure 11 shows the results of mass spectrometry analysis of peptides ribosomally synthesized using modified TrpRSs. Mass spectrum of the peptide translated using (a) wild-type TrpRS, (b) TrpRS04, (c) TrpRS05, or (d) TrpRS18. The MeW-containing peptide was observed as a main product when TrpRS04, 05, and 18 were used.

[Figure 12] Figure 12 shows the results of mass spectrometry analysis of peptides ribosomally synthesized using the modified LeuRS. Mass spectrum of the peptide translated using (a) wild-type LeuRS or (b) LeuRS02. The MeL-containing peptide was observed as a main product when LeuRS02 was used.

[Figure 13] Figure 13 shows aminoacylation reaction with MeVal using the modified ValRSs having mutation in the editing domain. The activities of three mutants, 13-11, 66, and 67 were not much different (lanes 17, 18, 19).

[Figure 14] Figure 14 shows aminoacylation reaction with Val using the modified ValRSs having mutation in the editing domain. The activities of the mutants 66 and 67 were attenuated compared to the activity of the mutant 13-11 (lane 17 vs 18, 19).

Mode for Carrying Out the Disclosure

[0034] Disclosed are mutant enzymes with altered enzyme-substrate specificity for aminoacyl-tRNA synthetase. Also disclosed is the preparation of mutant aminoacyl-tRNA synthetases that

can efficiently and selectively produce polypeptides containing N-methyl amino acids in large amounts, by altering the amino acid sequence of natural aminoacyl-tRNA synthetase.

[0035] Disclosed are polypeptides including ARSs that can react with N-methyl amino acids. More specifically, polypeptides are disclosed comprising modified ARSs that can react with N-methyl amino acids more efficiently than the original, natural ARSs. Further disclosed are polypeptides comprising modified ARSs that can incorporate N-methyl amino acids more efficiently than the original, natural ARSs. The polypeptides as described herein are polypeptides that have aminoacyl-tRNA synthetase activity and are modified to enhance the aminoacylation reaction with N-methyl amino acids. As used herein, the phrase "incorporate N-methyl amino acids" refers to, for example, aminoacylation of tRNAs with N-methyl amino acids corresponding to the tRNAs and may be attachment of the N-methyl amino acids to the tRNAs or incorporation of N-methyl amino acids to proteins synthesized in a translation reaction using the aminoacyl-tRNA produced in the acylation reaction. The phrase "a polypeptide has aminoacyl-tRNA synthetase activity" includes not only the case in which the polypeptide exhibits aminoacyl-tRNA synthetase activity by itself, but also the case in which the polypeptide exhibits aminoacyl-tRNA synthetase activity together with other factors. For example, when an aminoacyl-tRNA synthetase is composed of multiple subunits, the polypeptide as described herein may be one subunit or may exhibit aminoacyl-tRNA synthetase activity as a complex with other subunits. In such a case, when an aminoacyl-tRNA synthetase complex is formed between a modified polypeptide as described herein and other wild-type subunits, the aminoacyl-tRNA synthetase complex can enhance aminoacylation reaction with N-methyl amino acids more than an aminoacyl-tRNA synthetase complex consisting of wild-type subunits. That is to say, polypeptides with aminoacyl-tRNA synthetase activity modified to enhance aminoacylation reaction with N-methyl amino acids include a polypeptide that is one modified subunit in an aminoacyl-tRNA synthetase complex consisting of multiple subunits and is modified to enhance aminoacylation reaction with N-methyl amino acids by the aminoacyl-tRNA synthetase complex.

[0036] A polypeptide comprising a modified ARS refers to a polypeptide comprising the polypeptide chain of the modified ARS, and specifically the polypeptide is a polypeptide comprising the amino acid sequence of the modified ARS. The original, natural ARS refers to a natural ARS from which modified ARS is derived, may be for example a wild-type ARS, and includes a naturally-occurring polymorphism. The phrase "can react with N-methyl amino acids" means that modified ARSs can perform an enzymatic reaction with N-methyl amino acids as substrates. The reaction may be, for example, an acylation reaction of tRNAs with N-methyl amino acids, and specifically a reaction that catalyzes a coupling reaction between an N-methyl amino acid and a tRNA. For example, depending on an amino acid used as a substrate by an ARS, the reaction is performed in the presence of the appropriate N-methyl amino acid and the appropriate tRNA, and the coupling between the N-methyl amino acid and the tRNA may be detected. Alternatively, a reaction with N-methyl amino acids may be incorporation of N-methyl amino acids into polypeptides in translation. Production of N-methyl amino acid-tRNA by ARSs can be detected by, for example, performing translation in the presence of the modified ARSs and N-methyl amino acids and detecting incorporation of N-

methyl amino acids into the polypeptides produced by translation. The reactivity with N-methyl amino acids is considered to be higher when N-methyl amino acids are frequently incorporated into polypeptides.

[0037] The phrase "a modified ARS can react more efficiently than the original, natural ARS" may mean that the modified ARS reacts more efficiently than the original, natural ARS at least under a certain condition, or that a reaction or reaction product that cannot be observed when using the original ARS can be observed when using the modified ARS. For example, a modified ARS is considered to react with an N-methyl amino acid more efficiently than the original, natural ARS when the modified ARS produces polypeptides containing the N-methyl amino acid more than the original, natural ARS. A modified ARS is considered to react with its corresponding N-methyl amino acid more efficiently than the original, natural ARS when, for example, production of polypeptides containing the N-methyl amino acid that cannot be observed when using the original, natural ARS can be observed when using the modified ARS. For example, a modified ARS is considered to react with an N-methyl amino acid more efficiently than the original, natural ARS when, for example, production of polypeptides containing 2, 3, or more consecutive N-methyl amino acids cannot be observed when using the original, natural ARS, but can be observed when using the modified ARS.

[0038] The phrase "can react more efficiently than the original, natural ARS" may mean that a reactant of interest is purified at least under a certain condition with higher purity compared to the purity achieved with the original, natural ARS. For example, a modified ARS is considered to react with its corresponding N-methyl amino acid more efficiently than the original, natural ARS when the modified ARS produces polypeptides containing its corresponding N-methyl amino acid more than the original, natural ARS, relative to natural amino acids derived as a result of contamination. Alternatively, a modified ARS is considered to react with N-methyl amino acids more efficiently than the original ARS when it is confirmed that reaction efficiency of the modified ARS for the N-methyl amino acid remains unchanged and reaction efficiency of the modified ARS for natural amino acids decreases. For example, a modified ARS is considered to react with its corresponding N-methyl amino acid more efficiently than the original, natural ARS when reactivity of the modified ARS to the N-methyl amino acid is relatively higher than reactivity of the modified ARS to natural amino acids.

[0039] N-methyl amino acids are not particularly limited, but are appropriately selected based on ARSs. For example, when the modified ARS is valine ARS (ValRS), N-methyl amino acid is N-methylvaline; when the modified ARS is threonine ARS (Thr), N-methyl amino acid is N-methylthreonine; when the modified ARS is serine ARS (SerRS), N-methyl amino acid is N-methylserine; when the modified ARS is phenylalanine ARS α subunit (PheRS), N-methyl amino acid is N-methylphenylalanine; when the modified ARS is tryptophan ARS (TrpRS), N-methyl amino acid is N-methyltryptophan; when the modified ARS is leucine ARS (LeuRS), N-methyl amino acid is N-methylleucine.

[0040] For example in ValRS, modified sites of ARS are preferably the position(s) corresponding to asparagine at position 43 and/or threonine at position 45 and/or threonine at

position 279 of ValRS from *E. coli*. Modified sites of ARS are preferably a combination of any two positions selected from the positions corresponding to asparagine at position 43, threonine at position 45, and threonine at position 279 (e.g., a combination of position 43 and position 45, position 43 and position 279, or position 45 and position 279), and more preferably a combination of positions corresponding to asparagine at position 43, threonine at position 45, and threonine at position 279. SerRS can be modified at the position(s) corresponding to glutamic acid at position 239 and/or threonine at position 237 of SerRS from *E. coli*, and more preferably, can be modified at a combination of the positions corresponding to glutamic acid at position 239 and threonine at position 237. PheRS α subunit is preferably modified at the position corresponding to glutamine at position 169 of PheRS from *E. coli*. ThrRS can be modified at the position(s) corresponding to methionine at position 332 and/or histidine at position 511 of ThrRS from *E. coli*. TrpRS can be modified at the position(s) corresponding to methionine at position 132 and/or glutamine at position 150 and/or histidine at position 153 of TrpRS from *E. coli*, preferably can be modified at a combination of any two positions selected from positions corresponding to methionine at position 132, glutamine at position 150, and histidine at position 153 (e.g., a combination of position 132 and position 150, position 132 and position 153, or position 150 and position 153), and more preferably can be modified at a combination of the positions corresponding to methionine at position 132, glutamine at position 150, and histidine at position 153. LeuRS can be modified at the position corresponding to tyrosine at position 43 of LeuRS from *E. coli*. It should be noted that these modified ARSs may be further modified at other positions. The position numbers in each ARS are indicated taking the position number of the starting methionine in each ARS from *E. coli* as 1. Specifically, the position numbers in each ARS are indicated taking the position number of the first methionine in the sequences of P07118 (SEQ ID NO: 24) for ValRS, P08312 (SEQ ID NO: 28) for PheRS α subunit, P0A8M3 (SEQ ID NO: 29) for ThrRS, P0A8L1 (SEQ ID NO: 26) for SerRS, P00954 (SEQ ID NO: 188) for TrpRS, and P07813 (SEQ ID NO: 189) for LeuRS (UniProt (<http://www.uniprot.org/>)) as 1. In a certain ARS, a position corresponding to a certain amino acid in ARS from *E. coli* refers to the amino acid located in the site corresponding to the amino acid in ARS from *E. coli* and can be identified based on structural similarity between the certain ARS and ARS from *E. coli*. For example, the corresponding amino acid can be identified as the amino acid aligned at the position of the amino acid in ARS from *E. coli* when the amino acid sequences of an ARS of interest and ARS from *E. coli* are aligned. As used herein, a position corresponding to a certain amino acid in ARS from *E. coli* is preferably the position sterically corresponding to the certain amino acid in ARS from *E. coli*. The sterically corresponding position refers to the position of an amino acid corresponding to the position of the certain amino acid in ARS from *E. coli* in the conformation of ARS.

[0041] Those skilled in the art can easily identify the sterically corresponding position by aligning ARS from *E. coli* with all known ARSs from other biological species for example using Multiple Sequence Alignment with default parameters in ClustalW ver2.1 (<http://clustalw.ddbj.nig.ac.jp>). Particularly, ARSs of interest are not limited to those from prokaryotes. Generally, sequences of ARSs in eukaryotes comprise various functional domains in addition to catalytic domain, and the sequence identity between ARSs from eukaryotes and ARSs from prokaryotes is not always high. In contrast, sequences of catalytic sites, including

amino acid recognition site, and editing domain are highly conserved, and the sterically corresponding position in ARSs from eukaryotes can be easily identified using publicly-available alignment techniques.

[0042] For example, the sites corresponding to positions 43 and 45 of *E. coli* ValRS may be respectively amino acid sites of "N/Y/T" and "T/S" in PPP(N/Y/T)X(T/S)G motif (SEQ ID NO: 180; "N/Y/T" is preferably N; X is any amino acid, preferably V, I, or P, and more preferably V; "T/S" is preferably T) present in ValRS from other organisms. More preferably, the sites corresponding to positions 43 and 45 of *E. coli* ValRS may be amino acids at N and T respectively in PPPNXTG motif (SEQ ID NO: 181; X is any amino acid, preferably V, I, or P, and more preferably V) present in ValRS from other organisms. For example, the position corresponding to asparagine at position 43 of *E. coli* ValRS is asparagine at position 345 in human (Uniprot P26640) and asparagine at position 191 in *Saccharomyces cerevisiae* (Uniprot P07806).

[0043] Modifications of ARSs include preferably at least one substitution with an amino acid that causes a decrease of 10 or more in molecular weight. Such modifications include, for example, a modification of an amino acid selected from the group consisting of amino acids other than Thr (T), such as Gln (Q), Asn (N), Glu (E), Met (M), Tyr (Y), and His (H), into Ala (A) or Gly (G) (preferably into Gly), for example, a modification of Thr (T) into Ser (S), and for example, a modification of Met (M) into Val (V).

[0044] Amino acids to be modified may be appropriately selected. For example in ValRS, the position corresponding to asparagine at position 43 of ValRS from *E. coli* is preferably modified into glycine or alanine, the position corresponding to threonine at position 45 of ValRS from *E. coli* is preferably modified into serine, and the position corresponding to threonine at position 279 of ValRS from *E. coli* is preferably modified into glycine or alanine. In SerRS, the position corresponding to glutamic acid at position 239 of SerRS from *E. coli* is preferably modified into glycine or alanine, and the position corresponding to threonine at position 237 of SerRS from *E. coli* is preferably modified into serine. In PheRS, the position corresponding to glutamine at position 169 of PheRS α subunit from *E. coli* is preferably modified into glycine or alanine. In ThrRS, the position corresponding to methionine at position 332 of ThrRS from *E. coli* is preferably modified into glycine, and the position corresponding to histidine at position 511 of ThrRS from *E. coli* is preferably modified into glycine. In TrpRS, the position corresponding to methionine at position 132 of TrpRS from *E. coli* is preferably modified into valine or alanine, the position corresponding to glutamine at position 150 of TrpRS from *E. coli* is preferably modified into alanine, and the position corresponding to histidine at position 153 of TrpRS from *E. coli* is preferably modified into alanine. In LeuRS, the position corresponding to tyrosine at position 43 of LeuRS from *E. coli* is preferably modified into glycine.

[0045] Specifically, the following polypeptides are disclosed:

1. (a) a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-5, 182, and 183 (ValRS04, ValRS13, ValRS13-11, ValRS66, and ValRS67); and

2. (b) a polypeptide that has reactivity with N-methyl Val, has at least 90% identity to the amino acid sequence of any of SEQ ID NOs: 3-5, 182, and 183, and comprises at least one amino acid of the following (i) to (iii):
 1. (i) Gly or Ala at the amino acid position corresponding to position 43 of SEQ ID NOs: 3-5, 182, and 183;
 2. (ii) Ser at the amino acid position corresponding to position 45 of SEQ ID NOs: 3-5, 182, and 183; and
 3. (iii) Gly or Ala at the amino acid position corresponding to position 279 of SEQ ID NOs: 3-5, 182, and 183.

[0046] Furthermore, the above-mentioned reactivity is preferably higher than the reactivity of a polypeptide having (i) Asn at the amino acid position corresponding to position 43 and/or (ii) Thr at the amino acid position corresponding to position 45 and/or (iii) Thr at the amino acid position corresponding to position 279. For example, the ValRS as described herein preferably has a reactivity with N-methyl Val higher than the reactivity of a ValRS having Asn at the amino acid position corresponding to position 43, Thr at the amino acid position corresponding to position 45, and Thr at the amino acid position corresponding to position 279 of the ValRS.

[0047] Furthermore, the following polypeptides are disclosed:

1. (a) a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 6-9 (SerRS03, SerRS05, SerRS35, and SerRS37) and
2. (b) a polypeptide that has reactivity with N-methyl Ser, has at least 90% identity to the amino acid sequence of any of SEQ ID NOs: 6-9, and comprises at least one amino acid of the following (i) and (ii);
 1. (i) Ser at the amino acid position corresponding to position 237 of SEQ ID NOs: 6-9, and
 2. (ii) Gly or Ala at the amino acid position corresponding to position 239 of SEQ ID NOs: 6-9.

[0048] Furthermore, the above-mentioned reactivity is preferably higher than the reactivity of a polypeptide having (i) Thr at the amino acid position corresponding to position 237 and/or (ii) Glu at the amino acid position corresponding to position 239. For example, the SerRS as described herein preferably has a reactivity with N-methyl Ser higher than the reactivity of a SerRS having Thr at the amino acid position corresponding to position 237 of the SerRS and Glu at the amino acid position corresponding to position 239 of the SerRS.

[0049] In addition, the following polypeptides are disclosed:

1. (a) a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 1-2 (PheRS05 and PheRS04); and

2. (b) a polypeptide that has reactivity with N-methyl Phe, has at least 90% identity to the amino acid sequence of any of SEQ ID NOs: 1-2, and comprises an amino acid sequence in which the amino acid at the position corresponding to position 169 of any of SEQ ID NOs: 1-2 is Gly or Ala. Furthermore, the above-mentioned reactivity is preferably higher than the reactivity of a polypeptide in which the amino acid at the above-mentioned position is Gln. The polypeptides as described above represent ARS α subunit, and therefore the polypeptides can form a complex with β subunit to result in a functional ARS. β subunit is not particularly limited, but may be, for example, a desired wild-type subunit. As an example, β subunit that can be used is one comprising the amino acid sequence of NCBI Reference Sequence WP_000672380 (e.g., WP_000672380.1) (wherein the base sequence represents 1897337-1899721 of GenBank CP009685 (e.g., CP009685.1)).

[0050] In addition the following polypeptides are disclosed:

1. (a) a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 10-11 (ThrRS03 and ThrRS14); and
2. (b) a polypeptide that has reactivity with N-methyl Thr, has at least 90% identity to the amino acid sequence of any of SEQ ID NOs: 10-11, and comprises at least one amino acid of the following (i) and (ii):
 1. (i) Gly at the amino acid position corresponding to position 332 of SEQ ID NOs: 10-11; and
 2. (ii) Gly at the amino acid position corresponding to position 511 of SEQ ID NOs: 10-11.

[0051] Furthermore, the above-mentioned reactivity is preferably higher than the reactivity of a polypeptide having (i) Met at the amino acid position corresponding to position 332 and/or (ii) His at the amino acid position corresponding to position 511. For example, the ThrRS as described herein preferably has a reactivity with N-methyl Thr higher than the reactivity of a ThrRS having Met at the amino acid position corresponding to position 332 of the ThrRS and His at the amino acid position corresponding to position 511 of the ThrRS.

[0052] Furthermore, the following polypeptides are disclosed:

1. (a) a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 184-186 (TrpRS04, TrpRS05, and TrpRS18); and
2. (b) a polypeptide that has reactivity with N-methyl Trp, has at least 90% identity to the amino acid sequence of any of SEQ ID NOs: 184-186, and comprises at least one amino acid according to any of the following (i) to (iii):
 1. (i) Val or Ala at the amino acid position corresponding to position 132 of SEQ ID NOs: 184-186;

2. (ii) Ala at the amino acid position corresponding to position 150 of SEQ ID NOs: 184-186; and
3. (iii) Ala at the amino acid position corresponding to position 153 of SEQ ID NOs: 184-186.

[0053] Furthermore, the above-mentioned reactivity is preferably higher than the reactivity of a polypeptide having (i) Met at the amino acid position corresponding to position 132 and/or (ii) Gln at the amino acid position corresponding to position 150 and/or (iii) His at the amino acid position corresponding to position 153. For example, the TrpRS as described herein preferably has a reactivity with N-methyl Trp higher than the reactivity of a TrpRS having Met at the amino acid position corresponding to position 132 of the TrpRS, Gln at the amino acid position corresponding to position 150 of the TrpRS, and His at the amino acid position corresponding to position 153 of the TrpRS.

[0054] Additionally, the following polypeptides are disclosed:

1. (a) a polypeptide comprising the amino acid sequence of SEQ ID NO: 187 (LeuRS02); and
2. (b) a polypeptide that has reactivity with N-methyl Leu, has at least 90% identity to the amino acid sequence of SEQ ID NO: 187, and comprises an amino acid sequence in which the amino acid at the position corresponding to position 43 of SEQ ID NO: 187 is Gly.

[0055] Furthermore, the above-mentioned reactivity is preferably higher than the reactivity of a polypeptide having (i) Tyr at the amino acid position corresponding to position 43.

[0056] The identity of amino acid sequence is preferably 93% or more, more preferably 95% or more, more preferably 97% or more, 98% or more, or 99% or more.

[0057] The N-methyl aminoacyl-tRNA synthetases as described herein are characterized by the ability to efficiently acylate tRNAs with N-methyl amino acids, which are non-natural amino acids known to enhance drug-likeness of peptides. The N-methyl aminoacyl-tRNA synthetases as described herein may be derived from any organism including bacteria such as *E. coli*, yeast, animals, or plants, but, because of their versatility, preferable are those that have high sequence conservation with aminoacyl-tRNA synthetases exemplified in the Examples herein (SEQ ID NOs: 1-11, 182-187), and thus have mutation site(s) easily identified in other biological species. For example, polypeptides comprising the modified ARSs as described herein may be derived from ARSs from eukaryotes or prokaryotes. Eukaryotes include protists (including protozoan and unicellular green algae), fungi (including ascomycetes and basidiomycetes), plants (including bryophytes, pteridophytes, and seed plants (gymnosperms and angiosperms)), and animals (including invertebrates and vertebrates). Prokaryotes include

archaebacteria (including thermophiles and methane bacteria) and eubacteria (including cyanobacteria and *E. coli*). Polypeptides comprising the modified ARSs as described herein may be derived from mammals (such as human, mouse, rat, guinea pig, rabbit, sheep, monkey, goat, donkey, cattle, horse, and pig). Polypeptides comprising the modified ARSs as described herein may be derived from, for example, *E. coli* or yeast, and preferably derived from prokaryotes, for example bacteria. The polypeptides as described herein are derived from, for example, bacteria of the family Enterobacteriaceae, for example, *E. coli*, including, but are not limited to, for example, the genera *Escherichia* (including *E. coli*, *E. albertii*, *E. fergusonii*), *Shigella* (including *S. dysenteriae*, *S. flexneri*, *S. boydii*, *S. sonnei*, *S. enterica*, *S. bongori*), *Citrobacter* (including *C. rodentium*, *C. koseri*, *C. farmeri*, *C. youngae*), *Kluyvera* (including *K. ascorbata*), *Trabulsiella* (including *T. guamensis*), *Klebsiella*, and the like.

[0058] N-methyl aminoacyl-tRNA synthetases from *E. coli* were used in the Examples herein as one example, and therefore the modified positions indicated herein are positions in *E. coli*. When N-methyl aminoacyl-tRNA synthetases from other organisms are used, positions to be modified are amino acids at positions corresponding to those in the amino acid sequence of N-methyl aminoacyl-tRNA synthetases from *E. coli* based on sequence homology. ARSs are generally very highly conserved because they are enzymes playing an essential part in the translation mechanism that exists in all organisms. Accordingly, a desired ARS can be modified based on the method described herein to obtain a modified ARS with an increased ability to incorporate an N-methyl amino acid.

[0059] Amino acids to be newly introduced into N-methyl aminoacyl-tRNA synthetases are selected based on hydrophilicity, hydrogen bonds, and side chain size of amino acids in consideration of the distance to and interaction with N-methyl group. For example, when avoiding the steric repulsion between the amino acid at the position and an N-methyl group, the distance to the N-methyl amino group can be adjusted by, for example, substituting a high molecular weight amino acid with a low molecular weight amino acid. Specifically, the distance can be adjusted by reducing the molecular weight, for example, by substituting threonine (Thr) with serine (Ser).

[0060] For example, when an amino acid to be modified in an N-methyl aminoacyl-tRNA synthetase is asparagine, the low molecular weight amino acids can include, but are not limited to, for example, serine, valine, glycine, aspartic acid, and alanine. When an amino acid to be modified is glutamic acid, the low molecular weight amino acids can include, but are not limited to, for example, alanine, valine, serine, alanine, and aspartic acid. Moreover, for example, when an amino acid to be modified is Thr, the low molecular weight amino acids include preferably, for example, Ser. When an amino acid to be modified is an amino acid other than the amino acids as described above, the low molecular weight amino acids include preferably, for example, glycine (Gly) and alanine (Ala), and more preferably glycine (Gly). To give one specific example, Thr can be substituted with Ser; Gln, Glu, and Asn can be substituted with Gly or Ala; and Met, His, Gln, and the like can be substituted with Gly, but the substitutions are not limited thereto.

[0061] To give a more specific example, modification of valine aminoacyl-tRNA synthetase (ValRS) includes, for example, modification of amino acid(s) at position 43 and/or position 45 and/or position 279 of SEQ ID NO: 24 (natural ValRS), or amino acids at positions corresponding to these positions. Amino acids selected for substitution are not limited, but as mentioned above, for example, Thr can be substituted with Ser, Ala, or Gly, and amino acids (e.g., Asn) other than Thr can be substituted with Gly or Ala. For example, preferred are substitution of the amino acid at position 43 of SEQ ID NO: 24 or an amino acid at a position corresponding to position 43 of SEQ ID NO: 24 with Gly or Ala, and/or substitution of the amino acid at position 45 of SEQ ID NO: 24 or an amino acid at the position corresponding to position 45 of SEQ ID NO: 24 with Ser, and/or substitution of the amino acid at position 279 or an amino acid at a position corresponding to position 279 with Gly or Ala. These substitutions may be any one of the substitutions, a combination of any of these substitutions (e.g., substitutions at position 43 and position 45, substitutions at position 43 and position 279, or substitutions at position 45 and position 279), or all of the substitutions. Other substitutions may be further combined. As a more specific illustration, N43 and/or T45 and/or T279 of SEQ ID NO: 24, or amino acids at the positions corresponding to these, are preferably substituted, and preferably substituted to N43G and/or T45S and/or T279A.

[0062] Modification of serine aminoacyl-tRNA synthetase (SerRS) includes, for example, modification of amino acid(s) at position 237 and/or position 239 of SEQ ID NO: 26 (natural SerRS), or amino acids at the positions corresponding to these positions. Amino acids selected for substitution are not limited, but as mentioned above, for example, Thr can be substituted with Ser, and amino acids (e.g., Glu) other than Thr can be substituted with Gly or Ala. For example, preferred are substitution of the amino acid at position 237 of SEQ ID NO: 26 or an amino acid at the position corresponding to position 237 of SEQ ID NO: 26 with Ser, and/or substitution of the amino acid at position 239 of SEQ ID NO: 26 or an amino acid at the position corresponding to position 239 of SEQ ID NO: 26 with Gly or Ala. These substitutions may be any one or both of the substitutions. Other substitutions may be further combined. As a more specific illustration, T237 and/or E239 of SEQ ID NO: 26, or amino acids at positions corresponding to these, are preferably substituted, and preferably substituted to T237S and/or E239G (or E239A).

[0063] Modification of phenylalanine aminoacyl-tRNA synthetase α subunit (PheRS α) includes, for example, modification of amino acid at position 169 of SEQ ID NO: 28 (natural PheRS α subunit), or an amino acid at a position corresponding to position 169 of SEQ ID NO: 28. Amino acids selected for substitution are not limited, but as mentioned above, for example, Thr can be substituted with Ser, and amino acids other than Thr can be substituted with glycine (Gly) or alanine (Ala) (more preferably Gly). For example, preferred is substitution of the amino acid at position 169 of SEQ ID NO: 28 or an amino acid at the position corresponding to position 169 of SEQ ID NO: 28 with Gly. Other substitutions may be further combined. As a more specific illustration, Q169 of SEQ ID NO: 28, or an amino acid at the position corresponding to Q169 of SEQ ID NO: 28, is preferably substituted, and preferably substituted to Q169G (or Q169A).

[0064] Modification of threonine aminoacyl-tRNA synthetase (ThrRS) includes, for example, modification of amino acid(s) at position 332 and/or position 511 of SEQ ID NO: 29 (natural ThrRS), or amino acids at positions corresponding to these positions. Amino acids selected for substitution are not limited, but as mentioned above, for example, Thr can be substituted with Ser, and amino acids (*e.g.*, Met and His) other than Thr can be substituted with Gly. For example, preferred is/are substitution of the amino acid at position 332 of SEQ ID NO: 29 or an amino acid at a position corresponding to position 332 of SEQ ID NO: 29 with Gly and/or substitution of the amino acid at position 511 of SEQ ID NO: 29 or an amino acid at a position corresponding to position 511 of SEQ ID NO: 29 with Gly. These substitutions may be any one or both of the substitutions. Other substitutions may be further combined. As a more specific illustration, M332 and/or H511 of SEQ ID NO: 29, or amino acids at positions corresponding to these, are preferably substituted, and preferably substituted to M332G and/or H511G.

[0065] Modification of tryptophan aminoacyl-tRNA synthetase (TrpRS) includes, for example, modification of amino acid(s) at position 132 and/or position 150 and/or position 153 of SEQ ID NO: 188 (natural TrpRS), or amino acids at a position corresponding to these positions. Amino acids selected for substitution are not limited, but as mentioned above, for example, Met can be substituted with Val or Ala, and amino acids (*e.g.*, Gln) other than Met can be substituted with Ala. For example, preferred is/are substitution of the amino acid at position 132 of SEQ ID NO: 188 or an amino acid at a position corresponding to position 132 of SEQ ID NO: 188 with Val or Ala, and/or substitution of the amino acid at position 150 of SEQ ID NO: 188 or an amino acid at a position corresponding to position 150 of SEQ ID NO: 188 with Ala, and/or substitution of the amino acid at position 153 or an amino acid at a position corresponding to position 153 with Ala. These substitutions may be any one of the substitutions, a combination of any of these substitutions (*e.g.*, substitutions at position 132 and position 150, substitutions at position 132 and position 153, or substitutions at position 150 and position 153), or all of the substitutions. Other substitutions may be further combined. As a more specific illustration, M132 and/or Q150 and/or H153 of SEQ ID NO: 188, or amino acids at positions corresponding to these, are preferably substituted, and preferably substituted to M132V and/or Q150A and/or H153A.

[0066] Modification of leucine aminoacyl-tRNA synthetase (LeuRS) includes, for example, modification of the amino acid at position 43 of SEQ ID NO: 189 (natural LeuRS), or an amino acid at a position corresponding to position 43 of SEQ ID NO: 189. Amino acids selected for substitution are not limited, but as mentioned above, for example, Thr can be substituted with Gly. For example, preferred is substitution of the amino acid at position 43 of SEQ ID NO: 189 or an amino acid at a position corresponding to position 43 of SEQ ID NO: 189 with Gly. Other substitutions may be further combined. As a more specific illustration, Y43 of SEQ ID NO: 189, or an amino acid at a position corresponding to Y43 of SEQ ID NO: 189, is preferably substituted, and preferably substituted to Y43G.

[0067] A method for producing a mutant N-methyl aminoacyl-tRNA synthetase, which is modified by substituting an amino acid at a specific position with another amino acid, in accordance with the present disclosure, can be performed using any known genetic

engineering technique. For example, DNA fragments having base sequences encoding amino acid sequences comprising amino acids at positions of interest are amplified using primers having base sequences substituted with base sequences encoding amino acid sequences comprising modified amino acids, resulting in base sequences encoding amino acid sequences comprising modified amino acids. The amplified DNA fragments are ligated together to obtain a full-length DNA encoding the mutant aminoacyl-tRNA synthetase. This full-length DNA can be expressed using a host cell such as *E. coli* to easily produce the mutant N-methyl aminoacyl-tRNA synthetase. Primers used in the method are 20 to 70 bases in length, and preferably about 20 to 50 bases in length. The primers have 1 to 3 base mismatches with the original unmodified base sequence, and therefore relatively long primers, for example, primers of 20 bases or more in length are preferably used.

[0068] A method for producing a mutant N-methyl aminoacyl-tRNA synthetase, which is modified by substituting an amino acid at a specific position with another amino acid, according to the present disclosure, is not limited to the method as described above, and various genetic engineering techniques, such as known point mutation techniques and gene synthesis techniques, and methods for introducing modified fragments using restriction enzymes, can be utilized. Expression hosts are not limited to *E. coli*, and animal cells and cell-free translation systems may also be used.

[0069] Modified ARSs as described herein include a polypeptide comprising an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187 (PheRS05, PheRS04, ValRS04, ValRS13, ValRS13-11, SerRS03, SerRS05, SerRS35, SerRS37, ThrRS03, ThrRS14, ValRS66, ValRS67, TrpRS04, TrpRS05, TrpRS18, and LeuRS02) and polypeptides functionally equivalent to the polypeptide. "Functionally equivalent polypeptides" are ARSs with a high structural identity to a polypeptide comprising an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187 and have reactivity with N-methyl amino acids. More specifically, "functionally equivalent polypeptides" are polypeptides that have amino acids modified according to the above description in ARSs with a high structural identity to a polypeptide comprising an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187, thereby having increased reactivity with N-methyl amino acids compared to unmodified ARSs. Increased reactivity with N-methyl amino acids may be, for example, increased substrate specificity to N-methyl amino acids (e.g., increased value of reactivity with N-methyl amino acids / reactivity with unmodified amino acids).

[0070] Such polypeptides include, for example, a polypeptide in which one or more amino acids (preferably 1 to 20 amino acids, for example, 1 to 10 amino acids, 1 to 7 amino acids, 1 to 5 amino acids, 1 to 3 amino acids, 1 to 2 amino acids, or 1 amino acid) are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187. Such polypeptides may also be, for example, a polypeptide in which one to several amino acids are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187.

[0071] A polypeptide functionally equivalent to a modified ARS comprising an amino acid

sequence set forth in any of SEQ ID NOs: 1-11 and 182-187 typically has a high identity to an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187. A polynucleotide encoding a functionally equivalent polypeptide also typically has a high identity to a base sequence (e.g., SEQ ID NOs: 12-22 and 190-195) encoding an amino acid sequence set forth in any of SEQ ID NOs: 1-11 and 182-187. High identity (sequence identity) specifically refers to 70% or more, preferably 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, or 99% or more identity.

[0072] The identity of amino acid sequences or base sequences can be determined using the BLAST algorithm by Karlin and Altschul (Proc. Natl. Acad. Sci. USA (1993) 90: 5873-7). Programs called BLASTN and BLASTX have been developed based on the algorithm (Altschul et al., J. Mol. Biol. (1990)215: 403-10). When a base sequence is analyzed using BLAST-based BLASTN, parameters are set as, for example, score = 100 and wordlength = 12. When an amino acid sequence is analyzed using BLAST-based BLASTX, parameters are set as, for example, score = 50 and wordlength = 3. When BLAST and Gapped BLAST programs are used, default parameters of each program are used. Specific procedures of these analysis methods are known (see, information at the website of Basic Local Alignment Search Tool (BLAST) in National Center for Biotechnology Information (NCBI)).

[0073] The described polypeptides include any polypeptide of the following (a) to (d), wherein the amino acid corresponding to position 43 of SEQ ID NO: 24 (ValRS) is other than Asn, and/or the amino acid corresponding to position 45 of SEQ ID NO: 24 (ValRS) is other than Thr and/or the amino acid corresponding to position 279 of SEQ ID NO: 24 (ValRS) is other than Thr (*i.e.*, a polypeptide in which the amino acid at at least one of these 3 positions is other than the indicated respective amino acid) and wherein the polypeptide has increased reactivity to N-methyl Val compared to a polypeptide in which the amino acids corresponding to position 43, position 45, and position 279 are Asn, Thr, and Thr, respectively:

1. (a) a polypeptide comprising an amino acid sequence having a high identity to an amino acid sequence set forth in any of SEQ ID NOs: 3-5, 182, and 183 (ValRS04, ValRS13, ValRS13-11, ValRS66, and ValRS67);
2. (b) a polypeptide comprising an amino acid sequence in which one or more amino acids are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any one of SEQ ID NOs: 3-5, 182, and 183;
3. (c) a polypeptide encoded by a base sequence with high identity to a base sequence set forth in any one of SEQ ID NOs: 14-16, 190, and 191 (DNA encoding ValRS04, ValRS13, ValRS13-11, ValRS66, and ValRS67); or
4. (d) a polypeptide encoded by a DNA fragment hybridizing with a strand complementary to a base sequence set forth in any one of SEQ ID NOs: 14-16, 190, and 191 under stringent conditions.

[0074] The above-mentioned polypeptides preferably have (i) Gly or Ala at the amino acid position corresponding to position 43 and/or (ii) Ser at the amino acid position corresponding to

position 45 and/or (iii) Gly or Ala at the amino acid position corresponding to position 279. Such polypeptides include natural polypeptides and artificially-modified polypeptides, and preferably include polypeptides in which (i) the amino acid at the position corresponding to position 43 is substituted with Gly or Ala and/or (ii) the amino acid at the position corresponding to position 45 is substituted with Ser and/or (iii) the amino acid at the position corresponding to position 279 is substituted with Gly or Ala.

[0075] The described polypeptides include any polypeptide of the following (a) to (d), wherein the amino acid corresponding to position 237 of SEQ ID NO: 26 (SerRS) is other than Thr and/or the amino acid corresponding to position 239 of SEQ ID NO: 26 (SerRS) is other than Glu (*i.e.*, a polypeptide in which the amino acid at at least one of these 2 positions is other than the indicated respective amino acid) and wherein the polypeptide has increased reactivity with N-methyl Ser compared to a polypeptide in which the amino acids corresponding to position 237 and position 239 are Thr and Glu respectively:

1. (a) a polypeptide comprising an amino acid sequence having a high identity to an amino acid sequence set forth in any one of SEQ ID NOs: 6-9 (SerRS03, SerRS05, SerRS35, and SerRS37);
2. (b) a polypeptide comprising an amino acid sequence in which one or more amino acids are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any one of SEQ ID NOs: 6-9;
3. (c) a polypeptide encoded by a base sequence having a high identity to a base sequence set forth in any one of SEQ ID NOs: 17-20 (DNA encoding SerRS03, SerRS05, SerRS35, and SerRS37); or
4. (d) a polypeptide encoded by a DNA fragment hybridizing with a strand complementary to a base sequence set forth in any one of SEQ ID NOs: 17-20 under stringent conditions.

[0076] The above-mentioned polypeptides preferably have (i) Ser at the amino acid position corresponding to position 237 and/or (ii) Gly or Ala at the amino acid position corresponding to position 239. Such polypeptides include natural polypeptides and artificially-modified polypeptides, and preferably include polypeptides in which (i) the amino acid at the position corresponding to position 237 is substituted with Ser and/or (ii) the amino acid at the position corresponding to position 239 is substituted with Gly or Ala.

[0077] Such polypeptides include any polypeptide of the following (a) to (d), wherein the polypeptide has any amino acid other than Gln at the position corresponding to position 169 of SEQ ID NO: 28 (PheRS) and has increased reactivity with N-methyl Phe compared to a polypeptide in which the amino acid at position 169 of SEQ ID NO: 28 (PheRS) is Gln:

1. (a) a polypeptide comprising an amino acid sequence having a high identity to an amino acid sequence set forth in any one of SEQ ID NOs: 1-2 (PheRS05 and PheRS04);
2. (b) a polypeptide comprising an amino acid sequence in which one or more amino acids

are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any one of SEQ ID NOs: 1-2;

3. (c) a polypeptide encoded by a base sequence having a high identity to a base sequence set forth in any one of SEQ ID NOs: 12-13 (DNA encoding PheRS05 and PheRS04); or
4. (d) a polypeptide encoded by a DNA fragment hybridizing with a strand complementary to a base sequence set forth in any one of SEQ ID NOs: 12-13 under stringent conditions.

[0078] The above-mentioned polypeptides preferably have Gly or Ala at the amino acid position corresponding to position 169. Such polypeptides include natural polypeptides and artificially-modified polypeptides, and preferably include polypeptides in which the amino acid at the position corresponding to position 169 is substituted with Gly or Ala.

[0079] The described polypeptides include any polypeptide of the following (a) to (d), wherein the amino acid corresponding to position 332 of SEQ ID NO: 29 (ThrRS) is other than Met and/or the amino acid corresponding to position 511 of SEQ ID NO: 29 (ThrRS) is other than His (*i.e.*, a polypeptide in which the amino acid at at least one of these 2 positions is other than the indicated respective amino acid) and wherein the polypeptide has increased reactivity with N-methyl Thr compared to a polypeptide in which the amino acids corresponding to position 332 and position 511 are Met and His, respectively:

1. (a) a polypeptide comprising an amino acid sequence having a high identity to an amino acid sequence set forth in any one of SEQ ID NOs: 10-11 (ThrRS03 and ThrRS 14);
2. (b) a polypeptide comprising an amino acid sequence in which one or more amino acids are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any one of SEQ ID NOs: 10-11;
3. (c) a polypeptide encoded by a base sequence having a high identity to a base sequence set forth in any one of SEQ ID NOs: 21-22 (DNA encoding ThrRS03 and ThrRS 14); or
4. (d) a polypeptide encoded by a DNA fragment hybridizing with a strand complementary to a nucleotide sequence set forth in any one of SEQ ID NOs: 21-22 under stringent conditions.

[0080] The above-mentioned polypeptides preferably have (i) Gly at the amino acid position corresponding to 332 and/or (ii) Gly at the amino acid position corresponding to position 511. Such polypeptides include natural polypeptides and artificially-modified polypeptides, and preferably include polypeptides in which the amino acid at the position corresponding to position 332 is substituted with Gly and/or the amino acid at the position corresponding to position 511 is substituted with Gly.

[0081] The described polypeptides include a polypeptide that is any polypeptide of the following (a) to (d), wherein the amino acid corresponding to position 132 of SEQ ID NO: 188 (TrpRS) is other than Met and/or the amino acid corresponding to position 150 of SEQ ID NO: 188 (TrpRS) is other than Gln and/or the amino acid corresponding to position 153 of SEQ ID NO: 188 (TrpRS) is other than His (i.e., a polypeptide in which the amino acid at at least one of these 3 positions is other than the indicated respective amino acid) and wherein the polypeptide has increased reactivity with N-methyl Trp compared to a polypeptide in which the amino acids corresponding to position 132, position 150, and position 153 are Met, Gln, and His, respectively:

1. (a) a polypeptide comprising an amino acid sequence having a high identity to an amino acid sequence set forth in any one of SEQ ID NOs: 184-186 (TrpRS04, TrpRS05, and TrpRS 18);
2. (b) a polypeptide comprising an amino acid sequence in which one or more amino acids are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any one of SEQ ID NOs: 184-186;
3. (c) a polypeptide encoded by a base sequence having a high identity to a base sequence set forth in any one of SEQ ID NOs: 192-194 (DNA encoding TrpRS04, TrpRS05, and TrpRS18); or
4. (d) a polypeptide encoded by a DNA fragment hybridizing with a strand complementary to a base sequence set forth in any one of SEQ ID NOs: 192-194 under stringent conditions.

[0082] The above-mentioned polypeptides preferably have (i) Val or Ala at the amino acid position corresponding to position 132 and/or (ii) Ala at the amino acid position corresponding to position 150 and/or (iii) Ala at the amino acid position corresponding to position 153. Such polypeptides include natural polypeptides and artificially-modified polypeptides, and preferably include polypeptides in which (i) the amino acid at the position corresponding to position 132 is substituted with Val or Ala and/or (ii) the amino acid at the position corresponding to position 150 is substituted with Ala and/or (iii) the amino acid at the position corresponding to position 153 is substituted with Ala.

[0083] The described polypeptides include any polypeptide of the following (a) to (d), wherein the polypeptide has any amino acid other than Tyr at the position corresponding to position 43 of SEQ ID NO: 189 (LeuRS) and has increased reactivity with N-methyl Leu compared to a polypeptide in which the amino acid corresponding to position 43 is Tyr:

1. (a) a polypeptide comprising an amino acid sequence having a high identity to the amino acid sequence set forth in SEQ ID NO: 187 (LeuRS02);
2. (b) a polypeptide comprising an amino acid sequence in which one or more amino acids are substituted, deleted, inserted, and/or added in an amino acid sequence set forth in any one of SEQ ID NO: 187;
3. (c) a polypeptide encoded by a base sequence having a high identity to the base

- sequence set forth in SEQ ID NO: 195 (DNA encoding LeuRS02); or
4. (d) a polypeptide encoded by a DNA fragment hybridizing with a strand complementary to the base sequence set forth in SEQ ID NO: 195 under stringent conditions.

[0084] The above-mentioned polypeptides preferably have Gly at the amino acid position corresponding to position 43. Such polypeptides include natural polypeptides and artificially-modified polypeptides, and preferably include polypeptides in which the amino acid at the position corresponding to position 43 is substituted with Gly.

[0085] In the case of polypeptides obtained by modifying natural ARSs, polypeptides according to the present disclosure are those with increased reactivity to N-methyl amino acids compared to unmodified ARSs. When the polypeptides are natural ARSs or artificially produced polypeptides, the polypeptides in accordance with the present disclosure are those with reactivity to N-methyl amino acids, namely those having an activity to acylate tRNAs with N-methyl amino acids. When comparing the reactivities of modified ARSs and natural ARSs, and the ARSs are made up of multiple subunits, subunit(s) other than that/those compared is/are the same is/are used. These subunits may be natural (or wild-type) or a modified as long as they are the same in both ARSs.

[0086] High identity (or high sequence identity) refers to, as mentioned above, for example, 70% or more, preferably 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, or 99% or more identity. The number of amino acids substituted, deleted, inserted, and/or added may be one or several, for example, 1 to 20, preferably 1 to 15, more preferably 1 to 10, more preferably 1 to 8, more preferably 1 to 7, more preferably 1 to 6, more preferably 1 to 5, more preferably 1 to 4, more preferably 1 to 3, more preferably 1 to 2, and more preferably 1. Stringent hybridization conditions refer to, for example, conditions of about "1 X SSC, 0.1% SDS, at 37°C", more strictly conditions of about "0.5 X SSC, 0.1% SDS, at 42°C", further more strictly conditions of about "0.2 X SSC, 0.1% SDS, at 65°C", and yet more strictly conditions of about "0.1 X SSC, 0.1% SDS, at 65°C". It is noted that the conditions of SSC, SDS, and temperature as described above are only examples of combinations. Those skilled in the art can achieve stringency similar to those described above by appropriately combining the above-mentioned or other factors that determine hybridization stringency (such as probe concentration, probe length, and duration of hybridization).

[0087] Also disclosed polynucleotides encoding the polypeptides as described herein. The polynucleotides comprise any polynucleotide as long as it comprises a sequence encoding a polypeptide as described above. The polynucleotides as described herein also comprise any one of genomic DNA, cDNA, and DNAs artificially produced based on the genomic DNA or cDNA. Genomic DNA comprises exons and introns. That is to say, genomic DNA may or may not comprise introns, and may or may not comprise untranslated regions (5' UTR and/or 3' UTR), transcription control elements, and the like. cDNA is a nucleic acid sequence derived from a portion of intronic sequence and may comprise a nucleic acid sequence encoding an

amino acid sequence.

[0088] The polynucleotides also comprise degenerate polynucleotides comprising any codon that encodes the same amino acid. The described polynucleotides may also be a polynucleotide derived from desired organisms.

[0089] The described polynucleotides may be obtained using any method. For example, a complementary DNA (cDNA) prepared from mRNA, a DNA prepared from genomic DNA, a DNA obtained by chemical synthesis, a DNA obtained by amplifying RNA or DNA as a template using PCR, and a DNA constructed by appropriately combining these techniques are all included. The described polynucleotides can be produced by cloning the genomic DNA or RNA encoding the polypeptide as described herein according to conventional methods and introducing mutations into the cloned genomic DNA or RNA.

[0090] For example, in a method for cloning cDNA from mRNA encoding a polypeptide as described herein, first, the mRNA encoding the polypeptide as described herein is prepared according to conventional methods from any tissue or cell where the polypeptide as described herein is expressed and produced. For example, total RNA prepared using a method such as the guanidine thiocyanate method, hot-phenol method, or AGPC method can be subjected to affinity chromatography on oligo (dT) cellulose, poly U-Sepharose, or the like.

[0091] The obtained mRNA is then used as a template to synthesize a cDNA strand by using a known method (Mol. Cell. Biol., Vol. 2, p. 161, 1982; Mol. Cell. Biol., Vol. 3, p. 280, 1983; Gene, Vol. 25, p. 263, 1983), for example, by using reverse transcriptase. The cDNA strand is converted into a double-stranded cDNA and incorporated into a plasmid vector, phage vector, cosmid vector, or the like. The vector is used to transform *E. coli* or to perform *in vitro* packaging followed by transfection of *E. coli* to generate a cDNA library.

[0092] The cDNA library can be screened using a polynucleotide as described herein (*e.g.*, SEQ ID NOs: 12-22, 190-195) or a portion thereof as a probe to obtain a gene of interest. Alternatively, the cDNA library can be directly amplified by PCR using a polynucleotide as described herein (*e.g.*, SEQ ID NOs: 12-22, 190-195) or a portion thereof as a primer. The sites and lengths of probes and primers may be appropriately determined.

[0093] Further disclosed are vectors (recombinant vectors) comprising polynucleotides encoding the polypeptides as described above. The vectors are not particularly limited as long as they can replicate and can be maintained or can self-proliferate in any host prokaryotic and/or eukaryotic cell. The vectors include plasmid vectors, phage vectors, viral vectors, and the like.

[0094] Examples of vectors for cloning include, for example, pUC19, λ gt10, λ gt11, and the like. Furthermore, the vectors preferably have a promoter that can express the herein described polypeptide when cells that can express the polynucleotide in host cells are isolated.

[0095] The recombinant vectors as described herein can be prepared by simply ligating a polynucleotide encoding a herein disclosed polypeptide into a vector for recombination (a plasmid DNA and bacteriophage DNA) available in the art according to conventional methods.

[0096] Examples of recombinant vectors that can be used include, for example, plasmids from *E. coli* (such as pBR322, pBR325, pUC12, pUC13, and pUC19), plasmids from yeast (such as pSH19 and pSH15), and plasmids from *Bacillus subtilis* (such as pUB110, pTP5, and pC194).

[0097] Examples of phages include bacteriophage such as λ phage, and further animal and insect viruses such as retrovirus, vaccinia virus, nucleopolyhedrovirus, and lentivirus (pVL1393, from Invitrogen).

[0098] Expression vectors are useful for expressing polynucleotides encoding the herein described polypeptides and production of the herein described polypeptides. The expression vectors are not particularly limited as long as they have functions to express polynucleotides encoding the polypeptides as described herein and produce the polypeptides in any host prokaryotic and/or eukaryotic cell.

[0099] For example, expression vectors include pMAL C2, pEF-BOS (Nucleic Acid Research, Vol. 18, 1990, p. 5322, and the like), or pME18S ("Idenshi Kougaku Handbook (Genetic Engineering Handbook)", supplementary volume of Jikken Igaku (Experimental Medicine), 1992, and the like).

[0100] Also disclosed are fusion between polypeptides as described herein with another protein/other proteins. A fusion polypeptide according to the present disclosure is a fusion polypeptide between a polypeptide having reactivity with N-methyl amino acids as described herein and another polypeptide. The fusion polypeptide itself may have no reactivity with N-methyl amino acids as long as it comprises a polypeptide chain having reactivity with N-methyl amino acids as described herein. When prepared as a fusion protein with, for example, Glutathione S-transferase (GST), the fusion polypeptide as described herein can be prepared by subcloning a cDNA encoding the polypeptide as described herein into, for example, plasmid pGEX4T1 (from Pharmacia), transforming *E. coli* DH5 α or the like with the plasmid, and culturing the transformant.

[0101] Alternatively, a fusion polypeptide according to the present disclosure can be prepared as a fusion with HA (influenza agglutinin), immunoglobulin constant region, β -galactosidase, maltose-binding protein (MBP), or the like. Moreover, a fusion polypeptide can be prepared as a fusion with, for example, any known peptide such as FLAG (Hopp, T. P. et al., *BioTechnology* (1988) 6, 1204-1210), a tag consisting of several (e.g., six) histidine (His) residues (such as 6 x His, 10 x His), influenza agglutinin (HA), a human c-myc fragment, a VSV-GP fragment, a p18HIV fragment, T7-tag, HSV-tag, E-tag, a SV40T antigen fragment, Ick tag, an α -tubulin fragment, B-tag, a Protein C fragment, Stag, StrepTag, and HaloTag.

[0102] The vectors as described herein preferably comprise at least a promoter-operator

region, the initiation codon, a polynucleotide encoding the herein described polypeptide, a termination codon, a terminator region, and a replicable unit when bacteria, particularly *E. coli*, are used as host cells.

[0103] The expression vectors preferably comprise at least a promoter, the initiation codon, a polynucleotide encoding a polypeptide as described herein, and a termination codon when yeast, animal cells, or insect cells are used as hosts.

[0104] The vectors may also comprise DNA encoding a signal peptide, an enhancer sequence, 5' and 3' untranslated regions of the gene encoding a polypeptide as described herein, a splicing junction, a polyadenylation site, a selection marker region, a replicable unit, or the like.

[0105] The vectors may also comprise, if desired, a marker gene (such as a gene amplification gene, a drug-resistant gene) that allows selection of hosts in which gene amplification and transformation have been achieved.

[0106] Examples of marker genes include, for example, dihydrofolate reductase (DHFR) gene, thymidine kinase gene, neomycin resistant gene, glutamic acid synthetase gene, adenosine deaminase gene, ornithine decarboxylase gene, hygromycin B phosphotransferase gene, aspartate transcarbamylase gene, and the like.

[0107] The promoter-operator region for expressing a polypeptide in bacteria can include, a promoter, an operator, and Shine-Dalgarno (SD) sequence (such as AAGG).

[0108] An example of a promoter-operator region includes one comprising, for example, Trp promoter, lac promoter, recA promoter, λ PL promoter, lpp promoter, tac promoter, or the like when the host is a bacterium of the genus *Escherichia*, for example.

[0109] Promoters for expressing a polypeptide in yeast include PH05 promoter, PGK promoter, GAP promoter, and ADH promoter.

[0110] The promoters include SL01 promoter, SP02 promoter, penP promoter, and the like when the host is a bacterium of the genus *Bacillus*.

[0111] The promoters also include promoters derived from SV40, retroviral promoters, heat shock promoters, and the like when the hosts are eukaryotic cells such as mammalian cells. The promoters are preferably those derived from SV40 and retroviral promoters. However, the promoters are not particularly limited to those described above. Also, enhancers can be effectively used for expression.

[0112] An example of a suitable initiation codon includes the methionine codon (ATG). Examples of termination codons include common termination codons (e.g., TAG, TGA, and TAA). Terminator regions that can be used include natural or synthetic terminators generally used.

[0113] The replicable unit refers to DNA having an ability to replicate the full DNA sequence of the replicable unit in host cells, and includes natural plasmids, artificially modified plasmids (DNA fragments prepared from natural plasmids), synthetic plasmids, and the like. Suitable plasmids include plasmid pBR322 or an artificially modified pBR322 (a DNA fragment obtained by digesting pBR322 with a suitable restriction enzyme) for *E. coli*; yeast 2 μ plasmid or yeast chromosomal DNA for yeast; and plasmid pRSVneo (ATCC 37198), plasmid pSV2dhfr (ATCC 37145), plasmid pdBPV-MMTneo (ATCC 37224), plasmid pSV2neo (ATCC 37149), and the like for mammalian cells.

[0114] Enhancer sequence, polyadenylation site, and splicing junction that can be used are those commonly used by those skilled in the art, such as for example those derived from SV40.

[0115] The expression vectors can be prepared by ligating at least a promoter as described above, an initiation codon, a polynucleotide encoding a polypeptide, a termination codon, and a terminator region to a suitable replicable unit continuously and circularly. Suitable DNA fragments (e.g., a linker, other restriction enzyme cleavage sites, and the like) can also be used in conventional techniques such as digestion with restriction enzymes and ligation with T4 DNA ligase, if desired.

[0116] Also disclosed are recombinant cells transformed with the vectors as described above, and the recombinant cells can be prepared by introducing the expression vectors as described above into host cells.

[0117] Host cells used in accordance with the disclosure are not particularly limited as long as they are compatible with the expression vectors described above and can be transformed. Examples of host cells include various cells including natural cells and artificially-established recombinant cells commonly used in the technical field, for example, bacteria (*Escherichia* bacteria, *Bacillus* bacteria), yeast (such as *Saccharomyces*, *Pichia*), animal cells, insect cells, and the like.

[0118] The host cells are preferably *E. coli* or animal cells, and examples include, for example, *E. coli* (such as DH5 α , TB1, HB101), mouse-derived cells (such as COP, L, C127, Sp2/0, NS-1, or NIH3T3), rat-derived cells (PC12, PC12h), hamster-derived cells (such as BHK and CHO), monkey-derived cells (such as COS1, COS3, COS7, CV1, and Velo), human-derived cells (such as Hela, diploid fibroblast-derived cells, myeloma cells, and HepG2), and the like.

[0119] Introduction of expression vectors into host cells (transformation (transfection)) can be performed according to conventional methods ([for *E. coli*, *Bacillus subtilis*, and the like]: Proc. Natl. Acad. Sci. USA., Vol. 69, p. 2110, 1972; Mol. Gen. Genet., Vol. 168, p. 111, 1979; J. Mol. Biol., Vol. 56, p. 209, 1971; [for *Saccharomyces cerevisiae*]: Proc. Natl. Acad. Sci. USA., Vol. 75, p. 1927, 1978; J. Bacteriol., Vol. 153, p. 163, 1983); [for animal cells]: Virology, Vol. 52, p. 456, 1973; [for insect cells]: Mol. Cell. Biol., Vol. 3, p. 2156-2165, 1983).

[0120] The polypeptides can be produced by culturing recombinant transformed cells including expression vectors prepared as described above (hereinafter used to mean inclusion of an inclusion body) in a nutrient medium according to any conventional method.

[0121] The polypeptides can be produced such as by culturing recombinant cells as described above, particularly animal cells and allowing the recombinant cells to secrete the polypeptides into culture supernatant.

[0122] The obtained culture is subjected to filtration, centrifugation, or any other similar technique to obtain a culture filtrate (supernatant). The polypeptides are purified and isolated from the culture filtrate according to any conventional method commonly used to purify and isolate natural or synthetic proteins.

[0123] Isolation and purification methods include, for example, methods utilizing solubility such as salt precipitation and solvent precipitation, dialysis, ultrafiltration, gel filtration, methods utilizing the difference in molecular weight such as sodium dodecyl sulphate-polyacrylamide gel electrophoresis, methods utilizing electric charge such as ion exchange chromatography and hydroxylapatite chromatography, methods utilizing specific affinity such as affinity chromatography, methods utilizing the difference in hydrophobicity such as reversed-phase high-performance liquid chromatography, methods utilizing the difference in isoelectric point such as isoelectric focusing, and the like.

[0124] Meanwhile, when a polypeptide as described herein is present in periplasm or cytoplasm of cultured recombinant cells (e.g., *E. coli*), the culture is subjected to any conventional method such as filtration or centrifugation to collect bacterial pellets or cells. The collected bacterial pellets or cells are suspended in a suitable buffer, and cell wall and/or cytoplasmic membrane is/are disrupted by any method such as for example sonication, lysozyme, and freeze-thawing. Any method such as centrifugation or filtration is then performed to obtain membrane fraction containing the protein as described herein. The membrane fraction is solubilized with any surfactant such as TritonTM-X100 to give a crude solution. Then, the crude solution can be isolated and purified using any conventional method as described previously.

[0125] Also disclosed is a polynucleotide (cDNA or genomic DNA) encoding a polypeptide as described above or an oligonucleotide hybridizing with a strand complementary to the polynucleotide. For example, the oligonucleotide is a base sequence at least comprising modified sites of ARSs as described herein or an oligonucleotide hybridizing with a strand complementary to the base sequence. An oligonucleotide in accordance with the disclosure is also a partial fragment of polynucleotides encoding a modified ARS as described herein, and is preferably a fragment comprising bases in modified sites of the modified ARSs or a strand complementary to the fragment.

[0126] For example, for a polynucleotide encoding a polypeptide as described herein having reactivity with N-methyl Val, the oligonucleotide may be an oligonucleotide comprising bases

encoding the codon(s) for an amino acid corresponding to position 43, and/or an amino acid corresponding to position 45, and/or an amino acid corresponding to position 279 of the amino acid sequences set forth in SEQ ID NOs: 3-5, 182, and 183 (ValRS04, ValRS13, ValRS13-11, ValRS66, and ValRS67), or an oligonucleotide consisting of a sequence complementary to the oligonucleotide. In this case, an amino acid corresponding to position 43 is preferably Gly or Ala (more preferably Gly), and an amino acid corresponding to position 45 is preferably Ser, and an amino acid corresponding to position 279 is preferably Gly or Ala (more preferably Ala).

[0127] For example, for a polynucleotide encoding a polypeptide as described herein having reactivity with N-methyl Ser, the oligonucleotide may be an oligonucleotide comprising bases encoding the codon(s) for an amino acid corresponding to position 237 and/or an amino acid corresponding to position 239 of the amino acid sequences set forth in SEQ ID NOs: 6-9 (SerRS03, SerRS05, SerRS35, and SerRS37), or an oligonucleotide consisting of a sequence complementary to the oligonucleotide. In this case, an amino acid corresponding to position 237 is preferably Ser, and an amino acid corresponding to position 239 is preferably Gly or Ala (more preferably Gly).

[0128] For example, for a polynucleotide encoding a polypeptide as described herein having reactivity with N-methyl Phe, the oligonucleotide may be an oligonucleotide comprising bases encoding the codon for an amino acid corresponding to position 169 of the amino acid sequences set forth in SEQ ID NOs: 1-2 (PheRS04 and PheRS05) or an oligonucleotide consisting of a sequence complementary to the oligonucleotide. In this case, an amino acid corresponding to position 169 is preferably Gly or Ala.

[0129] For example, for a polynucleotide encoding a polypeptide as described herein having reactivity with N-methyl Thr, the oligonucleotide may be an oligonucleotide comprising bases encoding the codon(s) for an amino acid corresponding to position 332 and/or an amino acid corresponding to position 511 of the amino acid sequences set forth in SEQ ID NOs: 10-11 (ThrRS03 and ThrRS14), or an oligonucleotide consisting of a sequence complementary to the oligonucleotide. In this case, an amino acid corresponding to position 332 is preferably Gly, and an amino acid corresponding to position 511 is preferably Gly.

[0130] For example, for a polynucleotide encoding a polypeptide as described herein having reactivity with N-methyl Trp, the oligonucleotide may be an oligonucleotide comprising bases encoding the codon(s) for an amino acid corresponding to position 132, and/or an amino acid corresponding to position 150, and/or an amino acid corresponding to position 153 of the amino acid sequences set forth in SEQ ID NOs: 184-186 (TrpRS04, TrpRS05, and TrpRS18), or an oligonucleotide consisting of a sequence complementary to the oligonucleotide. In this case, an amino acid corresponding to position 132 is preferably Val or Ala (more preferably Val), an amino acid corresponding to position 150 is preferably Ala, and an amino acid corresponding to position 279 is preferably Ala.

[0131] For example, for a polynucleotide encoding a polypeptide as described herein having reactivity with N-methyl Leu, the oligonucleotide be an oligonucleotide comprising bases

encoding the codon for an amino acid corresponding to position 43 of the amino acid sequence set forth in SEQ ID NO: 187 (LeuRS02) or an oligonucleotide consisting of a sequence complementary to the oligonucleotide. In this case, an amino acid corresponding to position 43 is preferably Gly.

[0132] Length of a partial fragment of polynucleotides encoding the polypeptides as described herein having reactivity with N-methyl amino acids is not particularly limited, but is for example, at least 15 consecutive bases, preferably 16 or more consecutive bases, more preferably 17 or more consecutive bases, more preferably 18 or more consecutive bases, more preferably 20 or more consecutive bases, more preferably 25 or more consecutive bases, more preferably 28 or more consecutive bases, more preferably 30 or more consecutive bases, more preferably 32 or more consecutive bases, more preferably 35 or more consecutive bases, more preferably 40 or more consecutive bases, and more preferably 50 or more consecutive bases.

[0133] In addition to the partial fragment of polynucleotides encoding ARSs as described above, oligonucleotides may also further comprise (an) oligonucleotide(s) consisting of other sequences at its/their both or either end(s) (5' and/or 3' end). An oligonucleotide in accordance with the disclosure is, for example, 500 bases or less in length, more preferably 300 bases or less in length, more preferably 200 bases or less in length, more preferably 100 bases or less in length, more preferably 70 bases or less in length, more preferably 60 bases or less in length, and more preferably 50 bases or less in length.

[0134] The disclosed oligonucleotides are useful for producing nucleic acids encoding the polypeptides as described herein (e.g., useful for introducing mutations), and also useful for detecting nucleic acids encoding the polypeptides as described herein. For example, an oligonucleotide as described herein can also be used as a probe in DNA hybridization or RNA hybridization operations. An example of a DNA for the purpose of using as a probe include a partial base sequence of 20 or more consecutive bases hybridizing with a herein described polynucleotide, preferably, a partial base sequence of 30 or more consecutive bases, more preferably a partial base sequence of 40 or 50 or more consecutive bases, more preferably a partial base sequence of 100 or more consecutive bases, more preferably a partial base sequence of 200 or more consecutive bases, and particularly preferably a partial base sequence of 300 or more consecutive bases.

[0135] The herein described polypeptides, polynucleotides, and oligonucleotides can be included in compositions in combination with carriers or vehicles, appropriately. The compositions can be produced using any method known to those skilled in the art. The polypeptides, polynucleotides, and oligonucleotides as described herein can be appropriately combined with, for example, pharmacologically acceptable carriers or vehicles, specifically, sterile water, physiological saline, vegetable oil, emulsifying agents, suspending agents, surfactants, stabilizers, flavoring agents, excipients, vehicles, preservatives, bonding agents, or the like. These can be mixed together to be formulated in unit dosage form required by generally-accepted pharmaceutical practices.

[0136] Further disclosed are cells transformed with a polynucleotide (including DNA and RNA) encoding a mutant N-methyl aminoacyl-tRNA synthetase as described above. Such cells may be prokaryotic cells or eukaryotic cells.

[0137] When a herein described mutant N-methyl aminoacyl-tRNA synthetase expressed in cells is directly used for protein synthesis in the cells, cells depending on the intended use can be used. Any known method can be adopted for the transformation.

[0138] Further disclosed is a method for producing polypeptide containing N-methyl amino acids using the ARS with an altered amino acid sequence as described above.

[0139] The ARSs with altered amino acid sequences as described herein are used not only in cells but also *in vitro* (in a cell-free system).

[0140] In either case, unlike prior art using chemical synthesis such as the pdCpA method, tRNAs can be acylated repeatedly during a translation reaction, and therefore N-methyl aminoacyl-tRNAs can be supplied continuously, and addition of a large amount of tRNAs, which may inhibit translation, can be avoided.

[0141] Accordingly, disclosed is a method for producing non-natural amino acids efficiently, selectively, in particular regioselectively, and in large amounts.

<General method for preparing a modified ARS >

[0142] A modified ARS as described herein can be produced using any known genetic recombination technique as mentioned above, and generally can be prepared as follows: first, site-directed mutation is introduced into a specific position or a specific amino acid by PCR using a plasmid comprising the natural ARS gene as a template with appropriate primers. The template plasmid is digested with a restriction enzyme, and then *E. coli* or the like is transformed with the digested plasmid. The plasmid into which the desired mutation has been introduced is cloned.

[0143] If a second amino acid mutation is introduced, subsequently the introduction of site-directed mutation using the plasmid having the mutation introduced into the specific position as a template with appropriate primers is repeated as with the description above, which results in the construction of a plasmid DNA encoding the polypeptide in which the two amino acids are substituted. A similar procedure can be followed when introducing further amino acid mutations.

[0144] *E. coli* BL21 strain or the like is co-transformed with the constructed DNA and the plasmid pREP4 encoding lac repressor (LacI) or the like, and the obtained transformed strain is isolated and cultured followed by expression induction with IPTG. The obtained strain is then

disrupted, and the supernatant is passed through an affinity column for His-tag to purify the mutant ARS.

[0145] Alternatively, the following method can be used to prepare a mutant ARS. Specifically, a sequence of a certain mutant ARS is genetically synthesized and inserted into an expression vector, and proteins are allowed to be expressed. Affinity columns for different purification tags are used to purify the protein of interest.

[0146] Methods for producing a mutant N-methyl aminoacyl-tRNA synthetase as described herein, which is modified by substituting an amino acid at a specific position with another amino acid, are not limited to the methods as described above, and various genetic engineering techniques including known point mutation techniques and gene synthesis techniques, methods for introducing modified fragments using restriction enzymes, and the like may be used. Also, expression is not limited to expression in *E. coli*, and animal cells and cell-free translation systems may also be used. Purification methods are also not limited to methods in an affinity column for polyhistidine, and various peptide-tags and purification columns may be used.

<Method for confirming substrate specificity of the obtained modified ARS as described herein>

[0147] Substrate specificity of the obtained modified products can be confirmed by any of the following three assay methods, for example.

[0148] The first method can confirm by mass spectroscopy that N-methyl amino acids corresponding to codons on mRNA are introduced into a peptide by performing a translation reaction in a cell-free translation system reconstituted with the modified ARS rather than wild-type ARS and further N-methylated amino acids rather than natural amino acids, and performing aminoacylation with N-methyl amino acids in the translation system.

[0149] In the second method, peptides to be produced are labeled in a translation experiment using amino acids labeled with radioactive isotopes or fluorescent molecules, and the peptides are separated and visualized by electrophoresis or an analytical column to estimate yield of the peptides. In this method, translational introduction efficiency of corresponding N-methyl amino acids provided by using the modified ARS is increased compared to that provided by using wild-type ARS, and therefore the radioactivity or fluorescence, which is observed by electrophoresis or chromatogram, of the peptide produced using the modified ARS is observed more strongly than that of the peptide produced using wild-type ARS.

[0150] In the third method, three molecules, a modified ARS, a tRNA and an N-methyl amino acid corresponding to the modified ARS are reacted together *in vitro*, and the resulting N-methyl aminoacyl-tRNA is separated from unreacted tRNA by electrophoresis to quantify efficiency of the acylation reaction. The acylated tRNA produced using the modified ARS is

detected more than that produced using wild-type ARS.

[0151] As mentioned above, the modified ARSs as described herein have increased reactivity with N-methyl amino acids compared to unmodified ARSs. The increased reactivity with N-methyl amino acids may be an increase in reaction rate, an increase in the amount of reaction products acylated with N-methyl amino acids, or an increase in substrate specificity to N-methyl amino acids. The increased reactivity with N-methyl amino acids may also be qualitative or quantitative. For example, the reactivity with N-methyl amino acids is considered to have increased in a reaction performed under the same conditions except for the use of an unmodified ARS or a modified ARS when the reactivity (such as reaction rate or the amount of reaction products) of the modified ARS to the substrate N-methyl amino acid is significantly increased compared to the reactivity of the unmodified ARS to N-methyl amino acids. Even if the reaction rate or reaction product amount in a reaction using the modified ARS and an N-methyl amino acid as substrate does not significantly increase compared to that in a reaction using the unmodified ARS, the reactivity of the modified ARS to N-methyl amino acids is considered to have increased when substrate specificity of the modified ARS to the N-methyl amino acid is increased relative to that in a reaction using unmodified amino acids. Preferably, the modified ARSs as described herein increase reaction rate or reaction product amount in a reaction using N-methyl amino acids as substrates compared to unmodified ARSs.

[0152] For example, the modified ARSs as described herein significantly increase the amount of tRNA aminoacylated with N-methyl amino acids or the amount of peptides incorporating N-methyl amino acids, as measured under the same conditions except for using either unmodified ARS or modified ARS, preferably by at least 10%, preferably by 20%, preferably by 1.3 times or more, preferably by 1.5 times or more, preferably by 2 times or more, preferably by 3 times or more, further preferably by 5 times or more, further preferably by 10 times or more, further preferably by 20 times or more, further preferably by 30 times or more, further preferably by 50 times or more, and further preferably by 100 times or more. Alternatively, the modified ARSs as described herein significantly increase the amount ratio of "products of the reaction with N-methyl amino acids / products of the reaction with unmodified amino acids", as measured under the same conditions except for using either unmodified ARS or modified ARS, preferably by at least 10%, preferably by 20%, preferably by 1.3 times or more, preferably by 1.5 times or more, preferably by 2 times or more, preferably by 3 times or more, further preferably by 5 times or more, further preferably by 10 times or more, further preferably by 20 times or more, further preferably by 30 times or more, further preferably by 50 times or more, further preferably by 100 times or more compared to unmodified ARSs. Alternatively, when a nucleic acid encoding a polypeptide containing consecutive N-methyl amino acids is allowed to be translated, the modified ARSs as described herein significantly increase the production amount of the polypeptide containing consecutive N-methyl amino acids, as measured under the same conditions except for using either unmodified ARS or modified ARS, preferably by at least 10%, preferably by 20%, preferably by 1.3 times or more, preferably by 1.5 times or more, preferably by 2 times or more, preferably by 3 times or more compared to unmodified ARSs. The consecutive N-methyl amino acids may be, for example, two and/or three consecutive N-methyl amino acids.

[0153] The reactivity with N-methyl amino acids can be determined according to a confirmation method as described above. Specifically, for example, ribosomally-produced peptides are electrophoresed, and band intensity of peptides that have incorporated N-methyl amino acids and peptides that have not incorporated N-methyl amino acids can be measured qualitatively or quantitatively to determine the reactivity. Alternatively, mass spectral peaks can be measured, and peaks of peptides that have incorporated N-methyl amino acids and peaks of peptides that have not incorporated N-methyl amino acids can be measured to determine the reactivity.

[0154] For example, substrate specificity of a modified ARSs as described herein is significantly increased, preferably by at least 10%, preferably by 20%, preferably by 1.3 times or more, preferably by 1.5 times or more, preferably by 2 times or more, preferably by 3 times or more, further preferably by 5 times or more, further preferably by 10 times or more, and further preferably by 20 times or more compared to unmodified ARSs when peptides are synthesized in the presence of unmodified amino acids or N-methyl amino acids and the amount ratio of "products of the reaction with N-methyl amino acids / products of the reaction with unmodified amino acids" is measured.

[0155] The reaction conditions may be appropriately determined as long as the conditions used for unmodified ARSs and for modified ARSs are the same. The substrate concentration of unmodified amino acids and N-methyl amino acids at the reaction may be appropriately adjusted, and may be any concentration condition as long as reactivity with N-methyl amino acids is increased. Preferably, unmodified amino acids may not be added (may be only endogenous amino acids originally contained in a cell-free translation system), or may be appropriately adjusted in the range from 0.1 μM to 1 mM, for example, 0.1 μM to 500 μM , 0.1 μM to 250 μM , 0.1 μM to 100 μM , or 0.1 μM to 50 μM for reactions. N-methyl amino acids may be appropriately adjusted in the range from, for example, 50 μM to 10 mM, for example, 100 μM to 5 mM, 200 μM to 2 mM, or 500 μM to 1 mM for reactions.

[0156] Modified ARSs obtained by the preparation method as described above can be used to produce peptides and peptide-mRNA fusions that have incorporated N-methyl amino acids in a site-specific manner.

[0157] Moreover, ARSs used herein are highly conserved among biological species. Therefore, it is clear that the herein described method can be generally applied to modification of N-methyl amino acid-tRNA synthetases from other biological species.

[0158] The modified ARS as described herein can be used to produce peptides having a particular amino acid substituted with its non-natural, N-methylated amino acid, in a prokaryotic translation. Such peptides can be produced using ARSs that are derived from other organisms and modified in a similar way as disclosed herein.

<Acylation reaction of tRNAs using the modified ARSs >

[0159] Use of the modified ARSs as described herein makes it possible to employ N-methyl amino acids as substrates in acylation of tRNAs. The modified ARSs as described herein can acylate tRNAs with N-methyl amino acids and to ribosomally produce N-methyl amino acid-containing peptides.

[0160] In order to produce N-methyl aminoacyl-tRNAs using the modified ARSs as described herein, the modified ARSs, N-methyl amino acids corresponding to the modified ARSs, and tRNAs may be reacted *in vitro* as in chemical synthesis such as the pdCpA method. The products, N-methyl aminoacyl-tRNAs may be isolated using any known nucleic acid purification technique such as ethanol precipitation, and the isolated product may be added to a translation system. This leads to production of polypeptides or polypeptide-mRNA fusions that have introduced N-methyl amino acids at the intended positions.

[0161] Unlike chemical synthesis which requires reactions performed in a reaction solution comprising only a particular tRNA and the N-methyl amino acid corresponding to the tRNA as substrates, three molecules, a modified ARS, N-methyl amino acid, and tRNA are used to precisely and efficiently produce the intended N-methyl aminoacyl-tRNA even in a mixture of translation reaction solution comprising other tRNAs and other amino acids because the modified ARSs as described herein have high substrate specificity for tRNAs and amino acids. Accordingly, the isolation and purification steps as described above are not essential. A reaction solution in which tRNAs have been acylated with N-methyl amino acids can be directly used in a translation reaction, or a translation reaction can be performed at the same time as an acylation reaction of tRNAs with N-methyl amino acids. The modified ARSs as described herein are very convenient because the reactions using the modified ARSs require no chemical synthesis of substrates such as pdCpA amino acids and activated amino acids and can be performed using commercially available reagents.

[0162] The most important characteristics include no requirement of stoichiometric consideration for a tRNA of interest in performing peptide translation (peptide expression) using a modified ARS as described herein. In chemical synthesis in which aminoacyl-tRNAs are supplied, N-methyl aminoacyl-tRNAs are consumed during the reaction, decreasing the reaction efficiency. However, if ARSs are used, translational efficiency is good even in peptide synthesis in which multiple N-methyl amino acids are introduced. Aminoacyl-tRNAs are constantly deacylated in hydrolytic reactions or transpeptidation reactions in translation systems, but the modified ARSs as described herein recognize released tRNAs and newly acylate the tRNAs with N- α amino acids, which results in a constant supply of N- α aminoacyl-tRNAs in the translation system. Therefore, the amount of tRNA required is less than the amount of produced polypeptide. Addition of large amounts of tRNA itself contributes to reduction of peptide yield. Therefore, the modified ARSs as described herein are useful for producing N- α amino acid-containing peptides at a high translational efficiency and producing highly diverse peptide libraries.

[0163] Moreover, because the modified ARSs as described herein have increased reactivity with N-methyl amino acids compared to natural ARSs, the modified ARSs can acylate tRNAs in the presence of N-methyl amino acids at a concentration lower than a concentration of N-methyl amino acids used for natural ARSs. In other words, the absolute amount of N-methyl amino acids required for peptide expression using the modified ARSs is not as much as that of N-methyl amino acids required for peptide expression using natural ARSs.

[0164] In addition, the modified ARSs as described herein can be used to translate N-methyl amino acid-containing peptides at a concentration lower than a concentration of natural ARSs used for translating N-methyl amino acid-containing peptides. In other words, the absolute amount of the modified ARS as described herein required for peptide expression is significantly less than in translation of N-methyl amino acid-containing peptides using natural ARSs.

[0165] These characteristics are important for improving orthogonality of aminoacylation reaction. Namely, in order to translate N-methyl amino acid-containing peptides, the more the substrate N-methyl amino acid, or an ARS corresponding to the N-methyl amino acid, is needed at a concentration higher than other natural amino acids, or ARSs corresponding to the other natural amino acids, the more it makes it easier for non-specific reactions to happen, such non-specific reactions being: ARS that should be normally using an N-methyl amino acid as a substrate may use other natural amino acids for acylation reaction; or an N-methyl amino acid that is present in excess may be used in acylation reactions by ARSs corresponding to other natural amino acids. This results in the obtained translated peptides being a mixture of N-methyl amino acid-containing peptides and peptides containing corresponding natural amino acid, which may cause issues for peptide libraries. However, it is expected that this issue can be minimized by using the modified ARSs as described herein.

[0166] Accordingly, the characteristics of the modified ARSs disclosed herein can be described as follows:

modified ARSs that catalyze acylation of tRNAs and comprise

1. (a) a tRNA binding site;
2. (b) a binding site to an N-methyl amino acid substrate; and
3. (c) a catalytically active site having activity to catalyze a reaction in which an acyl group is transferred from the N-methyl amino acid substrate to 3' end of a tRNA,

wherein the modified ARSs are characterized in that they can correctly recognize tRNAs and N-methyl amino acids in a translation reaction mixture, acylates the tRNAs, and further reuse tRNAs released after the transacylation reaction to repeat acylation reactions again.

[0167] Moreover, the modified ARSs may comprise, in addition to (a), (b), and (c) as described above, (d) an editing site where a tRNA acylated with any undesired amino acid is hydrolyzed.

<Synthesis of acylated tRNAs using the modified ARSs >

[0168] The modified ARSs as described herein can be used to synthesize tRNAs acylated with desired N-methyl amino acid substrates.

[0169] A method for producing acylated tRNAs using the modified ARSs as described herein, comprising the following steps:

1. (a) providing one or more modified ARSs as described herein;
2. (b) providing tRNAs;
3. (c) providing N-methyl amino acids; and
4. (d) contacting the modified ARSs with the tRNAs and N-methyl amino acids to acylate the tRNAs.

[0170] In addition to the steps as described above, the method may further comprise the step of (e) collecting the reaction product comprising the acylated tRNAs. The acylated tRNAs require no purification in the collecting step, and the reaction mixture can be collected and used directly. By not separating or purifying the produced aminoacyl-tRNAs from the ARSs, deacylation can be prevented.

[0171] This method uses N-methyl amino acids as substrates. Particularly preferable N-methyl amino acids include N-methylphenylalanine, N-methylvaline, N-methylthreonine, N-methyltryptophan, N-methylleucine, and/or N-methylserine. A substrate of an ARS (an amino acid corresponding to the ARS) is appropriately selected as N-methyl amino acid used as a substrate.

[0172] In a method for producing an acylated tRNA using a modified ARS as described herein, a tRNA that can be used is a tRNA corresponding to an ARS of a corresponding natural amino acid. The term "corresponding natural amino acid" refers to an amino acid that is not N-methylated relative to an N-methyl amino acid. For example, for N-methylphenylalanine, the corresponding amino acid is phenylalanine, and the corresponding tRNA is a tRNA recognizing the codon UUU or UUC (having the anticodon corresponding to the codon). For N-methylvaline, the corresponding amino acid is valine, and the corresponding tRNA is a tRNA recognizing the codon GUU, GUC, GUA, or GUG (having the anticodon corresponding to the codon). For N-methylserine, the corresponding amino acid is serine, and the corresponding tRNA is a tRNA recognizing the codon UCU, UCC, UCA, UCG, AGU, or AGC (having the anticodon corresponding to the codon). For N-methylthreonine, the corresponding amino acid is threonine, and the corresponding tRNA is a tRNA recognizing the codon ACU, ACC, ACA, or ACG (having the anticodon corresponding to the codon). For N-methyltryptophan, the corresponding amino acid is tryptophan, and the corresponding tRNA is a tRNA recognizing the codon UGG (having the anticodon corresponding to the codon). For N-methylleucine, the corresponding amino acid is leucine, and the corresponding tRNA is a tRNA recognizing the codon UUA, UUG, CUU, CUC, CUA, or CUG (having the anticodon corresponding to the codon). Similarly, for other N-methyl amino acids, tRNAs recognizing codons corresponding to

the corresponding amino acid (having the anticodons corresponding to the codons) can be used.

[0173] When a tRNA is acylated by a modified ARS in a solution, a pellet obtained by ethanol precipitation of the reaction solution may be dissolved in a suitable buffer (such as 1 mM potassium acetate, pH 5) and added to a translation system. Typical reaction conditions include, for example, a reaction performed at 37°C for 5 minutes to 1 hour in 0.1 M reaction buffer (pH 7.5) containing a final concentration of 0.5 μ M to 40 μ M of a tRNA, 0.1 μ M to 10 μ M of a modified ARS as described herein, 0.1 mM to 10 mM of an N-methyl amino acid, 0.1 mM to 10 mM ATP, and 0.1 mM to 10 mM $MgCl_2$.

[0174] Furthermore, for an aminoacylation reaction, a tRNA can be refolded by, for example, heating 1 to 50 μ M tRNA, 10 to 200 (e.g., 50 to 200) mM HEPES-K (pH 7.0 to 8.0 (e.g., 7.6)), 1 to 100 (e.g., 10) mM KCl solution at 95°C for 2 minutes and then left at room temperature for 5 minutes or more. This tRNA solution can be added to an acylation buffer (a final concentration of 25 to 100 (e.g., 50) mM HEPES-K [pH 7.0 to 8.0 (e.g., 7.6)], 1 to 10 (e.g., 2) mM ATP, 10 to 100 (e.g., 100) mM potassium acetate, 1 to 20 (e.g., 10) mM magnesium acetate, 0.1 to 10 (e.g., 1) mM DTT, 0.1 mg/mL Bovine Serum Albumin) to a final concentration of 1 to 40 (e.g., 10) μ M, mixed with a modified ARS (a final concentration of 0.1 to 10 (e.g., 0.5) μ M) and an N-methyl amino acid (a final concentration of 0.1 to 10 (e.g., 1) mM), and incubated at 37°C for 5 to 60 (e.g., 10) minutes.

[0175] Thus, the acylation reaction using a modified ARS as described herein requires no activated amino acids that need substrates and need to be synthesized. The acylation reaction can be performed with commercially available N-methyl amino acids and therefore is convenient. The modified ARSs as described herein can be combined with their substrates to form a kit product for obtaining acylated tRNAs. The kit may at least comprise (a) one or more modified ARSs as described herein, (b) N-methyl amino acid(s), and (c) tRNA(s), and may further comprise a reaction buffer, a reaction vessel, instructions for use, and the like. In the kit, each of (b) N-methyl amino acid(s) and (c) tRNA(s) acts as a substrate for (a) modified ARSs. In other words, the tRNAs recognize codons corresponding to natural amino acids equivalent to the N-methyl amino acids (the tRNAs have anticodons corresponding to the codons).

<Synthesis of N-methyl amino acid-containing polypeptides using the modified ARSs as described herein>

[0176] N-methyl amino acid-bound tRNAs can be used to produce polypeptides with N-methyl amino acids introduced into desired sites. The method comprises translating a nucleic acid encoding a polypeptide of interest in the presence of a modified ARS as described herein.

[0177] More specifically, for example, a method for producing an N-methyl amino acid-containing polypeptide using a modified ARS(s) as described herein comprises (a) providing the modified ARS(s) as described herein, (b) establishing a cell-free translation system

reconstituted with the modified ARS(s) as described herein rather than wild-type ARS(s), (c) providing an mRNA having at a desired site(s) codon(s) corresponding to the anticodon(s) of a tRNA(s) that is the substrate(s) of the modified ARS(s), and (d) adding the mRNA to the cell-free translation system to produce a polypeptide with an N-methyl amino acid(s) introduced into a desired site(s). Matters particularly relevant to the production of polypeptide in (d) will be described below.

[0178] N-methyl amino acids preferably used in acylation of tRNAs with N-methyl amino acids using the modified ARSs as described herein include N-methylalanine, N-methylleucine, N-methyltryptophan, N-methylphenylalanine, N-methylvaline, N-methylthreonine, and/or N-methylserine, and particularly preferably N-methylphenylalanine, N-methylvaline, N-methylthreonine, N-methyltryptophan, N-methylleucine, and/or N-methylserine.

[0179] Specific methods for polypeptide synthesis may be essentially performed according to known methods, for example, performed as described in WO2013100132, and various modifications can be made. Generally, the methods can be performed according to the following description.

[0180] A suitable translation system that can be used is a cell-free translation system, typified by PURESYSYSTEM (Registered trademark) (BioComber, Japan), reconstituted with translation factors. In such a cell-free translation system, components of the translation system can be controlled flexibly. For example, phenylalanine and the ARS corresponding to phenylalanine are removed from the translation system and instead, N-methylphenylalanine and the modified phenylalanine-ARS as described herein can be added. This achieves introduction of N-methylphenylalanine into a codon, such as UUU and UUC, encoding phenylalanine in a site-specific manner.

[0181] In a cell-free translation system, ribonucleosides preferably used are ATP and GTP at 0.1 mM-10 mM. A buffer preferably used is HEPES-KOH at 5 mM-500 mM and pH 6.5-8.5. Examples of other buffers include, but are not limited to, Tris-HCl, phosphate, and the like. Salts that can be used are acetates such as potassium acetate and ammonium acetate, and glutamates such as potassium glutamate, which are preferably used at 10 mM-1000 mM. A magnesium component preferably used is magnesium acetate at 2 mM-200 mM. Examples of other magnesium components include, but are not limited to, magnesium chloride and the like. Components in an energy-regenerating system preferably used are creatine kinase at 0.4 µg/mL-40 µg/mL and creatine phosphate at 2 mM-200 mM. Other energy-regenerating systems, typified by pyruvate kinase and phosphoenol pyruvate, may also be used. A nucleoside converting enzyme preferably used is myokinase at 0.1 unit/mL-10 unit/mL or nucleoside diphosphate kinase at 0.2 µg/mL-20 µg/mL. A diphosphatase preferably used is inorganic pyrophosphatase at 0.2 unit/mL-20 unit/mL. A polyamine preferably used is spermidine at 0.2 mM-20 mM. Examples of other polyamines include, but are not limited to, spermine and the like. A reducing agent preferably used is dithiothreitol at 0.1 mM-10 mM. Examples of other reducing agents include, but are not limited to, β-mercaptoethanol and the like. A tRNA preferably used is, for example, *E. coli* MRE600 (RNase-negative)-derived tRNA

(Roche) at 0.5 mg/mL-50 mg/mL. Other tRNAs from *E. coli* may also be used. A formyl donor and an enzyme preferably used to synthesize formylmethionine used in a translational initiation reaction are 10-HCO-H4 folate at 0.1 mM-10 mM and methionyl-tRNA transformylase at 0.05 μ M-5 μ M. A translation initiation factor preferably used is IF1 at 0.5 μ M-50 μ M, IF2 at 0.1 μ M-50 μ M, or IF3 at 0.1 μ M-50 μ M. A translation elongation factor preferably used is EF-G at 0.1 μ M-50 μ M, EF-Tu at 1 μ M-200 μ M, or EF-Ts at 1 μ M-200 μ M. A translation termination factor preferably used is RF-2, RF3, or RRF at 0.1 μ M-10 μ M. Ribosome is preferably used at 1 μ M-100 μ M. There are 20 types of aminoacyl-tRNA synthetases, but only the enzymes corresponding to amino acids included in a peptide to be synthesized may be added. For example, ArgRS, AspRS, LysRS, MetRS, and TyrRS are all preferably used at 0.01 μ M-1 μ M. Amino acids used as substrates for peptide synthesis are natural 20 amino acids, which compose proteins, and derivatives thereof. Only amino acids included in a peptide to be synthesized are preferably used at 0.25 mM-10 mM. An mRNA as a template for peptide synthesis is preferably used at 0.1 μ M-10 μ M. When an mRNA is transcribed from a template DNA in a cell-free translation system, commercially available enzymes such as T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase can be used, and the enzymes suitable for a promoter sequence in the template DNA may be appropriately selected and are preferably used at 1 μ g/mL-100 μ g/mL. In this case, nucleosides CTP and UTP, which are substrates, are preferably used at 0.1 mM-10 mM. A solution containing mixed these components can be left, for example, at 37°C for 1 hour to achieve translational synthesis of peptides. Temperature and reaction time are not limited to those.

[0182] The modified ARSs as described herein can also be used in combination with other techniques for introducing non-natural amino acids, such as the pdCpA method or the Flexizyme method. For example, polypeptides containing both N-methylglycine and N-methylphenylalanine can be synthesized when the aminoacyl-tRNA to which N-methylglycine is attached by the pdCpA method is added to a translation system at the same time as addition of N-methylphenylalanine and the modified phenylalanine ARS.

[0183] Alternatively, polypeptides containing N-methyl amino acids can be expressed in cells by inserting the modified ARSs as described herein into an expression vector or genome, expressing the modified ARSs in cells, and using N-methyl amino acids added to a medium as substrates.

[0184] tRNAs that can be used include a tRNA for a natural amino acid corresponding to an N-methyl amino acid. The tRNAs may be tRNAs that contain modified bases and are purified from a living body, or tRNAs that do not contain modified bases and are produced using an *in vitro* transcription reaction. Mutant tRNAs, which have a mutation in a portion of tRNA other than the portion recognized by an ARS, can also be used as substrates.

Examples

Example 1: N-methylphenylalanine-accepting ARS

<Preparation of plasmids for wild-type and mutant PheRSs>

[0185] Using a plasmid (pQE-32(2) 2_wtPheRS) comprising the ORF sequence of wild-type PheRS α subunit gene of *E. coli* (SEQ ID NOs: 27, 28) as starting material, mutant PheRS plasmids (having His-tag (6 x His) at the N-terminus) listed in Table 1 were constructed by introducing site-directed mutations using PCR. Specifically, 2 μ L of 10 ng/ μ L template, 10 μ L of 2 x KOD Fx buffer (TOYOBO, KFX-101), 0.6 μ L of 10 μ M forward primer, 0.6 μ L of 10 μ M reverse primer, 4 μ L of 2 mM dNTP, 0.4 μ L of KOD FX (TOYOBO, KFX-101), and 2.4 μ L of H₂O were mixed together. Thereafter, the resulting reaction solution was heated at 94°C for 2 minutes and then subjected to 10 cycles, each consisting of heating at 98°C for 10 seconds and heating at 68°C for 7 minutes, to amplify the mutant gene. The combinations of the template plasmid, forward primer, and reverse primer used are listed in Table 2. Each sequence of primers "F.F02" through "F.F05" corresponds to SEQ ID NOs: 30-33 in ascending order, and each sequence of primers "R.F02" through "R.F05" corresponds to SEQ ID NOs: 34-37 in ascending order. 0.5 μ L of 10 U/ μ L *DpnI* was then added to the PCR reaction solution and further incubated at 37°C for 1.5 hours to digest the template DNA, and the resulting mutant DNA was purified. *E. coli* XL-1 Blue strain (STRATAGENE, 200236) was then co-transformed with the resulting mutant gene DNA and pREP4 (Invitrogen, V004-50) encoding the *lacI* gene. The transformants were seeded onto agar containing ampicillin and kanamycin.

[0186] The plasmids of interest were purified from the resulting clones. The mutations were confirmed to be introduced into the plasmids.

[Table 1]

Name	position 169	position 171
PheRS01(wt)	Q	S
PheRS02	Q	A
PheRS03	Q	G
PheRS04	A	S
PheRS05	G	S

[Table 2]

Name	Template	5' primer	3' primer
PheRS02	PheRS01 (wt)	F.F02	R.F02
PheRS03	PheRS01 (wt)	F.F03	R.F03
PheRS04	PheRS01 (wt)	F.F04	R.F04
PheRS05	PheRS01 (wt)	F.F05	R.F05

<Small scale expression of wild-type and mutant PheRSs>

[0187] Next, a plasmid comprising a resulting mutant gene and the gene encoding PheRS β subunit was introduced into *E. coli*, and a heterodimer comprising the mutant protein was expressed. First, *E. coli* BL21 strain transformed with the mutant plasmid and pREP4 (Invitrogen, V004-50) was cultured at 37°C in 4 mL of LB medium containing kanamycin, ampicillin, and 0.5% glucose. Subsequently, when the OD value at 600 nm reached 0.4 to 0.8, IPTG was added to a final concentration of 0.5 mM. After further culturing at 37°C for 4 hours, the bacterial pellets were collected using a centrifuge.

<Small scale purification of wild-type and mutant PheRSs>

[0188] Next, the resulting bacterial pellets were disrupted, and the mutant protein of interest was purified from the supernatant. Specifically, the bacterial pellets as described above were suspended in 600 μ L of CHAPS solution (0.5% CHAPS (DOJINDO: 349-04722), 50% TBS (TaKaRa, T903)) and mixed with 6 μ L of 30 U/ μ L rLysozyme (Novagen, 71110-3), and 2 μ L of 2.5 U/ μ L benzonase nuclease (Novagen, 70746-3) followed by incubation at room temperature for 30 minutes. Imidazole was then added to a final concentration of 15 mM, and an insoluble fraction was separated by centrifugation. Then, the mutant protein was purified from the resulting supernatant using QIAGEN Ni-NTA spin column kit (Qiagen, 31314) according to the product manual. Finally, excess imidazole was removed using a desalting column, PD miniTrap G-25 (GE Healthcare, 28-9180-07) according to the product manual.

<Large scale purification of wild-type and mutant PheRSs>

[0189] The mutant protein confirmed to have activity was prepared in large scale. Specifically, *E. coli* BL21 strain transformed with a plasmid comprising the mutant α subunit gene and wild-type β subunit gene and pREP4 (Invitrogen, V004-50) was cultured at 37°C in 3 L of LB medium containing kanamycin, ampicillin, and 0.5% glucose. Then, when the OD value at 600 nm reached 0.4, IPTG was added to a final concentration of 0.5 mM. After further culturing at 37°C for 4 hours, the bacterial pellets were collected with a centrifuge. The bacterial pellets as described above were suspended in 1 L of CHAPS solution (0.5% CHAPS (DOJINDO: 349-04722) and 50% TBS (TaKaRa, T903)), mixed with 10 μ L of 30 KU/ μ L rLysozyme (Novagen, 71110-3), and stirred at room temperature for 10 minutes. Next, 2 mL of 1 M $MgCl_2$ and 320 μ L of Benzonase Nuclease (Novagen, 70746-3) were added and stirred at room temperature for 20 minutes. Imidazole was then added to a final concentration of 20 mM, and an insoluble fraction was separated by centrifugation. Subsequently, the mutant protein was purified from the resulting supernatant using a column filled with 15 mL of Ni Sepharose High Performance (GE Healthcare) and AKTA10S (GE Healthcare) with an imidazole concentration gradient

(initial concentration 20 mM, final concentration 500 mM). Finally, dialysis was performed three times (2 hours x 2, overnight x 1) using a dialysis cassette (MWCO 10,000, Slide-A-Lyzer G2 Dialysis Cassettes 70mL, Thermo Scientific Pierce) and 3 L of stock solution (50 mM Hepes-KOH, 100 mM KCl, 10 mM MgCl₂, 1 mM DTT, pH 7.6) to obtain the mutant protein.

<Aminoacylation reaction with N-methylphenylalanine using wild-type and mutant PheRSs>

[Synthesis of E. coli tRNAPhe by *in vitro* transcription reaction]

[0190] E. coli tRNA (R-tRNAPhe (SEQ ID NO: 39)) was synthesized from a template DNA (D-tRNAPhe (SEQ ID NO: 38)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) in the presence of 7.5 mM GMP, and purified using RNeasy Mini kit (Qiagen).

[0191]

D-tRNAPhe (SEQ ID NO: 38)

tRNAPhe DNA sequence:

GGCCTAATACGACTCACTATAGCCCGGATAGCTCAGTCGGTAGAGCAGGGGATTGA
AAATCCCCGTGTCCTTGGTTCGATTCCGAGTCCGGGCACCA

R-tRNAPhe (SEQ ID NO: 39)

tRNAPhe RNA sequence:

GCCCGGAUAGCUCAGUCGGUAGAGCAGGGGAUUGAAAAUCCCCGUGUCCUUGGU
UCGAUUCCGAGUCCGGGCACCA

<Aminoacylation reaction>

[0192] For the aminoacylation reaction, the solution containing 40 μM transcribed tRNAPhe, 10 mM HEPES-K (pH 7.6), and 10 mM KCl solution was heated at 95°C for 2 minutes and then left at room temperature for 5 minutes or more to refold the tRNA. This tRNA solution was added to a final concentration of 10 μM to an acylation buffer (in final concentrations of 50 mM HEPES-K [pH 7.6], 2 mM ATP, 100 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 2 mM spermidine, 0.1 mg/mL Bovine Serum Albumin), mixed with wild-type or mutant PheRS (final concentration 0.5 μM) and phenylalanine (final concentration 0.25 mM, Watanabe Chemical Industries, Ltd., G00029) or N-methylphenylalanine (final concentration 1 mM, Watanabe Chemical Industries, Ltd., J00040), and incubated at 37°C for 10 minutes. Four volumes of a loading buffer (90 mM sodium acetate [pH 5.2], 10 mM EDTA, 95%(w/w)

formamide, 0.001%(w/v) xylene cyanol) was added to the reaction solution and analyzed with acidic PAGE (12 % (w/v) polyacrylamide gel, pH 5.2) containing 6 M urea, and aminoacylation activity was detected by separating unreacted tRNA and aminoacylated tRNA. RNA was stained with SYBR Gold (Life Technologies) and detected with LAS4000 (GE Healthcare).

[0193] The activity for the aminoacylation reaction was assessed, and mutants 04 and 05 had increased activity for aminoacylation with N-methyl-phenylalanine compared to the wildtype (Figure 1). The base sequence of mutant 04 is set forth in SEQ ID NO: 12, and the amino acid sequence of mutant 04 is set forth in SEQ ID NO: 1. The base sequence of mutant 05 is set forth in SEQ ID NO: 13, and the amino acid sequence of mutant 05 is set forth in SEQ ID NO: 2.

<Translational introduction of N-methylphenylalanine using wild-type and mutant PheRSs>

[Synthesis of template DNA-F by *in vitro* transcription reaction]

[0194] A template mRNA for translation (R-F (SEQ ID NO: 41)) was synthesized from a template DNA (D-F (SEQ ID NO: 40)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) and purified using RNeasy Mini kit (Qiagen).

[0195]

D-F (SEQ ID NO: 40)

DNA sequence:

GCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCGTTTC
CGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-F (SEQ ID NO: 41)

RNA sequence:

GGGUUAAACUUUAAACAAGGAGAAAAACAUGCGUUUCCGUGACUACAAGGACGACG
ACGACAAGUAAGCUUCG

<Cell-free translation system>

[0196] In order to confirm translational introduction of N-methylphenylalanine, a desired polypeptide containing N-methylphenylalanine was ribosomally synthesized by adding an N-methyl amino acid and PheRS to a cell-free translation system. The translation system used was PURE system, a reconstituted cell-free protein synthesis system from *E. coli*. Specifically,

wild-type or mutant PheRS and phenylalanine or N-methylphenylalanine were added to a solution containing a basic cell-free translation solution (1 mM GTP, 1 mM ATP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 9 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol, 0.5 mg/ml *E. coli* MRE600 (RNase negative)-derived tRNA (Roche), 4 µg/ml creatine kinase, 3 µg/ml myokinase, 2 unit/ml inorganic pyrophosphatase, 1.1 µg/ml nucleoside diphosphate kinase, 0.6 µM methionyl-tRNA transformylase, 0.26 µM EF-G, 0.24 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 40 µM EF-Tu, 44 µM EF-Ts, 1.2 µM ribosome, 0.03 µM ArgRS, 0.13 µM AspRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.02 µM TyrRS (wherein the proteins prepared by the inventors were essentially prepared as His-tagged proteins)), and 1 µM template mRNA, each 250 µM arginine, aspartic acid, lysine, methionine, and tyrosine, and left at 37°C for 1 hour to ribosomally synthesize the peptide.

<Detection by electrophoresis>

[0197] A peptide expression experiment was performed by using aspartic acid labeled with a radioisotope for detecting peptides into which N-methylphenylalanine is ribosomally introduced. Specifically, a solution containing 1 µM template mRNA (R-F (SEQ ID NO: 41)), arginine, lysine, methionine, and tyrosine (each final concentration 250 µM), and ¹⁴C-aspartic acid (final concentration 37 µM, Moravek Biochemicals, MC139) added to the cell-free translation system as described above was prepared. Wild-type PheRS or the mutant PheRS 05 (final concentration 0.1 µM) and phenylalanine (final concentration 250 µM) or N-methylphenylalanine (final concentration 1 mM or 250 µM) were added to the solution and incubated at 37°C for 60 minutes. An equal volume of 2× sample buffer (TEFCO, cat No. 06-323) was added to the resulting translation reaction solution and heated at 95°C for 3 minutes followed by electrophoresis (16% Peptide-PAGE mini, TEFCO, TB-162). After electrophoresis, the gel was dried using Clear Dry Quick Dry Starter KIT (TEFCO, 03-278), exposed to an imaging plate (GE Healthcare, 28-9564-75) for about 16 hours, detected using a Bioanalyzer System (Typhoon FLA 7000, GE Healthcare), and analyzed with ImageQuantTL (GE Healthcare).

[0198] Almost no bands of peptides that were ribosomally synthesized were observed in the presence of 0.1 µM wild-type PheRS when 0.25 mM or 1 mM N-methylphenylalanine was added (Figure 2). Meanwhile, peptide bands were observed in the presence of 0.1 µM mutant PheRS 05 even when 0.25 mM N-methylphenylalanine was added. The results can provide confirmation that the mutant PheRS 05 had increased aminoacylation activity with N-methylphenylalanine and that peptides containing N-methylphenylalanine were ribosomally synthesized at a high yield.

<Detection by mass spectroscopy>

[0199] Mass spectroscopy was performed using MALDI-TOF MS for detecting peptides into which N-methylphenylalanine is ribosomally introduced. Specifically, a solution containing 1 μM template mRNA (R-F (SEQ ID NO: 41)) and amino acids, arginine, lysine, methionine, tyrosine, and aspartic acid (each final concentration 250 μM) added to the cell-free translation system as described above was prepared. PheRS (final concentration 0.1 μM) and phenylalanine (final concentration 250 μM) or N-methylphenylalanine (final concentration 1 mM or 250 μM) were added to the solution and incubated at 37°C for 60 minutes. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS. α -Cyano-4-hydroxycinnamic acid was used as a matrix for translation products.

[0200] When a translation experiment was performed as a control experiment in the presence of 0.1 μM wild-type PheRS (SEQ ID NO: 28) or mutant PheRS 05 (SEQ ID NO: 2) with 250 μM phenylalanine added, the peak of the peptide into which phenylalanine was introduced (Calculated $[M+H]^+ = 1631.7$) was detected in both experiments using wild-type PheRS and the mutant PheRS 05 (Figure 3(a), (d)).

[0201] Translational synthesis was then performed by using wild-type PheRS and adding 0.25 mM N-methylphenylalanine, and the peaks of the peptide containing N-methylphenylalanine (Calculated $[M+H]^+ = 1645.7$) and the peptide containing phenylalanine were observed (Figure 3 (b)). This phenomenon is thought to be due to wild-type PheRS recognizing phenylalanine contaminated in N-methylphenylalanine, the subsequent occurrence of the aminoacylation reaction, and proceeding of translational synthesis. Furthermore, the peak intensity of the peptide containing N-methylphenylalanine was increased under the condition of 1 mM N-methylphenylalanine (Figure 3(c)). The ratio of peak intensities at each amino acid concentration was calculated (the calculating formula: the ratio of peak intensity = the peak intensity of the peptide containing N-methylphenylalanine / the peak intensity of the peptide containing phenylalanine) to be 0.8 at 0.25 mM and 2.6 at 1 mM.

[0202] Next, a translational synthesis was performed by using the mutant PheRS 05 and adding 0.25 mM or 1 mM N-methylphenylalanine, and the peak of the peptide containing N-methylphenylalanine was strongly detected, and the ratio of peak intensity of the peptide containing N-methylphenylalanine to the peak intensity of the peptide containing phenylalanine was 12.4 at 0.25 mM and 16.0 at 1 mM (Figure 3 (e), (f)). These results confirmed that the mutant PheRS 05 had increased aminoacylation activity with N-methylphenylalanine compared to wild-type PheRS, and consequently, translational synthesis of the peptide containing N-methylphenylalanine was promoted.

[0203]

Peptide sequence P-F1 (SEQ ID NO: 42) formylMetArgPheArgAspTyrLysAspAspAspAspLys

Peptide	sequence	P-MeF1	(SEQ	ID	NO:	42)
formylMetArg[MePhe]ArgAspTyrLysAspAspAspAspLys						

MALDI-TOF MS:

Calc. m/z : $[M+H]^+ = 1631.7$ (the peptide corresponding to the sequence P-F1)

Calc. m/z: $[H+M]^+ = 1051.7$ (the peptide corresponding to the sequence P-F1)

Calc. m/z: $[H+M]^+ = 1645.7$ (the peptide corresponding to the sequence P-MeF1)

<Comparison with the pdCpA method>

[0204] Incorporation efficiency of N-methylphenylalanine using an ARS as described herein was compared with that in the pdCpA method. Translation was performed using a sequence containing one, two consecutive, or three consecutive phenylalanines in a cell-free translation system, and the synthesized peptide band was detected by electrophoresis. The amount of the peptide synthesized using a mutant PheRS as described herein (PhrRS05 (SEQ ID NO: 2)) was compared with the amount of the peptide synthesized using the pdCpA method. It is confirmed that the amount of the peptide synthesized using the mutant PheRS was detected more than the amount of the peptide synthesized using the pdCpA method in all translations and that translational efficiency obtained by using the mutant PheRS was also higher than that obtained by using the pdCpA method. Particularly, it is revealed that the amount of the peptide produced by translating sequences containing two consecutive and three consecutive N-methylphenylalanines using the mutant PheRS (PhrRS05) was 4 to 8 times higher than the amount of the peptide produced using the pdCpA method.

Example 2: N-methylvaline-accepting ARS

<Preparation of plasmids for wild-type and mutant ValRSs>

[0205] The mutant ValRS plasmids (having His-tag (6 × His) at the N-terminus) listed in Table 3 were constructed by introducing site-directed mutations using PCR into the plasmid (PQE-32(2) 2_wtVALRS) comprising the ORF sequence of wild-type ValRS gene of *E. coli* (SEQ ID NOs: 23, 24), which was used as starting material. Specifically, 2 μL of 10 ng/μL template, 10 μL of 2 × KOD Fx buffer (TOYOBO, KFX-101), 0.6 μL of 10 μM forward primer, 0.6 μL of 10 μM reverse primer, 4 μL of 2 mM dNTP, 0.4 μL of KOD FX (TOYOBO, KFX-101), and 2.4 μL of H₂O were mixed together. Then, the resulting reaction solution was heated at 94°C for 2 minutes and then subjected to 10 cycles, each consisting of heating at 98°C for 10 seconds and heating at 68°C for 7 minutes, to amplify the mutant gene. The combinations of the template plasmid, forward primer, and reverse primer used are listed in Table 4. Each sequence of primers "F.V2" through "F.V19", "F.V46" through "F.V48", and "F.V13-01" through "F.V13-16" corresponds to SEQ ID NOs: 43-79 in ascending order. Each sequence of primers "R.V2" through "R.V19", "R.V46" through "R.V48", and "R.V13-01" through "R.V13-16" corresponds to SEQ ID NOs: 80-116 in ascending order. 0.5 μL of 10 U/μL *DpnI* was then added to the PCR reaction solution and further incubated at 37°C for 1.5 hours to digest the

template DNA, and the resulting mutant DNA was purified. *E. coli* XL-1 Blue strain (STRATAGENE, 200236) was then co-transformed with the resulting mutant gene DNA and pREP4 (Invitrogen, V004-50) encoding the lacI gene. The transformants were seeded onto agar containing ampicillin and kanamycin. The plasmids of interest were purified from the resulting clones. The mutations were confirmed to be introduced into the plasmids.

[0206] For plasmid construction requiring multistep mutation introduction, the procedure as described above was repeated to obtain plasmids into which mutations of interest were introduced. The combinations of the primers and template for such plasmid construction are listed in Table 4.

[Table 3]

Name	position 80	position 43	position 41
ValRS001(wt)	D	N	P
ValRS002			A
ValRS003			G
ValRS004		A	P
ValRS005			A
ValRS006			G
ValRS007		S	P
ValRS008			A
ValRS009			G
ValRS010		V	P
ValRS011			A
ValRS012			G
ValRS013		G	P
ValRS014			A
ValRS015			G
ValRS016		D	P
ValRS017			A
ValRS018			G
ValRS019	A	N	P
ValRS020			A
ValRS021			G
ValRS022		A	P
ValRS023			A
ValRS024			G
ValRS025		S	P
ValRS026			A
ValRS027			G
ValRS028		V	P
ValRS029			A
ValRS030			G
ValRS031		G	P
ValRS032			A
ValRS033			G
ValRS034		D	P
ValRS035			A
ValRS036			G
ValRS037	S	D	P
ValRS038			A
ValRS039			G
ValRS040	V	D	P
ValRS041			A
ValRS042			G
ValRS043	G	D	P
ValRS044			A
ValRS045			G
ValRS046	S	N	P
ValRS047	V	N	P
ValRS048	G	N	P

Name	position 43	position 45	position 85
ValRS13-01	G	T	Y
ValRS13-02	G	T	W
ValRS13-03	G	T	S
ValRS13-04	G	T	M
ValRS13-05	G	T	K
ValRS13-06	G	T	N
ValRS13-07	G	T	V
ValRS13-08	G	T	L
ValRS13-09	G	Y	A
ValRS13-10	G	W	A
ValRS13-11	G	S	A
ValRS13-12	G	M	A
ValRS13-13	G	K	A
ValRS13-14	G	N	A
ValRS13-15	G	V	A
ValRS13-16	G	L	A

[Table 4]

	forward primer	reverse primer	Template
VaIRS001			
VaIRS002	F. V2	R. V2	wild type
VaIRS003	F. V3	R. V3	wild type
VaIRS004	F. V4	R. V4	wild type
VaIRS005	F. V5	R. V5	wild type
VaIRS006	F. V6	R. V6	wild type
VaIRS007	F. V7	R. V7	wild type
VaIRS008	F. V8	R. V8	wild type
VaIRS009	F. V9	R. V9	wild type
VaIRS010	F. V10	R. V10	wild type
VaIRS011	F. V11	R. V11	wild type
VaIRS012	F. V12	R. V12	wild type
VaIRS013	F. V13	R. V13	wild type
VaIRS014	F. V14	R. V14	wild type
VaIRS015	F. V15	R. V15	wild type
VaIRS016	F. V16	R. V16	wild type
VaIRS017	F. V17	R. V17	wild type
VaIRS018	F. V18	R. V18	wild type
VaIRS019	F. V19	R. V19	wild type
VAIRS020	F. V2	R. V2	VaIRS019
VaIRS021	F. V3	R. V3	VaIRS019
VaIRS022	F. V4	R. V4	VaIRS019
VaIRS023	F. V5	R. V5	VaIRS019
VaIRS024	F. V6	R. V6	VaIRS019
VaIRS025	F. V7	R. V7	VaIRS019
VaIRS026	F. VB	R. V8	VaIRS019
VaIRS027	F. V9	R. V9	VaIRS019
VaIRS028	F. V10	R. V10	VaIRS019
VaIRS029	F. V11	R. V11	VaIRS019
VaIRS030	F. V12	R. V12	VaIRS019
VaIRS031	F. V13	R. V13	VaIRS019
VaIRS032	F. V14	R. V14	VaIRS019
VaIRS033	F. V15	R. V15	VaIRS019
VaIRS034	F. V16	R. V16	VaIRS019

	forward primer	reverse primer	Template
ValRS035	F. V17	R. V17	ValRS019
ValRS036	F. V18	R. V18	ValRS019
ValRS037	F. V16	R. V16	ValRS046
ValRS038	F. V17	R. V17	ValRS046
ValRS039	F. V18	R. V18	ValRS046
ValRS040	F. V16	R. V16	ValRS047
ValRS041	F. V17	R. V17	ValRS047
ValRS042	F. V18	R. V18	ValRS047
ValRS043	F. V16	R. V16	ValRS048
ValRS044	F. V17	R. V17	ValRS048
ValRS045	F. V18	R. V18	ValRS048
ValRS046	F. V46	R. V46	wt
ValRS047	F. V47	R. V47	wt
ValRS048	F. V48	R. V48	wt
ValRS13-01	F. V13-01	R. V13-01	ValRS13
ValRS13-02	F. V13-02	R. V13-02	ValRS13
ValRS13-03	F. V13-03	R. V13-03	ValRS13
ValRS13-04	F. V13-04	R. V13-04	ValRS13
ValRS13-05	F. V13-05	R. V13-05	ValRS13
ValRS13-06	F. V13-06	R. V13-06	ValRS13
ValRS13-07	F. V13-07	R. V13-07	ValRS13
ValRS13-08	F. V13-08	R. V13-08	ValRS13
ValRS13-09	F. V13-09	R. V13-09	ValRS13
ValRS13-10	F. V13-10	R. V13-10	ValRS13
ValRS13-11	F. V13-11	R. V13-11	ValRS13
ValRS13-12	F.V13-12	R. V13-12	ValRS13
ValRS13-13	F. V13-13	R. V13-13	ValRS13
ValRS13-14	F. V13-14	R. V13-14	ValRS13
VaoRS13-15	F. V13-15	R. V13-15	ValRS13
VaoRS13-16	F. V13-16	R. V13-16	ValRS13

<Small scale expression of wild-type and mutant ValRSs>

[0207] Next, a resulting mutant plasmid was introduced into *E. coli*, and the mutant protein was expressed. First, *E. coli* BL21 strain transformed with the mutant plasmid and pREP4 (Invitrogen, V004-50) was cultured at 37°C in 4 mL of LB medium containing kanamycin and ampicillin. Then, when the OD value at 600 nm reached 0.4 to 0.8, IPTG was added to a final concentration of 0.5 mM. After further culturing at 37°C for 4 hours, the bacterial pellets were collected using a centrifuge.

<Small scale purification of wild-type and mutant ValRSs>

[0208] Next, the resulting bacterial pellets were disrupted, and the mutant protein of interest was purified from the supernatant. Specifically, the bacterial pellets as described above were suspended in 600 µL of CHAPS solution (0.5% CHAPS (DOJINDO: 349-04722), 50% TBS (TaKaRa, T903)) and mixed with 6 µL of 30 U/µl rLysozyme (Novagen, 71110-3) followed by incubation at room temperature for 10 minutes. The reaction was further mixed with 2 µL of 2.5 U/µL benzonase nuclease (Novagen, 70746-3) followed by incubation at room temperature for 20 minutes, and an insoluble fraction was separated by centrifugation. The mutant protein was then purified from the resulting supernatant using QIAGEN Ni-NTA spin column kit (Qiagen, 31314) according to the product manual. Finally, excess imidazole was removed using a desalting column, PD miniTrap G-25 (GE Healthcare, 28-9180-07) according to the product manual.

<Aminoacylation reaction with N-methylvaline using wild-type and mutant ValRSs>

[Synthesis of *E. coli* tRNA^{Val} by *in vitro* transcription reaction]

[0209] *E. coli* tRNA (R-tRNA^{Val}2A (SEQ ID NO: 118)) was synthesized from a template DNA (D-tRNA^{Val}2A (SEQ ID NO: 117)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) in the presence of 7.5 mM GMP, and purified using RNeasy Mini kit (Qiagen).

[0210]

D-tRNA^{Val}2A (SEQ ID NO: 117)

tRNA^{Val}2A DNA sequence:

GGCGTAATACGACTCACTATAGCGTCCGTAGCTCAGTTGGTTAGAGCACCCACCTTGA
CATGGTGGGGGTCGGTGGTTCGAGTCCACTCGGACGCACCA

R-tRNA^{Val}2A (SEQ ID NO: 118)

tRNA^{Val}2A RNA sequence:

GCGUCCGUAGCUCAGUUGGUUAGAGCACCCACCUUGACAUGGUGGGGGUCGGUGG
UUCGAGUCCACUCGGACGCACCA

<Aminoacylation reaction>

[0211] For the aminoacylation reaction, the solution containing 40 μM transcribed tRNA, 10 mM HEPES-K (pH 7.6), and 10 mM KCl solution was heated at 95°C for 2 minutes and then left at room temperature for 5 minutes or more to refold the tRNA. This tRNA solution was added to a final concentration of 10 μM to an acylation buffer (in final concentrations of 50 mM HEPES-K [pH 7.6], 2 mM ATP, 100 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 2 mM spermidine, 0.1 mg/mL Bovine Serum Albumin), mixed with wild-type or mutant ValRS (final concentration 0.2 μM -1 μM) and N-methylvaline (final concentration 5 mM), and incubated at 37°C for 10 minutes. Four volumes of a loading buffer (90 mM sodium acetate [pH 5.2], 10 mM EDTA, 95%(w/w) formamide, 0.001%(w/v) xylene cyanol) were added to the reaction solution and analyzed with acidic PAGE containing 6 M urea, and aminoacylation activity was detected by separating unreacted tRNA and aminoacyl-tRNA. RNA was stained with SYBR Gold (Life Technologies) and detected with LAS4000 (GE Healthcare) (Figure 4).

[0212] As a result, tRNA acylated with N-methylvaline was observed when mutant 13 was used. It was demonstrated that mutant 13 had increased activity for aminoacylation with N-methylvaline compared to wild-type ValRS (Figure 4, lane 2 vs 10).

<Translational introduction of N-methylvaline using wild-type and mutant ValRSs>

[0213] Template mRNAs for translation (R-V, R-V2, and R-V3 (SEQ ID NOs: 120, 122, and 124, respectively)) were synthesized from template DNAs (D-V, D-V2, and D-V3 (SEQ ID NOs: 119, 121, and 123, respectively)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) and purified using RNeasy Mini kit (Qiagen).

[0214]

D-V (SEQ ID NO: 119)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTAACTTTAACAAGGAGAAAAACATGCGTGTC
CGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-V (SEQ ID NO: 120)

RNA sequence:

GGGUUAACUUUAACAAGGAGAAAAACAUGCGUGUCCGUGACUACAAGGACGACG
ACGACAAGUAAGCUUCG

D-V2 (SEQ ID NO: 121)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTAACTTTAACAAGGAGAAAAACATGCGTGTC

GTCCGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-V2 (SEQ ID NO: 122)

RNA sequence:

GGGUUAAACUUUAAACAAGGAGAAAAACAUGCGUGUCGUCGUGACUACAAGGACG
ACGACGACAAGUAAGCUUCG

D-V3 (SEQ ID NO: 123)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTAACTTTAACAAGGAGAAAAACATGCGTGTC
GTCGTCCGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-V3 (SEQ ID NO: 124)

RNA sequence:

GGGUUAAACUUUAAACAAGGAGAAAAACAUGCGUGUCGUGUCGUGACUACAAGG
ACGACGACGACAAGUAAGCUUCG

<Cell-free translation system>

[0215] In order to confirm translational introduction of N-methylvaline, a desired polypeptide containing N-methylvaline was ribosomally synthesized by adding an N-methylvaline and a mutant ValRS to a cell-free translation system. The translation system used was PURE system, a reconstituted cell-free protein synthesis system from *E. coli*. Specifically, wild-type or mutant ValRS and N-methylvaline were added to a solution containing a basic cell-free translation solution (1 mM GTP, 1 mM ATP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 9 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol, 1.5 mg/ml *E. coli* MRE600 (RNase negative)-derived tRNA (Roche), 0.1 mM 10-HCO-H4 folate, 4 µg/ml creatine kinase, 3 µg/ml myokinase, 2 unit/ml inorganic pyrophosphatase, 1.1 µg/ml nucleoside diphosphate kinase, 0.6 µM methionyl-tRNA transformylase, 0.26 µM EF-G, 0.24 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 40 µM EF-Tu, 84 µM EF-Ts, 1.2 µM ribosome, 0.03 µM ArgRS, 0.13 µM AspRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.02 µM TyrRS (wherein the proteins prepared by the inventors were essentially prepared as His-tagged proteins)), and 1 µM template mRNA, each 250 µM arginine, aspartic acid, lysine, methionine, and tyrosine, and left at 37°C for 1 hour to ribosomally synthesize the peptide.

<Detection by mass spectroscopy>

[0216] Mass spectroscopy was performed using MALDI-TOF MS for detecting peptides into which N-methylvaline is ribosomally introduced. Specifically, a solution containing 1 µM template mRNA (R-V (SEQ ID NO: 120)) and arginine, lysine, methionine, tyrosine, and aspartic acid (each final concentration 250 µM) added to the cell-free translation system as described above was prepared. ValRS (final concentration 0.1 µM-1 µM) and N-methylvaline

(final concentration 5 mM) were added to the solution and incubated at 37°C for 60 minutes. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS. Translation products were identified with MALDI-TOF MS spectrometry using α -cyano-4-hydroxycinnamic acid as a matrix.

[0217] As a result of the translation using wild-type ValRS (SEQ ID NO: 24), a peak corresponding to the peptide sequence P-V1 introduced with valine contaminated in the cell-free translation system ((Figure 5(a), Peak V1, m/z: [H+M]⁺ = 1583.6) was observed as main product. Similar experiments were performed using different mutant ValRSs. When ValRS04 (SEQ ID NO: 3) or ValRS13 (SEQ ID NO: 4) was used, a peak corresponding to the peptide sequence P-MeV1 with N-methylvaline introduced ((Figure 5(b), Peak MeV1, m/z: [H+M]⁺ = 1597.5, Figure 5(c), Peak MeV2, m/z: [H+M]⁺ = 1597.5) was observed as main product. This demonstrated, also from a view point of translation reaction that ValRS04 and ValRS13 have increased activity to N-methylvaline compared to wild-type ValRS. Although peaks corresponding to the peptide P-V1 introduced with valine contaminated in the translation system (Figure 5(b) Peak V2, (c) Peak V3) were also observed at the same time, the intensity of the peak in using ValRS13 was weaker than that in using ValRS04. Therefore, it was suggested that ValRS13 has a higher activity to N-methylvaline (Figure 5(b), (c)).

[0218]

Peptide sequence P-V1 (SEQ ID NO: 125)
formylMetArgValArgAspTyrLysAspAspAspLys

Peptide sequence P-MeV1 (SEQ ID NO: 125)
formylMetArg[MeVal]ArgAspTyrLysAspAspAspLys

MALDI-TOF MS:

Calc. m/z: [H+M]⁺ = 1583.7 (the peptide corresponding to the sequence P-V1)

Calc. m/z: [H+M]⁺ = 1597.7 (the peptide corresponding to the sequence P-MeV1)

<Production of mutant ValRS13-11 having further increased aminoacylation activity to N-methylvaline>

[0219] Improvement in activity to N-methylvaline was aimed by further introducing mutations into ValRS 13 as described above. Plasmids into which the mutations of interest are introduced were prepared as mentioned above, expressed in *E. coli*, and purified to prepare mutant ValRSs (Table 3). As a result of screening by performing aminoacylation reaction and translational synthesis, ValRS13-11 (SEQ ID NO: 5) exhibited higher activity to N-methylvaline compared to ValRS13.

[0220] When aminoacylation reactions with N-methylvaline using wild-type ValRS, ValRS13,

and ValRS13-11 were investigated, it was observed that the amount of aminoacyl-tRNA synthesized using ValRS13-11 was more than that synthesized using ValRS13 (Figure 6, lane 8 vs lane 9 and lane 12 vs lane 13). Particularly, the amounts of aminoacyl-tRNA synthesized using ValRS13 and ValRS13-11 with N-methylvaline at the low concentration of 1.25 mM were greatly different, demonstrating that ValRS13-11 had a high activity.

[0221] Next, mass spectroscopy was performed using MALDI-TOF MS for confirming translational synthesis of peptides having N-methylvaline using ValRS13 and ValRS13-11. Specifically, a solution containing 1 μ M template mRNA (R-V (SEQ ID NO: 120)) and arginine, lysine, methionine, tyrosine, and aspartic acid (each final concentration 250 μ M) added to the cell-free translation system as described above was prepared. ValRS (final concentration 4 μ M) and N-methylvaline (final concentration 5 mM) were added to the solution and incubated at 37°C for 60 minutes. Template mRNAs encoding peptide sequences containing two consecutive or three consecutive N-methylvalines (R-V2 (SEQ ID NO: 122), R-V3 (SEQ ID NO: 124)) were used to perform similar experiments. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS.

[0222] The intended synthesis of the peptide sequence P-MeV1 containing N-methylvaline using ValRS13 and ValRS13-11 was confirmed when the template mRNA containing one N-methylvaline (R-V (SEQ ID NO: 120)) was used (Figure 7(a) Peak MeV1, (b) Peak MeV2). A peak corresponding to the peptide sequence P-V1 introduced with valine contaminated in the cell-free translation system was observed weakly (Figure 7(a) Peak VI, (b) Peak V2).

[0223] Next, it was confirmed that the peptide sequence P-MeV2 containing two N-methylvaline residues was synthesized as main product by both the mutant ValRSs when a template mRNA containing two consecutive N-methylvalines (R-V2 (SEQ ID NO: 122)) was used (Figure 7(c) Peak MeV3, (d) Peak MeV5). Meanwhile, the peptide sequence P-MeV4 containing one N-methylvaline residue and one valine residue (Figure 7(c) Peak MeV4, (d) Peak MeV6) was observed, but the peak intensity was suppressed when using ValRS13-11 compared to ValRS13.

[0224] Finally, template mRNA containing three consecutive N-methylvalines (R-V3 (SEQ ID NO: 124)) was used to perform translation experiments. When ValRS13 was added, the synthesis of the target peptide sequence P-MeV3 containing three N-methylvaline residues was observed (Figure 7(e), Peak MeV7), but the peptide sequence P-MeV5 comprising two N-methylvaline residues and one valine residue was produced as main product (Figure 7(e), Peak MeV8). The peptide sequence P-MeV6 comprising one N-methylvaline residue and two valine residues was also observed (Figure 7(e), Peak MeV9). Meanwhile, when ValRS13-11 was added, it was observed that the target peptide sequence P-MeV3 containing three N-methylvaline residues was ribosomally synthesized as main product (Figure 7(f), Peak MeV10). The peptide sequences P-MeV5 and P-MeV6 containing valine were also observed, but the peak intensity of these peptide sequences was suppressed compared to that obtained when using ValRS13 (Figure 7(f), Peak MeV11 and MeV12). These results indicate that ValRS13-11 has increased aminoacylation activity to N-methylvaline compared to ValRS13, leading to the

increase of the ribosomally synthesized amount of the target peptide containing N-methylvaline.

[0225]

Peptide sequence P-MeV2 (SEQ ID NO: 126)

formylMetArg[MeVal][MeVal]ArgAspTyrLysAspAspAspAspLys

Peptide sequence P-MeV4 (SEQ ID NO: 126)

formylMetArg[MeVal]ValArgAspTyrLysAspAspAspAspLys or

formylMetArgVal[MeVal]ArgAspTyrLysAspAspAspAspLys

Peptide sequence P-MeV3 (SEQ ID NO: 127)

formylMetArg[MeVal][MeVal][MeVal]ArgAspTyrLysAspAspAspAspLys

Peptide sequence P-MeV5 (SEQ ID NO: 127)

formylMetArg[MeVal][MeVal]ValArgAspTyrLysAspAspAspAspLys or

formylMetArg[MeVal]Val[MeVal]ArgAspTyrLysAspAspAspAspLys or

formylMetArgVal[MeVal][MeVal]ArgAspTyrLysAspAspAspAspLys

Peptide sequence P-MeV6 (SEQ ID NO: 127)

formylMetArg[MeVal]ValValArgAspTyrLysAspAspAspAspLys or

formylMetArgVal[MeVal]ValArgAspTyrLysAspAspAspAspLys or

formylMetArgValVal[MeVal]ArgAspTyrLysAspAspAspAspLys

MALDI-TOF MS:

[0226]

Calc. m/z: $[H+M]^+ = 1710.8$ (the peptide corresponding to the sequence P-MeV2)

Calc. m/z: $[H+M]^+ = 1696.8$ (the peptide corresponding to the sequence P-MeV4)

Calc. m/z: $[H+M]^+ = 1823.9$ (the peptide corresponding to the sequence P-MeV3)

Calc. m/z: $[H+M]^+ = 1809.9$ (the peptide corresponding to the sequence P-MeV5)

Calc. m/z: $[H+M]^+ = 1795.9$ (the peptide corresponding to the sequence P-MeV6)

Example 3: Development of N-methylserine-accepting ARS**<Preparation of plasmids for wild-type and mutant SerRSs>**

[0227] The mutant SerRS plasmids (having His-tag (6 × His) at the N-terminus) listed in Table 5 were constructed by introducing site-directed mutations using PCR into a plasmid comprising the ORF sequence (SEQ ID NOs: 25, 26) of wild-type SerRS gene of *E. coli* (PQE-32(2) 2_wtSERRS), which was used as starting material. Specifically, 2.5 μL of 10 ng/μL template, 12.5 μL of 2 × KOD Fx buffer (TOYOBO, KFX-101), 0.75 μL of 10 μM forward primer, 0.75 μL of 10 μM reverse primer, 5 μL of 2 mM dNTP, 0.5 μL of KOD FX (TOYOBO, KFX-101), and 3 μL of H₂O were mixed together. Next, the resulting reaction solution was heated at 94°C for 2 minutes and then subjected to 10 cycles, each consisting of heating at 98°C for 10 seconds and heating at 68°C for 7 minutes, to amplify the mutant gene. The combinations of the template plasmid, forward primer, and reverse primer used are listed in Table 6. Each sequence of primers "F.S2" through "F.S8", "F.S15" through "F.S23", and "F.S33" through "F.S38" corresponds to SEQ ID NOs: 128-149 in ascending order. Each sequence of primers "R.S2" through "R.S8", "R.S15" through "R.S23", and "R.S33" through "R.S38" corresponds to SEQ ID NOs: 150-171 in ascending order. 0.5 μL of 10 U/μL *DpnI* was then added to the PCR reaction solution and further incubated at 37°C for 1.5 hours to digest the template DNA, and the resulting mutant DNA was purified. *E. coli* XL-1 Blue strain (STRATAGENE, 200236) was then co-transformed with the resulting mutant gene DNA and pREP4 (Invitrogen, V004-50) encoding the *lacI* gene. The transformants were seeded onto agar containing ampicillin and kanamycin. The plasmids of interest were purified from the resulting clones. The mutations were confirmed to be introduced into the plasmids.

[0228] For plasmid construction requiring multistep mutation introduction, the procedure as described above was repeated to obtain plasmids into which mutations of interest were introduced. The combinations of the primers and template for such plasmid construction are listed in Table 6.

[Table 5]

Name	position 237	position 239	position 389
SerRS01(wt)	T	E	N
SerRS02	A	E	N
SerRS03	S		
SerRS04	G	A	N
SerRS05	T		
SerRS06	T		
SerRS07	T	S	A
SerRS08	T	E	
SerRS09	A	E	A
SerRS10	S		
SerRS11	G		

SerRS12		A		
SerRS13	T	V		
SerRS14		S		
SerRS15		A	N	
SerRS16	A	V		
SerRS17		S		
SerRS18		A		
SerRS19	S	V		
SerRS20		S		
SerRS21		A		
SerRS22	G	V		
SerRS23		S		
SerRS24		A		A
SerRS25	A	V		
SerRS26		S		
SerRS27		A		
SerRS28	S	V		
SerRS29		S		
SerRS30		A		
SerRS31	G	V		
SerRS32		S		
SerRS33	E291A			
SerRS34	K289A			
SerRS35	T	G	N	
SerRS36	T	D	N	
SerRS37	S	G	N	
SerRS38	S	D	N	

[Table 6]

	forward primer	revers primer	Template
SerRS001			
SerRS002	F. S2	R. S2	wt
SerRS003	F. S3	R. S3	wt
SerRS004	F. S4	R. S4	wt
SerRS005	F. S5	R. S5	wt
SerRS006	F. S6	R. S6	wt
SerRS007	F. S7	R. S7	wt
SerRS008	F. S8	R. S8	wt
SerRS009	F. S2	R. S2	SerRS008
SerRS010	F. S3	R. S3	SerRS008
SerRS011	F. S4	R. S4	SerRS008
SerRS012	F. S5	R. S5	SerRS008
SerRS013	F. S6	R. S6	SerRS008
SerRS014	F. S7	R. S7	SerRS008
SerRS015	F. S15	R. S15	wt
SerRS016	F. S16	R. S16	wt

	forward primer	revers primer	Template
SerRS017	F. S17	R. S17	wt
SerRS018	F. S18	R. S18	wt
SerRS019	F. S19	R. S19	wt
SerRS020	F. S20	R. S20	wt
SerRS021	F. S21	R. S21	wt
SerRS022	F. S22	R. S22	wt
SerRS023	F. S23	R. S23	wt
SerRS024	F. S15	R. S15	SerRS008
SerRS025	F. S16	R. S16	SerRS008
SerRS026	F. S17	R. S17	SerRS008
SerRS027	F. S18	R. S18	SerRS008
SerRS02B	F. S19	R. S19	SerRS008
SerRS029	F. S20	R. S20	SerRS008
SerRS030	F. S21	R. S21	SerRS008
SerRS031	F. S22	R. S22	SerRS00B
SerRS032	F. S23	R. S23	SerRS008
SerRS033	F. S33	R. S33	wt
SerRS034	F. S34	R. S34	wt
SerRS035	F. S35	R. S35	wt
SerRS036	F. S36	R. S36	wt
SerRS037 F. S37 R. S37 wt			
SerRS038 F. S38 R. S38 wt			

<Small scale expression of wild-type and mutant SerRSs>

[0229] Next, the resulting mutant plasmid was introduced into *E. coli* to express the mutant protein. First, *E. coli* BL21 strain transformed with the mutant plasmid and pREP4 (Invitrogen, V004-50) was cultured at 37°C in 4 mL of LB medium containing kanamycin and ampicillin. Then, when the OD value at 600 nm reached 0.4 to 0.8, IPTG was added to a final concentration of 0.5 mM. After further culturing at 37°C for 4 hours, the bacterial pellets were collected with a centrifuge.

<Small scale purification of wild-type and mutant SerRSs>

[0230] Next, the resulting bacterial pellets were disrupted, and the mutant protein of interest was purified from the supernatant. Specifically, the bacterial pellets as described above were suspended in 600 μ L of CHAPS solution (0.5% CHAPS (DOJINDO: 349-04722), 50% TBS (TaKaRa, T903)) and mixed with 6 μ L of 30 U/ μ l rLysozyme (Novagen, 71110-3) followed by incubation at room temperature for 10 minutes. The reaction was further mixed with 2 μ L of 2.5 U/ μ L benzonase nuclease (Novagen, 70746-3) followed by incubation at room temperature for 20 minutes, and an insoluble fraction was separated by centrifugation. Then, the mutant protein was purified from the resulting supernatant using QIAGEN Ni-NTA spin column kit (Qiagen, 31314) according to the product manual. Finally, excess imidazole was removed using a desalting column, PD miniTrap G-25 (GE Healthcare, 28-9180-07) according to the product manual.

<Aminoacylation reaction with N-methylserine using wild-type and mutant SerRSs>

[Synthesis of E. coli tRNA^{Ser} by *in vitro* transcription reaction]

[0231] E. coli tRNA (R-tRNA^{Ser3} (SEQ ID NO: 173)) was synthesized from a template DNA (D-tRNA^{Ser3} (SEQ ID NO: 172)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) in the presence of 7.5 mM GMP, and purified using RNeasy Mini kit (Qiagen).

[0232]

D-tRNA^{Ser3} (SEQ ID NO: 172)

tRNA^{Ser3} DNA sequence:

GGCGTAATACGACTCACTATAGGTGAGGTGGCCGAGAGGCTGAAGGCGCTCCCCTG
CTAAGGGAGTATGCGGTCAAAGCTGCATCCGGGGTTCGAATCCCCGCCTCACCGC
CA

R-tRNA^{Ser3} (SEQ ID NO: 173)

tRNA^{Ser3} RNA sequence:

GGUGAGGUGGCCGAGAGGCUGAAGGCGCUCCCCUGCUAAGGGAGUAUGC GGUCA
AAAGCUGCAUCCGGGGUUCGAAUCCCCGCCUCACCGCCA

<Aminoacylation reaction>

[0233] For the aminoacylation reaction, the solution containing 40 μ M transcribed tRNA, 10 mM HEPES-K (pH 7.6), and 10 mM KCl solution was heated at 95°C for 2 minutes and then left at room temperature for 5 minutes or more to refold the tRNA. This tRNA solution was added to a final concentration of 10 μ M to an acylation buffer (final concentration 50 mM

HEPES-K [pH 7.6], 2 mM ATP, 100 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 2 mM spermidine, 0.1 mg/mL Bovine Serum Albumin), mixed with wild-type or mutant SerRS (final concentration 0.1 μ M-2 μ M) and N-methylserine (final concentration 1 mM), and incubated at 37°C for 10 minutes. Four volumes of a loading buffer (90 mM sodium acetate [pH 5.2], 10 mM EDTA, 95%(w/w) formamide, 0.001%(w/v) xylene cyanol) were added to the reaction solution and analyzed with acidic PAGE containing 6 M urea, and aminoacylation activity was detected by separating unreacted tRNA and aminoacyl-tRNA. RNA was stained with SYBR Gold (Life Technologies) and detected with LAS4000 (GE Healthcare) (Figure 8).

[0234] As a result, tRNA acylated with N-methylserine was observed when the mutants 03 (SEQ ID NO: 6), 35 (SEQ ID NO: 8), and 37 (SEQ ID NO: 9) were used. It is suggested that these mutants had increased activity for aminoacylation with N-methylserine compared to wild-type SerRS (Figure 8, lanes 3, 25, 27).

<Translational introduction of N-methylserine using wild-type and mutant SerRSs>

[0235] Template mRNA (R-S (SEQ ID NO: 175)) was synthesized from a template DNA (D-S (SEQ ID NO: 174)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) and purified using RNeasy Mini kit (Qiagen).

[0236]

D-S (CT21) (SEQ ID NO: 174)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCGTTCC
CGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-S (SEQ ID NO: 175)

RNA sequence:

GGGUUAACUUUAACAAGGAGAAAAACAUGCGUCCCGUGACUACAAGGACGACG
ACGACAAGUAAGCUUCG

<Cell-free translation system>

[0237] In order to confirm translational introduction of N-methylserine, a desired polypeptide containing N-methylserine was ribosomally synthesized by adding N-methylserine and SerRS to a cell-free translation system. The translation system used was PURE system, a reconstituted cell-free protein synthesis system from *E. coli*. Specifically, wild-type or mutant SerRS and N-methylserine were added to a solution containing a basic cell-free translation solution (1 mM GTP, 1 mM ATP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 9 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol, 1.5

mg/ml *E. coli* MRE600 (RNase negative)-derived tRNA (Roche), 0.1 mM 10-HCO-H4 folate, 4 µg/ml creatine kinase, 3 µg/ml myokinase, 2 unit/ml inorganic pyrophosphatase, 1.1 µg/ml nucleoside diphosphate kinase, 0.6 µM methionyl-tRNA transformylase, 0.26 µM EF-G, 0.24 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 40 µM EF-Tu, 84 µM EF-Ts, 1.2 µM ribosome, 0.03 µM ArgRS, 0.13 µM AspRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.02 µM TyrRS (wherein the proteins prepared by the inventors were essentially prepared as His-tagged proteins), and 1 µM template mRNA, each 250 µM arginine, aspartic acid, lysine, methionine, and tyrosine, and left at 37°C for 1 hour to ribosomally synthesize the peptide.

<Detection by mass spectroscopy>

[0238] Mass spectroscopy was performed using MALDI-TOF MS for detecting peptides into which N-methylserine is ribosomally introduced. Specifically, a solution containing 1 µM template mRNA (R-S (SEQ ID NO: 175)) and arginine, lysine, methionine, tyrosine, and aspartic acid (each final concentration 250 µM) added to the cell-free translation system as described above was prepared. SerRS (final concentration 0.1-2 µM) and N-methylserine (final concentration 5 mM) were added to the solution and incubated at 37°C for 60 minutes. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS. Translation products were identified with MALDI-TOF MS spectrometry using α-cyano-4-hydroxycinnamic acid as a matrix.

[0239] As a result of the translation using wild-type SerRS (SEQ ID NO: 25)), a peak corresponding to the target peptide peak P-CT21MeSer containing N-methylserine (Figure 9(a), Peak MeS1 m/z: [H+M]⁺ = 1585.6) was observed, but the peak corresponding to Ser-containing peptide P-CT21Ser probably derived from Ser contaminated in the translation system in trace amounts (Figure 9(a), Peak S1 m/z: [H+M]⁺ = 1571.6) was observed as main product (Figure 9(a)). On the other hand, when similar experiments were performed using the modified SerRS with mutations introduced (Ser03 (SEQ ID NO: 6), 05 (SEQ ID NO: 7)), the peak corresponding to P-CT21Ser (Figure 9(b) Peak S2, (c) Peak S3) was observed similarly but as a side product, and it was revealed that the main product was P-CT21MeSer (Figure 9(b) Peak MeS2, (c) Peak MeS3). Particularly, when SerRS35 and 37 were used, it was demonstrated that the peak corresponding to P-CT21Ser was not observed and CT21MeSer was synthesized with high purity (Figure 9(d) Peak MeS4, (e) Peak MeS5). Consequently, these modified SerRSs were demonstrated to have increased activity to MeSer compared to wild-type SerRS.

[0240]

Peptide sequence P-CT21Ser (SEQ ID NO: 176)
formylMetArgSerArgAspTyrLysAspAspAspAspLys

Peptide sequence P-CT21MeSer (SEQ ID NO: 176)
formylMetArg[MeSer]ArgAspTyrLysAspAspAspAspLys

MALDI-TOF MS:**[0241]**

Calc. m/z: [H+M]⁺ = 1571.7 (the peptide corresponding to the sequence P-CT21Ser)

Calc. m/z: [H+M]⁺ = 1585.7 (the peptide corresponding to the sequence P-CT21MeSer)

Example 4: Development of N-methylthreonine-accepting ARS**<Preparation of wild-type and mutant ThrRS proteins>**

[0242] Expression vectors having a polyhistidine sequence at the N-terminus and comprising mutations listed in Table 7 were constructed. Subsequently, an expression strain was transformed with the vectors, and the mutant proteins of interest were purified with a nickel column from supernatants obtained by disrupting cells.

[Table 7]

Name	position 332	position 334	position 385	position 462	position 484	position 511
ThrRS01(wt)	M	C	H	Y	Q	H
ThrRS02	A	C	H	Y	Q	H
ThrRS03	G	C	H	Y	Q	H
ThrRS04	M	A	H	Y	Q	H
ThrRS05	M	G	H	Y	Q	H
ThrRS06	M	C	A	Y	Q	H
ThrRS07	M	C	H	A	Q	H
ThrRS08	M	C	H	L	Q	H
ThrRS09	M	C	H	V	Q	H
ThrRS10	M	C	H	F	Q	H
ThrRS11	M	C	H	Y	A	H
ThrRS12	M	C	H	Y	N	H
ThrRS13	M	C	H	Y	Q	A
ThrRS14	M	C	H	Y	Q	G

<Translational introduction of N-methylthreonine using wild-type and mutant ThrRSs>

[0243] A template mRNA (R-T (SEQ ID NO: 178)) was synthesized from a template DNA (D-T (SEQ ID NO: 177)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) and purified using RNeasy Mini kit (Qiagen).

[0244]

D-T (3lib15#09) (SEQ ID NO: 177)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATGAAGG
CTGGTCCGGGTTTTATGACTAAGAGTGGTAGTGGTAGTTAAGCTTCG

R-T (SEQ ID NO: 178)

RNA sequence:

GGGUU AACUUUAAGAAGGAGAUAUACAUAUGAAGGCUGGUCCGGGUUUUAUGAC
UAAGAGUGGUAGUGGUAGUUAAGCUUCG

<Cell-free translation system>

[0245] In order to confirm translational introduction of N-methylthreonine, a desired polypeptide containing N-methylthreonine was ribosomally synthesized by adding N-methylthreonine and a mutant ThrRS to a cell-free translation system. The translation system used was PURE system, a reconstituted cell-free protein synthesis system from *E. coli*. Specifically, wild-type or mutant ThrRS and N-methylthreonine were added to a solution containing a basic cell-free translation solution (1 mM GTP, 1 mM ATP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 9 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol, 1.5 mg/ml *E. coli* MRE600 (RNase negative)-derived tRNA (Roche), 0.1 mM 10-HCO-H4 folate, 4 µg/ml creatine kinase, 3 µg/ml myokinase, 2 unit/ml inorganic pyrophosphatase, 1.1 µg/ml nucleoside diphosphate kinase, 0.6 µM methionyl-tRNA transformylase, 0.26 µM EF-G, 0.24 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 40 µM EF-Tu, 93 µM EF-Ts, 1.2 µM ribosome, 2.73 µM AlaRS, 0.13 µM AspRS, 0.09 µM GlyRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.68 µM PheRS, 0.16 µM ProRS, 0.25 µM SerRS (wherein the proteins prepared by the inventors were essentially prepared as His-tagged proteins)), and 1 µM template mRNA, each 250 µM glycine, proline, alanine, phenylalanine, lysine, methionine, and serine, and left at 37°C for 1 hour to ribosomally synthesize the peptide.

<Detection by mass spectroscopy>

[0246] Mass spectroscopy was performed using MALDI-TOF MS for detecting peptides into which N-methylthreonine is ribosomally introduced. Specifically, a solution containing 1 μ M template mRNA (R-T (SEQ ID NO: 178)) and glycine, proline, alanine, phenylalanine, lysine, methionine, and serine (each final concentration 250 μ M) added to the cell-free translation system as described above was prepared. ThrRS (final concentration 2 μ M) and N-methylthreonine (final concentration 5 mM) were added to the solution and incubated at 37°C for 60 minutes. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS. Translation products were identified with MALDI-TOF MS spectrometry using α -cyano-4-hydroxycinnamic acid as a matrix.

[0247] As a result of the translation using wild-type ThrRS (SEQ ID NO: 29), a peak corresponding to the peptide peak P-3lib15MeThr of interest containing N-methylthreonine ((Figure 10(a), Peak MeT1) and a peak corresponding to a potassium salt of the peptide (Figure 10(a), Peak MeT2) were observed as main products, but at the same time, a peak corresponding to the peptide P-3lib15Thr derived from Thr contaminated in the translation system in trace amounts (Figure 10(a), PeakT1) and a peak corresponding to a potassium salt of the peptide (Figure 10(a), PeakT2) were also observed, revealing that the purity of the translation products was adversely affected (Figure 10(a)). On the other hand, when experiments were performed using the modified ThrRS 03 (SEQ ID NO: 10) and the modified ThrRS 14 (SEQ ID NO: 11) with mutations introduced, the peaks derived from P-CT21MeThr (Peak MeT3-6, Figure 10) were similarly observed, and the peak corresponding to the peptide sequence P-3lib15Thr was only slightly observed (Figure 10(b), (c)). In other words, it was demonstrated that the peptide with MeThr introduced was synthesized with higher purity using the modified ThrRSs compared to using wild-type ThrRS, and it was suggested that the modified ThrRS 03 and 14 are ARSs that can introduce MeThr into peptides more efficiently compared to wild-type ThrRS.

[0248]

Peptide sequence P-3lib15Thr (SEQ ID NO: 179)
 formylMetLysAlaGlyProGlyPheMetThrLysSerGlySerGlySer

Peptide sequence P-3lib15MeThr (SEQ ID NO: 179)
 formylMetLysAlaGlyProGlyPheMet[MeThr]LysSerGlySerGlySer

MALDI-TOF MS:

[0249]

Calc. m/z: [H+M]⁺ = 1470.7, [K+M]⁺ = 1508.8 (the peptide corresponding to the sequence P-3lib15Thr)

Calc. m/z: [H+M]⁺ = 1470.7, [K+M]⁺ = 1508.8 (the peptide corresponding to the sequence P-3lib15Thr)

Calc. m/z: [H+M]⁺ = 1484.7, [K+M]⁺ = 1522.8 (the peptide corresponding to the sequence P-3lib15MeThr)

Example 5: Development of N-methyltryptophan-accepting ARS

<Preparation of wild-type and mutant TrpRS proteins>

[0250] Expression vectors having a polyhistidine sequence at the N-terminus and containing mutations listed in Table 8 were constructed. Then, an expression strain was transformed with the vectors, and the mutant proteins of interest were purified with a nickel column from supernatants obtained by disrupting cells.

[Table 8]

Name	position 128	position 132	position 150	position 153
TrpRS01(wt)	Y	M	Q	H
TrpRS02	Y	M	Q	A
TrpRS03	Y	M	A	H
TrpRS04	Y	M	A	A
TrpRS05	Y	A	Q	H
TrpRS06	Y	A	Q	A
TrpRS07	Y	A	A	H
TrpRS08	Y	A	A	A
TrpRS09	A	M	Q	H
TrpRS10	Y	M	S	H
TrpRS11	Y	M	G	H
TrpRS12	Y	M	Q	G
TrpRS13	Y	M	G	A
TrpRS14	Y	M	A	G
TrpRS15	Y	M	G	G
TrpRS16	Y	G	A	A
TrpRS17	Y	I	A	A
TrpRS18	Y	V	A	A
TrpRS19	Y	Q	A	A
TrpRS20	Y	L	A	A

<Translational introduction of N-methyltryptophan using wild-type and mutant TrpRSs>

[0251] A template mRNA(R-W (SEQ ID NO: 197)) was synthesized from a template DNA (D-W (SEQ ID NO: 196)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) and purified using RNeasy Mini kit (Qiagen).

[0252]

D-W (CT29) (SEQ ID NO: 196)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCGTTGG
CGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-W (SEQ ID NO: 197)

RNA sequence:

GGGUUAAUUUACAAGGAGAAAAACAUGCGUUGGCGUGACUACAAGGACGACG
ACGACAAGUAAGCUUCG

<Cell-free translation system>

[0253] In order to confirm translational introduction of N-methyltryptophan, a desired polypeptide containing N-methyltryptophan was ribosomally synthesized by adding N-methyltryptophan and a mutant TrpRS to a cell-free translation system. The translation system used was PURE system, a reconstituted cell-free protein synthesis system from *E. coli*. Specifically, wild-type or mutant TrpRS and N-methyltryptophan were added to a solution containing a basic cell-free translation solution (1 mM GTP, 1 mM ATP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 9 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol, 1.5 mg/ml *E. coli* MRE600 (RNase negative)-derived tRNA (Roche), 0.1 mM 10-HCO-H4 folate, 4 µg/ml creatine kinase, 3 µg/ml myokinase, 2 unit/ml inorganic pyrophosphatase, 1.1 µg/ml nucleoside diphosphate kinase, 0.6 µM methionyl-tRNA transformylase, 0.26 µM EF-G, 0.24 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 40 µM EF-Tu, 84 µM EF-Ts, 1.2 µM ribosome, 0.03 µM ArgRS, 0.13 µM AspRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.02 µM TyrRS (wherein the proteins prepared by the inventors were essentially prepared as His-tagged proteins)), and 1 µM template mRNA, each 250 µM arginine, aspartic acid, lysine, methionine, and tyrosine, and left at 37°C for 1 hour to ribosomally synthesize the peptide.

<Detection by mass spectroscopy>

[0254] Mass spectroscopy was performed using MALDI-TOF MS for detecting peptides into which N-methyltryptophan is ribosomally introduced. Specifically, a solution containing 1 µM template mRNA (R-W (SEQ ID NO: 197)) and arginine, lysine, methionine, tyrosine, and

aspartic acid (each final concentration 250 μ M) added to the cell-free translation system as described above was prepared. TrpRS (final concentration 5 μ M) and N-methyltryptophan (final concentration 5 mM) were added to the solution and incubated at 37°C for 60 minutes. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS. Translation products were identified with MALDI-TOF MS spectrometry using α -cyano-4-hydroxycinnamic acid as a matrix.

[0255] As a result of the translation using wild-type TrpRS (SEQ ID NO: 188), a peak corresponding to the target peptide peak P-CT29MeTrp containing N-methyltryptophan (Figure 11(a), Peak MeW1) was observed, but a peak corresponding to peptide P-CT29Trp derived from Trp contaminated in the translation system in trace amounts was observed as main product (Figure 11(a), PeakW1). On the other hand, when experiments were performed using the modified TrpRS 04 (SEQ ID NO: 184), the modified TrpRS 05 (SEQ ID NO: 185), and the modified TrpRS 18 (SEQ ID NO: 186) with mutations introduced, peaks derived from P-CT29MeTrp (Peak MeW2-4, Figure 11(b)-(d)) were observed as main product. In other words, it is demonstrated that the peptide introduced with MeTrp was synthesized at a higher purity using modified TrpRSs compared to using wild-type TrpRS, and it was suggested that these modified TrpRSs are ARSs that can introduce MeTrp into peptides more efficiently compared to wild-type TrpRS.

[0256]

Peptide sequence P-CT29Trp (SEQ ID NO: 198)
formylMetArgTrpArgAspTyrLysAspAspAspAspLys

Peptide sequence P-CT29MeTrp (SEQ ID NO: 199)
formylMetArg[MeTrp]ArgAspTyrLysAspAspAspAspLys

MALDI-TOF MS:

[0257]

Calc. m/z: $[H+M]^+$ = 1670.7 (the peptide corresponding to the sequence P-CT29Trp)

Calc. m/z: $[H+M]^+$ = 1684.7 (the peptide corresponding to the sequence P-CT29MeTrp)

Example 6: Development of N-methylleucine-accepting ARS

<Preparation of wild-type and mutant LeuRS proteins>

[0258] Expression vectors having a polyhistidine sequence at the N-terminus and containing mutations listed in Table 9 were constructed. Subsequently, an expression strain was transformed with the vectors, and the mutant proteins of interest were purified with a nickel column from supernatants obtained by disrupting cells.

[Table 9]

Name	position 43	position 80	position 252
LeuRS01(wt)	Y	D	T
LeuRS02	G	D	T
LeuRS03	A	D	T
LeuRS04	E	D	T
LeuRS06	F	D	T
LeuRS07	Y	G	T
LeuRS08	Y	A	T
LeuRS09	E	G	T
LeuRS10	E	A	T
LeuRS11	D	G	T
LeuRS12	D	A	T
LeuRS13	A	G	T
LeuRS14	G	G	T
LeuRS15	Y	D	A
LeuRS16	A	G	A
LeuRS17	G	G	A
LeuRS18	G	D	A
LeuRS19	Y	G	A
LeuRS21	E	G	A
LeuRS22	E	A	A
LeuRS23	D	G	A

<Translational introduction of N-methylleucine using wild-type and mutant LeuRSs>

[0259] A template mRNA(R-L (SEQ ID NO: 201)) was synthesized from a template DNA (D-L (SEQ ID NO: 200)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) and purified using RNeasy Mini kit (Qiagen).

[0260]

D-L (CT23) (SEQ ID NO: 200)

DNA sequence:

GGCGTAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCGTCTC
CGTGACTACAAGGACGACGACGACAAGTAAGCTTCG

R-L (SEQ ID NO: 201)

RNA sequence:

GGGUUAAACUUUACAAGGAGAAAAACAUGCGUCUCCGUGACUACAAGGACGACG
ACGACAAGUAAGCUUCG

<Cell-free translation system>

[0261] In order to confirm translational introduction of N-methylleucine, a desired polypeptide containing N-methylleucine was ribosomally synthesized by adding N-methylleucine and a mutant LeuRS to a cell-free translation system. The translation system used was PURE system, a reconstituted cell-free protein synthesis system from *E. coli*. Specifically, wild-type or mutant LeuRS and N-methylleucine were added to a solution containing a basic cell-free translation solution (1 mM GTP, 1 mM ATP, 20 mM creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM potassium acetate, 9 mM magnesium acetate, 2 mM spermidine, 1 mM dithiothreitol, 1.5 mg/ml *E. coli* MRE600 (RNase negative)-derived tRNA (Roche), 0.1 mM 10-HCO-H4 folate, 4 µg/ml creatine kinase, 3 µg/ml myokinase, 2 unit/ml inorganic pyrophosphatase, 1.1 µg/ml nucleoside diphosphate kinase, 0.6 µM methionyl-tRNA transformylase, 0.26 µM EF-G, 0.24 µM RF2, 0.17 µM RF3, 0.5 µM RRF, 2.7 µM IF1, 0.4 µM IF2, 1.5 µM IF3, 40 µM EF-Tu, 84 µM EF-Ts, 1.2 µM ribosome, 0.03 µM ArgRS, 0.13 µM AspRS, 0.11 µM LysRS, 0.03 µM MetRS, 0.02 µM TyrRS (wherein the proteins prepared by the inventors were essentially prepared as His-tagged proteins)), and 1 µM template mRNA, each 250 µM arginine, aspartic acid, lysine, methionine, and tyrosine, and left at 37°C for 1 hour to ribosomally synthesize the peptide.

<Detection by mass spectroscopy>

[0262] Mass spectroscopy was performed using MALDI-TOF MS for detecting peptides into which N-methylleucine is ribosomally introduced. Specifically, a solution containing 1 µM template mRNA (R-L (SEQ ID NO: 201)) and arginine, lysine, methionine, tyrosine, and aspartic acid (each final concentration 250 µM) added to the cell-free translation system as described above was prepared. LeuRS (final concentration 0.4-2 µM) and N-methylleucine (final concentration 5 mM) were added to the solution and incubated at 37°C for 60 minutes. The resulting translation reaction products were purified with SPE C-TIP (Nikkyo Technos Co., Ltd) and analyzed with MALDI-TOF MS. Translation products were identified with MALDI-TOF MS spectrometry using α-cyano-4-hydroxycinnamic acid as a matrix.

[0263] As a result of the translation using wild-type LeuRS (SEQ ID NO: 189), a peak

corresponding to the target peptide peak P-CT23MeLeu containing N-methylleucine was not observed, but the peptide P-CT23Leu derived from Leu contaminated in the translation system in trace amounts (Figure 12(a), PeakL1) was observed as an almost single product. On the other hand, when experiments were performed using the modified LeuRS 02 (SEQ ID NO: 187) with mutations introduced, the peak derived from P-CT29MeLeu (Peak MeL1, Figure 12(b)) was observed, and the intensity of the peak was as strong as the intensity of the peak of the peptide P-CT23Leu containing Leu (Peak L2, Figure 12(b)). In other words, it is demonstrated that the MeLeu-introduced peptide synthesized using the modified LeuRS 02 was more than that using wild-type LeuRS, suggesting that this modified LeuRS is an ARS that can introduce MeLeu into peptides more efficiently compared to wild-type LeuRS.

[0264]

Peptide sequence P-CT23Leu (SEQ ID NO: 202)
formylMetArgLeuArgAspTyrLysAspAspAspAspLys

Peptide sequence P-CT23MeLeu (SEQ ID NO: 203)
formylMetArg[MeLeu]ArgAspTyrLysAspAspAspAspLys

MALDI-TOF MS:

[0265]

Calc. m/z: [H+M]⁺ = 1597.7 (the peptide corresponding to the sequence P-CT23Leu)

Calc. m/z: [H+M]⁺ = 1611.7 (the peptide corresponding to the sequence P-CT23MeLeu)

Example 7: Development of modified ValRSs having increased selectivity to N-methylvaline achieved by enhancing valine-hydrolyzing ability in the editing domain

<Preparation of wild-type and mutant ValRS proteins>

[0266] Expression vectors for modified ValRSs that have a polyhistidine sequence at the N-terminus and containing mutations (N44G,T45S) in the catalytic domain and mutation T279A(G) in the editing domain, which mutations increase activity to N-methylvaline, were constructed (Table 10). Then, an expression strain was transformed with the vectors, and the mutant proteins of interest were purified using a nickel column from supernatants obtained by disrupting cells.

[Table 10]

--	--	--

Name	Catalytic domain		Editing domain
	position 43	position 45	position 279
ValRS01(wt)	N	T	T
ValRS13-11	G	S	T
ValRS66	G	S	A
ValRS67	G	S	G

<Aminoacylation reaction with valine and N-methylvaline using mutant ValRSs>

[Synthesis of E. coli tRNA^{Val} by *in vitro* transcription reaction]

[0267] E. coli tRNA (R-tRNA^{Val}1 (SEQ ID NO: 205)) was synthesized from a template DNA (D-tRNA^{Val}1 (SEQ ID NO: 204)) by *in vitro* transcription reaction using RiboMAX Large Scale RNA production System T7 (Promega, P1300) in the presence of 7.5 mM GMP, and purified using RNeasy Mini kit (Qiagen).

[0268]

D-tRNA^{Val}1 (SEQ ID NO: 204)

tRNA^{Val}1 DNA sequence:

GGCGTAATACGACTCACTATAGGGTGATTAGCTCAGCTGGGAGAGCACCTCCCTTAC
AAGGAGGGGGTCGGCGGTTTCGATCCCGTCATCACCCACCA

R-tRNA^{Val}1 (SEQ ID NO: 205)

tRNA^{Val}1 RNA sequence:

GGGUGAUUAGCUCAGCUGGGAGAGCACCUCUUACAAGGAGGGGGUCGGCGGU
UCGAUCCCGUCAUCACCCACCA

<Aminoacylation reaction>

[0269] For the aminoacylation reaction, the solution containing 50 μ M transcribed tRNA, 10 mM HEPES-K (pH 7.6), and 10 mM KCl solution was heated at 95°C for 2 minutes and then left at room temperature for 5 minutes or more to refold the tRNA. This tRNA solution was added to a final concentration of 10 μ M to an acylation buffer (in final concentrations of 50 mM

HEPES-K [pH 7.6], 2 mM ATP, 100 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 2 mM spermidine, 0.1 mg/mL Bovine Serum Albumin), mixed with a mutant ValRS (final concentration 2 μ M) and N-methylvaline (final concentration 0.08 mM-5 mM) or valine (final concentration 0.031 mM-0.25 mM), and incubated at 37°C for 10 minutes. Four volumes of a loading buffer (90 mM sodium acetate [pH 5.2], 10 mM EDTA, 95%(w/w) formamide, 0.1%(w/v) xylene cyanol) were added to the reaction solution and analyzed with acidic PAGE containing 6 M urea, and aminoacylation activity was detected by separating unreacted tRNA and aminoacyl-tRNA. RNA was stained with SYBR Gold (Life Technologies) and detected with LAS4000 (GE Healthcare).

[0270] This result proved that the acylation abilities of the mutants 13-11, 66 (SEQ ID NO: 182), and 67 (SEQ ID NO: 183) were not very different at each N-methylvaline concentration when N-methylvaline was used as substrate (Figure 13, for example, lanes 17-19). On the other hand, it was proved that when valine was used as a substrate, the acylation abilities of mutants 66 and 67 reduced compared to mutants 13-11 (Figure 14, for example, lanes 17-19). This demonstrated that mutants 66 and 67 having mutations newly introduced into the editing domain are modified ValRSs that have reduced aminoacylation activity to valine and therefore has increased selectivity to N-methylvaline.

ABSTRACT

[0271] The present invention provides modified valyl-tRNA synthetases (ValRSs) having increased reactivity with N-methyl valine compared to natural ValRSs. The present invention enables a more efficient production of polypeptides containing N-methyl valine.

SEQUENCE LISTING

[0272]

<110> CHUGAI SEIYAKU KABUSHIKI KAISHA

<120> Modified aminoacyl tRNA synthetases and their use

<130> AA2504 EP S3

<140> EP 16 76 4874.0

<141> 2016-03-11

<150> PCT/JP2016/057707

<151> 2015-03-13

<160> 205

<170> PatentIn version 3.5

<210> 1

<211> 327

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 1

```

Met Ser His Leu Ala Glu Leu Val Ala Ser Ala Lys Ala Ala Ile Ser
1          5          10          15

Gln Ala Ser Asp Val Ala Ala Leu Asp Asn Val Arg Val Glu Tyr Leu
20          25          30

Gly Lys Lys Gly His Leu Thr Leu Gln Met Thr Thr Leu Arg Glu Leu
35          40          45

Pro Pro Glu Glu Arg Pro Ala Ala Gly Ala Val Ile Asn Glu Ala Lys
50          55          60

Glu Gln Val Gln Gln Ala Leu Asn Ala Arg Lys Ala Glu Leu Glu Ser
65          70          75          80

Ala Ala Leu Asn Ala Arg Leu Ala Ala Glu Thr Ile Asp Val Ser Leu
85          90          95

Pro Gly Arg Arg Ile Glu Asn Gly Gly Leu His Pro Val Thr Arg Thr
100          105          110

Ile Asp Arg Ile Glu Ser Phe Phe Gly Glu Leu Gly Phe Thr Val Ala
115          120          125

Thr Gly Pro Glu Ile Glu Asp Asp Tyr His Asn Phe Asp Ala Leu Asn
130          135          140

Ile Pro Gly His His Pro Ala Arg Ala Asp His Asp Thr Phe Trp Phe
145          150          155          160

Asp Thr Thr Arg Leu Leu Arg Thr Ala Thr Ser Gly Val Gln Ile Arg
165          170          175

Thr Met Lys Ala Gln Gln Pro Pro Ile Arg Ile Ile Ala Pro Gly Arg
180          185          190

Val Tyr Arg Asn Asp Tyr Asp Gln Thr His Thr Pro Met Phe His Gln
195          200          205

Met Glu Gly Leu Ile Val Asp Thr Asn Ile Ser Phe Thr Asn Leu Lys
210          215          220

Gly Thr Leu His Asp Phe Leu Arg Asn Phe Phe Glu Glu Asp Leu Gln
225          230          235          240

```

Ile Arg Phe Arg Pro Ser Tyr Phe Pro Phe Thr Glu Pro Ser Ala Glu
 245 250 255
 Val Asp Val Met Gly Lys Asn Gly Lys Trp Leu Glu Val Leu Gly Cys
 260 265 270
 Gly Met Val His Pro Asn Val Leu Arg Asn Val Gly Ile Asp Pro Glu
 275 280 285
 Val Tyr Ser Gly Phe Ala Phe Gly Met Gly Met Glu Arg Leu Thr Met
 290 295 300
 Leu Arg Tyr Gly Val Thr Asp Leu Arg Ser Phe Phe Glu Asn Asp Leu
 305 310 315 320
 Arg Phe Leu Lys Gln Phe Lys
 325

<210> 2

<211> 327

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 2

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

Ile Pro Gly His His Pro Ala Arg Ala Asp His Asp Thr Phe Trp Phe
145 150 155 160

Asp Thr Thr Arg Leu Leu Arg Thr Gly Thr Ser Gly Val Gln Ile Arg
165 170 175

Thr Met Lys Ala Gln Gln Pro Pro Ile Arg Ile Ile Ala Pro Gly Arg
180 185 190

Val Tyr Arg Asn Asp Tyr Asp Gln Thr His Thr Pro Met Phe His Gln
195 200 205

Met Glu Gly Leu Ile Val Asp Thr Asn Ile Ser Phe Thr Asn Leu Lys
210 215 220

Gly Thr Leu His Asp Phe Leu Arg Asn Phe Phe Glu Glu Asp Leu Gln
225 230 235 240

Ile Arg Phe Arg Pro Ser Tyr Phe Pro Phe Thr Glu Pro Ser Ala Glu
245 250 255

Val Asp Val Met Gly Lys Asn Gly Lys Trp Leu Glu Val Leu Gly Cys
260 265 270

Gly Met Val His Pro Asn Val Leu Arg Asn Val Gly Ile Asp Pro Glu
275 280 285

Val Tyr Ser Gly Phe Ala Phe Gly Met Gly Met Glu Arg Leu Thr Met
290 295 300

Leu Arg Tyr Gly Val Thr Asp Leu Arg Ser Phe Phe Glu Asn Asp Leu
305 310 315 320

Arg Phe Leu Lys Gln Phe Lys
325

<210> 3

<211> 951

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 3

Met Glu Lys Thr Tyr Asn Pro Gln Asp Ile Glu Gln Pro Leu Tyr Glu
1 5 10 15

His Trp Glu Lys Gln Gly Tyr Phe Lys Pro Asn Gly Asp Glu Ser Gln
20 25 30

Glu Ser Phe Cys Ile Met Ile Pro Pro Pro Ala Val Thr Gly Ser Leu
35 40 45

His Met Gly His Ala Phe Gln Gln Thr Ile Met Asp Thr Met Ile Arg
50 55 60

Tyr Gln Arg Met Gln Gly Lys Asn Thr Leu Trp Gln Val Gly Thr Asp
65 70 75 80

His Ala Gly Ile Ala Thr Gln Met Val Val Glu Arg Lys Ile Ala Ala
85 90 95

Glu Glu Gly Lys Thr Arg His Asp Tyr Gly Arg Glu Ala Phe Ile Asp
100 105 110

Lys Ile Trp Glu Trp Lys Ala Glu Ser Gly Gly Thr Ile Thr Arg Gln
115 120 125

Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr
130 135 140

Met Asp Glu Gly Leu Ser Asn Ala Val Lys Glu Val Phe Val Arg Leu
145 150 155 160

Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp
165 170 175

Pro Lys Leu Arg Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu
180 185 190

Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala
195 200 205

Lys Thr Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro
210 215 220

Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
225 230 235 240

Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn
245 250 255

Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu Lys Gly
260 265 270

Thr Gly Cys Val Lys Ile Thr Pro Ala His Asp Phe Asn Asp Tyr Glu
275 280 285

Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile Leu Thr Phe Asp
290 295 300

Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu
305 310 315 320

Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala Glu Phe Gln Lys Leu Glu
325 330 335

Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly
340 345 350

Leu Leu Glu Glu Ile Lys Pro His Asp Leu Thr Val Pro Tyr Gly Asp
355 360 365

Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr
 675 680 685

Arg Glu Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu
 690 695 700

Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr
 705 710 715 720

Lys Pro Val Met Asn Gly Gly Thr Glu Ala Glu Leu Arg Gly Thr Arg
 725 730 735

His Thr Leu Val Thr Val Leu Glu Gly Leu Leu Arg Leu Ala His Pro
 740 745 750

Ile Ile Pro Phe Ile Thr Glu Thr Ile Trp Gln Arg Val Lys Val Leu
 755 760 765

Cys Gly Ile Thr Ala Asp Thr Ile Met Leu Gln Pro Phe Pro Gln Tyr
 770 775 780

Asp Ala Ser Gln Val Asp Glu Ala Ala Leu Ala Asp Thr Glu Trp Leu
 785 790 795 800

Lys Gln Ala Ile Val Ala Val Arg Asn Ile Arg Ala Glu Met Asn Ile
 805 810 815

Ala Pro Gly Lys Pro Leu Glu Leu Leu Leu Arg Gly Cys Ser Ala Asp
 820 825 830

Ala Glu Arg Arg Val Asn Glu Asn Arg Gly Phe Leu Gln Thr Leu Ala
 835 840 845

Arg Leu Glu Ser Ile Thr Val Leu Pro Ala Asp Asp Lys Gly Pro Val
 850 855 860

Ser Val Thr Lys Ile Ile Asp Gly Ala Glu Leu Leu Ile Pro Met Ala
 865 870 875 880

Gly Leu Ile Asn Lys Glu Asp Glu Leu Ala Arg Leu Ala Lys Glu Val
 885 890 895

Ala Lys Ile Glu Gly Glu Ile Ser Arg Ile Glu Asn Lys Leu Ala Asn
 900 905 910

Glu Gly Phe Val Ala Arg Ala Pro Glu Ala Val Ile Ala Lys Glu Arg
 915 920 925

Glu Lys Leu Glu Gly Tyr Ala Glu Ala Lys Ala Lys Leu Ile Glu Gln
 930 935 940

Gln Ala Val Ile Ala Ala Leu
 945 950

<210> 4

<211> 951

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 4

Met Glu Lys Thr Tyr Asn Pro Gln Asp Ile Glu Gln Pro Leu Tyr Glu
 1 5 10 15

His Trp Glu Lys Gln Gly Tyr Phe Lys Pro Asn Gly Asp Glu Ser Gln
 20 25 30

Glu Ser Phe Cys Ile Met Ile Pro Pro Pro Gly Val Thr Gly Ser Leu
 35 40 45

His Met Gly His Ala Phe Gln Gln Thr Ile Met Asp Thr Met Ile Arg
 50 55 60

Tyr Gln Arg Met Gln Gly Lys Asn Thr Leu Trp Gln Val Gly Thr Asp
 65 70 75 80

His Ala Gly Ile Ala Thr Gln Met Val Val Glu Arg Lys Ile Ala Ala
 85 90 95

Glu Glu Gly Lys Thr Arg His Asp Tyr Gly Arg Glu Ala Phe Ile Asp
 100 105 110

Lys Ile Trp Glu Trp Lys Ala Glu Ser Gly Gly Thr Ile Thr Arg Gln
 115 120 125

Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr
 130 135 140

Met Asp Glu Gly Leu Ser Asn Ala Val Lys Glu Val Phe Val Arg Leu
 145 150 155 160

Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp
 165 170 175

Pro Lys Leu Arg Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu
 180 185 190

Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala
 195 200 205

Lys Thr Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro
 210 215 220

Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
 225 230 235 240

Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn

245 250 255
 Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu Lys Gly
 260 265 270
 Thr Gly Cys Val Lys Ile Thr Pro Ala His Asp Phe Asn Asp Tyr Glu
 275 280 285
 Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile Leu Thr Phe Asp
 290 295 300
 Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu
 305 310 315 320
 Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala Glu Phe Gln Lys Leu Glu
 325 330 335
 Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly
 340 345 350
 Leu Leu Glu Glu Ile Lys Pro His Asp Leu Thr Val Pro Tyr Gly Asp
 355 360 365
 Arg Gly Gly Val Val Ile Glu Pro Met Leu Thr Asp Gln Trp Tyr Val
 370 375 380
 Arg Ala Asp Val Leu Ala Lys Pro Ala Val Glu Ala Val Glu Asn Gly
 385 390 395 400
 Asp Ile Gln Phe Val Pro Lys Gln Tyr Glu Asn Met Tyr Phe Ser Trp
 405 410 415
 Met Arg Asp Ile Gln Asp Trp Cys Ile Ser Arg Gln Leu Trp Trp Gly
 420 425 430
 His Arg Ile Pro Ala Trp Tyr Asp Glu Ala Gly Asn Val Tyr Val Gly
 435 440 445
 Arg Asn Glu Asp Glu Val Arg Lys Glu Asn Asn Leu Gly Ala Asp Val
 450 455 460
 Val Leu Arg Gln Asp Glu Asp Val Leu Asp Thr Trp Phe Ser Ser Ala
 465 470 475 480
 Leu Trp Thr Phe Ser Thr Leu Gly Trp Pro Glu Asn Thr Asp Ala Leu
 485 490 495
 Arg Gln Phe His Pro Thr Ser Val Met Val Ser Gly Phe Asp Ile Ile
 500 505 510
 Phe Phe Trp Ile Ala Arg Met Ile Met Met Thr Met His Phe Ile Lys
 515 520 525
 Asp Glu Asn Gly Lys Pro Gln Val Pro Phe His Thr Val Tyr Met Thr
 530 535 540

Gly Leu Ile Arg Asp Asp Glu Gly Gln Lys Met Ser Lys Ser Lys Gly
 545 550 555 560

Asn Val Ile Asp Pro Leu Asp Met Val Asp Gly Ile Ser Leu Pro Glu
 565 570 575

Leu Leu Glu Lys Arg Thr Gly Asn Met Met Gln Pro Gln Leu Ala Asp
 580 585 590

Lys Ile Arg Lys Arg Thr Glu Lys Gln Phe Pro Asn Gly Ile Glu Pro
 595 600 605

His Gly Thr Asp Ala Leu Arg Phe Thr Leu Ala Ala Leu Ala Ser Thr
 610 615 620

Gly Arg Asp Ile Asn Trp Asp Met Lys Arg Leu Glu Gly Tyr Arg Asn
 625 630 635 640

Phe Cys Asn Lys Leu Trp Asn Ala Ser Arg Phe Val Leu Met Asn Thr
 645 650 655

Glu Gly Gln Asp Cys Gly Phe Asn Gly Gly Glu Met Thr Leu Ser Leu
 660 665 670

Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr
 675 680 685

Arg Glu Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu
 690 695 700

Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr
 705 710 715 720

Lys Pro Val Met Asn Gly Gly Thr Glu Ala Glu Leu Arg Gly Thr Arg
 725 730 735

His Thr Leu Val Thr Val Leu Glu Gly Leu Leu Arg Leu Ala His Pro
 740 745 750

Ile Ile Pro Phe Ile Thr Glu Thr Ile Trp Gln Arg Val Lys Val Leu
 755 760 765

Cys Gly Ile Thr Ala Asp Thr Ile Met Leu Gln Pro Phe Pro Gln Tyr
 770 775 780

Asp Ala Ser Gln Val Asp Glu Ala Ala Leu Ala Asp Thr Glu Trp Leu
 785 790 795 800

Lys Gln Ala Ile Val Ala Val Arg Asn Ile Arg Ala Glu Met Asn Ile
 805 810 815

Ala Pro Gly Lys Pro Leu Glu Leu Leu Leu Arg Gly Cys Ser Ala Asp
 820 825 830

Ala Glu Arg Arg Val Asn Glu Asn Arg Gly Phe Leu Gln Thr Leu Ala
 835 840 845

Arg Leu Glu Ser Ile Thr Val Leu Pro Ala Asp Asp Lys Gly Pro Val
850 855 860

Ser Val Thr Lys Ile Ile Asp Gly Ala Glu Leu Leu Ile Pro Met Ala
865 870 875 880

Gly Leu Ile Asn Lys Glu Asp Glu Leu Ala Arg Leu Ala Lys Glu Val
885 890 895

Ala Lys Ile Glu Gly Glu Ile Ser Arg Ile Glu Asn Lys Leu Ala Asn
900 905 910

Glu Gly Phe Val Ala Arg Ala Pro Glu Ala Val Ile Ala Lys Glu Arg
915 920 925

Glu Lys Leu Glu Gly Tyr Ala Glu Ala Lys Ala Lys Leu Ile Glu Gln
930 935 940

Gln Ala Val Ile Ala Ala Leu
945 950

<210> 5

<211> 951

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 5

Met Glu Lys Thr Tyr Asn Pro Gln Asp Ile Glu Gln Pro Leu Tyr Glu
1 5 10 15

His Trp Glu Lys Gln Gly Tyr Phe Lys Pro Asn Gly Asp Glu Ser Gln
20 25 30

Glu Ser Phe Cys Ile Met Ile Pro Pro Pro Gly Val Ser Gly Ser Leu
35 40 45

His Met Gly His Ala Phe Gln Gln Thr Ile Met Asp Thr Met Ile Arg
50 55 60

Tyr Gln Arg Met Gln Gly Lys Asn Thr Leu Trp Gln Val Gly Thr Asp
65 70 75 80

His Ala Gly Ile Ala Thr Gln Met Val Val Glu Arg Lys Ile Ala Ala
85 90 95

Glu Glu Gly Lys Thr Arg His Asp Tyr Gly Arg Glu Ala Phe Ile Asp
100 105 110

Lys Ile Trp Glu Trp Lys Ala Glu Ser Gly Gly Thr Ile Thr Arg Gln
115 120 125

Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr
 130 135 140

Met Asp Glu Gly Leu Ser Asn Ala Val Lys Glu Val Phe Val Arg Leu
 145 150 155 160

Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp
 165 170 175

Pro Lys Leu Arg Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu
 180 185 190

Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala
 195 200 205

Lys Thr Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro
 210 215 220

Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
 225 230 235 240

Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn
 245 250 255

Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu Lys Gly
 260 265 270

Thr Gly Cys Val Lys Ile Thr Pro Ala His Asp Phe Asn Asp Tyr Glu
 275 280 285

Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile Leu Thr Phe Asp
 290 295 300

Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu
 305 310 315 320

Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala Glu Phe Gln Lys Leu Glu
 325 330 335

Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly
 340 345 350

Leu Leu Glu Glu Ile Lys Pro His Asp Leu Thr Val Pro Tyr Gly Asp
 355 360 365

Arg Gly Gly Val Val Ile Glu Pro Met Leu Thr Asp Gln Trp Tyr Val
 370 375 380

Arg Ala Asp Val Leu Ala Lys Pro Ala Val Glu Ala Val Glu Asn Gly
 385 390 395 400

Asp Ile Gln Phe Val Pro Lys Gln Tyr Glu Asn Met Tyr Phe Ser Trp
 405 410 415

Met Arg Asp Ile Gln Asp Trp Cys Ile Ser Arg Gln Leu Trp Trp Gly
 420 425 430

His Arg Ile Pro Ala Trp Tyr Asp Glu Ala Gly Asn Val Tyr Val Gly
 435 440 445
 Arg Asn Glu Asp Glu Val Arg Lys Glu Asn Asn Leu Gly Ala Asp Val
 450 455 460
 Val Leu Arg Gln Asp Glu Asp Val Leu Asp Thr Trp Phe Ser Ser Ala
 465 470 475 480
 Leu Trp Thr Phe Ser Thr Leu Gly Trp Pro Glu Asn Thr Asp Ala Leu
 485 490 495
 Arg Gln Phe His Pro Thr Ser Val Met Val Ser Gly Phe Asp Ile Ile
 500 505 510
 Phe Phe Trp Ile Ala Arg Met Ile Met Met Thr Met His Phe Ile Lys
 515 520 525
 Asp Glu Asn Gly Lys Pro Gln Val Pro Phe His Thr Val Tyr Met Thr
 530 535 540
 Gly Leu Ile Arg Asp Asp Glu Gly Gln Lys Met Ser Lys Ser Lys Gly
 545 550 555 560
 Asn Val Ile Asp Pro Leu Asp Met Val Asp Gly Ile Ser Leu Pro Glu
 565 570 575
 Leu Leu Glu Lys Arg Thr Gly Asn Met Met Gln Pro Gln Leu Ala Asp
 580 585 590
 Lys Ile Arg Lys Arg Thr Glu Lys Gln Phe Pro Asn Gly Ile Glu Pro
 595 600 605
 His Gly Thr Asp Ala Leu Arg Phe Thr Leu Ala Ala Leu Ala Ser Thr
 610 615 620
 Gly Arg Asp Ile Asn Trp Asp Met Lys Arg Leu Glu Gly Tyr Arg Asn
 625 630 635 640
 Phe Cys Asn Lys Leu Trp Asn Ala Ser Arg Phe Val Leu Met Asn Thr
 645 650 655
 Glu Gly Gln Asp Cys Gly Phe Asn Gly Gly Glu Met Thr Leu Ser Leu
 660 665 670
 Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr
 675 680 685
 Arg Glu Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu
 690 695 700
 Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr
 705 710 715 720
 Lys Pro Val Met Asn Gly Gly Thr Glu Ala Glu Leu Arg Gly Thr Arg
 725 730 735
 His Phe Leu Val Phe Val Leu Glu Gly Leu Leu Asp Leu Ala His Asp

His Thr Leu Val Thr Val Leu Glu Gly Leu Leu Arg Leu Ala His Pro
740 745 750

Ile Ile Pro Phe Ile Thr Glu Thr Ile Trp Gln Arg Val Lys Val Leu
755 760 765

Cys Gly Ile Thr Ala Asp Thr Ile Met Leu Gln Pro Phe Pro Gln Tyr
770 775 780

Asp Ala Ser Gln Val Asp Glu Ala Ala Leu Ala Asp Thr Glu Trp Leu
785 790 795 800

Lys Gln Ala Ile Val Ala Val Arg Asn Ile Arg Ala Glu Met Asn Ile
805 810 815

Ala Pro Gly Lys Pro Leu Glu Leu Leu Leu Arg Gly Cys Ser Ala Asp
820 825 830

Ala Glu Arg Arg Val Asn Glu Asn Arg Gly Phe Leu Gln Thr Leu Ala
835 840 845

Arg Leu Glu Ser Ile Thr Val Leu Pro Ala Asp Asp Lys Gly Pro Val
850 855 860

Ser Val Thr Lys Ile Ile Asp Gly Ala Glu Leu Leu Ile Pro Met Ala
865 870 875 880

Gly Leu Ile Asn Lys Glu Asp Glu Leu Ala Arg Leu Ala Lys Glu Val
885 890 895

Ala Lys Ile Glu Gly Glu Ile Ser Arg Ile Glu Asn Lys Leu Ala Asn
900 905 910

Glu Gly Phe Val Ala Arg Ala Pro Glu Ala Val Ile Ala Lys Glu Arg

915

920

925

Glu Lys Leu Glu Gly Tyr Ala Glu Ala Lys Ala Lys Leu Ile Glu Gln
930 935 940

Gln Ala Val Ile Ala Ala Leu
945 950

<210> 6

<211> 430

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 6

Met Leu Asp Pro Asn Leu Leu Arg Asn Glu Pro Asp Ala Val Ala Glu
1 5 10 15

Lys Leu Ala Arg Arg Gly Phe Lys Leu Asp Val Asp Lys Leu Gly Ala
 20 25 30

Leu Glu Glu Arg Arg Lys Val Leu Gln Val Lys Thr Glu Asn Leu Gln
 35 40 45

Ala Glu Arg Asn Ser Arg Ser Lys Ser Ile Gly Gln Ala Lys Ala Arg
 50 55 60

Gly Glu Asp Ile Glu Pro Leu Arg Leu Glu Val Asn Lys Leu Gly Glu
 65 70 75 80

Glu Leu Asp Ala Ala Lys Ala Glu Leu Asp Ala Leu Gln Ala Glu Ile
 85 90 95

Arg Asp Ile Ala Leu Thr Ile Pro Asn Leu Pro Ala Asp Glu Val Pro
 100 105 110

Val Gly Lys Asp Glu Asn Asp Asn Val Glu Val Ser Arg Trp Gly Thr
 115 120 125

Pro Arg Glu Phe Asp Phe Glu Val Arg Asp His Val Thr Leu Gly Glu
 130 135 140

Met His Ser Gly Leu Asp Phe Ala Ala Ala Val Lys Leu Thr Gly Ser
 145 150 155 160

Arg Phe Val Val Met Lys Gly Gln Ile Ala Arg Met His Arg Ala Leu
 165 170 175

Ser Gln Phe Met Leu Asp Leu His Thr Glu Gln His Gly Tyr Ser Glu
 180 185 190

Asn Tyr Val Pro Tyr Leu Val Asn Gln Asp Thr Leu Tyr Gly Thr Gly
 195 200 205

Gln Leu Pro Lys Phe Ala Gly Asp Leu Phe His Thr Arg Pro Leu Glu
 210 215 220

Glu Glu Ala Asp Thr Ser Asn Tyr Ala Leu Ile Pro Ser Ala Glu Val
 225 230 235 240

Pro Leu Thr Asn Leu Val Arg Gly Glu Ile Ile Asp Glu Asp Asp Leu
 245 250 255

Pro Ile Lys Met Thr Ala His Thr Pro Cys Phe Arg Ser Glu Ala Gly
 260 265 270

Ser Tyr Gly Arg Asp Thr Arg Gly Leu Ile Arg Met His Gln Phe Asp
 275 280 285

Lys Val Glu Met Val Gln Ile Val Arg Pro Glu Asp Ser Met Ala Ala
 290 295 300

Leu Glu Glu Met Thr Gly His Ala Glu Lys Val Leu Gln Leu Leu Gly
 305 310 315 320

Leu Pro Tyr Arg Lys Ile Ile Leu Cys Thr Gly Asp Met Gly Phe Gly
 325 330 335

Ala Cys Lys Thr Tyr Asp Leu Glu Val Trp Ile Pro Ala Gln Asn Thr
 340 345 350

Tyr Arg Glu Ile Ser Ser Cys Ser Asn Val Trp Asp Phe Gln Ala Arg
 355 360 365

Arg Met Gln Ala Arg Cys Arg Ser Lys Ser Asp Lys Lys Thr Arg Leu
 370 375 380

Val His Thr Leu Asn Gly Ser Gly Leu Ala Val Gly Arg Thr Leu Val
 385 390 395 400

Ala Val Met Glu Asn Tyr Gln Gln Ala Asp Gly Arg Ile Glu Val Pro
 405 410 415

Glu Val Leu Arg Pro Tyr Met Asn Gly Leu Glu Tyr Ile Gly

420 425 430

<210> 7

<211> 430

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 7

Met Leu Asp Pro Asn Leu Leu Arg Asn Glu Pro Asp Ala Val Ala Glu
 1 5 10 15

Lys Leu Ala Arg Arg Gly Phe Lys Leu Asp Val Asp Lys Leu Gly Ala
 20 25 30

Leu Glu Glu Arg Arg Lys Val Leu Gln Val Lys Thr Glu Asn Leu Gln
 35 40 45

Ala Glu Arg Asn Ser Arg Ser Lys Ser Ile Gly Gln Ala Lys Ala Arg
 50 55 60

Gly Glu Asp Ile Glu Pro Leu Arg Leu Glu Val Asn Lys Leu Gly Glu
 65 70 75 80

Glu Leu Asp Ala Ala Lys Ala Glu Leu Asp Ala Leu Gln Ala Glu Ile
 85 90 95

Arg Asp Ile Ala Leu Thr Ile Pro Asn Leu Pro Ala Asp Glu Val Pro
 100 105 110

Val Gly Lys Asp Glu Asn Asp Asn Val Glu Val Ser Arg Trp Gly Thr
 115 120 125

Glu Val Leu Arg Pro Tyr Met Asn Gly Leu Glu Tyr Ile Gly
 420 425 430

<210> 8

<211> 430

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 8

Met Leu Asp Pro Asn Leu Leu Arg Asn Glu Pro Asp Ala Val Ala Glu
 1 5 10 15

Lys Leu Ala Arg Arg Gly Phe Lys Leu Asp Val Asp Lys Leu Gly Ala
 20 25 30

Leu Glu Glu Arg Arg Lys Val Leu Gln Val Lys Thr Glu Asn Leu Gln
 35 40 45

Ala Glu Arg Asn Ser Arg Ser Lys Ser Ile Gly Gln Ala Lys Ala Arg
 50 55 60

Gly Glu Asp Ile Glu Pro Leu Arg Leu Glu Val Asn Lys Leu Gly Glu
 65 70 75 80

Glu Leu Asp Ala Ala Lys Ala Glu Leu Asp Ala Leu Gln Ala Glu Ile
 85 90 95

Arg Asp Ile Ala Leu Thr Ile Pro Asn Leu Pro Ala Asp Glu Val Pro
 100 105 110

Val Gly Lys Asp Glu Asn Asp Asn Val Glu Val Ser Arg Trp Gly Thr
 115 120 125

Pro Arg Glu Phe Asp Phe Glu Val Arg Asp His Val Thr Leu Gly Glu
 130 135 140

Met His Ser Gly Leu Asp Phe Ala Ala Val Lys Leu Thr Gly Ser
 145 150 155 160

Arg Phe Val Val Met Lys Gly Gln Ile Ala Arg Met His Arg Ala Leu
 165 170 175

Ser Gln Phe Met Leu Asp Leu His Thr Glu Gln His Gly Tyr Ser Glu
 180 185 190

Asn Tyr Val Pro Tyr Leu Val Asn Gln Asp Thr Leu Tyr Gly Thr Gly
 195 200 205

Gln Leu Pro Lys Phe Ala Gly Asp Leu Phe His Thr Arg Pro Leu Glu
 210 215 220

Glu Glu Ala Asp Thr Ser Asn Tyr Ala Leu Ile Pro Thr Ala Gly Val

Ala Glu Arg Asn Ser Arg Ser Lys Ser Ile Gly Gln Ala Lys Ala Arg
50 55 60

Gly Glu Asp Ile Glu Pro Leu Arg Leu Glu Val Asn Lys Leu Gly Glu
65 70 75 80

Glu Leu Asp Ala Ala Lys Ala Glu Leu Asp Ala Leu Gln Ala Glu Ile
85 90 95

Arg Asp Ile Ala Leu Thr Ile Pro Asn Leu Pro Ala Asp Glu Val Pro
100 105 110

Val Gly Lys Asp Glu Asn Asp Asn Val Glu Val Ser Arg Trp Gly Thr
115 120 125

Pro Arg Glu Phe Asp Phe Glu Val Arg Asp His Val Thr Leu Gly Glu
130 135 140

Met His Ser Gly Leu Asp Phe Ala Ala Ala Val Lys Leu Thr Gly Ser
145 150 155 160

Arg Phe Val Val Met Lys Gly Gln Ile Ala Arg Met His Arg Ala Leu
165 170 175

Ser Gln Phe Met Leu Asp Leu His Thr Glu Gln His Gly Tyr Ser Glu
180 185 190

Asn Tyr Val Pro Tyr Leu Val Asn Gln Asp Thr Leu Tyr Gly Thr Gly
195 200 205

Gln Leu Pro Lys Phe Ala Gly Asp Leu Phe His Thr Arg Pro Leu Glu
210 215 220

Glu Glu Ala Asp Thr Ser Asn Tyr Ala Leu Ile Pro Ser Ala Gly Val
225 230 235 240

Pro Leu Thr Asn Leu Val Arg Gly Glu Ile Ile Asp Glu Asp Asp Leu
245 250 255

Pro Ile Lys Met Thr Ala His Thr Pro Cys Phe Arg Ser Glu Ala Gly
260 265 270

Ser Tyr Gly Arg Asp Thr Arg Gly Leu Ile Arg Met His Gln Phe Asp
275 280 285

Lys Val Glu Met Val Gln Ile Val Arg Pro Glu Asp Ser Met Ala Ala
290 295 300

Leu Glu Glu Met Thr Gly His Ala Glu Lys Val Leu Gln Leu Leu Gly
305 310 315 320

Leu Pro Tyr Arg Lys Ile Ile Leu Cys Thr Gly Asp Met Gly Phe Gly
325 330 335

Ala Cys Lys Thr Tyr Asp Leu Glu Val Trp Ile Pro Ala Gln Asn Thr
340 345 350

Tyr Arg Glu Ile Ser Ser Cys Ser Asn Val Trp Asp Phe Gln Ala Arg
355 360 365

Arg Met Gln Ala Arg Cys Arg Ser Lys Ser Asp Lys Lys Thr Arg Leu
370 375 380

Val His Thr Leu Asn Gly Ser Gly Leu Ala Val Gly Arg Thr Leu Val
385 390 395 400

Ala Val Met Glu Asn Tyr Gln Gln Ala Asp Gly Arg Ile Glu Val Pro
405 410 415

Glu Val Leu Arg Pro Tyr Met Asn Gly Leu Glu Tyr Ile Gly
420 425 430

<210> 10

<211> 642

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 10

Met Pro Val Ile Thr Leu Pro Asp Gly Ser Gln Arg His Tyr Asp His
1 5 10 15

Ala Val Ser Pro Met Asp Val Ala Leu Asp Ile Gly Pro Gly Leu Ala
20 25 30

Lys Ala Cys Ile Ala Gly Arg Val Asn Gly Glu Leu Val Asp Ala Cys
35 40 45

Asp Leu Ile Glu Asn Asp Ala Gln Leu Ser Ile Ile Thr Ala Lys Asp
50 55 60

Glu Glu Gly Leu Glu Ile Ile Arg His Ser Cys Ala His Leu Leu Gly
65 70 75 80

His Ala Ile Lys Gln Leu Trp Pro His Thr Lys Met Ala Ile Gly Pro
85 90 95

Val Ile Asp Asn Gly Phe Tyr Tyr Asp Val Asp Leu Asp Arg Thr Leu
100 105 110

Thr Gln Glu Asp Val Glu Ala Leu Glu Lys Arg Met His Glu Leu Ala
115 120 125

Glu Lys Asn Tyr Asp Val Ile Lys Lys Lys Val Ser Trp His Glu Ala
130 135 140

Arg Glu Thr Phe Ala Asn Arg Gly Glu Ser Tyr Lys Val Ser Ile Leu
145 150 155 160

Asp Glu Asn Ile Ala His Asp Asp Lys Pro Glu Leu Tyr Phe His Glu

```

-----
                165                               170                               175
-----
Glu Tyr Val Asp Met Cys Arg Gly Pro His Val Pro Asn Met Arg Phe
      180                               185                               190

Cys His His Phe Lys Leu Met Lys Thr Ala Gly Ala Tyr Trp Arg Gly
      195                               200                               205

Asp Ser Asn Asn Lys Met Leu Gln Arg Ile Tyr Gly Thr Ala Trp Ala
      210                               215                               220

Asp Lys Lys Ala Leu Asn Ala Tyr Leu Gln Arg Leu Glu Glu Ala Ala
      225                               230                               235                               240

Lys Arg Asp His Arg Lys Ile Gly Lys Gln Leu Asp Leu Tyr His Met
      245                               250                               255

Gln Glu Glu Ala Pro Gly Met Val Phe Trp His Asn Asp Gly Trp Thr
      260                               265                               270

Ile Phe Arg Glu Leu Glu Val Phe Val Arg Ser Lys Leu Lys Glu Tyr
      275                               280                               285

Gln Tyr Gln Glu Val Lys Gly Pro Phe Met Met Asp Arg Val Leu Trp
      290                               295                               300

Glu Lys Thr Gly His Trp Asp Asn Tyr Lys Asp Ala Met Phe Thr Thr
      305                               310                               315                               320

Ser Ser Glu Asn Arg Glu Tyr Cys Ile Lys Pro Gly Asn Cys Pro Gly
      325                               330                               335

His Val Gln Ile Phe Asn Gln Gly Leu Lys Ser Tyr Arg Asp Leu Pro
      340                               345                               350

Leu Arg Met Ala Glu Phe Gly Ser Cys His Arg Asn Glu Pro Ser Gly
      355                               360                               365

Ser Leu His Gly Leu Met Arg Val Arg Gly Phe Thr Gln Asp Asp Ala
      370                               375                               380

His Ile Phe Cys Thr Glu Glu Gln Ile Arg Asp Glu Val Asn Gly Cys
      385                               390                               395                               400

Ile Arg Leu Val Tyr Asp Met Tyr Ser Thr Phe Gly Phe Glu Lys Ile
      405                               410                               415

Val Val Lys Leu Ser Thr Arg Pro Glu Lys Arg Ile Gly Ser Asp Glu
      420                               425                               430

Met Trp Asp Arg Ala Glu Ala Asp Leu Ala Val Ala Leu Glu Glu Asn
      435                               440                               445

Asn Ile Pro Phe Glu Tyr Gln Leu Gly Glu Gly Ala Phe Tyr Gly Pro
      450                               455                               460

Lys Ile Glu Phe Thr Leu Tyr Asp Cys Leu Asp Arg Ala Trp Gln Cys

```

Lys Ile Glu Phe Thr Leu Tyr Asp Cys Met Asp Arg Ala Asp Gln Cys
 465 470 475 480
 Gly Thr Val Gln Leu Asp Phe Ser Leu Pro Ser Arg Leu Ser Ala Ser
 485 490 495
 Tyr Val Gly Glu Asp Asn Glu Arg Lys Val Pro Val Met Ile His Arg
 500 505 510
 Ala Ile Leu Gly Ser Met Glu Arg Phe Ile Gly Ile Leu Thr Glu Glu
 515 520 525
 Phe Ala Gly Phe Phe Pro Thr Trp Leu Ala Pro Val Gln Val Val Ile
 530 535 540
 Met Asn Ile Thr Asp Ser Gln Ser Glu Tyr Val Asn Glu Leu Thr Gln
 545 550 555 560
 Lys Leu Ser Asn Ala Gly Ile Arg Val Lys Ala Asp Leu Arg Asn Glu
 565 570 575
 Lys Ile Gly Phe Lys Ile Arg Glu His Thr Leu Arg Arg Val Pro Tyr
 580 585 590
 Met Leu Val Cys Gly Asp Lys Glu Val Glu Ser Gly Lys Val Ala Val
 595 600 605
 Arg Thr Arg Arg Gly Lys Asp Leu Gly Ser Met Asp Val Asn Glu Val
 610 615 620
 Ile Glu Lys Leu Gln Gln Glu Ile Arg Ser Arg Ser Leu Lys Gln Leu
 625 630 635 640
 Glu Glu
 <210> 11
 <211> 642
 <212> PRT
 <213> Artificial
 <220>
 <223> artificially synthesized sequence
 <400> 11
 Met Pro Val Ile Thr Leu Pro Asp Gly Ser Gln Arg His Tyr Asp His
 1 5 10 15
 Ala Val Ser Pro Met Asp Val Ala Leu Asp Ile Gly Pro Gly Leu Ala
 20 25 30
 Lys Ala Cys Ile Ala Gly Arg Val Asn Gly Glu Leu Val Asp Ala Cys
 35 40 45
 Asp Leu Ile Glu Asn Asp Ala Gln Leu Ser Ile Ile Thr Ala Lys Asp

Leu Arg Met Ala Glu Phe Gly Ser Cys His Arg Asn Glu Pro Ser Gly
 355 360 365
 Ser Leu His Gly Leu Met Arg Val Arg Gly Phe Thr Gln Asp Asp Ala
 370 375 380
 His Ile Phe Cys Thr Glu Glu Gln Ile Arg Asp Glu Val Asn Gly Cys
 385 390 395 400
 Ile Arg Leu Val Tyr Asp Met Tyr Ser Thr Phe Gly Phe Glu Lys Ile
 405 410 415
 Val Val Lys Leu Ser Thr Arg Pro Glu Lys Arg Ile Gly Ser Asp Glu
 420 425 430
 Met Trp Asp Arg Ala Glu Ala Asp Leu Ala Val Ala Leu Glu Glu Asn
 435 440 445
 Asn Ile Pro Phe Glu Tyr Gln Leu Gly Glu Gly Ala Phe Tyr Gly Pro
 450 455 460
 Lys Ile Glu Phe Thr Leu Tyr Asp Cys Leu Asp Arg Ala Trp Gln Cys
 465 470 475 480
 Gly Thr Val Gln Leu Asp Phe Ser Leu Pro Ser Arg Leu Ser Ala Ser
 485 490 495
 Tyr Val Gly Glu Asp Asn Glu Arg Lys Val Pro Val Met Ile Gly Arg
 500 505 510
 Ala Ile Leu Gly Ser Met Glu Arg Phe Ile Gly Ile Leu Thr Glu Glu
 515 520 525
 Phe Ala Gly Phe Phe Pro Thr Trp Leu Ala Pro Val Gln Val Val Ile
 530 535 540
 Met Asn Ile Thr Asp Ser Gln Ser Glu Tyr Val Asn Glu Leu Thr Gln
 545 550 555 560
 Lys Leu Ser Asn Ala Gly Ile Arg Val Lys Ala Asp Leu Arg Asn Glu
 565 570 575
 Lys Ile Gly Phe Lys Ile Arg Glu His Thr Leu Arg Arg Val Pro Tyr
 580 585 590
 Met Leu Val Cys Gly Asp Lys Glu Val Glu Ser Gly Lys Val Ala Val
 595 600 605
 Arg Thr Arg Arg Gly Lys Asp Leu Gly Ser Met Asp Val Asn Glu Val
 610 615 620
 Ile Glu Lys Leu Gln Gln Glu Ile Arg Ser Arg Ser Leu Lys Gln Leu
 625 630 635 640
 Glu Glu

<210> 12
 <211> 981
 <212> DNA
 <213> Artificial

<220>
 <223> artificially synthesized sequence

<400> 12
 atgtcacatc tcgcagaact ggttgccagt gcgaaggcgg ccattagcca ggcgtcagat 60
 gttgccgcgt tagataatgt gcgcgtcgaa tatttgggta aaaaagggca cttaccctt 120
 cagatgacga ccctgcgtga gctgccgcca gaagagcgtc cggcagctgg tgcggtatc 180
 aacgaagcga aagagcaggt tcagcaggcg ctgaatgcmc gtaaagcggga actggaaagc 240
 gctgcaactga atgcgcgtct ggccgcccga acgattgatg tctctctgcc aggtcgtcmc 300
 attgaaaacg gcggtctgca tccggttacc cgtaccatcg accgtatcga aagtttcttc 360
 ggtgagcttg gctttaccgt ggcaaccggg ccggaatcga aagacgatta tcataacttc 420
 gatgctctga acattcctgg tcaccaccgg gcgcgcgctg accacgacac tttctggttt 480
 gacactaccg gcctgctgcm taccgcgacc tctggcgtac agatccgcac catgaaagcc 540
 cagcagccac cgattcgtat catcgcgcct ggccgtgttt atcgtaacga ctacgaccag 600
 actcacacgc cgatggtcca tcagatggaa ggtctgattg ttgataccaa catcagcttt 660
 accaacctga aaggcacgct gcacgacttc ctgcgtaact tctttgagga agatttgcag 720
 attcgcctcc gtccttccta cttcccgttt accgaacctt ctgcagaagt ggacgtcatg 780
 ggtaaaaacg gtaaatggct ggaagtgctg ggctgcggga tggatcatcc gaacgtggtg 840
 cgtaacgttg gcatcgaccc ggaagtttac tctggtttcg ccttcgggat ggggatggag 900
 cgtctgacta tgttgcgtta cggcgtcacc gacctgcggt cattcttcga aaacgatctg 960
 cgtttctca aacagtttaa a 981

<210> 13
 <211> 981
 <212> DNA
 <213> Artificial

<220>
 <223> artificially synthesized sequence

<400> 13
 atgtcacatc tcgcagaact ggttgccagt gcgaaggcgg ccattagcca ggcgtcagat 60
 gttgccgcgt tagataatgt gcgcgtcgaa tatttgggta aaaaagggca cttaccctt 120
 cagatgacga ccctgcgtga gctgccgcca gaagagcgtc cggcagctgg tgcggtatc 180
 aacgaagcga aagagcaggt tcagcaggcg ctgaatgcmc gtaaagcggga actggaaagc 240
 gctgcaactga atgcgcgtct ggccgcccga acgattgatg tctctctgcc aggtcgtcmc 300
 attgaaaacg gcggtctgca tccggttacc cgtaccatcg accgtatcga aagtttcttc 360

```

ggtgagcttg gctttaccgt ggcaaccggg ccggaatcg aagacgatta tcataacttc 420
gatgctctga acattcctgg tcaccaccgg gcgcgcgctg accacgacac tttctggttt 480
gacactaccg gcctgctgcg taccggcacc tctggcgtag agatccgcac catgaaagcc 540
cagcagccac cgattcgtat catcgcgctt ggccgtgttt atcgtaacga ctacgaccag 600
actcacacgc cgatgttcca tcagatggaa ggtctgattg ttgataccaa catcagcttt 660
accaacctga aaggcacgct gcacgacttc ctgcgtaact tctttgagga agatttgcag 720
attcgcctcc gtccttccta cttcccgttt accgaacctt ctgcagaagt ggacgtcatg 780
ggtaaaaacg gtaaatggct ggaagtgctg ggctgcggga tgggtgatcc gaacgtggtg 840
cgtaacgttg gcatcgaccc ggaagtttac tctggtttcg ccttcgggat ggggatggag 900
cgtctgacta tgttgcttta cggcgtcacc gacctcgctt cattcttoga aaacgatctg 960
cgtttcctca aacagtttaa a 981

```

<210> 14

<211> 2853

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 14

```

atggaaaaga catataaccc acaagatcgc gaacagccgc ttacgagca ctgggaaaag 60
cagggctact ttaagcctaa tggcgatgaa agccaggaaa gtttctgcat catgatcccg 120
ccgccggctg tcaccggcag tttgcatatg ggtcacgcct tccagcaaac catcatggat 180
accatgatcc gctatcagcg catgcagggc aaaaacaccc tgtggcaggt cggtagctgac 240
cacgccggga tcgctaccca gatggtcggt gagcgcaaga ttgccgcaga agaaggtaaa 300
accgctcagc actacggccg cgaagcttcc atcgacaaaa tctgggaatg gaaagcggaa 360
tctggcggca ccattaccgg tcagatgcgc cgtctcggca actccgtcga ctgggagcgt 420
gaacgcttca ccattggacga aggcctgtcc aatgcgggta aagaagtttt cgttcgtctg 480
tataaagaag acctgattta ccgtggcaaa cgcctggtaa actgggatcc gaaactgcgc 540
accgctatct ctgacctgga agtcgaaaac cgcgaatcga aaggttcgat gtggcacatc 600
cgctatccgc tggctgacgg tgcgaaaacc gcagacggtg aagattatct ggtggtcgcg 660
actaccgctc cagaaacctt gctggcgcat actggcgtag ccgttaaccc ggaagatccg 720
cgttacaaag atctgattgg caaatatgtc attctgcgcg tggttaaccg tcgtattccg 780
atcgttggcg acgaacacgc cgacatggaa aaaggcaccg gctgcgtgaa aatcactccg 840
gcgcacgact ttaacgacta tgaagtgggt aaacgtcacg ccctgccgat gatcaacatc 900
ctgaccttgg acggcgatat ccgtgaaagc gccaggtgt tcgataccaa aggtaacgaa 960
tctgacgttt attccagcga aatccctgca gagttccaga aactggagcg ttttgctgca 1020
cgtaaagcag tcgttgccgc agttgacgcg cttggcctgc tggaagaaat taaaccgcac 1080

```

gacctgaccg ttccttacgg cgaccgtggc ggcgtagtta tcgaaccaat gctgaccgac 1140
cagtggtaac tgcgtgccga tgcctctggc aaaccggcgg ttgaagcggg tgagaacggc 1200
gacattcagt tcgtaccgaa gcagtagcaa aacatgtact tctcctggat gcgcgatatt 1260
caggactggg gtatctctcg tcagttgtgg tggggtcacc gtatcccggc atgggatgac 1320
gaagcgggta acgtttatgt tggccgcaac gaagacgaag tgcgtaaaga aaataacctc 1380
ggtgctgatg ttgtcctgcg tcaggacgaa gacgttctcg atacctgggt ctcttctgcg 1440
ctgtggacct tctctaccct tggctggcgg gaaaataacc acgccctgcg tcagttccac 1500
ccaaccagcg tgatggatc tggtttcgac atcattttct tctggattgc ccgcatgac 1560
atgatgacca tgcacttcat caaagatgaa aatggcaaac cgcaggtgcc gttccacacc 1620
gtttacatga ccggcctgat tcgtgatgac gaaggccaga agatgtcca atccaagggt 1680
aacgttatcg acccactgga tatggtgac ggtatttcgc tgcagaact gctggaaaa 1740
cgtaccggca atatgatgca gccgcagctg gcggacaaa tccgtaagcg caccgagaag 1800
cagttcccga acggtattga gccgcacggg actgacgcg tgcgcttcac cctggcggcg 1860
ctggcgtcta ccggctgta catcaactgg gatatgaagc gtctggaagg ttaccgtaac 1920
ttctgtaaca agctgtggaa cgccagccgc tttgtgctga tgaacacaga aggtcaggat 1980
tgcggcttca acggcggcga aatgacgctg tcgctggcgg accgctggat tctggcggag 2040
ttcaaccaga ccatacaagc gtaccgcgaa gcgctggaca gcttccgctt cgatatcgcc 2100
gcaggcattc tgtatgagtt cacctggaac cagttctgtg actggtatct cgagctgacc 2160
aagccggtaa tgaacgggtg caccgaagca gaactgcccg gtactcgcca tacgctggtg 2220
actgtactgg aaggtctgct gcgcctcggc catccgatca ttcggttcat caccgaaacc 2280
atctggcagc gtgtgaaagt actttgcggt atcaactgcc acaccatcat gctgcagccg 2340
ttcccgcagt acgatgcatc tcaggtgat gaagccgcac tggccgacac cgaatggctg 2400

aacaggcga tcgttgccgt acgtaacatc cgtgcagaaa tgaacatcgc gccgggcaaa 2460
ccgctggagc tgcgtctgcg tggttgcagc gcggatgcag aacgtcgcgt aatgaaaac 2520
cgtggcttcc tgcaaacctt ggcgcgtctg gaaagtatca ccgtgctgcc tgccgatgac 2580
aaaggctccg tttccgttac gaagatcacc gacggtgcag agctgctgat cccgatggct 2640
ggcctcatca acaagaaga tgagctggcg cgtctggcga aagaagtggc gaagattgaa 2700
ggtgaaatca gccgatcga gaacaaactg gcgaacgaag gctttgtcgc ccgcgaccg 2760
gaagcggta tcgcgaaaga gcgtgagaag ctggaaggct atgcggaagc gaaagcgaaa 2820
ctgattgaac agcaggctgt tatcgccgcg ctg 2853

<210> 15

<211> 2853

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 15

atggaaaaga	catataaccc	acaagatata	gaacagccgc	tttacgagca	ctgggaaaag	60
cagggctact	ttaagcctaa	tggcgatgaa	agccaggaaa	gtttctgcat	catgatcccg	120
cgcgcgggtg	tcaccggcag	tttgcataatg	ggtcacgcct	tccagcaaac	catcatggat	180
accatgatcc	gctatcagcg	catgcagggc	aaaaacaccc	tgtggcaggt	cggtactgac	240
cacgccggga	tcgctaccca	gatggtcggt	gagcgcaaga	ttgccgcaga	agaaggtaaa	300
accgctcacg	actacggccg	cgaagcttcc	atcgacaaaa	tctgggaatg	gaaagcggaa	360
tctggcggca	ccattacccg	tcagatgcgc	cgtctcggca	actccgtcga	ctgggagcgt	420
gaacgcttca	ccatggacga	aggcctgtcc	aatgcgggtga	aagaagtttt	cgttcgtctg	480
tataaagaag	acctgattta	ccgtggcaaa	cgccctggtaa	actgggatcc	gaaactgcgc	540
accgctatct	ctgacctgga	agtcgaaaac	cgcgaaatcga	aaggctcgat	gtggcacatc	600
cgctatccgc	tggctgacgg	tgcgaaaacc	gcagacggta	aagattatct	ggtggtcgcg	660
actaccgctc	cagaaaccct	gctgggcgat	actggcgtag	ccgttaacc	ggaagatccg	720
cgttacaag	atctgattgg	caaatatgtc	attctgcgc	tggtaaccg	tcgtattccg	780
atcgttggtg	acgaacacgc	cgacatggaa	aaaggcaccg	gctgcgtgaa	aatcactccg	840
gcgcacgact	ttaacgacta	tgaagtgggt	aaacgtcacg	ccctgccgat	gatcaacatc	900
ctgacctttg	acggcगतat	ccgtgaaaagc	gcccagggtgt	tcgataccaa	aggtaacgaa	960
tctgacgttt	attccagcga	aatccctgca	gagttccaga	aactggagcg	ttttgctgca	1020
cgtaaagcag	tcgttgccgc	agttgacgcg	cttgccctgc	tggaagaaat	taaaccgcac	1080
gacctgaccg	ttccttacgg	cgaccgtggc	ggcgtagtta	tcgaaccaat	gctgaccgac	1140
cagtggtaag	tgcgtgccga	tgtcctggcg	aaaccggcgg	ttgaagcgggt	tgagaacggc	1200
gacattcagt	tcgtaccgaa	gcagtagcaa	aacatgtact	tctcctggat	gcgcgatatt	1260
caggactgggt	gtatctctcg	tcagttgtgg	tggggtcacc	gtatcccggc	atggatgac	1320
gaagcgggta	acgtttatgt	tggccgcaac	gaagacgaag	tgcgtaaaga	aaataacctc	1380
ggtgctgatg	ttgtcctgcg	tcaggacgaa	gacgttctcg	atacctggtt	ctcttctgcg	1440
ctgtggacct	tctctaccct	tggctggccg	gaaaataaccg	acgccctgcg	tcagttccac	1500
ccaaccagcg	tgatggatc	tggtttcgac	atcattttct	tctggattgc	ccgcatgac	1560
atgatgacca	tgcacttcat	caaagatgaa	aatggcaaac	cgcagggtgcc	gttccacacc	1620
gtttacatga	ccggcctgat	tcgtgatgac	gaaggccaga	agatgtccaa	atccaaggggt	1680
aacgttatcg	accactgga	tatggttgac	ggtatttctc	tgccagaact	gctggaaaaa	1740
cgtaccggca	atatgatgca	gccgcagctg	gcggacaaaa	tccgtaagcg	caccgagaag	1800
cagttcccga	acggtattga	gcgcacgggt	actgacgcgc	tgcgcttcac	octggcggcg	1860
ctggcgtcta	ccggtcgtga	catcaactgg	gatatgaagc	gtctggaagg	ttaccgtaac	1920
ttctgtaaca	agctgtggaa	cgccagccgc	tttgtgctga	tgaacacaga	aggtcaggat	1980
tgcggcttca	acggcggcga	aatgacgctg	tcgctggcgg	accgctggat	tctggcggag	2040
ttcaaccaga	ccatcaaagc	gtaccgcgaa	gcgctggaca	gcttccgctt	cgatatcgcc	2100
gcaggcattc	tgtatgagtt	cacctggaac	cagttctgtg	actggtatct	cgagctgacc	2160

aagccggtaa tgaacggtgg caccgaagca gaactgcgcg gtactcgcca tacgctggtg	2220
actgtactgg aaggtctgct gcgcctcgcg catccgatca ttccggtcat caccgaaacc	2280
atctggcagc gtgtgaaagt actttgcggt atcactgccc acaccatcat gctgcagccg	2340
ttcccgagc acgatgcatc tcaggttgat gaagccgac tggccgacac cgaatggctg	2400
aaacaggcga tcggtgaggc acgtaacatc cgtgcagaaa tgaacatcgc gccgggcaaa	2460
ccgctggagc tgctgctgcg tggttgcagc gcggatgcag aacgtcgcgt aatgaaaac	2520
cgtggcttcc tgcaaacctt ggcgcgtctg gaaagtatca ccgtgctgcc tgccgatgac	2580
aaaggtccgg tttccgttac gaagatcatc gacggtgcag agctgctgat cccgatggct	2640
ggcctcatca acaaagaaga tgagctggcg cgtctggcga aagaagtggc gaagattgaa	2700
ggtgaaatca gccgatcga gaacaaactg gcgaacgaag gctttgtcgc ccgcgaccg	2760
gaagcggcga tcgcgaaaga gcgtgagaag ctggaaggct atgcggaagc gaaagcgaaa	2820
ctgattgaac agcaggctgt tatcgccgcg ctg	2853

<210> 16

<211> 2853

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 16

atggaaaaga catataacct acaagatata gaacagccgc ttacgagca ctgggaaaag	60
cagggctact ttaagcctaa tggcgatgaa agccaggaaa gtttctgcat catgatcccg	120
ccgccgggtg tctcaggcag tttgcatatg ggtcacgcct tccagcaaac catcatggat	180
accatgatcc gctatcagcg catgcagggc aaaaacacct tgtggcaggc cggactgac	240
cacgccggga tcgctaccga gatggctcgt gagcgaaga ttgccgcaga agaaggtaaa	300
accgctcacg actacggccg cgaagcttcc atcgacaaaa tctgggaatg gaaagcggaa	360
tctggcggca ccattacctg tcagatgcgc cgtctcggca actccgtcga ctgggagcgt	420
gaacgcttca ccatggacga aggcctgtcc aatgcggtga aagaagttt cgttcgtctg	480
tataaagaag acctgattta ccgtggcaaa cgcctggtta actgggatcc gaaactgcgc	540
accgctatct ctgacctgga agtcgaaaac cgcgaatcga aaggttcgat gtggcacatc	600
cgctatccgc tggtgacggc tgcgaaaacc gcagacggta aagattatct ggtggtcggc	660
actaccgctc cagaaacctt gctgggcgat actggcgtag ccgtaacctt ggaagatccc	720
cgttacaagc atctgattgg caaatatgtc attctgccgc tggtaaccg tcgtattccg	780
atcgttggcg acgaacacgc cgacatgtaa aaagcaccg gctgcgtgaa aatcactccc	840
ggcacgact ttaacgacta tgaagtgggt aaacgtcacg ccctgccgat gatcaacatc	900
ctgacctttg acggcgatat ccgtgaaagc gccaggtgt tcgataccaa aggtaacgaa	960
tctgacgttt attccagcga aatccctgca gagttccaga aactggagcg ttttgcctga	1020
cgtaaaacac tcgttcccgc agttgacgca cttgacctgc tggaaqaaat taaaccgac	1080

<400> 17

```

atgctcgatc ccaatctgct gcgtaatgag ccagacgcag tcgctgaaaa actggcacgc      60
cggggcttta agctggatgt agataagctg ggcgctcttg aagagcgtcg taaagtattg      120
caggtcaaaa cggaaaacct gcaagcggag cgtaactccc gatcgaaatc cattggccag      180
gcgaaagcgc gcggggaaga tatcgagcct ttacgtctgg aagtgaacaa actgggcgaa      240
gagctggatg cagcaaaagc cgagctggat gctttacagg ctgaaattcg cgatatcgcg      300
ctgaccatcc ctaacctgcc tgcagatgaa gtgccggtag gtaaagacga aaatgacaac      360
gttgaagtca gccgctgggg taccocgcgt gagtttgact ttgaagttcg tgaccacgtg      420
acgctgggtg aatgcactc tggcctcgac tttgcagctg cagttaagct gactggttcc      480
cgctttgtgg taatgaaagg gcagattgct cgcatgcacc gcgcactgtc gcagtttatg      540
ctggatctgc ataccgaaca gcatggctac agtgagaact atgttccgta cctggttaac      600
caggacacgc tgtacggtac ggttcaactg ccgaaatttg ctggcgatct gttccatact      660
cgtccgctgg aagaagaagc agacaccagt aactatgcgc tgatcccaag tgcagaagtt      720
ccgctgacta acctggtgcg cggtgaaatc atcgatgaag atgatctgcc aattaagatg      780
accgcccaca ccccatgctt ccgttctgaa gccggttcat atggtcgtga caccctggtt      840
ctgatccgta tgcaccagtt cgacaaagtt gaaatggtgc agatcgtgcg cccagaagac      900
tcaatggcgg cgctggaaga gatgactggt catgcagaaa aagtctgca gttgctgggc      960
ctgccgtacc gtaaaatcat cctttgcact ggcgacatgg gctttggcgc ttgcaaaact     1020
tacgacctgg aagtatggat cccggcacag aacacctacc gtgagatctc ttcctgctcc     1080
aacgtttggg atttccaggc acgtcgtatg caggcacggt gccgcagcaa gtcggacaag     1140
aaaaccctgc tggttcatac cctgaacggt tctggtctgg ctggttggtcg tacgctggtt     1200
gcagtaatgg aaaactatca gcaggctgat ggtcgtattg aagtaccaga agttctgcgt     1260
cogtatatga acggactgga atatattggc                                     1290

```

<210> 18

<211> 1290

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 18

```

atgctcgatc ccaatctgct gcgtaatgag ccagacgcag tcgctgaaaa actggcacgc      60
cggggcttta agctggatgt agataagctg ggcgctcttg aagagcgtcg taaagtattg      120
caggtcaaaa cggaaaacct gcaagcggag cgtaactccc gatcgaaatc cattggccag      180
gcgaaagcgc gcggggaaga tatcgagcct ttacgtctgg aagtgaacaa actgggcgaa      240
gagctggatg cagcaaaagc cgagctggat gctttacagg ctgaaattcg cgatatcgcg      300
ctgaccatcc ctaacctgcc tgcagatgaa gtgccggtag gtaaagacga aaatgacaac      360
cttccatccc cggcctcccc tccccccgat ccttttccat tcccccttcc tccccccctc      420

```

gttgaagtca gccgctgggg taccctcgct gagtttgact ttgaagttcg tgaccacgtg 420
 acgctgggtg aaatgcactc tggcctcgac tttgcagctg cagttaagct gactggttcc 480
 cgctttgtgg taatgaaagg gcagattgct cgcatgcacc gcgcactgtc gcagtttatg 540
 ctggatctgc ataccgaaca gcatggctac agtgagaact atgttccgta cctggttaac 600
 caggacacgc tgtacggtac ggttcaactg ccgaaatttg ctggcgatct gttccatact 660
 cgtccgctgg aagaagaagc agacaccagt aactatgccc tgatcccaac ggcagctggt 720
 ccgctgacta acctggtgcg cggtgaaatc atcgatgaag atgatctgcc aattaagatg 780
 accgcccaca ccccatgctt ccgttctgaa gccggttcat atggtcgtga caccctggtt 840
 ctgatccgta tgcaccagtt cgacaaagt gaaatggtgc agatcgtgcg ccagaagac 900
 tcaatggcgg cgctggaaga gatgactggt catgcagaaa aagtcctgca gttgctgggc 960
 ctgccgtacc gtaaaatcat cctttgcact ggcgacatgg gctttggcgc ttgcaaaact 1020
 tacgacctgg aagtatggat cccggcacag aacacctacc gtgagatctc ttcctgctcc 1080
 aacgtttggg atttccagcc acgtcgtatg caggcacgtt gccgcagcaa gtcggacaag 1140
 aaaaccctgc tggttcatac cctgaacggt tctggtctgg ctggtggtcg tacgctggtt 1200
 gcagtaatgg aaaactatca gcaggctgat ggtcgtattg aagtaccaga agttctgcgt 1260
 ccgtatatga acggaactgga atatattggc 1290

<210> 19

<211> 1290

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 19

atgctcgatc ccaatctgct gcgtaatgag ccagacgcag tcgctgaaaa actggcacgc 60
 cggggcttta agctggatgt agataagctg ggcgctcttg aagagcgtcg taaagtattg 120
 caggtcaaaa cggaaaacct gcaagcggag cgtaactccc gatcgaaatc cattggccag 180
 gcgaaagcgc gcggggaaga tatcgagcct ttacgtctgg aagtgaacaa actgggcgaa 240
 gagctggatg cagcaaaagc cgagctggat gctttacagg ctgaaattcg cgatatcgcg 300
 ctgaccatcc ctaacctgcc tgcagatgaa gtgccggtag gtaaagacga aatgacaac 360
 gttgaagtca gccgctgggg taccctcgct gagtttgact ttgaagttcg tgaccacgtg 420
 acgctgggtg aaatgcactc tggcctcgac tttgcagctg cagttaagct gactggttcc 480
 cgctttgtgg taatgaaagg gcagattgct cgcatgcacc gcgcactgtc gcagtttatg 540
 ctggatctgc ataccgaaca gcatggctac agtgagaact atgttccgta cctggttaac 600
 caggacacgc tgtacggtac ggttcaactg ccgaaatttg ctggcgatct gttccatact 660
 cgtccgctgg aagaagaagc agacaccagt aactatgccc tgatcccaac ggcaggtggt 720
 ccgctgacta acctggtgcg cggtgaaatc atcgatgaag atgatctgcc aattaagatg 780
 accgcccaca ccccatgctt ccgttctgaa gccggttcat atggtcgtga caccctggtt 840

ctgatccgta tgcaccagtt cgacaaagtt gaaatggtgc agatcgtgcg cccagaagac 900
 tcaatggcgg cgctggaaga gatgactggt catgcagaaa aagtccctgca gttgctgggc 960
 ctgccgtacc gtaaaatcat cctttgcaact ggcgacatgg gctttggcgc ttgcaaaact 1020
 tacgacctgg aagtatggat cccggcacag aacacctacc gtgagatctc ttcctgctcc 1080
 aacgtttggg atttccaggc acgtcgtatg caggcacggt gccgcagcaa gtcggacaag 1140
 aaaaccgctc tggttcatac cctgaacggt tctggtctgg ctggttggtcg tacgctggtt 1200
 gcagtaatgg aaaactatca gcaggctgat ggtcgtattg aagtaccaga agttctgcgt 1260
 ccgtatatga acggactgga atatattggc 1290

<210> 20

<211> 1290

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 20

atgctcgatc ccaatctgct gcgtaatgag ccagacgcag tcgctgaaaa actggcacgc 60
 cggggcttta agctggatgt agataagctg ggcgctcttg aagagcgtcg taaagtattg 120
 caggcctaaa cggaaaacct gcaagcggag cgtaactccc gatcgaaatc cattggccag 180
 gcgaaagcgc gcggggaaga tatcgagcct ttacgtctgg aagtgaacaa actgggcgaa 240
 gagctggatg cagcaaaagc cgagctggat gctttacagg ctgaaattcg cgatatcgcg 300
 ctgaccatcc ctaacctgcc tgcagatgaa gtgccggtag gtaaagacga aaatgacaac 360
 gttgaagtca gccgctgggg taccctcgct gagtttgact ttgaagttcg tgaccacgtg 420
 acgctgggtg aaatgcactc tggcctcgac tttgcagctg cagttaagct gactggttcc 480
 cgctttgtgg taatgaaagg gcagattgct cgcatgcacc gcgcactgtc gcagtttatg 540
 ctggatctgc ataccgaaca gcatggctac agtgagaact atgttccgta cctggttaac 600
 caggacacgc tgtacggtac ggttcaactg ccgaaatttg ctggcgatct gttccatact 660
 cgtccgctgg aagaagaagc agacaccagt aactatgcgc tgatcccaag tgcaggtggt 720
 ccgctgacta acctgggtgcg cggtgaaatc atcgatgaag atgatctgcc aattaagatg 780
 accgcccaca ccccatgctt ccgttctgaa gccggttcat atggtcgtga caccctggt 840
 ctgatccgta tgcaccagtt cgacaaagtt gaaatggtgc agatcgtgcg cccagaagac 900
 tcaatggcgg cgctggaaga gatgactggt catgcagaaa aagtccctgca gttgctgggc 960
 ctgccgtacc gtaaaatcat cctttgcaact ggcgacatgg gctttggcgc ttgcaaaact 1020
 tacgacctgg aagtatggat cccggcacag aacacctacc gtgagatctc ttcctgctcc 1080
 aacgtttggg atttccaggc acgtcgtatg caggcacggt gccgcagcaa gtcggacaag 1140
 aaaaccgctc tggttcatac cctgaacggt tctggtctgg ctggttggtcg tacgctggtt 1200
 gcagtaatgg aaaactatca gcaggctgat ggtcgtattg aagtaccaga agttctgcgt 1260

cogtatatga acggactgga atatattggc

1290

<210> 21

<211> 1926

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 21

atgccggtta ttaccctgcc ggacggttcg caacgtcatt acgatcacgc tgttagcccg	60
atggatgtcg cactggatat tggcccgggt ctggcaaaag cttgcattgc aggccgcgtt	120
aatggtgaac tggtcgacgc ttgtgatctg atcgaaaacg acgcgagct gagcattatc	180
accgccaaaag atgaagaagg cctggaaatt atccgtcatt cttgcgcgca cctgctgggt	240
catgccatta aacagctgtg gccgcatacc aaaatggcga ttggcccgggt catcgacaat	300
ggtttttatt acgatgtgga cctggatcgc accctgacgc aggaagatgt ggaagcgcctg	360
gaaaaacgta tgcacgaact ggccgagaaa aactacgacg tgattaagaa aaaagttagt	420
tggcatgaag cacgtgaaac ctttgctaat cgcggcgaat catacaaagt ttcgattctg	480
gatgaaaaaca tcgccacga tgacaaaaccg ggcctgtatt tccatgaaga atacgtggat	540
atgtgccgtg gtccgcacgt tccgaatatg cgcttttctc atcacttcaa actgatgaaa	600
accgcaggtg catattggcg tggtgacagc aacaataaaa tgctgcagcg tatttatggt	660
acggcatggg ctgataaaaa agcgcctgaat gcctacctgc aacgcctgga agaagcggcc	720
aaacgtgacc accgcaaaat tggcaaacag ctggatctgt atcatatgca agaagaagcg	780
ccgggcatgg tgttttgcca taacgatggt tggaccatct ttcgtgaact ggaagttttc	840
gtccgcagca aactgaaaga atatcagtac caagaagtta aaggcccgtt tatgatggat	900
cgcgtcctgt gggaaaaaac cggtcactgg gacaactata aagatgccat gttcaccacg	960
agctctgaaa accgtgaata ctgcattaaa ccgggcaatt gtccgggtca tgtgcagatc	1020
tttaaccaag gcctgaaaag ttatcgtgat ctgccgctgc gcatggcaga attcggttcc	1080
tgccaccgca atgaaccgag tggctccctg catggtctga tgcgtgttcg cggttttacc	1140
caggatgacg cccatatttt ctgcacggaa gaacaaatcc gtgacgaagt taacggctgt	1200
attcgcctgg tctatgatat gtactctacc tttggtttcg aaaaaattgt ggttaaactg	1260
agcacgcgtc cggaaaaacg catcggctct gacgaaatgt gggatcgtgc ggaagccgat	1320
ctggcagtggt ctctggaaga aaacaatatt ccgtttgaat atcagctggg cgagggtgcg	1380
ttttacggcc cgaaaatcga attcaccctg tatgactgcc tggatcgcgc ctggcagtgt	1440
ggtacgggtcc aactggactt cagtctgccg tcccgtctgt cagcatcgta cgtgggcgaa	1500
gataatgaac gtaaagtgcc ggttatgatt caccgcgcta tcctgggcag tatggaacgt	1560
tttattggta tcctgaccga agaattcgca ggctttttcc cgacgtggct ggctccgggt	1620
caggctcgtga ttatgaatat caccgatagc cagtctgaat atgtgaacga actgacgcaa	1680

aaactgtcca atgcaggcat tcgtgttaaa gctgatctgc gcaacgaaaa aatcggtttc 1740
 aaaatccgcy aacataccct gcgtcgcgtc ccgtacatgc tgggtgtgtgg cgataaagaa 1800
 gttgaaagcy gtaaagtcgc ggtgcgtacg cgtcgcggca aagacctggg ttcgatggat 1860
 gtcaacgaag tcattgaaaa actgcaacaa gaaatccgta gccgctcgtc gaaacaactg 1920
 gaagaa 1926

<210> 22

<211> 1926

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 22

atgccggtca ttaccctgcc ggatggctct cagcgtcact acgatcatgc ggtctcgcgc 60
 atggatgttg ctctggatat tggcccgggt ctggcaaaag cttgcattgc aggccgcggt 120
 aatgggtgaa tggtcgcagc ttgtgatctg atcgaaaacg acgcgcagct gagcattatc 180
 accgccaaaag atgaagaagg cctggaaatt atccgtcatt cttgcgcgca cctgctgggt 240
 catgccatta agcagctgtg gccgcatacc aaaatggcga ttggcccgggt catcgacaat 300
 ggtttttatt acgatgtgga cctggatcgc accctgacgc aggaagatgt ggaagcgcctg 360
 gaaaagcgtg tgcacgaact ggccgaaaag aactatgacg tgatcaaaaa gaaagttagt 420
 tggcatgaag cacgtgaaac ctttgctaata cgcggcgaat catacaaggt ttcgattctg 480
 gatgaaaaca tcgccacgca tgacaaaccg ggcctgtatt tccatgaaga atacgtggat 540
 atgtgccgtg gtcccacagt tccgaatatg cgcttttctc atcaactcaa gctgatgaaa 600
 accgcaggtg catattggcg tggtgacagc aacaataaaa tgctgcagcg tatttatggt 660
 acggcatggg ctgataagaa agcgcgtgaat gcctacctgc aacgcctgga agaagcggcc 720
 aaacgtgacc accgcaagat tggcaaacag ctggatctgt atcatatgca agaagaagcg 780
 ccgggcatgg tgttttgga taacgatggt tggaccatct ttcgtgaact ggaagttttc 840
 gtccgcagca agctgaaaga atatcagtac caagaagtta aaggcccgtt tatgatggat 900
 cgcgtcctgt gggaaaagac cggtcactgg gacaactata aagatgccat gttcaccacg 960
 agctctgaaa accgtgaata ctgcattaag ccgatgaatt gtccgggcca tgtgcagatc 1020
 tttaaccaag gtctgaaaag ttatcgtgat ctgccgctgc gcatggcaga attcggttcc 1080
 tgccaccgca atgaaccgag tggctccctg catggtctga tgcgtgttcg cggctttacc 1140
 caggatgacg cccacatddd ctgcacggaa gaacaaatcc gtgacgaagt taacggctgt 1200
 attcgcctgg tctatgatat gtactctacc tttggtttcg aaaagattgt ggttaaactg 1260
 agcacgcgtc cggaaaaacg catcggttct gacgaaatgt gggatcgtgc ggaagccgat 1320
 ctggcagtgg ctctggaaga aaacaatatt ccgtttgaat atcagctggg cgagggtgcy 1380
 ttttacggcc cgaaaatcga attcaccctg tatgactgcc tggatcgcgc ctggcagtgt 1440
 catcacatcc aactcaactt caatctacca tcccatctat caccatcata catcaaccaa 1500

```

gataatgaac gtaaagtgcc ggttatgatt ggccgcgcta tcctgggtag tatggaacgt 1560
tttattggca tcctgaccga agaattcgca ggctttttcc cgacgtggct ggctccggtt 1620
caggctcgtga ttatgaatat caccgatagc cagtctgaat atgtgaacga actgacgcaa 1680
aagctgtcca atgcaggcat tcgtgttaaa gctgatctgc gcaacgaaaa gattggtttt 1740
aaaatccgcg aacataccct gcgtcgcgtc ccgtacatgc tgggtgtgtg cgataaggaa 1800
gttgaaagcg gtaaagtccg ggtgcgtacg cgtcgcggca aagacctggg ttcgatggat 1860
gtcaacgaag tcattgaaaa actgcaacaa gaaatccgct cacgctcgct gaaacaactg 1920
gaagaa 1926

```

<210> 23

<211> 2853

<212> DNA

<213> E. coli

<400> 23

```

atggaaaaga catataaccc acaagatatic gaacagccgc tttacgagca ctgggaaaag 60
cagggctact ttaagcctaa tggcgatgaa agccaggaaa gtttctgcat catgatcccg 120
ccgccgaacg tcaccggcag tttgcatatg ggtcacgcct tccagcaaac catcatggat 180
accatgatcc gctatcagcg catgcagggc aaaaacaccc tgtggcaggt cggtagctgac 240
cacgccggga tcgctaccga gatggtcggt gagcgaaga ttgccgcaga agaaggtaaa 300
accgctcagc actacggccc cgaagcttcc atcgacaaaa tctgggaatg gaaagcggaa 360
tctggcggca ccattaccgg tcagatgcgc cgtctcggca actccgctga ctgggagcgt 420
gaacgcttca ccatggacga aggcctgtcc aatgcgggtga aagaagtfff cgttcgtctg 480
tataaagaag acctgattta ccgtggcaaa cgcctggtaa actgggatcc gaaactgcgc 540
accgctatct ctgacctgga agtcgaaaaac cgcgaatcga aaggttcgat gtggcacatc 600
cgctatccgc tggctgacgg tgcgaaaacc gcagacggta aagattatct ggtggtcgcg 660
actaccgctc cagaaaccct gctgggcgat actggcgtag ccgttaaccc ggaagatccg 720
cgttacaaag atctgattgg caaatatgtc attctgccgc tggtaaccg tcgtattccg 780
atcgttggcg acgaacacgc cgacatggaa aaaggcaccg gctgcgtgaa aatcactccg 840
gogcagcact ttaacgacta tgaagtgggt aaacgtcagc ccctgccgat gatcaacatc 900
ctgacctttg acggcgatat ccgtgaaagc gccagggtgt tcgataccaa aggtaacgaa 960
tctgacgttt attccagcga aatccctgca gagttccaga aactggagcg ttttgctgca 1020
cgtaaagcag tcgttgccgc agttgacgcg cttggcctgc tggaaagaaat taaaccgcac 1080
gacctgaccg ttcccttacgg cgaccgtggc ggcgtagtta tcgaaccaat gctgaccgac 1140
cagtggtacg tgcgtgccga tgtcctggcg aaaccggcgg ttgaagcggg tgagaacggc 1200
gacattcagt tcgtaccgaa gcagtacgaa aacatgtact tctcctggat gcgcgatatt 1260
caggactggg gtatctctcg tcagttgtgg tggggtcacc gtatccggc atggtatgac 1320
gaagcgggta acgtttatgt tggccgcaac gaagacgaag tgcgtaaga aaataacctc 1380
cctgctgata tttctctggc tcaggacgaa gacgttctcg atagctcctt ctcttctgac 1440

```

```

gggctgaty ttgctctgcy tcaagacgaa gacgtctcty atacctggct cctctctgcy 1470
ctgtggacct tctctaccct tggctggccg gaaaataccg acgcctcgcy tcagttccac 1500
ccaaccagcg tgatggatc tggtttcgac atcattttct tctggattgc ccgcatgatc 1560
atgatgacca tgcacttcat caaagatgaa aatggcaaac cgcaggtgcc gttccacacc 1620
gtttacatga cggcctgat tctgtgatgac gaaggccaga agatgtccaa atccaagggt 1680
aacgttatcg acccactgga tatggttgac ggtatttcgc tgccagaact gctggaaaaa 1740
cgtaccggca atatgatgca gccgcagctg gcggacaaaa tccgtaagcg caccgagaag 1800
cagttcccga acggtattga gccgcacggt actgacgcgc tgcgcttcac cctggcggcg 1860
ctggcgctca cggctcgtga catcaactgg gatatgaagc gtctggaagg ttaccgtaac 1920
ttctgtaaca agctgtggaa cgcagccgc tttgtgctga tgaacacaga aggtcaggat 1980
tgcggcttca acggcggcga aatgacgctg tgcgtggcgg accgctggat tctggcggag 2040
ttcaaccaga ccatcaaagc gtaccgcgaa gcgctggaca gcttccgctt cgatatcgcc 2100
gcaggcattc tgtatgagtt cacctggaac cagttctgtg actggtatct cgagctgacc 2160
aagccggtaa tgaacgggtg caccgaagca gaactgcgcg gtactcgcca tacgctggtg 2220
actgtactgg aaggctcgtc gcgcctcgcy catccgatca ttccgttcat caccgaaacc 2280
atctggcagc gtgtgaaagt actttgcggt atcaactgcc acaccatcat gctgcagccg 2340
ttcccgcagt acgatgcatc tcaggttgat gaagccgcac tggccgacac cgaatggctg 2400
aaacaggcga tcgttgccgt acgtaacatc cgtgcagaaa tgaacatcgc gccgggcaaa 2460
ccgctggagc tgctgctgcy tggttgcagc gcgatgcag aacgtcgcgt aatgaaaac 2520
cgtggcttcc tgcaaaccct ggcgcgtctg gaaagtatca ccgtgctgcc tgccgatgac 2580
aaaggtccgg tttccgttac gaagatcatc gacggtgcag agctgctgat cccgatggct 2640
ggcctcatca acaagaaga tgagctggcg cgtctggcga aagaagtggc gaagattgaa 2700

ggtgaaatca gccgtatcga gaacaaactg gcgaacgaag gctttgtcgc ccgcgcaccg 2760
gaagcggtea tgcgaaaaga gcgtgagaag ctggaaggct atgcggaagc gaaagcgaaa 2820
ctgattgaac agcaggctgt tatcgccgcy ctg 2853

```

<210> 24

<211> 951

<212> PRT

<213> E. coli

<400> 24

```

Met Glu Lys Thr Tyr Asn Pro Gln Asp Ile Glu Gln Pro Leu Tyr Glu
1          5          10          15

```

```

His Trp Glu Lys Gln Gly Tyr Phe Lys Pro Asn Gly Asp Glu Ser Gln
20          25          30

```

```

Glu Ser Phe Cys Ile Met Ile Pro Pro Asn Val Thr Gly Ser Leu
35          40          45

```

```

His Met Gly His Ala Phe Gln Gln Thr Ile Met Asp Thr Met Ile Arg
50          55          60

```

Tyr Gln Arg Met Gln Gly Lys Asn Thr Leu Trp Gln Val Gly Thr Asp
 65 70 75 80

His Ala Gly Ile Ala Thr Gln Met Val Val Glu Arg Lys Ile Ala Ala
 85 90 95

Glu Glu Gly Lys Thr Arg His Asp Tyr Gly Arg Glu Ala Phe Ile Asp
 100 105 110

Lys Ile Trp Glu Trp Lys Ala Glu Ser Gly Gly Thr Ile Thr Arg Gln
 115 120 125

Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr
 130 135 140

Met Asp Glu Gly Leu Ser Asn Ala Val Lys Glu Val Phe Val Arg Leu
 145 150 155 160

Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp
 165 170 175

Pro Lys Leu Arg Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu
 180 185 190

Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala
 195 200 205

Lys Thr Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro
 210 215 220

Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
 225 230 235 240

Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn
 245 250 255

Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu Lys Gly
 260 265 270

Thr Gly Cys Val Lys Ile Thr Pro Ala His Asp Phe Asn Asp Tyr Glu
 275 280 285

Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile Leu Thr Phe Asp
 290 295 300

Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu
 305 310 315 320

Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala Glu Phe Gln Lys Leu Glu
 325 330 335

Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly
 340 345 350

Leu Leu Glu Glu Ile Lys Pro His Asp Leu Thr Val Pro Tyr Gly Asp
 355 360 365

Arg Gly Gly Val Val Ile Glu Pro Met Leu Thr Asp Gln Trp Tyr Val
 370 375 380

Arg Ala Asp Val Leu Ala Lys Pro Ala Val Glu Ala Val Glu Asn Gly
 385 390 395 400

Asp Ile Gln Phe Val Pro Lys Gln Tyr Glu Asn Met Tyr Phe Ser Trp
 405 410 415

Met Arg Asp Ile Gln Asp Trp Cys Ile Ser Arg Gln Leu Trp Trp Gly
 420 425 430

His Arg Ile Pro Ala Trp Tyr Asp Glu Ala Gly Asn Val Tyr Val Gly
 435 440 445

Arg Asn Glu Asp Glu Val Arg Lys Glu Asn Asn Leu Gly Ala Asp Val
 450 455 460

Val Leu Arg Gln Asp Glu Asp Val Leu Asp Thr Trp Phe Ser Ser Ala
 465 470 475 480

Leu Trp Thr Phe Ser Thr Leu Gly Trp Pro Glu Asn Thr Asp Ala Leu
 485 490 495

Arg Gln Phe His Pro Thr Ser Val Met Val Ser Gly Phe Asp Ile Ile
 500 505 510

Phe Phe Trp Ile Ala Arg Met Ile Met Met Thr Met His Phe Ile Lys
 515 520 525

Asp Glu Asn Gly Lys Pro Gln Val Pro Phe His Thr Val Tyr Met Thr
 530 535 540

Gly Leu Ile Arg Asp Asp Glu Gly Gln Lys Met Ser Lys Ser Lys Gly
 545 550 555 560

Asn Val Ile Asp Pro Leu Asp Met Val Asp Gly Ile Ser Leu Pro Glu
 565 570 575

Leu Leu Glu Lys Arg Thr Gly Asn Met Met Gln Pro Gln Leu Ala Asp
 580 585 590

Lys Ile Arg Lys Arg Thr Glu Lys Gln Phe Pro Asn Gly Ile Glu Pro
 595 600 605

His Gly Thr Asp Ala Leu Arg Phe Thr Leu Ala Ala Leu Ala Ser Thr
 610 615 620

Gly Arg Asp Ile Asn Trp Asp Met Lys Arg Leu Glu Gly Tyr Arg Asn
 625 630 635 640

Phe Cys Asn Lys Leu Trp Asn Ala Ser Arg Phe Val Leu Met Asn Thr
 645 650 655

Glu Gly Gln Asp Cys Gly Phe Asn Gly Gly Glu Met Thr Leu Ser Leu
 660 665 670

Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr
 675 680 685

Arg Glu Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu
 690 695 700

Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr
 705 710 715 720

Lys Pro Val Met Asn Gly Gly Thr Glu Ala Glu Leu Arg Gly Thr Arg
 725 730 735

His Thr Leu Val Thr Val Leu Glu Gly Leu Leu Arg Leu Ala His Pro
 740 745 750

Ile Ile Pro Phe Ile Thr Glu Thr Ile Trp Gln Arg Val Lys Val Leu
 755 760 765

Cys Gly Ile Thr Ala Asp Thr Ile Met Leu Gln Pro Phe Pro Gln Tyr
 770 775 780

Asp Ala Ser Gln Val Asp Glu Ala Ala Leu Ala Asp Thr Glu Trp Leu
 785 790 795 800

Lys Gln Ala Ile Val Ala Val Arg Asn Ile Arg Ala Glu Met Asn Ile
 805 810 815

Ala Pro Gly Lys Pro Leu Glu Leu Leu Leu Arg Gly Cys Ser Ala Asp
 820 825 830

Ala Glu Arg Arg Val Asn Glu Asn Arg Gly Phe Leu Gln Thr Leu Ala
 835 840 845

Arg Leu Glu Ser Ile Thr Val Leu Pro Ala Asp Asp Lys Gly Pro Val
 850 855 860

Ser Val Thr Lys Ile Ile Asp Gly Ala Glu Leu Leu Ile Pro Met Ala
 865 870 875 880

Gly Leu Ile Asn Lys Glu Asp Glu Leu Ala Arg Leu Ala Lys Glu Val
 885 890 895

Ala Lys Ile Glu Gly Glu Ile Ser Arg Ile Glu Asn Lys Leu Ala Asn
 900 905 910

Glu Gly Phe Val Ala Arg Ala Pro Glu Ala Val Ile Ala Lys Glu Arg
 915 920 925

Glu Lys Leu Glu Gly Tyr Ala Glu Ala Lys Ala Lys Leu Ile Glu Gln
 930 935 940

Gln Ala Val Ile Ala Ala Leu
 945 950

<210> 25
 <211> 1290
 <212> DNA
 <213> E. coli

<400> 25
 atgctcgatc ccaatctgct gcgtaatgag ccagacgcag tcgctgaaaa actggcacgc 60
 cggggcttta agctggatgt agataagctg ggcgctcttg aagagcgtcg taaagtattg 120
 caggtcaaaa cggaaaacct gcaagcggag cgtaactccc gatcgaaatc cattggccag 180
 gcgaaagcgc gcggggaaga tatcgagcct ttacgtctgg aagtgaacaa actgggcgaa 240
 gagctggatg cagcaaaagc cgagctggat gctttacagg ctgaaattcg cgatatcgcg 300
 ctgaccatcc ctaacctgcc tgcagatgaa gtgccggtag gtaaagacga aatgacaac 360
 gttgaagtca gccgctgggg taccctcgct gagtttgact ttgaagttcg tgaccacgtg 420
 acgctgggtg aatgcaactc tggcctcgac tttgcagctg cagttaagct gactggttcc 480
 cgctttgtgg taatgaaagg gcagattgct cgcatgcacc gcgcactgtc gcagtttatg 540
 ctggatctgc ataccgaaca gcatggctac agtgagaact atgttccgta cctggttaac 600
 caggacacgc tgtacggtac gggtaactg ccgaaatttg ctggcgatct gttccatact 660
 cgtccgctgg aagaagaagc agacaccagt aactatgcgc tgatcccaac ggcagaagtt 720
 ccgctgacta acctgggtgc cggtgaaatc atcgatgaag atgatctgcc aattaagatg 780
 accgcccaca ccccatgctt ccgttctgaa gccggttcat atggtcgtga caccctggtt 840
 ctgatccgta tgcaccagtt cgacaaagtt gaaatggtgc agatcgtgcg ccgagaagac 900
 tcaatggcgg cgctggaaga gatgactggt catgcagaaa aagtcctgca gttgctgggc 960
 ctgccgtacc gtaaaatcat cctttgcact ggcgacatgg gctttggcgc ttgcaaaact 1020
 tacgacctgg aagtatggat cccggcacag aacacctacc gtgagatctc ttctgctcc 1080
 aacgtttggg atttccaggc acgtcgtatg caggcacggt gccgcagcaa gtcggacaag 1140
 aaaaccctgc tggttcatac cctgaacggt tctggtctgg ctggtggtcg tacgctggtt 1200
 gcagtaatgg aaaactatca gcaggctgat ggtcgtattg aagtaccaga agttctgcgt 1260
 ccgtatatga acggactgga atatattggc 1290

<210> 26
 <211> 430
 <212> PRT
 <213> E. coli

<400> 26
 Met Leu Asp Pro Asn Leu Leu Arg Asn Glu Pro Asp Ala Val Ala Glu
 1 5 10 15
 Lys Leu Ala Arg Arg Gly Phe Lys Leu Asp Val Asp Lys Leu Gly Ala
 20 25 30
 Leu Glu Glu Arg Arg Lys Val Leu Gln Val Lys Thr Glu Asn Leu Gln
 35 40 45

Ala Glu Arg Asn Ser Arg Ser Lys Ser Ile Gly Gln Ala Lys Ala Arg
50 55 60

Gly Glu Asp Ile Glu Pro Leu Arg Leu Glu Val Asn Lys Leu Gly Glu
65 70 75 80

Glu Leu Asp Ala Ala Lys Ala Glu Leu Asp Ala Leu Gln Ala Glu Ile
85 90 95

Arg Asp Ile Ala Leu Thr Ile Pro Asn Leu Pro Ala Asp Glu Val Pro
100 105 110

Val Gly Lys Asp Glu Asn Asp Asn Val Glu Val Ser Arg Trp Gly Thr
115 120 125

Pro Arg Glu Phe Asp Phe Glu Val Arg Asp His Val Thr Leu Gly Glu
130 135 140

Met His Ser Gly Leu Asp Phe Ala Ala Ala Val Lys Leu Thr Gly Ser
145 150 155 160

Arg Phe Val Val Met Lys Gly Gln Ile Ala Arg Met His Arg Ala Leu
165 170 175

Ser Gln Phe Met Leu Asp Leu His Thr Glu Gln His Gly Tyr Ser Glu
180 185 190

Asn Tyr Val Pro Tyr Leu Val Asn Gln Asp Thr Leu Tyr Gly Thr Gly
195 200 205

Gln Leu Pro Lys Phe Ala Gly Asp Leu Phe His Thr Arg Pro Leu Glu
210 215 220

Glu Glu Ala Asp Thr Ser Asn Tyr Ala Leu Ile Pro Thr Ala Glu Val
225 230 235 240

Pro Leu Thr Asn Leu Val Arg Gly Glu Ile Ile Asp Glu Asp Asp Leu
245 250 255

Pro Ile Lys Met Thr Ala His Thr Pro Cys Phe Arg Ser Glu Ala Gly
260 265 270

Ser Tyr Gly Arg Asp Thr Arg Gly Leu Ile Arg Met His Gln Phe Asp
275 280 285

Lys Val Glu Met Val Gln Ile Val Arg Pro Glu Asp Ser Met Ala Ala
290 295 300

Leu Glu Glu Met Thr Gly His Ala Glu Lys Val Leu Gln Leu Leu Gly
305 310 315 320

Leu Pro Tyr Arg Lys Ile Ile Leu Cys Thr Gly Asp Met Gly Phe Gly
325 330 335

Ala Cys Lys Thr Tyr Asp Leu Glu Val Trp Ile Pro Ala Gln Asn Thr
340 345 350

Tyr Arg Glu Ile Ser Ser Cys Ser Asn Val Trp Asp Phe Gln Ala Arg
 355 360 365

Arg Met Gln Ala Arg Cys Arg Ser Lys Ser Asp Lys Lys Thr Arg Leu
 370 375 380

Val His Thr Leu Asn Gly Ser Gly Leu Ala Val Gly Arg Thr Leu Val
 385 390 395 400

Ala Val Met Glu Asn Tyr Gln Gln Ala Asp Gly Arg Ile Glu Val Pro
 405 410 415

Glu Val Leu Arg Pro Tyr Met Asn Gly Leu Glu Tyr Ile Gly
 420 425 430

<210> 27
 <211> 981
 <212> DNA
 <213> E. coli

<400> 27
 atgtcacatc tgcgagaact ggttgccagt gcgaaggcgg ccattagcca ggcgtcagat 60
 gttgccgcgt tagataatgt gcgcgtcgaa tatttgggta aaaaagggca cttaccctt 120
 cagatgacga ccttgcgtga gctgccgcca gaagagcgtc cggcagctgg tgcggttacc 180
 aacgaagcga aagagcaggt tcagcaggcg ctgaatgcgc gtaaagcggg actggaagc 240
 gctgcactga atgcgcgtct ggccggcgaa acgattgatg tctctctgcc aggtcgtcgc 300
 attgaaaacg gcggtctgca tccggttacc cgtaccatcg accgtatcga aagtttcttc 360
 ggtgagcttg gctttaccgt ggcaaccggg ccggaatcg aagacgatta tcataacttc 420
 gatgctctga acattcctg tcaccacccg gcgcgcgctg accacgacac tttctggttt 480
 gacactaccg gctgctgctg taccagacc tctggcgtac agatccgcac catgaaagcc 540
 cagcagccac cgattogtat catcgcgcct ggccgtggtt atcgtaacga ctacgaccag 600
 actcacacgc cgatgttcca tcagatggaa ggtctgattg ttgataccta catcagcttt 660
 accaacctga aaggcagcgt gcacgacttc ctgcgtaact tctttgagga agatttgcag 720
 attcgttcc gtccttctta cttcccgttt accgaacctt ctgcagaagt ggacgtcatg 780
 ggtaaaaacg gtaaatggct ggaagtgctg ggctgcggga tggatcatcc gaacgtggtg 840
 cgtaacgttg gcatcgaccc ggaagtttac tctggtttcg ccttcgggat ggggatggag 900
 cgtctgacta tgttgcgtta cggcgtcacc gacctgcgtt cattcttcga aaacgatctg 960
 cgtttctca aacagtttaa a 981

<210> 28
 <211> 327
 <212> PRT
 <213> E. coli

<400> 28

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

Leu Arg Tyr Gly Val Thr Asp Leu Arg Ser Phe Phe Glu Asn Asp Leu
 305 310 315 320

Arg Phe Leu Lys Gln Phe Lys
 325

<210> 29

<211> 642

<212> PRT

<213> E. coli

<400> 29

Met Pro Val Ile Thr Leu Pro Asp Gly Ser Gln Arg His Tyr Asp His
 1 5 10 15

Ala Val Ser Pro Met Asp Val Ala Leu Asp Ile Gly Pro Gly Leu Ala
 20 25 30

Lys Ala Cys Ile Ala Gly Arg Val Asn Gly Glu Leu Val Asp Ala Cys
 35 40 45

Asp Leu Ile Glu Asn Asp Ala Gln Leu Ser Ile Ile Thr Ala Lys Asp
 50 55 60

Glu Glu Gly Leu Glu Ile Ile Arg His Ser Cys Ala His Leu Leu Gly
 65 70 75 80

His Ala Ile Lys Gln Leu Trp Pro His Thr Lys Met Ala Ile Gly Pro
 85 90 95

Val Ile Asp Asn Gly Phe Tyr Tyr Asp Val Asp Leu Asp Arg Thr Leu
 100 105 110

Thr Gln Glu Asp Val Glu Ala Leu Glu Lys Arg Met His Glu Leu Ala
 115 120 125

Glu Lys Asn Tyr Asp Val Ile Lys Lys Lys Val Ser Trp His Glu Ala
 130 135 140

Arg Glu Thr Phe Ala Asn Arg Gly Glu Ser Tyr Lys Val Ser Ile Leu
 145 150 155 160

Asp Glu Asn Ile Ala His Asp Asp Lys Pro Gly Leu Tyr Phe His Glu
 165 170 175

Glu Tyr Val Asp Met Cys Arg Gly Pro His Val Pro Asn Met Arg Phe
 180 185 190

Cys His His Phe Lys Leu Met Lys Thr Ala Gly Ala Tyr Trp Arg Gly
 195 200 205

Asp Ser Asn Asn Lys Met Leu Gln Arg Ile Tyr Gly Thr Ala Trp Ala
 210 215 220

Asp Lys Lys Ala Leu Asn Ala Tyr Leu Gln Arg Leu Glu Glu Ala Ala

Met Asn Ile Thr Asp Ser Gln Ser Glu Tyr Val Asn Glu Leu Thr Gln
545 550 555 560

Lys Leu Ser Asn Ala Gly Ile Arg Val Lys Ala Asp Leu Arg Asn Glu
565 570 575

Lys Ile Gly Phe Lys Ile Arg Glu His Thr Leu Arg Arg Val Pro Tyr
580 585 590

Met Leu Val Cys Gly Asp Lys Glu Val Glu Ser Gly Lys Val Ala Val
595 600 605

Arg Thr Arg Arg Gly Lys Asp Leu Gly Ser Met Asp Val Asn Glu Val
610 615 620

Ile Glu Lys Leu Gln Gln Glu Ile Arg Ser Arg Ser Leu Lys Gln Leu
625 630 635 640
Glu Glu

<210> 30

<211> 41

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 30

gcctgctgcg taccagacc gcgggcgtac agatccgcac c 41

<210> 31

<211> 41

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 31

gcctgctgcg taccagacc gcgggcgtac agatccgcac c 41

<210> 32

<211> 42

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 32

ctacccgcct gctgcgtacc gcgacctctg gcgtacagat cc 42

<210> 33

<211> 42

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 33

ctacccgcct gctgcgtacc ggcacctctg gcgtacagat cc 42

<210> 34

<211> 41

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 34

ggtgcggatc tgtacgcccg cggctctgggt acgcagcagg c 41

<210> 35

<211> 41

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 35

ggtgcggatc tgtacgcccg cggctctgggt acgcagcagg c 41

<210> 36

<211> 42

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 36

ggatctgtac gccagaggtc gcggtacgca gcaggcgggt ag 42

<210> 37

<211> 42

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 37

ggatctgtac gccagagtg ccggtacgca gcaggcgggt ag 42

<210> 38

<211> 97

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 38

ggcgtaatac gactcactat agcccggata gctcagtcgg tagagcagg gattgaaaat 60

ccccgtgtcc ttggttcgat tccgagtccg ggcacca 97

<210> 39

<211> 76

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 39

gcccggauag cucagucggu agagcagggg auugaaaauc cccguguccu ugguucgauu 60

ccgaguccgg gcacca 76

<210> 40

<211> 92

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 40

ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgtttccgtg 60

actacaagga cgacgacgac aagtaagctt cg 92

<210> 41

<211> 71

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 41

```
ggguuaacuu uaacaaggag aaaaacaugc guuuccguga cuacaaggac gacgacgaca 60
aguaagcuuc g 71
```

<210> 42

<211> 12

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> MISC_FEATURE

<222> (3)..(3)

<223> Xaa is phenylalanine or methyl phenylalanine

<400> 42

```
Met Arg Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
1           5           10
```

<210> 43

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 43

```
gtttctgcat catgatcccg gctccgaacg tcaccggcag tttgcatat 49
```

<210> 44

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 44

```
gtttctgcat catgatcccg ggtccgaacg tcaccggcag tttgcatat 49
```

<210> 45

<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 45
gtttctgcat catgatcccg ccgccggctg tcaccggcag tttgcatat 49

<210> 46
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 46
gtttctgcat catgatcccg gctccggctg tcaccggcag tttgcatat 49

<210> 47
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 47
gtttctgcat catgatcccg ggtccggctg tcaccggcag tttgcatat 49

<210> 48
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 48
gtttctgcat catgatcccg ccgccgagtg tcaccggcag tttgcatat 49

<210> 49
<211> 49
<212> DNA
<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 49

gtttctgcat catgatcccg gctccgagtg tcaccggcag tttgcatat 49

<210> 50

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 50

gtttctgcat catgatcccg ggtccgagtg tcaccggcag tttgcatat 49

<210> 51

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 51

gtttctgcat catgatcccg ccgccggtag tcaccggcag tttgcatat 49

<210> 52

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 52

gtttctgcat catgatcccg gctccggtag tcaccggcag tttgcatat 49

<210> 53

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 53

gtttctgcat catgatcccg ggtccggtag tcaccggcag tttgcatat 49

<210> 54

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 54

gtttctgcat catgatcccg ccgccgggtg tcaccggcag tttgcatat 49

<210> 55

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 55

gtttctgcat catgatcccg gctccgggtg tcaccggcag tttgcatat 49

<210> 56

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 56

gtttctgcat catgatcccg ggtccgggtg tcaccggcag tttgcatat 49

<210> 57

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 57

gtttctgcat catgatcccg ccgccggatg tcaccggcag tttgcatat 49

<210> 58

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 58

gtttctgcat catgatcccg gctccggatg tcaccggcag tttgcatat 49

<210> 59

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 59

gtttctgcat catgatcccg ggtccggatg tcaccggcag tttgcatat 49

<210> 60

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 60

tggcaggtcg gtactgctca cgccgggatc gct 33

<210> 61

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 61

tggcaggtcg gtactagtca cgccgggatc gct 33

<210> 62

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 62

tggcaggtcg gtactgtaca cgccgggatc gct 33

<210> 63

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 63

tggcaggtcg gtactgttca cgccgggatc gct 33

<210> 64

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 64

gtactgacca cgccgggatc tataaccaga tggtcgttga gcg 43

<210> 65

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 65

gtactgacca cgccgggatc tggaccaga tggtcgttga gcg 43

<210> 66

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 66

gtactgacca cgccgggatc tcaaccaga tggtcgttga gcg 43

<210> 67
<211> 43
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 67
gtactgacca cgccgggatc atgaccaga tggcgttga gcg 43

<210> 68
<211> 43
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 68
gtactgacca cgccgggatc aaaaccaga tggcgttga gcg 43

<210> 69
<211> 43
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 69
gtactgacca cgccgggatc aacaccaga tggcgttga gcg 43

<210> 70
<211> 43
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 70
gtactgacca cgccgggatc gtgaccaga tggcgttga gcg 43

<210> 71
<211> 43
<212> DNA
<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 71

gtactgacca cgccgggatc ctgaccaga tggctgtga gcg 43

<210> 72

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 72

gatcccgccg ccgggtgtct atggcagttt gcatatgggt cac 43

<210> 73

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 73

gatcccgccg ccgggtgtct gggcagttt gcatatgggt cac 43

<210> 74

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 74

gatcccgccg ccgggtgtct caggcagttt gcatatgggt cac 43

<210> 75

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 75

gatcccgccg ccgggtgtca tgggcagttt gcatatgggt cac 43

<210> 76

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 76

gatcccgccg ccgggtgtca aaggcagttt gcatatgggt cac 43

<210> 77

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 77

gatcccgccg ccgggtgtca acggcagttt gcatatgggt cac 43

<210> 78

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 78

gatcccgccg ccgggtgtcg tgggcagttt gcatatgggt cac 43

<210> 79

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 79

gatcccgccg ccgggtgtcc tgggcagttt gcatatgggt cac 43

<210> 80

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 80

atatgcaaac tgccggtgac gttcggagcc gggatcatga tgcagaaac 49

<210> 81

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 81

atatgcaaac tgccggtgac gttcggaccc gggatcatga tgcagaaac 49

<210> 82

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 82

atatgcaaac tgccggtgac agccggcggc gggatcatga tgcagaaac 49

<210> 83

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 83

atatgcaaac tgccggtgac agccggagcc gggatcatga tgcagaaac 49

<210> 84

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 84

atatgcaaac tgccggtgac agccggaccc gggatcatga tgcagaaac 49

<210> 85

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 85

atatgcaaac tgccggtgac actcggcggc gggatcatga tgcagaaac 49

<210> 86

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 86

atatgcaaac tgccggtgac actcggagcc gggatcatga tgcagaaac 49

<210> 87

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 87

atatgcaaac tgccggtgac actcggaccc gggatcatga tgcagaaac 49

<210> 88

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 88

atatgcaaac tgccggtgac taccggcggc gggatcatga tgcagaaac 49

<210> 89
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 89
atatgcaaac tgccggtgac taccggagcc gggatcatga tgcagaaac 49

<210> 90
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 90
atatgcaaac tgccggtgac taccggaccc gggatcatga tgcagaaac 49

<210> 91
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 91
atatgcaaac tgccggtgac acccggcggc gggatcatga tgcagaaac 49

<210> 92
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 92
atatgcaaac tgccggtgac acccggagcc gggatcatga tgcagaaac 49

<210> 93
<211> 49
<212> DNA
<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 93

atatgcaaac tgccggtgac acccggaccc gggatcatga tgcagaaac 49

<210> 94

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 94

atatgcaaac tgccggtgac atccggcggc gggatcatga tgcagaaac 49

<210> 95

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 95

atatgcaaac tgccggtgac atccggagcc gggatcatga tgcagaaac 49

<210> 96

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 96

atatgcaaac tgccggtgac atccggaccc gggatcatga tgcagaaac 49

<210> 97

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 97

agcgatcccg gcgtgagcag taccgacctg cca 33

<210> 98

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 98

agcgatcccg gcgtgactag taccgacctg cca 33

<210> 99

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 99

agcgatcccg gcgtgtacag taccgacctg cca 33

<210> 100

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 100

agcgatcccg gcgtgaccag taccgacctg cca 33

<210> 101

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 101

cgctcaacga ccatctgggt atagatcccg gcgtggtcag tac 43

<210> 102

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 102

cgctcaacga ccatctgggt ccagatcccg gcgtggtcag tac 43

<210> 103

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 103

cgctcaacga ccatctgggt tgagatcccg gcgtggtcag tac 43

<210> 104

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 104

cgctcaacga ccatctgggt catgatcccg gcgtggtcag tac 43

<210> 105

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 105

cgctcaacga ccatctgggt ttgatcccg gcgtggtcag tac 43

<210> 106

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 106

cgctcaacga ccatctgggt gttgatcccg gcgtggtcag tac 43

<210> 107

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 107

cgctcaacga ccatctgggt cacgatcccg gcgtggtcag tac 43

<210> 108

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 108

cgctcaacga ccatctgggt caggatcccg gcgtggtcag tac 43

<210> 109

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 109

gtgaccata tgcaaactgc catagacacc cggcggcggg atc 43

<210> 110

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 110

gtgaccata tgcaaactgc cccagacacc cggcggcggg atc 43

<210> 111

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 111

gtgaccata tgcaaactgc ctgagacacc cggcggcggg atc 43

<210> 112

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 112

gtgaccata tgcaaactgc ccatgacacc cggcggcggg atc 43

<210> 113

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 113

gtgaccata tgcaaactgc cttgacacc cggcggcggg atc 43

<210> 114

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 114

gtgaccata tgcaaactgc cgttgacacc cggcggcggg atc 43

<210> 115

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 115

gtgaccata tgcaaactgc ccacgacacc cggcggcggg atc 43

<210> 116

<211> 43

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 116

gtgaccata tgcaaactgc ccaggacacc cggcggcggg atc 43

<210> 117

<211> 98

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 117

ggcgtaatac gactcactat agcgtccgta gctcagttgg ttagagcacc accttgacat 60

ggtgggggtc ggtggttcga gtccactcgg acgcacca 98

<210> 118

<211> 77

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 118

gcguccguag cucaguuggu uagagcacca ccuugacaug gugggggucg gugguucgag 60

uccacucgga cgcacca 77

<210> 119

<211> 92

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 119
 ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgtgtccgtg 60
 actacaagga cgacgacgac aagtaagctt cg 92

<210> 120

<211> 71

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 120
 ggguaaacuu uaacaaggag aaaaacaugc guguccguga cuacaaggac gacgacgaca 60
 aguaagcuuc g 71

<210> 121

<211> 95

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 121
 ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgtgtcgtcc 60
 gtgactacaa ggacgacgac gacaagtaag ctctg 95

<210> 122

<211> 74

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 122
 ggguaaacuu uaacaaggag aaaaacaugc gugucguccg ugacuacaag gacgacgacg 60
 acaaguaagc uucg 74

<210> 123

<211> 98

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 123
 ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgtgtcgtcg 60
 tccgtgacta caaggacgac gacgacaagt aagcttcg 98

<210> 124
 <211> 77
 <212> RNA
 <213> Artificial

<220>
 <223> artificially synthesized sequence

<400> 124
 ggguaaacuu uaacaaggag aaaaacaugc gugucgucgu ccgugacuac aaggacgacg 60
 acgacaagua agcuucg 77

<210> 125
 <211> 12
 <212> PRT
 <213> Artificial

<220>
 <223> artificially synthesized sequence

<220>
 <221> MISC_FEATURE
 <222> (3)..(3)
 <223> Xaa is valine or methyl valine

<400> 125
 Met Arg Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
 1 5 10

<210> 126
 <211> 13
 <212> PRT
 <213> Artificial

<220>
 <223> artificially synthesized sequence

<220>
 <221> MISC_FEATURE
 <222> (3)..(4)
 <223> Xaa is valine or methyl valine

<400> 126
 Met Arg Xaa Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
 1 5 10

<210> 127

<211> 14

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> MISC_FEATURE

<222> (3)..(5)

<223> Xaa is valine or methyl valine

<400> 127

Met Arg Xaa Xaa Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
1 5 10

<210> 128

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 128

gtaactatgc gctgatccca gctgcagaag ttccgctgac taacctggt 49

<210> 129

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 129

gtaactatgc gctgatccca agtgcagaag ttccgctgac taacctggt 49

<210> 130

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 130

gtaactatgc gctgatccca ggtgcagaag ttccgctgac taacctggt 49

<210> 131

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 131

gtaactatgc gctgatcca acggcagctg ttccgctgac taacctggt 49

<210> 132

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 132

gtaactatgc gctgatcca acggcagtag ttccgctgac taacctggt 49

<210> 133

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 133

gtaactatgc gctgatcca acggcaagtg ttccgctgac taacctggt 49

<210> 134

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 134

ctggtcata cctggctgg ttctggtctg gct 33

<210> 135

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 135

gtaactatgc gctgatccca gctgcagctg ttccgctgac taacctggt 49

<210> 136

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 136

gtaactatgc gctgatccca gctgcagtag ttccgctgac taacctggt 49

<210> 137

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 137

gtaactatgc gctgatccca gctgcaagtg ttccgctgac taacctggt 49

<210> 138

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 138

gtaactatgc gctgatccca agtgcagctg ttccgctgac taacctggt 49

<210> 139

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 139

gtaactatgc gctgatccca agtgcaagtg ttccgctgac taacctggt 49

<210> 140

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 140

gtaactatgc gctgatccca agtgcaagtg ttccgctgac taacctggt 49

<210> 141

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 141

gtaactatgc gctgatccca ggtgcagctg ttccgctgac taacctggt 49

<210> 142

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 142

gtaactatgc gctgatccca ggtgcagtag ttccgctgac taacctggt 49

<210> 143

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 143

gtaactatgc gctgatccca ggtgcaagtg ttccgctgac taacctggt 49

<210> 144

<211> 33
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 144
cagttcgaca aagttgctat ggtgcagatc gtg 33

<210> 145
<211> 33
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 145
atgcaccagt tcgacgctgt tgaatggtg cag 33

<210> 146
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 146
gtaactatgc gctgatcca acggcaggtg ttccgctgac taacctggt 49

<210> 147
<211> 49
<212> DNA
<213> Artificial

<220>
<223> artificially synthesized sequence

<400> 147
gtaactatgc gctgatcca acggcagacg ttccgctgac taacctggt 49

<210> 148
<211> 49
<212> DNA
<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 148

gtaactatgc gctgatccca agtgcaggtg tccgctgac taacctggt 49

<210> 149

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 149

gtaactatgc gctgatccca agtgcagacg tccgctgac taacctggt 49

<210> 150

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 150

accaggtag tcagcggaac ttctgcagct gggatcagcg catagttac 49

<210> 151

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 151

accaggtag tcagcggaac ttctgcactt gggatcagcg catagttac 49

<210> 152

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 152

accaggtag tcagcggaac ttctgacct gggatcagcg catagttac 49

<210> 153

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 153

accaggtag tcagcggaac agctgccgtt gggatcagcg catagttac 49

<210> 154

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 154

accaggtag tcagcggaac tactgccgtt gggatcagcg catagttac 49

<210> 155

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 155

accaggtag tcagcggaac actgccgtt gggatcagcg catagttac 49

<210> 156

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 156

agccagacca gaaccagcca gggatgaac cag 33

<210> 157

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 157

accaggttag tcagcggaac agctgcagct gggatcagcg catagttac 49

<210> 158

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 158

accaggttag tcagcggaac tactgcagct gggatcagcg catagttac 49

<210> 159

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 159

accaggttag tcagcggaac acttgcagct gggatcagcg catagttac 49

<210> 160

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 160

accaggttag tcagcggaac agctgcactt gggatcagcg catagttac 49

<210> 161

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 161

accaggttag tcagcggaac tactgcactt gggatcagcg catagttac 49

<210> 162

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 162

accaggttag tcagcggaac actgcactt gggatcagcg catagttac 49

<210> 163

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 163

accaggttag tcagcggaac agctgcacct gggatcagcg catagttac 49

<210> 164

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 164

accaggttag tcagcggaac tactgcacct gggatcagcg catagttac 49

<210> 165

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 165

accaggttag tcagcggaac actgcacct gggatcagcg catagttac 49

<210> 166

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 166

cacgatctgc accatagcaa cttgtcgaa ctg 33

<210> 167

<211> 33

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 167

ctgcaccatt tcaacagcgt cgaactggtg cat 33

<210> 168

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 168

accaggttag tcagcggaac acctgccgtt gggatcagcg catagttac 49

<210> 169

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 169

accaggttag tcagcggaac gtctgccgtt gggatcagcg catagttac 49

<210> 170

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 170

accaggttag tcagcggaac acctgcactt gggatcagcg catagttac 49

<210> 171

<211> 49

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 171

accaggttag tcagcggaac gtctgcactt gggatcagcg catagttac 49

<210> 172

<211> 114

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 172

ggcgtaatac gactcactat aggtgaggtg gccgagaggg tgaaggcgct cccctgctaa 60

gggagtatgc ggtcaaaagc tgcattccggg gttcgaatcc ccgcctcacc gcca 114

<210> 173

<211> 93

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 173

ggugaggugg ccgagaggcu gaaggcgcuc ccugcuaag ggaguangcg gucaaaagcu 60

gcauccgggg uucgaauccc cgccucaccg cca 93

<210> 174

<211> 92

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 174
 ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgttcccgtg 60
 actacaagga cgacgacgac aagtaagctt cg 92

<210> 175

<211> 71

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 175
 ggguaaacuu uaacaaggag aaaaacaugc guucccguga cuacaaggac gacgacgaca 60
 aguaagcuuc g 71

<210> 176

<211> 12

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> MISC_FEATURE

<222> (3)..(3)

<223> Xaa is serine or methyl serine

<400> 176
 Met Arg Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
 1 5 10

<210> 177

<211> 103

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 177
 ggcgtaatac gactcactat agggttaact ttaagaagga gatatacata tgaaggctgg 60
 tccgggtttt atgactaaga gtggtagtgg tagttaagct tcg 103

<210> 178

<211> 82

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 178

ggguuaacuu uaagaaggag auauacauau gaaggcuggu ccggguuuua ugacuaagag 60

ugguaguggu aguuagcuu cg 82

<210> 179

<211> 15

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> MISC_FEATURE

<222> (9)..(9)

<223> Xaa is threonine or methyl threonine

<400> 179

Met	Lys	Ala	Gly	Pro	Gly	Phe	Met	Xaa	Lys	Ser	Gly	Ser	Gly	Ser
1				5					10					15

<210> 180

<211> 7

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> MISC_FEATURE

<222> (4)..(4)

<223> Xaa is Asn (N), Tyr (Y), or Thr (T).

<220>

<221> MISC_FEATURE

<222> (5)..(5)

<223> Xaa is any amino acid.

<220>

<221> MISC_FEATURE

<222> (6)..(6)

<223> Xaa is Thr (T) or Ser (S).

Lys Ile Trp Glu Trp Lys Ala Glu Ser Gly Gly Thr Ile Thr Arg Gln
 115 120 125

Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr
 130 135 140

Met Asp Glu Gly Leu Ser Asn Ala Val Lys Glu Val Phe Val Arg Leu
 145 150 155 160

Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp
 165 170 175

Pro Lys Leu Arg Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu
 180 185 190

Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala
 195 200 205

Lys Thr Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro
 210 215 220

Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
 225 230 235 240

Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn
 245 250 255

Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu Lys Gly
 260 265 270

Thr Gly Cys Val Lys Ile Ala Pro Ala His Asp Phe Asn Asp Tyr Glu
 275 280 285

Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile Leu Thr Phe Asp
 290 295 300

Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu
 305 310 315 320

Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala Glu Phe Gln Lys Leu Glu
 325 330 335

Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly
 340 345 350

Leu Leu Glu Glu Ile Lys Pro His Asp Leu Thr Val Pro Tyr Gly Asp
 355 360 365

Arg Gly Gly Val Val Ile Glu Pro Met Leu Thr Asp Gln Trp Tyr Val
 370 375 380

Arg Ala Asp Val Leu Ala Lys Pro Ala Val Glu Ala Val Glu Asn Gly
 385 390 395 400

Asp Ile Gln Phe Val Pro Lys Gln Tyr Glu Asn Met Tyr Phe Ser Trp
 405 410 415

Met Arg Asp Ile Gln Asp Trp Cys Ile Ser Arg Gln Leu Trp Trp Gly
 420 425 430

His Arg Ile Pro Ala Trp Tyr Asp Glu Ala Gly Asn Val Tyr Val Gly
 435 440 445

Arg Asn Glu Asp Glu Val Arg Lys Glu Asn Asn Leu Gly Ala Asp Val
 450 455 460

Val Leu Arg Gln Asp Glu Asp Val Leu Asp Thr Trp Phe Ser Ser Ala
 465 470 475 480

Leu Trp Thr Phe Ser Thr Leu Gly Trp Pro Glu Asn Thr Asp Ala Leu
 485 490 495

Arg Gln Phe His Pro Thr Ser Val Met Val Ser Gly Phe Asp Ile Ile
 500 505 510

Phe Phe Trp Ile Ala Arg Met Ile Met Met Thr Met His Phe Ile Lys
 515 520 525

Asp Glu Asn Gly Lys Pro Gln Val Pro Phe His Thr Val Tyr Met Thr
 530 535 540

Gly Leu Ile Arg Asp Asp Glu Gly Gln Lys Met Ser Lys Ser Lys Gly
 545 550 555 560

Asn Val Ile Asp Pro Leu Asp Met Val Asp Gly Ile Ser Leu Pro Glu
 565 570 575

Leu Leu Glu Lys Arg Thr Gly Asn Met Met Gln Pro Gln Leu Ala Asp
 580 585 590

Lys Ile Arg Lys Arg Thr Glu Lys Gln Phe Pro Asn Gly Ile Glu Pro
 595 600 605

His Gly Thr Asp Ala Leu Arg Phe Thr Leu Ala Ala Leu Ala Ser Thr
 610 615 620

Gly Arg Asp Ile Asn Trp Asp Met Lys Arg Leu Glu Gly Tyr Arg Asn
 625 630 635 640

Phe Cys Asn Lys Leu Trp Asn Ala Ser Arg Phe Val Leu Met Asn Thr
 645 650 655

Glu Gly Gln Asp Cys Gly Phe Asn Gly Gly Glu Met Thr Leu Ser Leu
 660 665 670

Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr
 675 680 685

Arg Glu Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu
 690 695 700

Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr

Met Glu Lys Thr Tyr Asn Pro Gln Asp Ile Glu Gln Pro Leu Tyr Glu
 1 5 10 15

His Trp Glu Lys Gln Gly Tyr Phe Lys Pro Asn Gly Asp Glu Ser Gln
 20 25 30

Glu Ser Phe Cys Ile Met Ile Pro Pro Pro Gly Val Ser Gly Ser Leu
 35 40 45

His Met Gly His Ala Phe Gln Gln Thr Ile Met Asp Thr Met Ile Arg
 50 55 60

Tyr Gln Arg Met Gln Gly Lys Asn Thr Leu Trp Gln Val Gly Thr Asp
 65 70 75 80

His Ala Gly Ile Ala Thr Gln Met Val Val Glu Arg Lys Ile Ala Ala
 85 90 95

Glu Glu Gly Lys Thr Arg His Asp Tyr Gly Arg Glu Ala Phe Ile Asp
 100 105 110

Lys Ile Trp Glu Trp Lys Ala Glu Ser Gly Gly Thr Ile Thr Arg Gln
 115 120 125

Met Arg Arg Leu Gly Asn Ser Val Asp Trp Glu Arg Glu Arg Phe Thr
 130 135 140

Met Asp Glu Gly Leu Ser Asn Ala Val Lys Glu Val Phe Val Arg Leu
 145 150 155 160

Tyr Lys Glu Asp Leu Ile Tyr Arg Gly Lys Arg Leu Val Asn Trp Asp
 165 170 175

Pro Lys Leu Arg Thr Ala Ile Ser Asp Leu Glu Val Glu Asn Arg Glu
 180 185 190

Ser Lys Gly Ser Met Trp His Ile Arg Tyr Pro Leu Ala Asp Gly Ala
 195 200 205

Lys Thr Ala Asp Gly Lys Asp Tyr Leu Val Val Ala Thr Thr Arg Pro
 210 215 220

Glu Thr Leu Leu Gly Asp Thr Gly Val Ala Val Asn Pro Glu Asp Pro
 225 230 235 240

Arg Tyr Lys Asp Leu Ile Gly Lys Tyr Val Ile Leu Pro Leu Val Asn
 245 250 255

Arg Arg Ile Pro Ile Val Gly Asp Glu His Ala Asp Met Glu Lys Gly
 260 265 270

Thr Gly Cys Val Lys Ile Gly Pro Ala His Asp Phe Asn Asp Tyr Glu
 275 280 285

Val Gly Lys Arg His Ala Leu Pro Met Ile Asn Ile Leu Thr Phe Asp
 290 295 300

Gly Asp Ile Arg Glu Ser Ala Gln Val Phe Asp Thr Lys Gly Asn Glu
 305 310 315 320

Ser Asp Val Tyr Ser Ser Glu Ile Pro Ala Glu Phe Gln Lys Leu Glu
 325 330 335

Arg Phe Ala Ala Arg Lys Ala Val Val Ala Ala Val Asp Ala Leu Gly
 340 345 350

Leu Leu Glu Glu Ile Lys Pro His Asp Leu Thr Val Pro Tyr Gly Asp
 355 360 365

Arg Gly Gly Val Val Ile Glu Pro Met Leu Thr Asp Gln Trp Tyr Val
 370 375 380

Arg Ala Asp Val Leu Ala Lys Pro Ala Val Glu Ala Val Glu Asn Gly
 385 390 395 400

Asp Ile Gln Phe Val Pro Lys Gln Tyr Glu Asn Met Tyr Phe Ser Trp
 405 410 415

Met Arg Asp Ile Gln Asp Trp Cys Ile Ser Arg Gln Leu Trp Trp Gly
 420 425 430

His Arg Ile Pro Ala Trp Tyr Asp Glu Ala Gly Asn Val Tyr Val Gly
 435 440 445

Arg Asn Glu Asp Glu Val Arg Lys Glu Asn Asn Leu Gly Ala Asp Val
 450 455 460

Val Leu Arg Gln Asp Glu Asp Val Leu Asp Thr Trp Phe Ser Ser Ala
 465 470 475 480

Leu Trp Thr Phe Ser Thr Leu Gly Trp Pro Glu Asn Thr Asp Ala Leu
 485 490 495

Arg Gln Phe His Pro Thr Ser Val Met Val Ser Gly Phe Asp Ile Ile
 500 505 510

Phe Phe Trp Ile Ala Arg Met Ile Met Met Thr Met His Phe Ile Lys
 515 520 525

Asp Glu Asn Gly Lys Pro Gln Val Pro Phe His Thr Val Tyr Met Thr
 530 535 540

Gly Leu Ile Arg Asp Asp Glu Gly Gln Lys Met Ser Lys Ser Lys Gly
 545 550 555 560

Asn Val Ile Asp Pro Leu Asp Met Val Asp Gly Ile Ser Leu Pro Glu
 565 570 575

Leu Leu Glu Lys Arg Thr Gly Asn Met Met Gln Pro Gln Leu Ala Asp
 580 585 590

Lys Ile Arg Lys Arg Thr Glu Lys Gln Phe Pro Asn Gly Ile Glu Pro
 595 600 605

His Gly Thr Asp Ala Leu Arg Phe Thr Leu Ala Ala Leu Ala Ser Thr
 610 615 620

Gly Arg Asp Ile Asn Trp Asp Met Lys Arg Leu Glu Gly Tyr Arg Asn
 625 630 635 640

Phe Cys Asn Lys Leu Trp Asn Ala Ser Arg Phe Val Leu Met Asn Thr
 645 650 655

Glu Gly Gln Asp Cys Gly Phe Asn Gly Gly Glu Met Thr Leu Ser Leu
 660 665 670

Ala Asp Arg Trp Ile Leu Ala Glu Phe Asn Gln Thr Ile Lys Ala Tyr
 675 680 685

Arg Glu Ala Leu Asp Ser Phe Arg Phe Asp Ile Ala Ala Gly Ile Leu
 690 695 700

Tyr Glu Phe Thr Trp Asn Gln Phe Cys Asp Trp Tyr Leu Glu Leu Thr
 705 710 715 720

Lys Pro Val Met Asn Gly Gly Thr Glu Ala Glu Leu Arg Gly Thr Arg
 725 730 735

His Thr Leu Val Thr Val Leu Glu Gly Leu Leu Arg Leu Ala His Pro
 740 745 750

Ile Ile Pro Phe Ile Thr Glu Thr Ile Trp Gln Arg Val Lys Val Leu
 755 760 765

Cys Gly Ile Thr Ala Asp Thr Ile Met Leu Gln Pro Phe Pro Gln Tyr
 770 775 780

Asp Ala Ser Gln Val Asp Glu Ala Ala Leu Ala Asp Thr Glu Trp Leu
 785 790 795 800

Lys Gln Ala Ile Val Ala Val Arg Asn Ile Arg Ala Glu Met Asn Ile
 805 810 815

Ala Pro Gly Lys Pro Leu Glu Leu Leu Arg Gly Cys Ser Ala Asp
 820 825 830

Ala Glu Arg Arg Val Asn Glu Asn Arg Gly Phe Leu Gln Thr Leu Ala
 835 840 845

Arg Leu Glu Ser Ile Thr Val Leu Pro Ala Asp Asp Lys Gly Pro Val
 850 855 860

Ser Val Thr Lys Ile Ile Asp Gly Ala Glu Leu Leu Ile Pro Met Ala
 865 870 875 880

Gly Leu Ile Asn Lys Glu Asp Glu Leu Ala Arg Leu Ala Lys Glu Val
 885 890 895

Ala Lys Ile Glu Gly Glu Ile Ser Arg Ile Glu Asn Lys Leu Ala Asn
 900 905 910

Glu Gly Phe Val Ala Arg Ala Pro Glu Ala Val Ile Ala Lys Glu Arg
 915 920 925

Glu Lys Leu Glu Gly Tyr Ala Glu Ala Lys Ala Lys Leu Ile Glu Gln
 930 935 940

Gln Ala Val Ile Ala Ala Leu
 945 950

<210> 184

<211> 334

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 184

Met Thr Lys Pro Ile Val Phe Ser Gly Ala Gln Pro Ser Gly Glu Leu
 1 5 10 15

Thr Ile Gly Asn Tyr Met Gly Ala Leu Arg Gln Trp Val Asn Met Gln
 20 25 30

Asp Asp Tyr His Cys Ile Tyr Cys Ile Val Asp Gln His Ala Ile Thr
 35 40 45

Val Arg Gln Asp Ala Gln Lys Leu Arg Lys Ala Thr Leu Asp Thr Leu
 50 55 60

Ala Leu Tyr Leu Ala Cys Gly Ile Asp Pro Glu Lys Ser Thr Ile Phe
 65 70 75 80

Val Gln Ser His Val Pro Glu His Ala Gln Leu Gly Trp Ala Leu Asn
 85 90 95

Cys Tyr Thr Tyr Phe Gly Glu Leu Ser Arg Met Thr Gln Phe Lys Asp
 100 105 110

Lys Ser Ala Arg Tyr Ala Glu Asn Ile Asn Ala Gly Leu Phe Asp Tyr
 115 120 125

Pro Val Leu Met Ala Ala Asp Ile Leu Leu Tyr Gln Thr Asn Leu Val
 130 135 140

Pro Val Gly Glu Asp Ala Lys Gln Ala Leu Glu Leu Ser Arg Asp Ile
 145 150 155 160

Ala Gln Arg Phe Asn Ala Leu Tyr Gly Glu Ile Phe Lys Val Pro Glu
 165 170 175

Pro Phe Ile Pro Lys Ser Gly Ala Arg Val Met Ser Leu Leu Glu Pro
 180 185 190

Thr Lys Lys Met Ser Lys Ser Asp Asp Asn Arg Asn Asn Val Ile Gly
 195 200 205

Leu Leu Glu Asp Pro Lys Ser Val Val Lys Lys Ile Lys Arg Ala Val
 210 215 220

Thr Asp Ser Asp Glu Pro Pro Val Val Arg Tyr Asp Val Gln Asn Lys
 225 230 235 240

Ala Gly Val Ser Asn Leu Leu Asp Ile Leu Ser Ala Val Thr Gly Gln
 245 250 255

Ser Ile Pro Glu Leu Glu Lys Gln Phe Glu Gly Lys Met Tyr Gly His
 260 265 270

Leu Lys Gly Glu Val Ala Asp Ala Val Ser Gly Met Leu Thr Glu Leu
 275 280 285

Gln Glu Arg Tyr His Arg Phe Arg Asn Asp Glu Ala Phe Leu Gln Gln
 290 295 300

Val Met Lys Asp Gly Ala Glu Lys Ala Ser Ala His Ala Ser Arg Thr
 305 310 315 320

Leu Lys Ala Val Tyr Glu Ala Ile Gly Phe Val Ala Lys Pro
 325 330

<210> 185

<211> 334

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 185

Met Thr Lys Pro Ile Val Phe Ser Gly Ala Gln Pro Ser Gly Glu Leu
 1 5 10 15

Thr Ile Gly Asn Tyr Met Gly Ala Leu Arg Gln Trp Val Asn Met Gln
 20 25 30

Asp Asp Tyr His Cys Ile Tyr Cys Ile Val Asp Gln His Ala Ile Thr
 35 40 45

Val Arg Gln Asp Ala Gln Lys Leu Arg Lys Ala Thr Leu Asp Thr Leu
 50 55 60

Ala Leu Tyr Leu Ala Cys Gly Ile Asp Pro Glu Lys Ser Thr Ile Phe
 65 70 75 80

Val Gln Ser His Val Pro Glu His Ala Gln Leu Gly Trp Ala Leu Asn
 85 90 95

Cys Tyr Thr Tyr Phe Gly Glu Leu Ser Arg Met Thr Gln Phe Lys Asp
 100 105 110

Lys Ser Ala Arg Tyr Ala Glu Asn Ile Asn Ala Gly Leu Phe Asp Tyr
 115 120 125

Pro Val Leu Ala Ala Ala Asp Ile Leu Leu Tyr Gln Thr Asn Leu Val
 130 135 140

Pro Val Gly Glu Asp Gln Lys Gln His Leu Glu Leu Ser Arg Asp Ile
 145 150 155 160

Ala Gln Arg Phe Asn Ala Leu Tyr Gly Glu Ile Phe Lys Val Pro Glu
 165 170 175

Pro Phe Ile Pro Lys Ser Gly Ala Arg Val Met Ser Leu Leu Glu Pro
 180 185 190

Thr Lys Lys Met Ser Lys Ser Asp Asp Asn Arg Asn Asn Val Ile Gly
 195 200 205

Leu Leu Glu Asp Pro Lys Ser Val Val Lys Lys Ile Lys Arg Ala Val
 210 215 220

Thr Asp Ser Asp Glu Pro Pro Val Val Arg Tyr Asp Val Gln Asn Lys
 225 230 235 240

Ala Gly Val Ser Asn Leu Leu Asp Ile Leu Ser Ala Val Thr Gly Gln
 245 250 255

Ser Ile Pro Glu Leu Glu Lys Gln Phe Glu Gly Lys Met Tyr Gly His
 260 265 270

Leu Lys Gly Glu Val Ala Asp Ala Val Ser Gly Met Leu Thr Glu Leu
 275 280 285

Gln Glu Arg Tyr His Arg Phe Arg Asn Asp Glu Ala Phe Leu Gln Gln
 290 295 300

Val Met Lys Asp Gly Ala Glu Lys Ala Ser Ala His Ala Ser Arg Thr
 305 310 315 320

Leu Lys Ala Val Tyr Glu Ala Ile Gly Phe Val Ala Lys Pro
 325 330

<210> 186

<211> 334

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 186

Met Thr Lys Pro Ile Val Phe Ser Gly Ala Gln Pro Ser Gly Glu Leu
 1 5 10 15

Thr Ile Gly Asn Tyr Met Gly Ala Leu Arg Gln Trp Val Asn Met Gln
 20 25 30

Asp Asp Tyr His Cys Ile Tyr Cys Ile Val Asp Gln His Ala Ile Thr
 35 40 45

Val Arg Gln Asp Ala Gln Lys Leu Arg Lys Ala Thr Leu Asp Thr Leu
 50 55 60

Ala Leu Tyr Leu Ala Cys Gly Ile Asp Pro Glu Lys Ser Thr Ile Phe
 65 70 75 80

Val Gln Ser His Val Pro Glu His Ala Gln Leu Gly Trp Ala Leu Asn
 85 90 95

Cys Tyr Thr Tyr Phe Gly Glu Leu Ser Arg Met Thr Gln Phe Lys Asp
 100 105 110

Lys Ser Ala Arg Tyr Ala Glu Asn Ile Asn Ala Gly Leu Phe Asp Tyr
 115 120 125

Pro Val Leu Val Ala Ala Asp Ile Leu Leu Tyr Gln Thr Asn Leu Val
 130 135 140

Pro Val Gly Glu Asp Ala Lys Gln Ala Leu Glu Leu Ser Arg Asp Ile
 145 150 155 160

Ala Gln Arg Phe Asn Ala Leu Tyr Gly Glu Ile Phe Lys Val Pro Glu
 165 170 175

Pro Phe Ile Pro Lys Ser Gly Ala Arg Val Met Ser Leu Leu Glu Pro
 180 185 190

Thr Lys Lys Met Ser Lys Ser Asp Asp Asn Arg Asn Asn Val Ile Gly
 195 200 205

Leu Leu Glu Asp Pro Lys Ser Val Val Lys Lys Ile Lys Arg Ala Val
 210 215 220

Thr Asp Ser Asp Glu Pro Pro Val Val Arg Tyr Asp Val Gln Asn Lys
 225 230 235 240

Ala Gly Val Ser Asn Leu Leu Asp Ile Leu Ser Ala Val Thr Gly Gln
 245 250 255

Ser Ile Pro Glu Leu Glu Lys Gln Phe Glu Gly Lys Met Tyr Gly His
 260 265 270

Leu Lys Gly Glu Val Ala Asp Ala Val Ser Gly Met Leu Thr Glu Leu
 275 280 285

Gln Glu Arg Tyr His Arg Phe Arg Asn Asp Glu Ala Phe Leu Gln Gln
 290 295 300

Val Met Lys Asn Gly Ala Glu Lys Ala Ser Ala His Ala Ser Arg Thr

305 310 315 320

Leu Lys Ala Val Tyr Glu Ala Ile Gly Phe Val Ala Lys Pro
 325 330

<210> 187

<211> 860

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 187

Met Gln Glu Gln Tyr Arg Pro Glu Glu Ile Glu Ser Lys Val Gln Leu
 1 5 10 15

His Trp Asp Glu Lys Arg Thr Phe Glu Val Thr Glu Asp Glu Ser Lys
 20 25 30

Glu Lys Tyr Tyr Cys Leu Ser Met Leu Pro Gly Pro Ser Gly Arg Leu
 35 40 45

His Met Gly His Val Arg Asn Tyr Thr Ile Gly Asp Val Ile Ala Arg
 50 55 60

Tyr Gln Arg Met Leu Gly Lys Asn Val Leu Gln Pro Ile Gly Trp Asp
 65 70 75 80

Ala Phe Gly Leu Pro Ala Glu Gly Ala Ala Val Lys Asn Asn Thr Ala
 85 90 95

Pro Ala Pro Trp Thr Tyr Asp Asn Ile Ala Tyr Met Lys Asn Gln Leu
 100 105 110

Lys Met Leu Gly Phe Gly Tyr Asp Trp Ser Arg Glu Leu Ala Thr Cys
 115 120 125

Thr Pro Glu Tyr Tyr Arg Trp Glu Gln Lys Phe Phe Thr Glu Leu Tyr
 130 135 140

Lys Lys Gly Leu Val Tyr Lys Lys Thr Ser Ala Val Asn Trp Cys Pro
 145 150 155 160

Asn Asp Gln Thr Val Leu Ala Asn Glu Gln Val Ile Asp Gly Cys Cys
 165 170 175

Trp Arg Cys Asp Thr Lys Val Glu Arg Lys Glu Ile Pro Gln Trp Phe
 180 185 190

Ile Lys Ile Thr Ala Tyr Ala Asp Glu Leu Leu Asn Asp Leu Asp Lys
 195 200 205

Leu Asp His Trp Pro Asp Thr Val Lys Thr Met Gln Arg Asn Trp Ile
 210 215 220

Gly Arg Ser Glu Gly Val Glu Ile Thr Phe Asn Val Asn Asp Tyr Asp
 225 230 235 240

Asn Thr Leu Thr Val Tyr Thr Thr Arg Pro Asp Thr Phe Met Gly Cys
 245 250 255

Thr Tyr Leu Ala Val Ala Ala Gly His Pro Leu Ala Gln Lys Ala Ala
 260 265 270

Glu Asn Asn Pro Glu Leu Ala Ala Phe Ile Asp Glu Cys Arg Asn Thr
 275 280 285

Lys Val Ala Glu Ala Glu Met Ala Thr Met Glu Lys Lys Gly Val Asp
 290 295 300

Thr Gly Phe Lys Ala Val His Pro Leu Thr Gly Glu Glu Ile Pro Val
 305 310 315 320

Trp Ala Ala Asn Phe Val Leu Met Glu Tyr Gly Thr Gly Ala Val Met
 325 330 335

Ala Val Pro Gly His Asp Gln Arg Asp Tyr Glu Phe Ala Ser Lys Tyr
 340 345 350

Gly Leu Asn Ile Lys Pro Val Ile Leu Ala Ala Asp Gly Ser Glu Pro
 355 360 365

Asp Leu Ser Gln Gln Ala Leu Thr Glu Lys Gly Val Leu Phe Asn Ser
 370 375 380

Gly Glu Phe Asn Gly Leu Asp His Glu Ala Ala Phe Asn Ala Ile Ala
 385 390 395 400

Asp Lys Leu Thr Ala Met Gly Val Gly Glu Arg Lys Val Asn Tyr Arg
 405 410 415

Leu Arg Asp Trp Gly Val Ser Arg Gln Arg Tyr Trp Gly Ala Pro Ile
 420 425 430

Pro Met Val Thr Leu Glu Asp Gly Thr Val Met Pro Thr Pro Asp Asp
 435 440 445

Gln Leu Pro Val Ile Leu Pro Glu Asp Val Val Met Asp Gly Ile Thr
 450 455 460

Ser Pro Ile Lys Ala Asp Pro Glu Trp Ala Lys Thr Thr Val Asn Gly
 465 470 475 480

Met Pro Ala Leu Arg Glu Thr Asp Thr Phe Asp Thr Phe Met Glu Ser
 485 490 495

Ser Trp Tyr Tyr Ala Arg Tyr Thr Cys Pro Gln Tyr Lys Glu Gly Met
 500 505 510

Leu Asp Ser Glu Ala Ala Asn Tyr Trp Leu Pro Val Asp Ile Tyr Ile
 515 520 525

Gly Gly Ile Glu His Ala Ile Met His Leu Leu Tyr Phe Arg Phe Phe
 530 535 540

His Lys Leu Met Arg Asp Ala Gly Met Val Asn Ser Asp Glu Pro Ala
 545 550 555 560

Lys Gln Leu Leu Cys Gln Gly Met Val Leu Ala Asp Ala Phe Tyr Tyr
 565 570 575

Val Gly Glu Asn Gly Glu Arg Asn Trp Val Ser Pro Val Asp Ala Ile
 580 585 590

Val Glu Arg Asp Glu Lys Gly Arg Ile Val Lys Ala Lys Asp Ala Ala
 595 600 605

Gly His Glu Leu Val Tyr Thr Gly Met Ser Lys Met Ser Lys Ser Lys
 610 615 620

Asn Asn Gly Ile Asp Pro Gln Val Met Val Glu Arg Tyr Gly Ala Asp
 625 630 635 640

Thr Val Arg Leu Phe Met Met Phe Ala Ser Pro Ala Asp Met Thr Leu
 645 650 655

Glu Trp Gln Glu Ser Gly Val Glu Gly Ala Asn Arg Phe Leu Lys Arg
 660 665 670

Val Trp Lys Leu Val Tyr Glu His Thr Ala Lys Gly Asp Val Ala Ala
 675 680 685

Leu Asn Val Asp Ala Leu Thr Glu Asn Gln Lys Ala Leu Arg Arg Asp
 690 695 700

Val His Lys Thr Ile Ala Lys Val Thr Asp Asp Ile Gly Arg Arg Gln
 705 710 715 720

Thr Phe Asn Thr Ala Ile Ala Ala Ile Met Glu Leu Met Asn Lys Leu
 725 730 735

Ala Lys Ala Pro Thr Asp Gly Glu Gln Asp Arg Ala Leu Met Gln Glu
 740 745 750

Ala Leu Leu Ala Val Val Arg Met Leu Asn Pro Phe Thr Pro His Ile
 755 760 765

Cys Phe Thr Leu Trp Gln Glu Leu Lys Gly Glu Gly Asp Ile Asp Asn
 770 775 780

Ala Pro Trp Pro Val Ala Asp Glu Lys Ala Met Val Glu Asp Ser Thr
 785 790 795 800

Leu Val Val Val Gln Val Asn Gly Lys Val Arg Ala Lys Ile Thr Val
 805 810 815

Pro Val Asp Ala Thr Glu Glu Gln Val Arg Glu Arg Ala Gly Gln Glu
 820 825 830


```

-----
      210                215                220
Thr Asp Ser Asp Glu Pro Pro Val Val Arg Tyr Asp Val Gln Asn Lys
225                230                235                240

Ala Gly Val Ser Asn Leu Leu Asp Ile Leu Ser Ala Val Thr Gly Gln
                245                250                255

Ser Ile Pro Glu Leu Glu Lys Gln Phe Glu Gly Lys Met Tyr Gly His
                260                265                270

Leu Lys Gly Glu Val Ala Asp Ala Val Ser Gly Met Leu Thr Glu Leu
                275                280                285

Gln Glu Arg Tyr His Arg Phe Arg Asn Asp Glu Ala Phe Leu Gln Gln
                290                295                300

Val Met Lys Asp Gly Ala Glu Lys Ala Ser Ala His Ala Ser Arg Thr
305                310                315                320

Leu Lys Ala Val Tyr Glu Ala Ile Gly Phe Val Ala Lys Pro
                325                330

```

<210> 189

<211> 860

<212> PRT

<213> E. coli

<400> 189

```

Met Gln Glu Gln Tyr Arg Pro Glu Glu Ile Glu Ser Lys Val Gln Leu
1                5                10                15

His Trp Asp Glu Lys Arg Thr Phe Glu Val Thr Glu Asp Glu Ser Lys
                20                25                30

Glu Lys Tyr Tyr Cys Leu Ser Met Leu Pro Tyr Pro Ser Gly Arg Leu
                35                40                45

His Met Gly His Val Arg Asn Tyr Thr Ile Gly Asp Val Ile Ala Arg
50                55                60

Tyr Gln Arg Met Leu Gly Lys Asn Val Leu Gln Pro Ile Gly Trp Asp
65                70                75                80

Ala Phe Gly Leu Pro Ala Glu Gly Ala Ala Val Lys Asn Asn Thr Ala
                85                90                95

Pro Ala Pro Trp Thr Tyr Asp Asn Ile Ala Tyr Met Lys Asn Gln Leu
                100                105                110

Lys Met Leu Gly Phe Gly Tyr Asp Trp Ser Arg Glu Leu Ala Thr Cys
115                120                125

Thr Pro Glu Tyr Tyr Arg Tyr Glu Glu Lys Phe Phe Thr Glu Leu Tyr

```

Thr Phe Glu Tyr Tyr Arg Asp Glu Gln Lys Phe Phe Thr Glu Leu Tyr
 130 135 140

Lys Lys Gly Leu Val Tyr Lys Lys Thr Ser Ala Val Asn Trp Cys Pro
 145 150 155 160

Asn Asp Gln Thr Val Leu Ala Asn Glu Gln Val Ile Asp Gly Cys Cys
 165 170 175

Trp Arg Cys Asp Thr Lys Val Glu Arg Lys Glu Ile Pro Gln Trp Phe
 180 185 190

Ile Lys Ile Thr Ala Tyr Ala Asp Glu Leu Leu Asn Asp Leu Asp Lys
 195 200 205

Leu Asp His Trp Pro Asp Thr Val Lys Thr Met Gln Arg Asn Trp Ile
 210 215 220

Gly Arg Ser Glu Gly Val Glu Ile Thr Phe Asn Val Asn Asp Tyr Asp
 225 230 235 240

Asn Thr Leu Thr Val Tyr Thr Thr Arg Pro Asp Thr Phe Met Gly Cys
 245 250 255

Thr Tyr Leu Ala Val Ala Ala Gly His Pro Leu Ala Gln Lys Ala Ala
 260 265 270

Glu Asn Asn Pro Glu Leu Ala Ala Phe Ile Asp Glu Cys Arg Asn Thr
 275 280 285

Lys Val Ala Glu Ala Glu Met Ala Thr Met Glu Lys Lys Gly Val Asp
 290 295 300

Thr Gly Phe Lys Ala Val His Pro Leu Thr Gly Glu Glu Ile Pro Val
 305 310 315 320

Trp Ala Ala Asn Phe Val Leu Met Glu Tyr Gly Thr Gly Ala Val Met
 325 330 335

Ala Val Pro Gly His Asp Gln Arg Asp Tyr Glu Phe Ala Ser Lys Tyr
 340 345 350

Gly Leu Asn Ile Lys Pro Val Ile Leu Ala Ala Asp Gly Ser Glu Pro
 355 360 365

Asp Leu Ser Gln Gln Ala Leu Thr Glu Lys Gly Val Leu Phe Asn Ser
 370 375 380

Gly Glu Phe Asn Gly Leu Asp His Glu Ala Ala Phe Asn Ala Ile Ala
 385 390 395 400

Asp Lys Leu Thr Ala Met Gly Val Gly Glu Arg Lys Val Asn Tyr Arg
 405 410 415

Leu Arg Asp Trp Gly Val Ser Arg Gln Arg Tyr Trp Gly Ala Pro Ile
 420 425 430

Pro Met Val Thr Leu Glu Asp Gly Thr Val Met Pro Thr Pro Asp Asp
 435 440 445

Gln Leu Pro Val Ile Leu Pro Glu Asp Val Val Met Asp Gly Ile Thr
 450 455 460

Ser Pro Ile Lys Ala Asp Pro Glu Trp Ala Lys Thr Thr Val Asn Gly
 465 470 475 480

Met Pro Ala Leu Arg Glu Thr Asp Thr Phe Asp Thr Phe Met Glu Ser
 485 490 495

Ser Trp Tyr Tyr Ala Arg Tyr Thr Cys Pro Gln Tyr Lys Glu Gly Met
 500 505 510

Leu Asp Ser Glu Ala Ala Asn Tyr Trp Leu Pro Val Asp Ile Tyr Ile
 515 520 525

Gly Gly Ile Glu His Ala Ile Met His Leu Leu Tyr Phe Arg Phe Phe
 530 535 540

His Lys Leu Met Arg Asp Ala Gly Met Val Asn Ser Asp Glu Pro Ala
 545 550 555 560

Lys Gln Leu Leu Cys Gln Gly Met Val Leu Ala Asp Ala Phe Tyr Tyr
 565 570 575

Val Gly Glu Asn Gly Glu Arg Asn Trp Val Ser Pro Val Asp Ala Ile
 580 585 590

Val Glu Arg Asp Glu Lys Gly Arg Ile Val Lys Ala Lys Asp Ala Ala
 595 600 605

Gly His Glu Leu Val Tyr Thr Gly Met Ser Lys Met Ser Lys Ser Lys
 610 615 620

Asn Asn Gly Ile Asp Pro Gln Val Met Val Glu Arg Tyr Gly Ala Asp
 625 630 635 640

Thr Val Arg Leu Phe Met Met Phe Ala Ser Pro Ala Asp Met Thr Leu
 645 650 655

Glu Trp Gln Glu Ser Gly Val Glu Gly Ala Asn Arg Phe Leu Lys Arg
 660 665 670

Val Trp Lys Leu Val Tyr Glu His Thr Ala Lys Gly Asp Val Ala Ala
 675 680 685

Leu Asn Val Asp Ala Leu Thr Glu Asn Gln Lys Ala Leu Arg Arg Asp
 690 695 700

Val His Lys Thr Ile Ala Lys Val Thr Asp Asp Ile Gly Arg Arg Gln
 705 710 715 720

Thr Phe Asn Thr Ala Ile Ala Ala Ile Met Glu Leu Met Asn Lys Leu
 725 730 735

Ala Lys Ala Pro Thr Asn Gly Glu Gln Asp Arg Ala Leu Met Gln Glu

Ala Leu Leu Ala Val Val Arg Met Leu Asn Pro Phe Thr Pro His Ile
 740 745 750

Ala Leu Leu Ala Val Val Arg Met Leu Asn Pro Phe Thr Pro His Ile
 755 760 765

Cys Phe Thr Leu Trp Gln Glu Leu Lys Gly Glu Gly Asp Ile Asp Asn
 770 775 780

Ala Pro Trp Pro Val Ala Asp Glu Lys Ala Met Val Glu Asp Ser Thr
 785 790 795 800

Leu Val Val Val Gln Val Asn Gly Lys Val Arg Ala Lys Ile Thr Val
 805 810 815

Pro Val Asp Ala Thr Glu Glu Gln Val Arg Glu Arg Ala Gly Gln Glu
 820 825 830

His Leu Val Ala Lys Tyr Leu Asp Gly Val Thr Val Arg Lys Val Ile
 835 840 845

Tyr Val Pro Gly Lys Leu Leu Asn Leu Val Val Gly
 850 855 860

<210> 190

<211> 2853

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 190

```

atggaaaaaa cctataatcc gcaagacatt gaacaaccgc tgtatgaaca ctgggaaaaa    60
caaggctact ttaaaccgaa cggcgatgaa agtcaggaat ccttttgcac tatgattccg    120
ccgccgggag tgctcaggctc gctgcatacg ggccacgcgt tccagcaaac gattatggat    180
accatgatcc gttatcagcg catgcaaggt aaaaacacgc tgtggcaggt tggcaccgat    240
catgcaggta tcgctacgca aatggtggtt gaacgtaaaa ttgcccga agaaggcaaa    300
accgctcagc actacggctc cgaagcgttt attgataaaa tctgggaatg gaaagccgaa    360
agtggcggta ccattacgag tcagatgcgt cgcctgggta actccgtgga ctgggaacgt    420
gaacgcttta ccattgatga aggtctgagc aatgcagtta aagaagtttt cgtccgtctg    480
tataaagaag atctgatcta ccgtggcaaa cgcctgggta actgggacct gaaactgcgc    540
accgccattt ctgatctgga agttgaaaat cgtgaatcaa aaggttcgat gtggcatatc    600
cgctatccgc tggcggatgg cgcacaaacc gcagatggta aagactacct ggtcgtggca    660
accacgcgtc cggaaacgct gctgggagat accggtgtgg ccgttaacct ggaagaccgg    720
cgctataaag atctgattgg caaatcgtg atcctgccgc tggttaatcg tcgattccg    780
atcgtgggtg acgaacacgc ggatattgaa aaaggcaccg gttgcgttaa aattgcgccg    840
gccatgatt ttaacgacta tgaagtcgac aaacgtcagc ccctgccgat gattaatac    900
atgagcttgc atggagagat tggagagact ggaagctctc tggagagagc cctatctgag    960
    
```

```

ctgacgtctg atygcgacat tgcgaaaat gcaacagctt tgcacacca yggtaatgaa 300
tccgatgtgt acagctctga aatcccggcg gaatttcaaa aactggaacg tttcgcagct 1020
cgcaaagcag ttgtcgcagc agtcgatgca ctgggtctgc tgaagaaat taaaccgcat 1080
gacctgacgg ttccgtatgg tgatcgtggc ggtgtgggta tccaaccgat gctgaccgac 1140
cagtggatg tgcgcgccga tgcctcggct aaaccggcgg tcaagccgt ggaaaacggt 1200
gatattcagt ttgtgccgaa acaatacгаа aacatgtact tctcatggat gcgtgatatt 1260
caggactggt gtatctcgcg tcaactgtgg tggggccatc gcattccggc atggtatgat 1320
gaagctggca acgtctacgt gggtcgtaat gaagacgaag tgcgcaaaga aaacaatctg 1380
ggtgcggatg tcgtgctgcg ccaggatгаа gacgttctgg acacctggtt tagttccgca 1440
ctgtggacct tcagtacgct gggctggccg gaaaacacgg atgcgctgcg tcagtttcac 1500
ccgaccagcg ttatggtctc tggtttcgac attatctttt tctggattgc ccgcatgac 1560
atgatgacca tgcatttcat caaagatгаа aacggcaaac cgcaggtccc gttccacacg 1620
gtgtatatga ccggcctgat ccgtgatgac gaaggtcaaa aatgagcaa atctaaaggc 1680
aacgttatcg atccgctgga catggtcgat ggtattagcc tgcgggaact gctggaaaaa 1740
cgcacgggta acatgatgca gccgcaactg gcggataaaa tccgtaaacg caccgaaaaa 1800
cagtttccga atggcattga accgcatggt acggatgcbc tgcgtttcac cctggcagct 1860
ctggcatcaa ccggtcgtga cattaattgg gatatgaaac gcctggaagg ttatcgcaac 1920
ttttgcaata aactgtggaa cgcctcacgt ttcgtgctga tgaatacggg aggccaggat 1980
tgtggtttta acggcgggta aatgaccctg tcgctggcag accgctggat cctggctgaa 2040
ttcaatcaaa cgattaaagc gtaccgtгаа gccctggact cttttcgtt cgatattgcg 2100
gccggcatcc tgtatgaatt tacctggaac cagttctgcg attggtacct ggaactgacg 2160
aaaccggtga tgaatggcgg taccgaagcg gaactcgtg gcacccgcca tacgctggtg 2220
accgttctgg aaggctcgtc gcgtctggcc caccgatta tcccgtttat taccgaaacg 2280
atctggcagc gcgtcaaagt gctgtgtggt attacggcgg ataccatcat gctgcagccg 2340
ttcccgcaat atgacgcaag tcaggttgat gaagcagctc tggctgatac cgaatggctg 2400
aaacaagcaa ttgttctgtt ccgtaacatc cgcgcggaaa tgaatattgc cccgggcaaa 2460
ccgctggaac tgctgctgcg tggttgttcc gcagatgctg aacgtcgcgt gaacgaaaat 2520
cgtggctttc tgcagacgct ggcacgcctg gaaagcatta ccgttctgcc ggctgatgac 2580
aaaggctccg tgtctgttac caaaattatc gatggcgcag aactgctgat cccgatggct 2640
ggtctgatta acaaagaaga cgaactggcg cgtctggcca aagaagtggc aaaaattгаа 2700
ggcgaatta gccgcatcga aaacaaactg gctaatagag gttttgtggc gcgtgccccg 2760
gaagcagtta tcgctaaaga acgcgaaaaa ctggaagggt acgcagaagc aaaagccaaa 2820
ctgattgaac aacaagcggg tatcgcagca ctg 2853

```

<210> 191

<211> 2853

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 191

```

atggaaaaaa cctataatcc gcaagacatt gaacaaccgc tgtatgaaca ctgggaaaaa    60

caaggctact ttaaaccgaa cggcgatgaa agtcaggaat ccttttgcat tatgattccg    120

ccgccgggcg tgtcaggttc gctgcatatg ggccacgcgt tccagcaaac gattatggat    180

accatgatcc gttatcagcg catgcaaggt aaaaacacgc tgtggcaggt tggcaccgat    240

catgcaggta tcgctacgca aatggtgggt gaacgtaaaa ttgcggccga agaaggcaaa    300

acccgtcacg actacggtcg cgaagcgttt attgataaaa tctgggaatg gaaagccgaa    360

agtggcggta ccattacgcg tcagatgcgt cgctgggta actccgtgga ctgggaacgt    420

gaacgcttta ccattgatga aggtctgagc aatgcagtta aagaagtttt cgtccgtctg    480

tataaagaag atctgatcta ccgtggcaaa cgctgggta actgggaccc gaaactgcgc    540

accgccattt ctgatctgga agttgaaaat cgtgaatcaa aaggttcgat gtggcatatc    600

cgctatccgc tggcgggatg cgcctgggta aagactacct ggtcgtggca    660

accacgcgct cggaaacgct gctgggcgat accggtgtgg ccgttaaccc ggaagaccgc    720

cgctataaag atctgatttg caaatacgtg atcctgccgc tggttaatcg tcgcattccg    780

atcgtgggtg acgaacacgc ggatatggaa aaaggcaccg gttgcgttaa aattggtccg    840

gcccatgatt ttaacgacta tgaagtcggc aaacgtcacg ccctgccgat gattaatatc    900

ctgacgtttg atggcgacat tcgcgaaagt gcacaggtct tcgacaccaa gggtaatgaa    960

tccgatgtgt acagctctga aatcccggcg gaatttcaaa aactggaacg tttcgcagct   1020

cgcaaagcag ttgtcgcagc agtcgatgca ctgggtctgc tggaaagaaat taaaccgcat   1080

gacctgacgg ttccgtatgg tgatcgtggc ggtgtggta tcgaaccgat gctgaccgac   1140

cagtggatg tgcgcgccga tgtcctggct aaaccggcgg tcgaagccgt ggaaaacggt   1200

gatattcagt ttgtgccgaa acaatacga aacatgtact tctcatggat gcgtgatatt   1260

caggactggt gtatctcgcg tcaactgtgg tggggccatc gcattccggc atggtatgat   1320

gaagctggca acgtctacgt gggtcgtaat gaagacgaag tgcgcaaaga aaacaatctg   1380

ggtgcggatg tcgtgctgcg ccaggatgaa gacgttctgg acacctggtt tagttccgca   1440

ctgtggacct tcagtacgct gggctggccg gaaaacacgg atgcgctgcg tcagtttcac   1500

ccgaccagcg ttatggtctc tggtttcgac attatctttt tctggattgc ccgcatgatc   1560

atgatgacca tgcatttcat caaagatgaa aacggcaaac cgcaggtccc gttccacacg   1620

gtgtatatga ccggcctgat ccgtgatgac gaaggtcaaa aatgagcaa atctaaaggc   1680

aacgttatcg atccgctgga catggtcgat ggtattagcc tgcgggaact gctggaaaaa   1740

cgcacgggta acatgatgca gccgcaactg gcggataaaa tccgtaaacg caccgaaaaa   1800

cagtttccga atggcattga accgcatggt acggatgcgc tgcgtttcac cctggcagct   1860

ctggcatcaa ccggtcgtga cattaattgg gatatgaaac gcctggaagg ttatcgcaac   1920

ttttgcaata aactgtggaa cgctcaccgt ttcgtgctga tgaatacggg aggccaggat   1980

tgtggtttta acggcgggta aatgaccctg tcgctggcag accgctggat cctggctgaa   2040

```

```

- - - - -
ttcaatcaaa cgattaaagc gtaccgtgaa gccctggact cttttcgctt cgatattgcg 2100
gccggcatcc tgtatgaatt tacctggaac cagttctgcg attggtacct ggaactgacg 2160
aaaccggtga tgaatggcgg tacccaagcg gaactgcgtg gcacccgcca tacgctggtg 2220
accgttctgg aaggtctgct gcgtctggcc caccgatta tcccgtttat taccgaaacg 2280
atctggcagc gcgtcaaagt gctgtgtggt attacggcgg ataccatcat gctgcagccg 2340
ttcccgcaat atgacgcaag tcaggttgat gaagcagctc tggctgatac cgaatggctg 2400
aaacaagcaa ttgttgctgt ccgtaacatc cgcgcggaaa tgaatattgc cccgggcaaa 2460
ccgctggaac tgctgctgcg tggttgttcc gcagatgctg aacgtcgcgt gaacgaaaat 2520
cgtggcttcc tgcagacgct ggcacgcctg gaaagcatta ccgttctgcc ggctgatgac 2580
aaaggtccgg tgtctgttac caaattatc gatggcgcag aactgctgat cccgatggct 2640
ggtctgatta acaaagaaga cgaactggcg cgtctggcca aagaagtggc aaaaattgaa 2700
ggcgaaatta gccgcatcga aaacaaactg gctaataaag gttttgtggc gcgtgccccg 2760
gaagcagtta tcgctaaaga acgcgaaaaa ctggaaggtt acgcagaagc aaaagccaaa 2820
ctgattgaac aacaagcggg tatcgagca ctg 2853

```

<210> 192

<211> 1002

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 192

```

atgacgaaac cgattgtctt ctctgggtct caaccgtctg gtgaactgac gattgtaac 60
tacatgggtg ctctcgccca gtgggtgaa atgcaggatg actatcattg catttactgt 120
atcgtggatc aacacgccat taccgctcgt caggacgctc aaaaactgcg caaagcgacc 180
ctggatacgc tggccctgta tctggcatgc ggtattgacc cggaaaaatc aacgatcttt 240
gtgcagtcgc atgttccgga acacgctcaa ctgggttggg cgctgaactg ttatacctac 300
tttggcgaac tgtcacgtat gacgcagttc aaagataaat cgctcgcta tgcgaaaaac 360
attaatgcgg gtctgtttga ttaccgggtg ctgatggcgg ccgacatcct gctgtatcag 420
accaacctgg ttccggctcg cgaagatgca aaacaagcac tggaaactgag ccgtgacatc 480
gcccagcgtc tcaatgcact gtacgggtgaa atttttaaag tcccggaacc gttcatcccg 540
aaaagcggcg cccgtgtgat gtctctgctg gaaccgacca agaaaatgag caaatctgat 600
gacaaccgca acaatgttat tggcctgctg gaagatccga aaagcgtggt gaagaaaatt 660
aaacgtgcgg tcaccgattc tgacgaaccg ccggtcgtgc gctatgatgt tcagaacaaa 720
gctggtgtct caaatctgct ggatattctg agtgcagtga cgggccagtc catcccggaa 780
ctggaaaaac aattcgaagg caaaatgtac ggtcatctga aaggcgaagt cgcagatgct 840
gtgtccggtg tgctgaccga actgcaggaa cgttatcacc gttttcgcaa tgatgaagcc 900

```

ttcctgcagc aagttatgaa agacggcgca gaaaaagcga gtgccacgc atcccgcacg 960
 ctgaaagcgg tctacgaagc aatcggcttt gtggcgaaac cg 1002

<210> 193

<211> 1002

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 193

atgacgaaac cgattgtcct ctctgggtgct caaccgtctg gtgaactgac gattggtaac 60
 tacatgggtg ctctgcgccca gtgggtgaaac atgcaggatg actatcattg catttactgt 120
 atcgtggatc aacacgccat taccgttcgt caggacgctc aaaaactgcg caaagcgacc 180
 ctggatacgc tggcctgta tctggcatgc ggtattgacc cggaaaaatc aacgatcttt 240
 gtgcagtcgc atgttcocga acacgctcaa ctgggttggg cgctgaactg ttatacctac 300
 tttggcgaac tgtcacgtat gacgcagttc aaagataaat cgctcgcta tgcgaaaaac 360
 attaatgcgg gtctgtttga ttaccgggtg ctggcggcgg ccgacatcct gctgtatcag 420
 accaacctgg ttccggtcgg cgaagatcag aaacaacatc tggaaactgag ccgtgacatc 480
 gccagcgcct tcaatgcact gtacggtgaa atttttaag tcccggaacc gttcatcccc 540
 aaaagcggcg cccgtgtgat gtctctgctg gaaccgacca agaaaatgag caaatctgat 600
 gacaaccgca acaatgttat tggcctgctg gaagatccga aaagcgtggt gaagaaaatt 660
 aaacgtgcgg tcaccgatcc tgacgaaccg ccggtcgtgc gctatgatgt tcagaacaaa 720
 gctgggtgct caaatctgct ggatattctg agtgcagtga cgggccagtc catcccggaa 780
 ctggaaaaac aattcgaagg caaaatgtac ggtcatctga aaggcgaagt cgcagatgct 840
 gtgtccggta tgctgaccga actgcaggaa cgttatcacc gttttcgsaa tgatgaagcc 900
 ttcctgcagc aagttatgaa agacggcgca gaaaaagcga gtgccacgc atcccgcacg 960
 ctgaaagcgg tctacgaagc aatcggcttt gtggcgaaac cg 1002

<210> 194

<211> 1002

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 194

atgacgaaac cgattgtcct ctctgggtgct caaccgtctg gtgaactgac gattggtaac 60
 tacatgggtg ctctgcgccca gtgggtgaaac atgcaggatg actatcattg catttactgt 120
 atcgtggatc aacacgccat taccgttcgt caggacgctc aaaaactgcg caaagcgacc 180
 ctggatacgc tggcctgta tctggcatgc ggtattgacc cggaaaaatc aacgatcttt 240

gtgcagtcgc atgttccgga acacgctcaa ctgggttggg cgctgaactg ttatacctac 300
 tttggcgaac tgtcacgtat gacgcagttc aaagataaat cggctcgtta tgcggaaaac 360
 attaatgcgg gtctgtttga ttaccggtg ctgggtggcg cgcacatcct gctgtatcag 420
 accaacctgg ttccggtcgg cgaagatgca aaacaagccc tggaaactgag ccgtgacatc 480
 gccagcgct tcaatgcact gtacggtgaa atttttaag tcccggaacc gttcatcccg 540
 aaaagcggcg cccgtgtgat gtctctgctg gaaccgacca agaaaatgag caaatctgat 600
 gacaaccgca acaatgttat tggcctgctg gaagatccga aaagcgtggt gaagaaaatt 660
 aaacgtgcgg tcaccgattc tgacgaaccg ccggtcgtgc gctatgatgt tcagaacaaa 720
 gctggtgtct caaatctgct ggatattctg agtgcagtga cgggccagt ccatcccgaa 780
 ctggaaaaac aattcgaagg caaaatgtac ggtcatctga aaggcgaagt cgcagatgct 840
 gtgtccggta tgctgaccga actgcaggaa cgttatcacc gttttcgcaa tgatgaagcc 900
 ttctgcagc aagttatgaa agacggcgca gaaaaagcga gtgccacgc atcccgcacg 960
 ctgaaagcgg tctacgaagc aatcggcttt gtggcgaaac cg 1002

<210> 195

<211> 2580

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 195

atgcaagaac aataccgccc ggaagaaatc gaaagtaaag tccaactgca ctgggatgaa 60
 aaacgcacct ttgaagtcac ggaagacgaa tctaaagaaa aatattactg cctgagcatg 120
 ctgccgggtc cgtctggtcg tctgcatatg ggtcacgtgc gcaactatac cattggcgat 180
 gttatcgcgc gttaccagcg catgctgggt aaaaatgttc tgcaaccgat tggctgggac 240
 gcatttggtc tgccggctga aggcgcggcc gtgaaaaaca ataccgcacc ggctccgtgg 300
 acgtatgata acatcgcata catgaaaaat cagctgaaaa tgctgggctt cggttatgac 360
 tggctctctg aactggcgac ctgtacgccg gaatattacc gctgggaaca aaaatttttc 420
 accgaactgt acaaaaaagc tctggtctac aagaaaacca gcgcggtgaa ctggtgcccg 480
 aatgatcaaa cggctctggc caacgaacaa gtgattgatg gctgctggtg gcgttgtgac 540
 accaaagtgt aacgcaaaga aatcccgcag tggtttatta aatcacggc gtatgccgat 600
 gaactgctga acgatctgga caaactggat cattggcccg acaccgtgaa aacgatgcag 660
 cgtaattgga ttggccgcag cgaaggtgtt gaaatcacct ttaacgtcaa tgattatgac 720
 aacaccctga cggctctacac cacgcgtccg gataccttca tgggttgcac gtacctggca 780
 gtggcagctg gtcacccgct ggctcagaaa gcagcagaaa acaatccgga actggcagct 840
 tttattgatg aatgtcgtca taccaaagtg gcagaagctg aatggcaac catggagaaa 900
 aaagcgtggt atacgggttt taaagctgtc catccgctga ccggtgaaga aatcccgtg 960

tgggcggcca acttcgttct gatggaatat ggtacgggtg cagtcatggc tgtgccgggt 1020
 cacgatcagc gtgactatga atttgcctcc aaatacggtc tgaatattaa accggttatc 1080
 ctggcagctg atggotcaga accggacctg tcgcagcaag cactgaccga aaaagggtgtg 1140
 ctgtttaact caggogaatt caatggtctg gatcatgaag cggccttcaa cgcgattgcc 1200
 gacaaaactga ccgcgatggg cgttggtgaa cgcгааgtca attatcgtct gcgcgattgg 1260
 ggtgtgtcgc gtcagcgtta ctggggtgca ccgattccga tggtgaccct ggaagatggc 1320
 accgttatgc cgacgccgga tgaccaactg ccggttatcc tgccggaaga tgtggttatg 1380
 gacggtatta ccagcccgat caaagctgat ccggaatggg ccaaaaaccac ggtgaatggc 1440
 atgccggcgc tcgctgaaac cgatacgttt gacaccttca tggaaagctc ttggtattac 1500
 gcccgctata cgtgcccgca gtacaaagaa ggcgatgctgg atagcgaagc agctaactat 1560
 tggctgccgg ttgacattta catcggcggg attgaacatg cgatcatgca cctgctgtat 1620
 tttcgtttct ttcataaact gatgcgcgat gccggtatgg tgaattcaga cgaaccggct 1680
 aaacagctgc tgtgtcaagg catggtcctg gcagatgctt tctattacgt gggcgaaaac 1740
 ggtgaacgta attgggtttc gccggtcgat gcgattgttg aacgtgacga aaaaggctgc 1800
 atcgtcaaag ccaaagatgc ggccggtcac gaactggtgt ataccggcat gagcaaatg 1860
 agtaaatcca aaaacaacgg tatcgatccg caggttatgg tcgaacgta cggcgcggac 1920
 accgttcgcc tgtttatgat gttcgcgagt ccggccgata tgacgctgga atggcaagaa 1980
 tccggcgtgg aagggtgcaa ccgttttctg aaacgcgtgt ggaaactggt ttatgaacat 2040
 accgcaaaaag cgcgatgtgc agctctgaac gttgacgcac tgacggaaaa tcagaaagct 2100
 ctgcgtcgcg atgtccacaa aaccattgca aaagtgacgg atgacatcgg tcgtcgccaa 2160
 accttaaca cggcgattgc ggccatcatg gaactgatga ataaactggc gaaagcaccg 2220
 accgatggtg aacaggaccg tgcactgatg caagaagcac tgctggctgt cgtgcgcatg 2280
 ctgaaccctt taccccgcga tatttgcttc acgctgtggc aggaactgaa aggcgaaggt 2340

 gatatcgaca atgccccgtg gccggttgcg gatgaaaaag ccatggtcga agacagtacc 2400
 ctggttgcg tgcagggtcaa tggtaaagtg cgcgcgaaaa ttaccgtgcc ggttgatgcc 2460
 acggaagaac aagtgcgtga acgcgcgggc caagaacacc tggttgcaa atatctggat 2520
 ggcgtcacgg tccgcaaagt catctatgtt ccgggtaaac tgctgaatct ggtcgttgg 2580

<210> 196

<211> 92

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 196

ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgttggcgtg 60
 actacaagga cgacgacgac aagtaagctt cg 92

<210> 197

<211> 71

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 197

ggguuaacuu uaacaaggag aaaaacaugc guuggcguga cuacaaggac gacgacgaca 60

aguaagcuuc g 71

<210> 198

<211> 12

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 198

Met Arg Trp Arg Asp Tyr Lys Asp Asp Asp Asp Lys
1 5 10

<210> 199

<211> 12

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> misc_feature

<222> (3)..(3)

<223> Xaa is methyl tryptophan

<400> 199

Met Arg Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
1 5 10

<210> 200

<211> 92

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 200

ggcgtaatac gactcactat agggttaact ttaacaagga gaaaaacatg cgtctccgtg 60
actacaagga cgacgacgac aagtaagctt cg 92

<210> 201

<211> 71

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 201

ggguuaacuu uaacaaggag aaaaacaugc gucuccguga cuacaaggac gacgacgaca 60
aguaagcuuc g 71

<210> 202

<211> 12

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 202

Met Arg Leu Arg Asp Tyr Lys Asp Asp Asp Asp Lys
1 5 10

<210> 203

<211> 12

<212> PRT

<213> Artificial

<220>

<223> artificially synthesized sequence

<220>

<221> misc_feature

<222> (3)..(3)

<223> Xaa is methyl leucine

<400> 203

Met Arg Xaa Arg Asp Tyr Lys Asp Asp Asp Asp Lys
1 5 10

<210> 204

<211> 97

<212> DNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 204

ggcgtaatac gactcactat aggggtgatta gctcagctgg gagagcacct cccttacaag 60

gagggggtcg gcggttcgat cccgtcatca cccacca 97

<210> 205

<211> 76

<212> RNA

<213> Artificial

<220>

<223> artificially synthesized sequence

<400> 205

gggugauuag cucagcuggg agagcaccuc ccuacaagg agggggucgg cgguucgauc 60

ccgucaucac ccacca 76

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [WO2013100132A \[0016\] \[0179\]](#)
- [WO2007066627A \[0016\]](#)
- [WO2003014354A \[0016\]](#)
- [WO2007103307A \[0016\]](#)
- [WO2002085923A \[0016\]](#)
- [EP16764874A \[0272\]](#)
- [JP2016057707W \[0272\]](#)

Non-patent literature cited in the description

- R. S. LOKEY et al. Nat. Chem. Biol., 2011, vol. 7, 11810-817 [0017]
- J. W. SZOSTAK et al. J. Am. Chem. Soc., 2008, vol. 130, 6131-6136 [0017]
- T. KAWAKAMI et al. Chemistry & Biology, 2008, vol. 15, 32-42 [0017]
- H. KESSLER et al. J. Am. Chem. Soc., 2012, vol. 134, 12125-12133 [0017]
- S.M. HECHT et al. J. Biol. Chem., 1978, vol. 253, 4517-4520 [0017]
- Z. TAN et al. J. Am. Chem. Soc., 2004, vol. 126, 12752-12753 [0017]
- A. O. SUBTELNY et al. Angew Chem Int Ed, 2011, vol. 50, 3164- [0017]
- M. C. T. HARTMAN et al. PLoS one, 2007, vol. 10, e972- [0017]
- S. W. MILLWARD et al. J. Am. Chem. Soc., 2005, vol. 127, 14142-14143 [0017]
- Y. YAMAGISHI et al. Chem. Biol., 2011, vol. 18, 1562-1570 [0017]
- H. R. HOOGENBOOM Nature Biotechnol., 2005, vol. 23, 1105-1116 [0017]
- J. R. PEACOCK et al. RNA, 2014, vol. 20, 758-64 [0017]
- T. KAWAKAMI ACS Chem Biol., 2013, vol. 8, 1205-1214 [0017]
- M. C. T. HARTMAN et al. Proc Natl Acad Sci USA., 2006, vol. 103, 4356-4361 [0017]
- KARLINALTSCHUL Proc. Natl. Acad. Sci. USA, 1993, vol. 90, 5873-7 [0072]
- ALTSCHUL et al. J. Mol. Biol., 1990, vol. 215, 403-10 [0072]
- Nucleic Acid Research, 1990, vol. 18, 5322- [0099]
- Idenshi Kougaku Handbook (Genetic Engineering Handbook) Jikken Igaku (Experimental Medicine) 19920000 [0099]
- HOPP, T. P. et al. BioTechnology, 1988, vol. 6, 1204-1210 [0101]
- Proc. Natl. Acad. Sci. USA., 1972, vol. 69, 2110- [0119]
- Mol. Gen. Genet., 1979, vol. 168, 111- [0119]
- J. Mol. Biol., 1971, vol. 56, 209- [0119]
- Proc. Natl. Acad. Sci. USA., 1978, vol. 75, 1927- [0119]
- J. Bacteriol., 1983, vol. 153, 163- [0119]
- Virology, 1973, vol. 52, 456- [0119]
- Mol. Cell. Biol., 1983, vol. 3, 2156-2165 [0119]

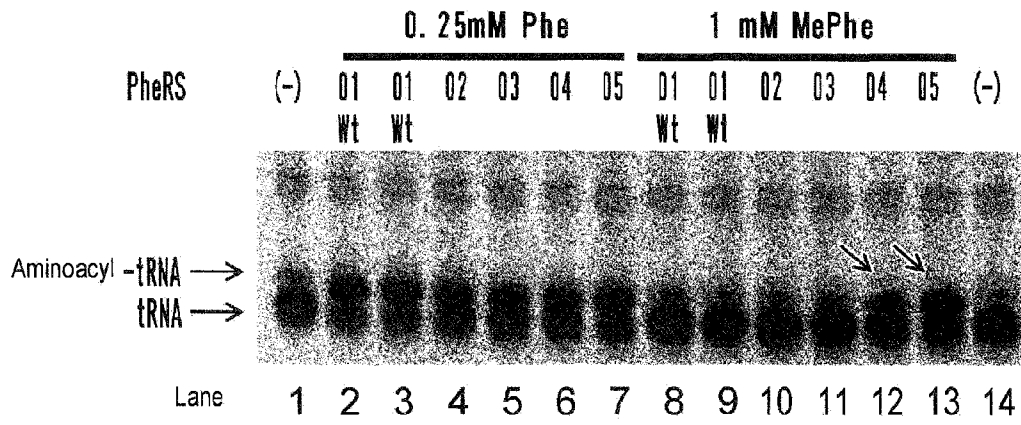
Patentkrav

- 1.** Modificeret valyl-tRNA-syntetase (ValRS) som inkorporerer en N-methylvalin mere effektivt end ValRS'en med aminosyresekvensen SEQ ID NO:24, hvor nævnte modificerede ValRS er valgt fra følgende (a) og (b):
- 5 (a) en ValRS modificeret ved en position(er) svarende til asparagin ved position 43 og/eller threonin ved position 45 og/eller threonin ved position 279 af ValRS fra *Escherichia coli* med aminosyresekvensen SEQ ID NO:24, og
- 10 (b) en ValRS med (i) glycin eller alanin ved en position svarende til asparagin ved position 43 og/eller (ii) serin ved en position svarende til threonin ved position 45 og/eller (iii) glycin eller alanin ved en position svarende til threonin ved position 279 af ValRS fra *Escherichia coli* med aminosyresekvensen SEQ ID NO:24,
- 15 hvor den modificerede ValRS af (a) og (b) omfatter en aminosyresekvens med mindst 90% identitet med en aminosyresekvens valgt fra gruppen bestående af SEQ ID NOs: 3 til 5 og 182 og 183.
- 2.** Den modificerede ValRS ifølge krav 1, hvor modifikationen omfatter mindst en aminosyresubstitution, som forårsager en reduktion på 10 eller mere i
- 20 molekylvægt.
- 3.** Den modificerede ValRS ifølge kravene 1 eller 2, hvor ValRS'en omfatter aminosyrer valgt fra gruppen bestående af SEQ ID NOs: 3 til 5 og 182 og 183.
- 25 **4.** Polynukleotid som koder for den modificerede ValRS ifølge et hvilket som helst af kravene 1 til 3.
- 5.** Vektor omfattende polynukleotidet ifølge krav 4.
- 30 **6.** Værtscelle omfattende polynukleotidet ifølge krav 4 eller vektoren ifølge krav 5.

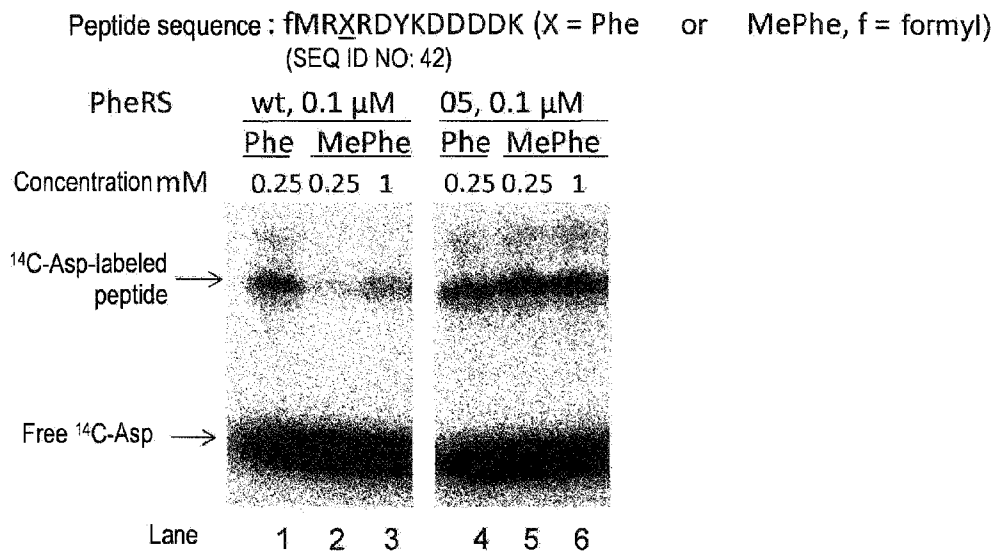
- 7.** Fremgangsmåde til fremstilling af den modificerede ValRS ifølge et hvilket som helst af kravene 1 til 3, omfattende trinnet til dyrkning af værtscellen ifølge krav 6.
- 5 **8.** Fremgangsmåde til fremstilling af et tRNA acyleret med en N-methylvalin, omfattende trinnet at bringe N-methylvalinen i kontakt med et tRNA i tilstedeværelsen af den modificerede ValRS ifølge et hvilket som helst af kravene 1 til 3.
- 10 **9.** Fremgangsmåde til fremstilling af et polypeptid omfattende en N-methylvalin, omfattende trinnet at udføre translation i tilstedeværelsen af den modificerede ValRS ifølge et hvilket som helst af kravene 1 til 3 og N-methylvalinen.
- 10.** Fremgangsmåden ifølge krav 9, hvor trinnet at udføre translation udføres i et
15 cellefrit translationssystem.

DRAWINGS

[Figure 1]



[Figure 2]

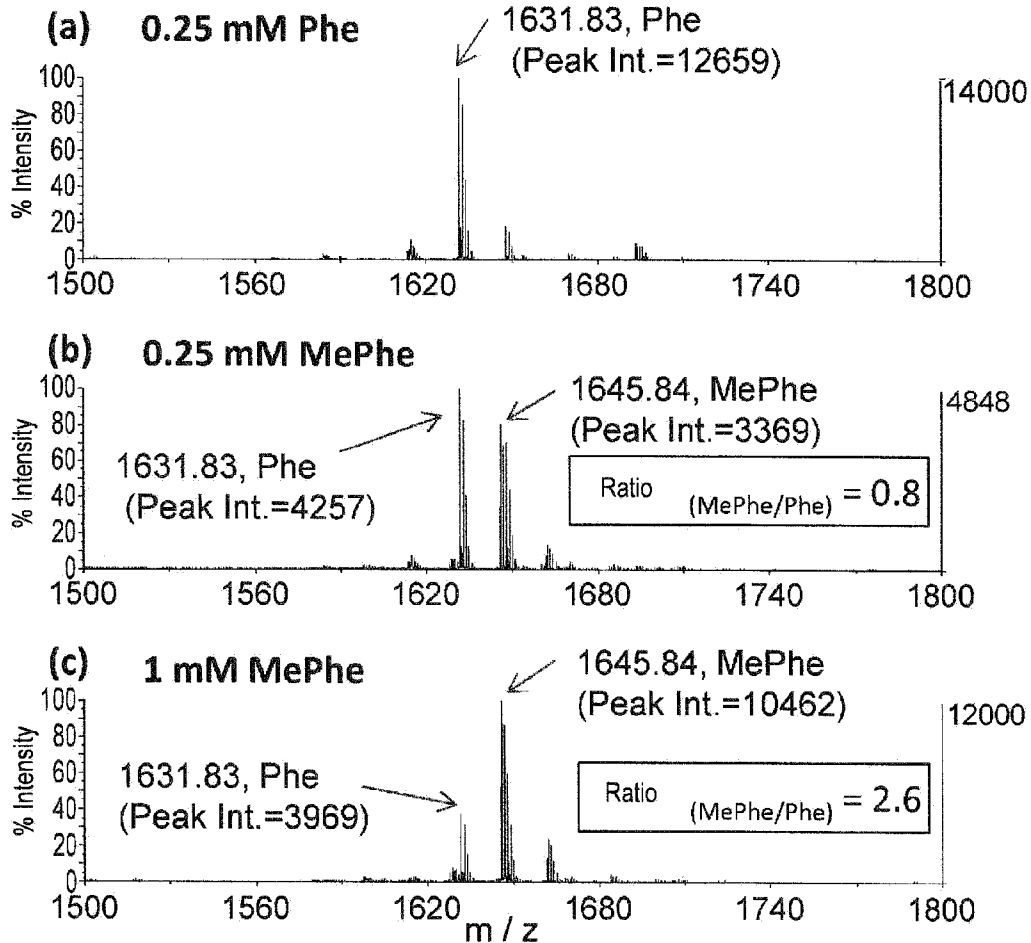


[Figure 3-1]

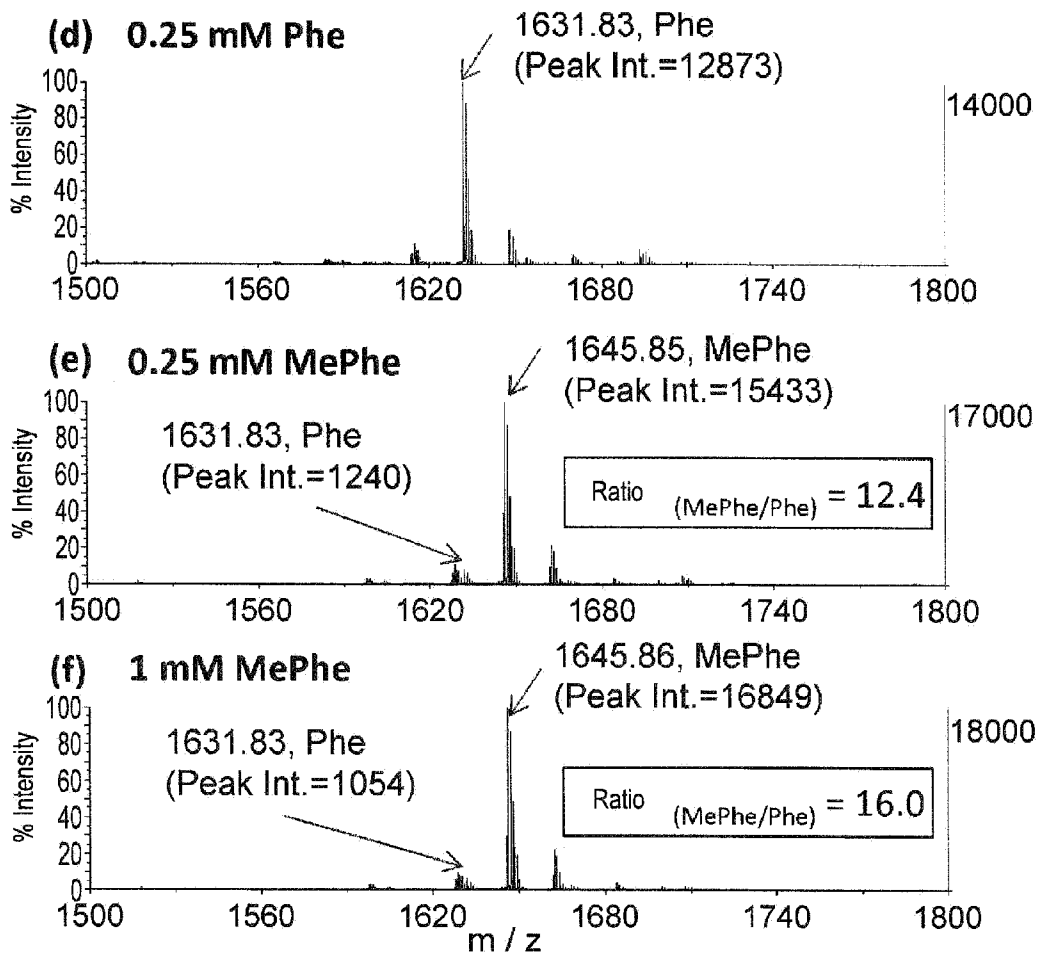
Peptide sequence : fMRPheRDYKDDDDK (f = formyl) 1631.72
 (SEQ ID NO: 42) fMRMePheRDYKDDDDK (f = formyl) 1645.73

Calculated [M+H]⁺

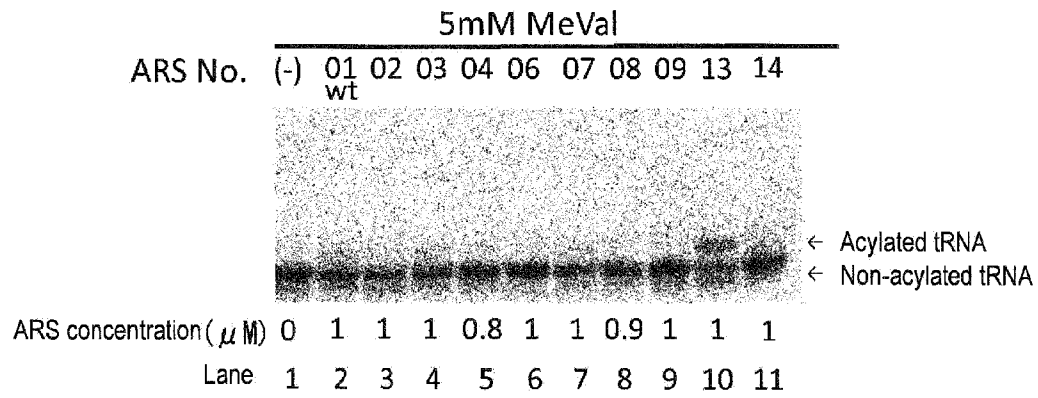
0.1 μ M PheRS wt



[Figure 3-2]

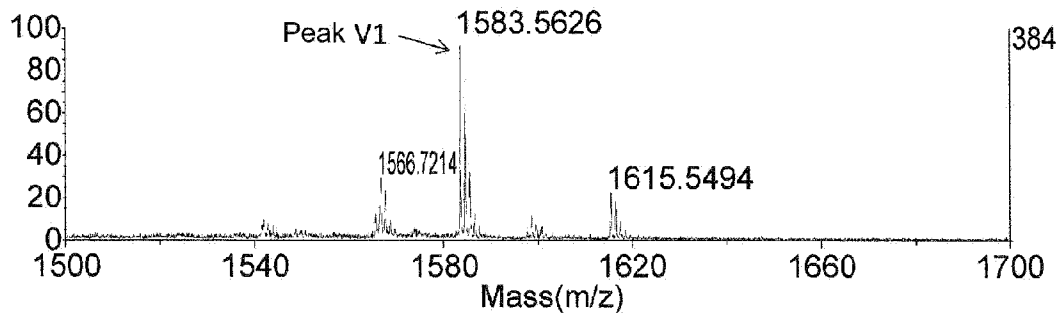
0.1 μ M PheRS05

[Figure 4]

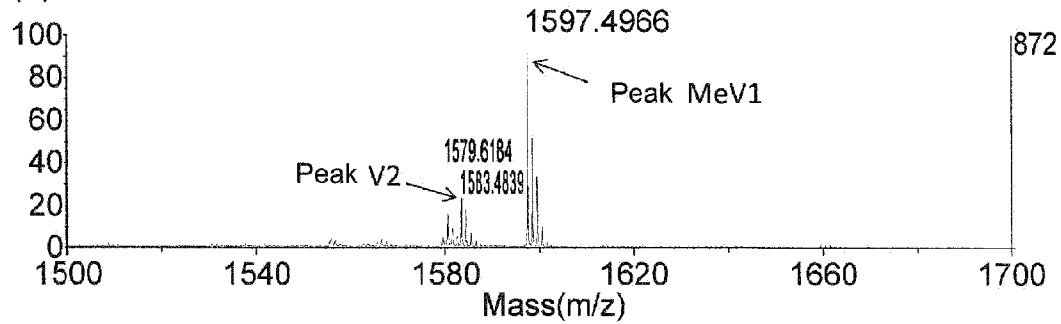


[Figure 5]

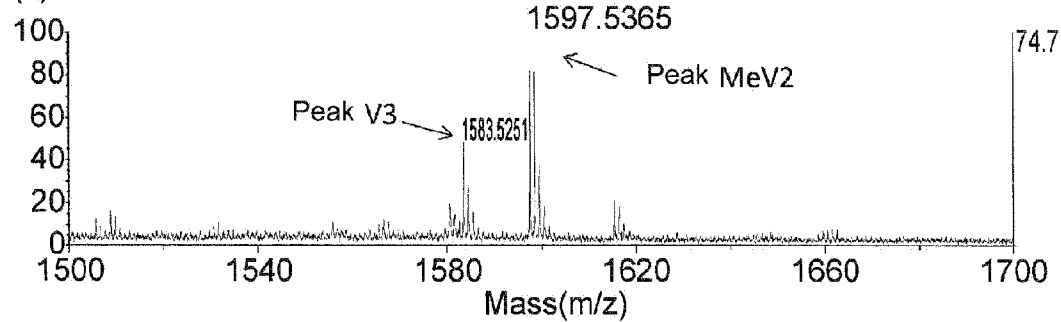
(a) ValRS01(wt)



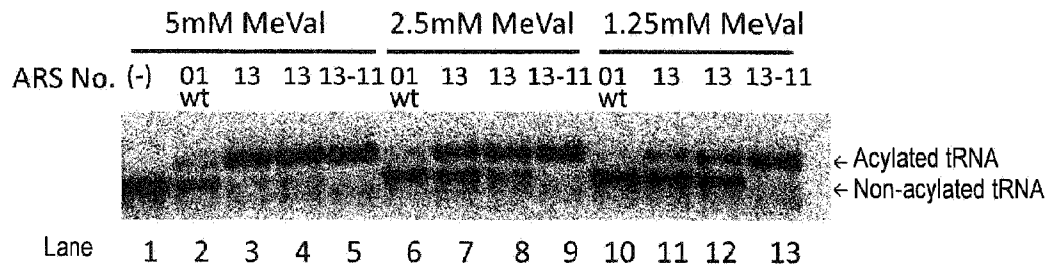
(b) ValRS13



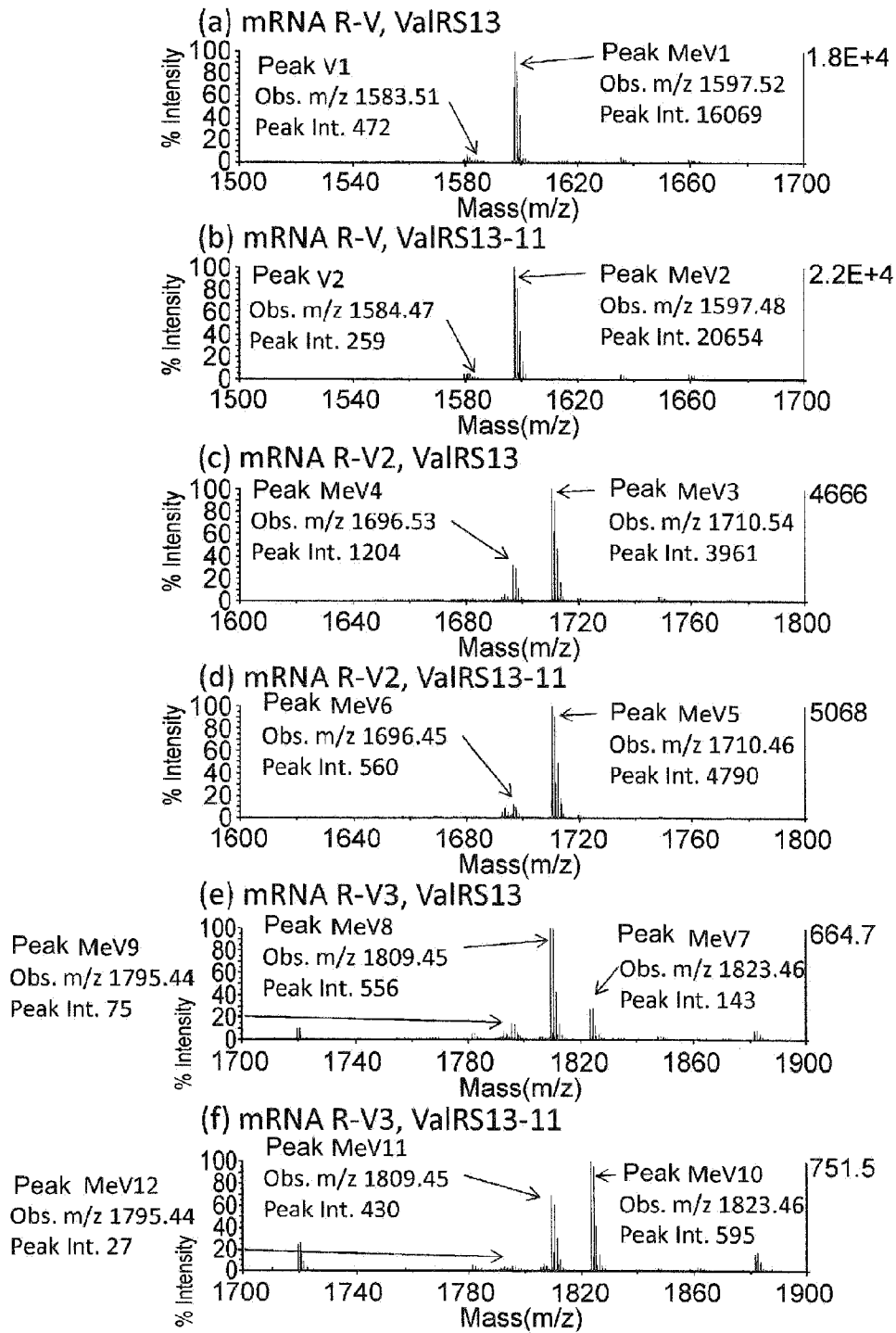
(c) ValRS04



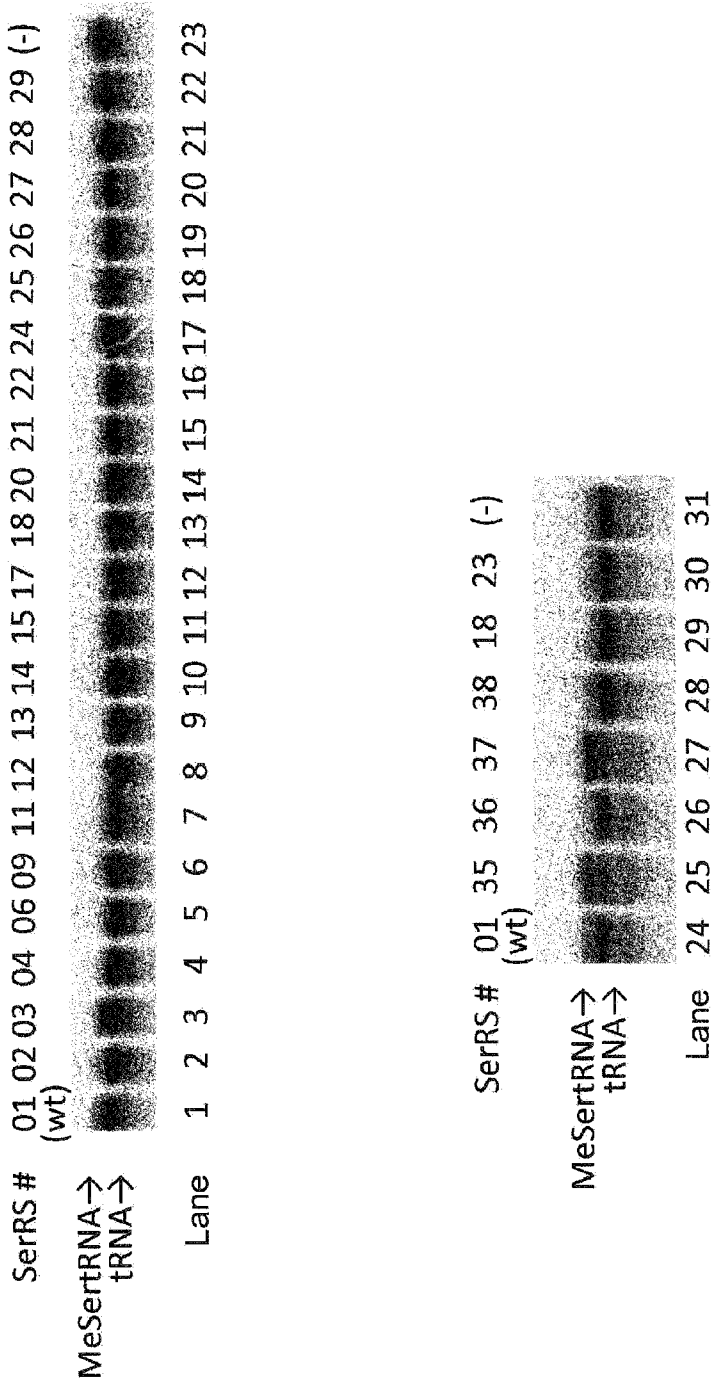
[Figure 6]



[Figure 7]

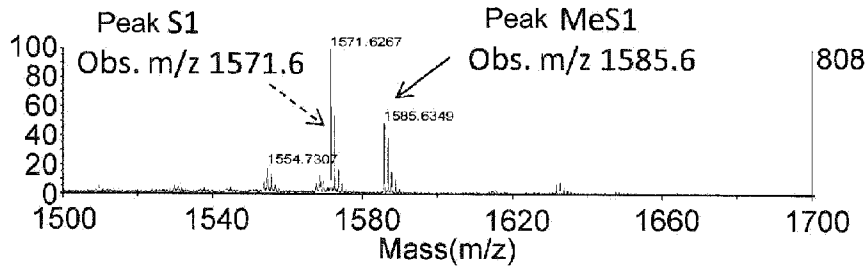


[Figure 8]

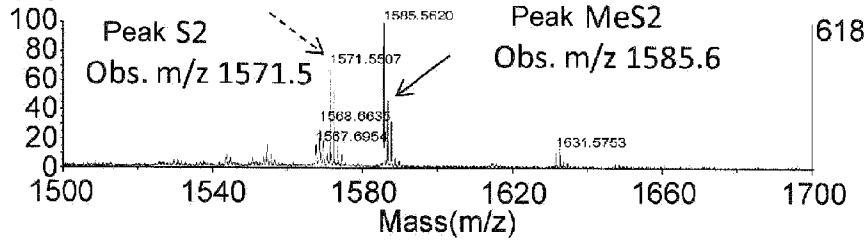


[Figure 9]

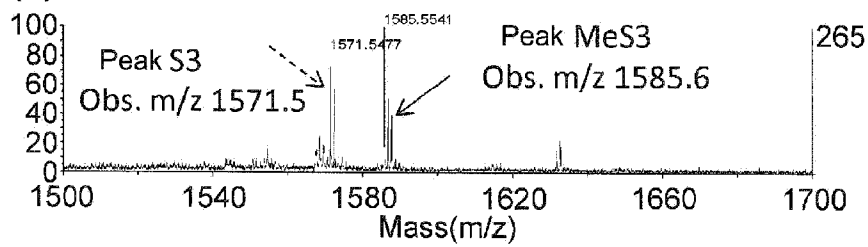
(a) SerRS01(wt)



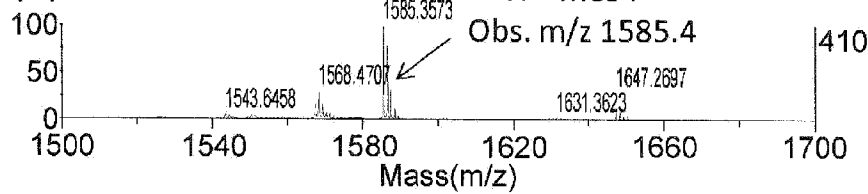
(b) SerRS03



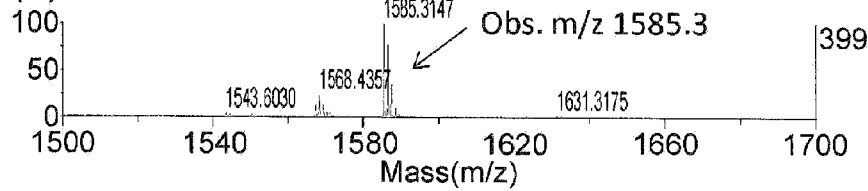
(c) SerRS05



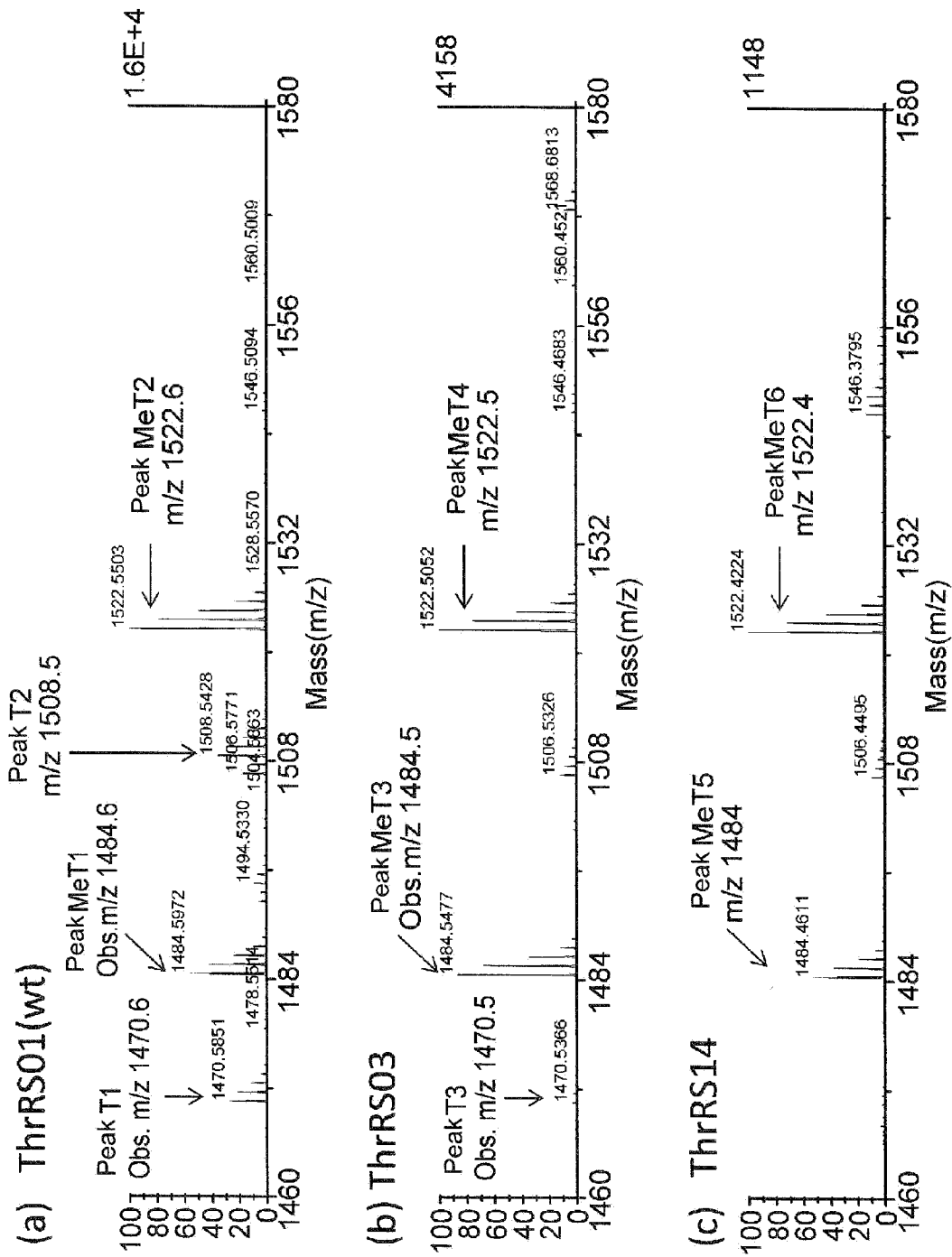
(d) SerRS35



(e) SerRS37

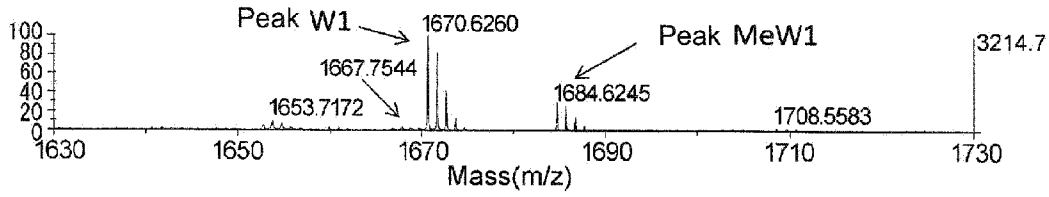


[Figure 10]

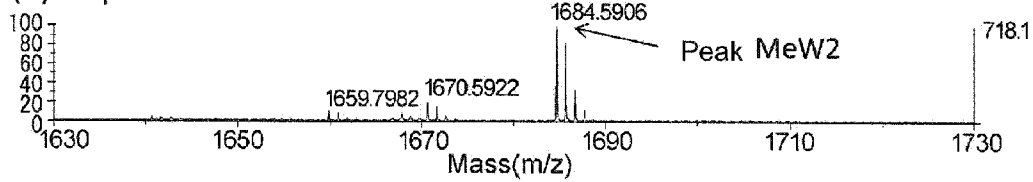


[Figure 11]

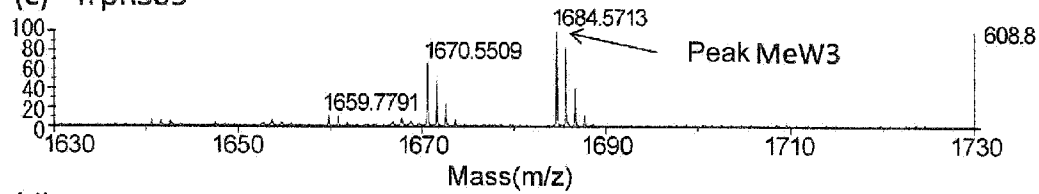
(a) TrpRS01(wt)



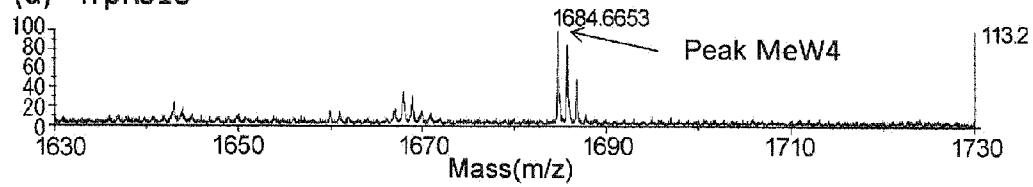
(b) TrpRS04



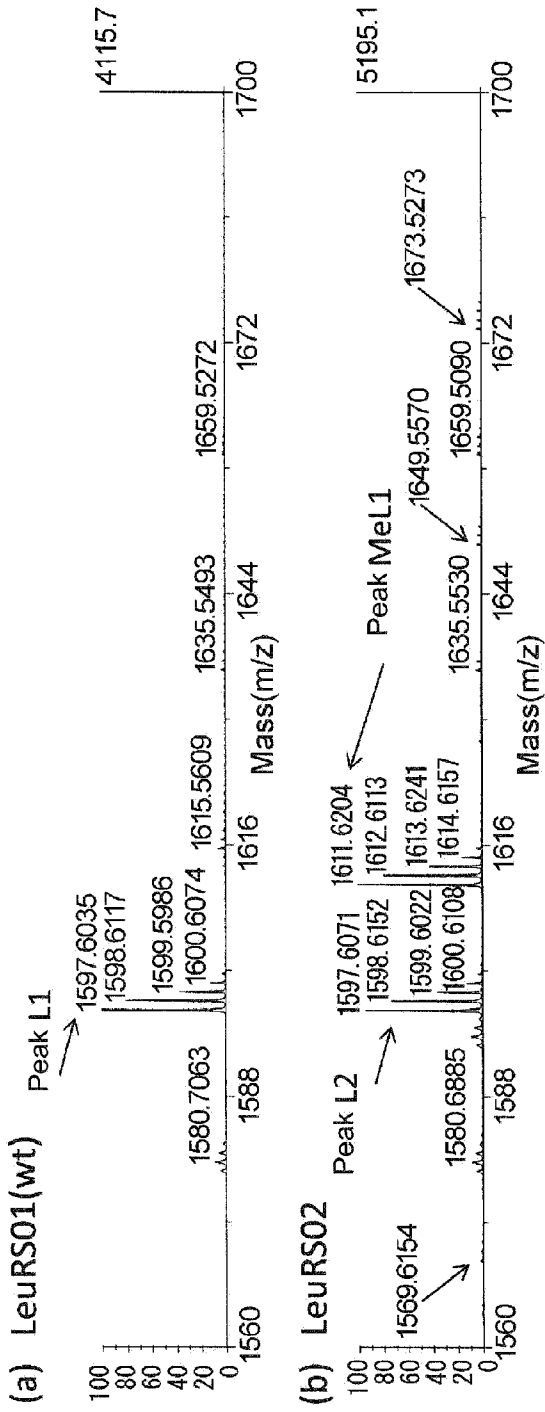
(c) TrpRS05



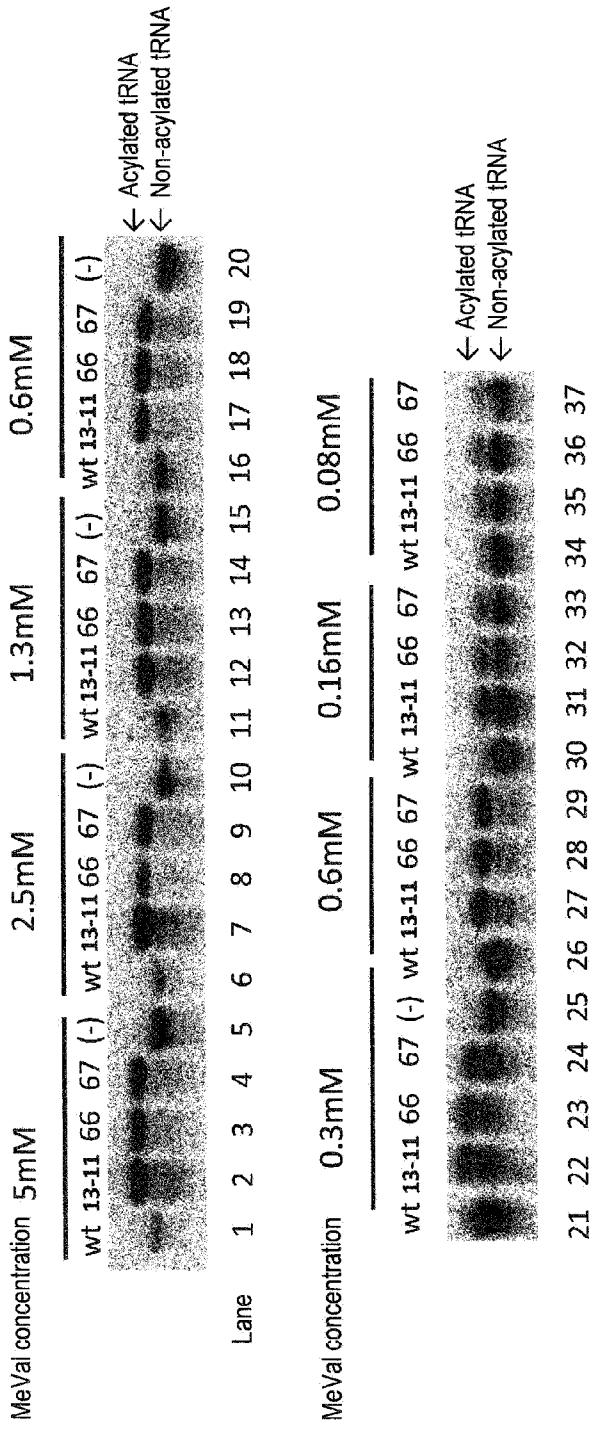
(d) TrpRS18



[Figure 12]



[Figure 13]



[Figure 14]

