US010385367B2

(12) **United States Patent**
Way et al.

(10) **Patent No.: US 10,385,367 B2**
(45) **Date of Patent: Aug. 20, 2019**

(54) **METHODS AND MOLECULES FOR YIELD IMPROVEMENT INVOLVING METABOLIC ENGINEERING**

(75) Inventors: **Jeffrey C. Way**, Cambridge, MA (US); **Joseph H. Davis**, Cambridge, MA (US)

(73) Assignee: **Ginkgo Bioworks, Inc.**

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 39 days.

(21) Appl. No.: **13/322,383**

(22) PCT Filed: **Jun. 1, 2010**

(86) PCT No.: **PCT/US2010/036902**

§ 371 (c)(1),
(2), (4) Date: **Nov. 23, 2011**

(87) PCT Pub. No.: **WO2010/141468**

PCT Pub. Date: **Dec. 9, 2010**

(65) **Prior Publication Data**

US 2012/0070870 A1 Mar. 22, 2012

Related U.S. Application Data

(60) Provisional application No. 61/182,839, filed on Jun. 1, 2009.

(51) **Int. Cl.**

C12N 1/00	(2006.01)
C12N 5/00	(2006.01)
C12P 13/00	(2006.01)
C12P 13/04	(2006.01)
C12P 7/42	(2006.01)
C12N 9/12	(2006.01)
C12N 15/10	(2006.01)
C12N 15/67	(2006.01)

(52) **U.S. Cl.**
CPC ***C12P 7/42*** (2013.01); ***C12N 9/1205*** (2013.01); ***C12N 15/10*** (2013.01); ***C12N 15/67*** (2013.01); *C07K 2319/95* (2013.01)

(58) **Field of Classification Search**
CPC A61K 2039/5256; A61K 2039/5156; A61K 2039/523; C12N 2710/24161; C12N 15/63; C12N 1/21; C07K 2319/00; C12R 1/19; C12P 13/22; C12P 13/24; C12P 13/04
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2006/0073577	A1*	4/2006	Ka-Yiu	C12P 7/40 435/106
2007/0118916	A1	5/2007	Puzio et al.	
2008/0038779	A1	2/2008	Miasnikov et al.	
2009/0000047	A1	1/2009	Trimbur et al.	
2009/0137010	A1*	5/2009	Shakulov et al.	435/107

FOREIGN PATENT DOCUMENTS

EP	WO2009065777	* 5/2009
WO	WO2003089614	* 4/2003
WO	2010/141468	12/2010

OTHER PUBLICATIONS

Dosselaere et al., A Metabolic Node in Action: Chorismate-Utilizing Enzymes in Microorganisms, Critical Reviews in Microbiology (2001), vol. 27, No. 2 , pp. 75-131.*
Hersch et al., SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags., PNAS (2004), vol. 101, pp. 12136-12141.*
Whitten et al., The Structure of the KinA-Sda Complex Suggests an Allosteric Mechanism of Histidine Kinase Inhibition., The Journal of Molecular Biology (2007), vol. 368, Issue 2, pp. 407-420.*
Wiegert et al., SsrA-Mediated Tagging in *Bacillus subtilis*., Journal of Bacteriology (2001), vol. 183, pp. 3885-3889.*
Gottesman, Proteolysis in Bacterial Regulatory Circuits, Annual Review of Cell and Developmental Biology (2003), vol. 19, pp. 565-587.*
Andexer et al., Emerging Enzymes for ATP Regeneration in Biocatlytic Processes., ChemBioChem (2015), vol. 16, pp. 380-386.*
metK_BACSU (last viewed on Feb. 1, 2016).*
SSPB_ECOLI (last viewed on Feb. 1, 2016).*
Meada et al., The Shikimate Pathway and Aromatic Amino acid Biosynthesis in Plants., Annu. Rev. Plant Biol. (2012), vol. 63, pp. 73-105.*
glycolysis-10-steps-explained (last viewed on Aug. 2, 2016).*
Pedrido et al Spo0A links de novo fatty acid synthesis to sporulation and biofilm development in *Bacillus subtilis.,* Molecular Microbiology (2013), vol. 87(2), pp. 348-367.*
Stack et al. A ubiquitin-based tagging system for controlled modulation of protein stability., Nature Biotechnology (2000), vol. 18, pp. 1298-1302.*
pcDNA3 Invitrogen (last viewed on Jul. 13, 2017).*
Kinoshita et al. Regulation of CMV promoter-driven exogenous gene expression with doxorubicin in genetically modified cells. J Pharm Pharmacol. (2008), vol. 60(12), pp. 1659-1665.*
Smith et al. ATP Binding to PAN or the 26S ATPases Causes Association with the 20S Proteasome, Gate Opening, and Translocation of Unfolded Proteins., Molecular Cell (2005), vol. 20, Issue 5, pp. 687-698.*
Kang et al. Functional Proteolytic Complexes of the Human Mitochondrial ATP-dependent Protease, hClpXP., The Journal of Biological Chemistry (2002), vol. 277, pp. 21095-21102.*
Zhang et al. The inhibition of CMV promoter by heat shock factor 4b is regulated by Daxx., The International Journal of Biochemistry & Cell Biology (2010), vol. 42, Issue 10, pp. 1698-1707.*

(Continued)

Primary Examiner — Maryam Monshipouri
(74) *Attorney, Agent, or Firm* — Wolf, Greenfield & Sacks, P.C.

(57) **ABSTRACT**

The invention features methods and compositions relating to cells that have been engineered to reduce or eliminate proteins having enzymatic activity that interfere with the expression of a metabolic product.

27 Claims, 5 Drawing Sheets

Specification includes a Sequence Listing.

(56) **References Cited**

OTHER PUBLICATIONS

Definition of metabolic-pathway (last viewed on Jul. 14, 2017).*
pGEM vector system Promega (last viewed on Jul. 17, 2017).*
Atsumi et al., "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels." Nature. 451: (2008). 86-90.
Burton et al., "Nucleotide-dependent substrate recognition by the AAA+ HsIUV protease," Nat. Struct. Mol. Bio. 12(3): 245-251 (2005).
Chien et al., "Structure and Substrate Specificity of an SspB Ortholog: Design Implications for AAA+ Adaptors," Structure. 15(10):1296-1305 (2007).
Chowdhury et al., "Versatile modes of peptide recognition by the CIpX N domain mediate alternative adaptor-binding specificities in different bacterial species," Protein Science 19(2): 242-254 (2010).
Flynn et al., "Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis," PNAS 98:10584-10589 (2001).
Flynn et al., "Proteomic Discovery of Cellular Substrates of the ClpXP Protease Reveals Five Classes of ClpXP-Recognition Signals," Mol. Cell 11(3): 671-683 (2003).
Griffith et al., "Inducible protein degradation in *Bacillus subtilis* using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP," Mol. Microbiol. 70(4): 1012-1025 (2008).
Grilly et al., "A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae,*" Mol. Syst. Biol. 3 (127): 1-5 (2007).
Gur et al., "Biological Sciences—Biochemistry," PNAS. 105(42): 16113-16118 (2008).
Gur et al., "Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine," PNAS. 106(44): 18503-18508 (2009).
Janssen et al., "Metabolic pathways and energetics of the acetone-oxidizing, sulfate-reducing bacterium, Desulfobacteriumcetonicum," *March. Microbiol.* 163(3): 188-194 (1995).
Kramer et al., "Metabolic engineering for microbial production of shikimic acid,"Metabolic Engineering. 5:277-283 (2003).
McGinness et al.,"Engineering Controllable Protein Degradation," Mol. Cell. 22(5): 701-707 (2006).
Park et al., "Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation," PNAS. 104(19):7797-7802 (2007).
Park and Lee. "Fermentative production of branched chain amino acids: a focus on metabolic engineering," Appl. Micribiol. Biotechnol. 85:491-596 (2010).
Sauer et al., "Sculpting the Proteome with AAA+ Proteases and Disassembly Machines," Cell. 119(1): 9-18 (2004).
Voelker and Davies. "Alteration of the specificity and regulation of fatty acid synthesis of *Escherichia coli* by expression of a plant medium-chain acyl-acyl carrier protein thioesterase," J. Bact. 176 (23): 7320-7327 (1994).
Yuan et al., "Molecular basis for the accumulation of acidic isozymes of triosephosphate isomerase on aging," Mech. Ageing Dev. 17(2): 151-162 (1981).
Zhou et al., "Harnessing the Ubiquitination Machinery to Target the Degradation of Specific Cellular Proteins," Molecular Cell. 6(3): 751-756 (2000).
International Search Report for International Patent Application No. PCT/US2010/36902 dated Oct. 28, 2010.
Written Opinion for PCT/US2010/036902, 6 pages dated (Oct. 28, 2010).

* cited by examiner

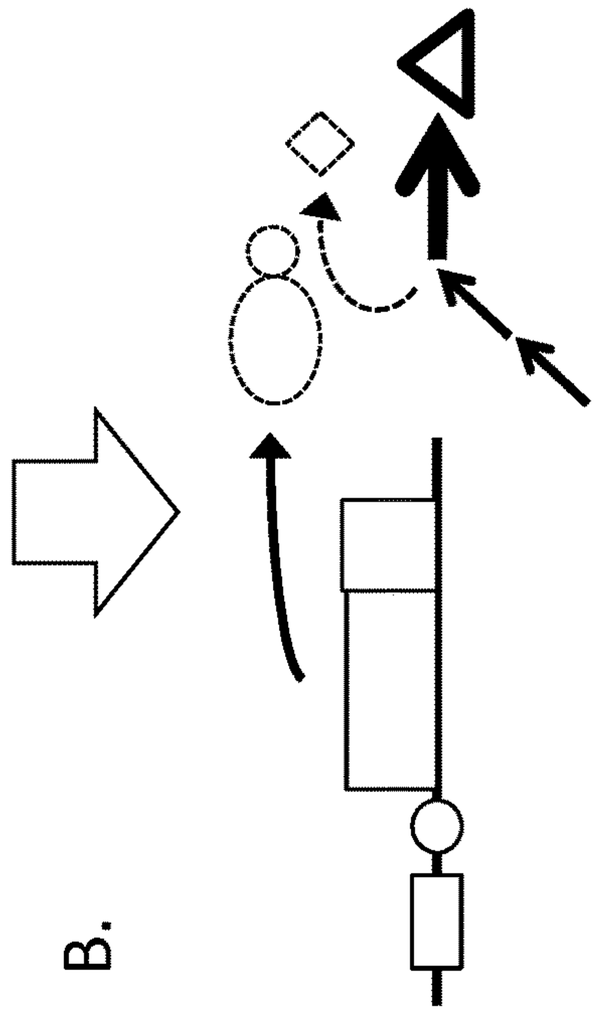


FIG. 1

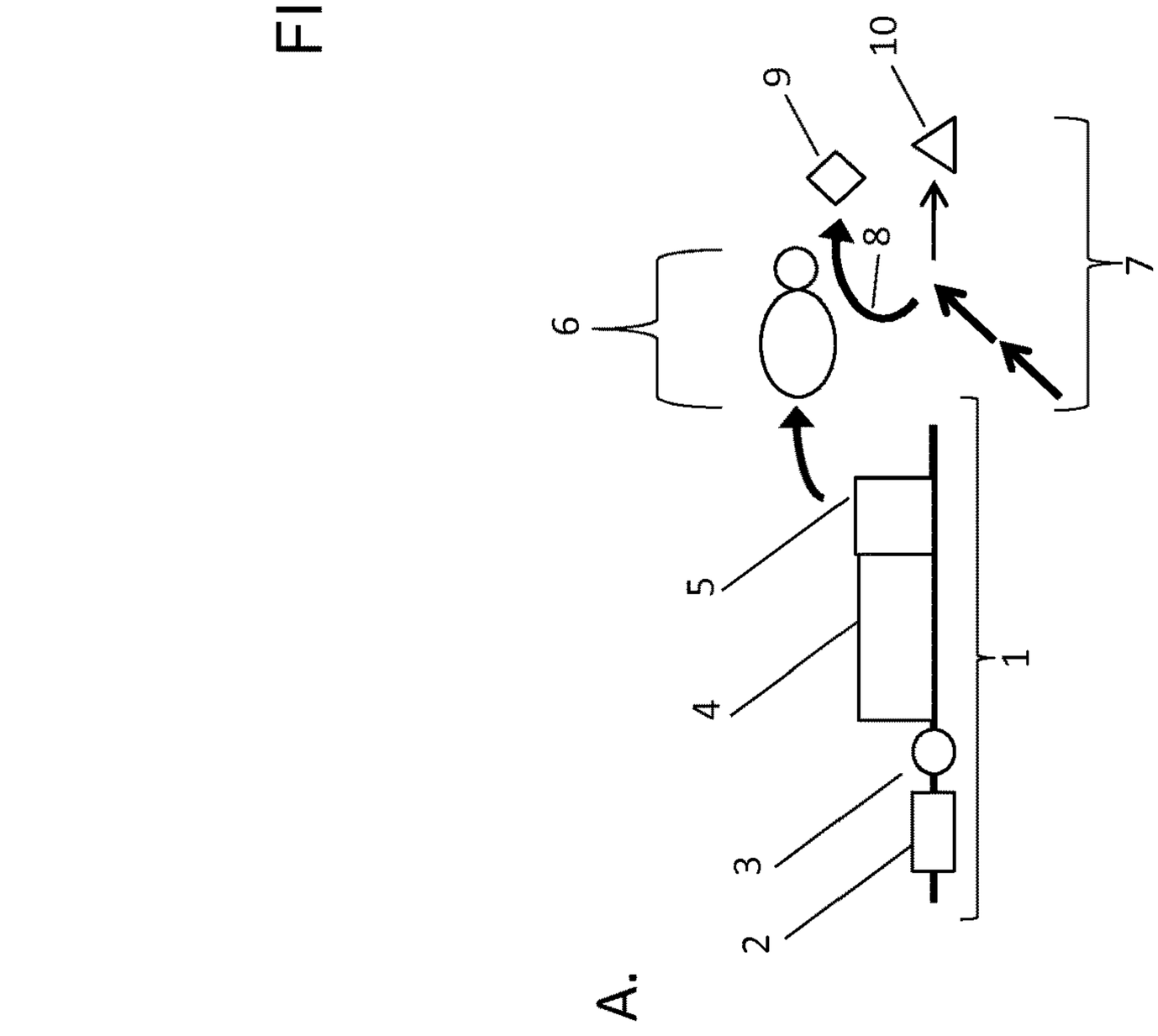
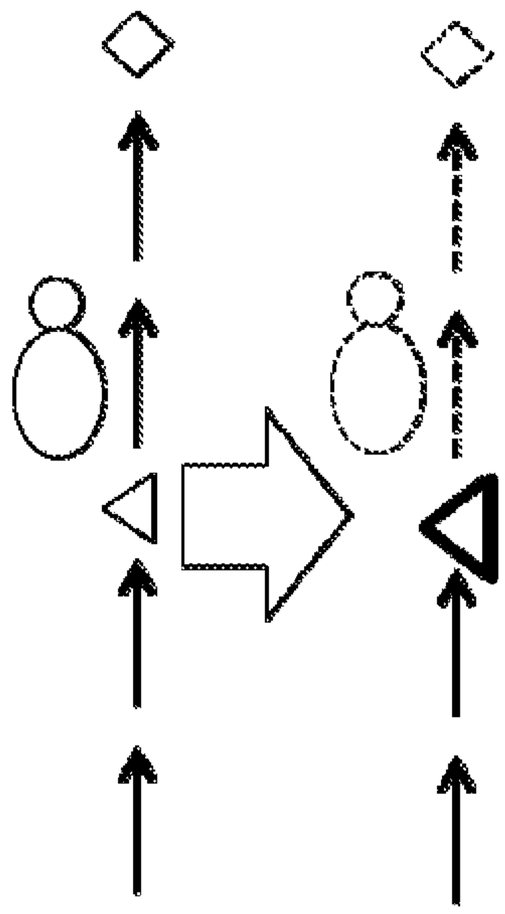


FIG. 2



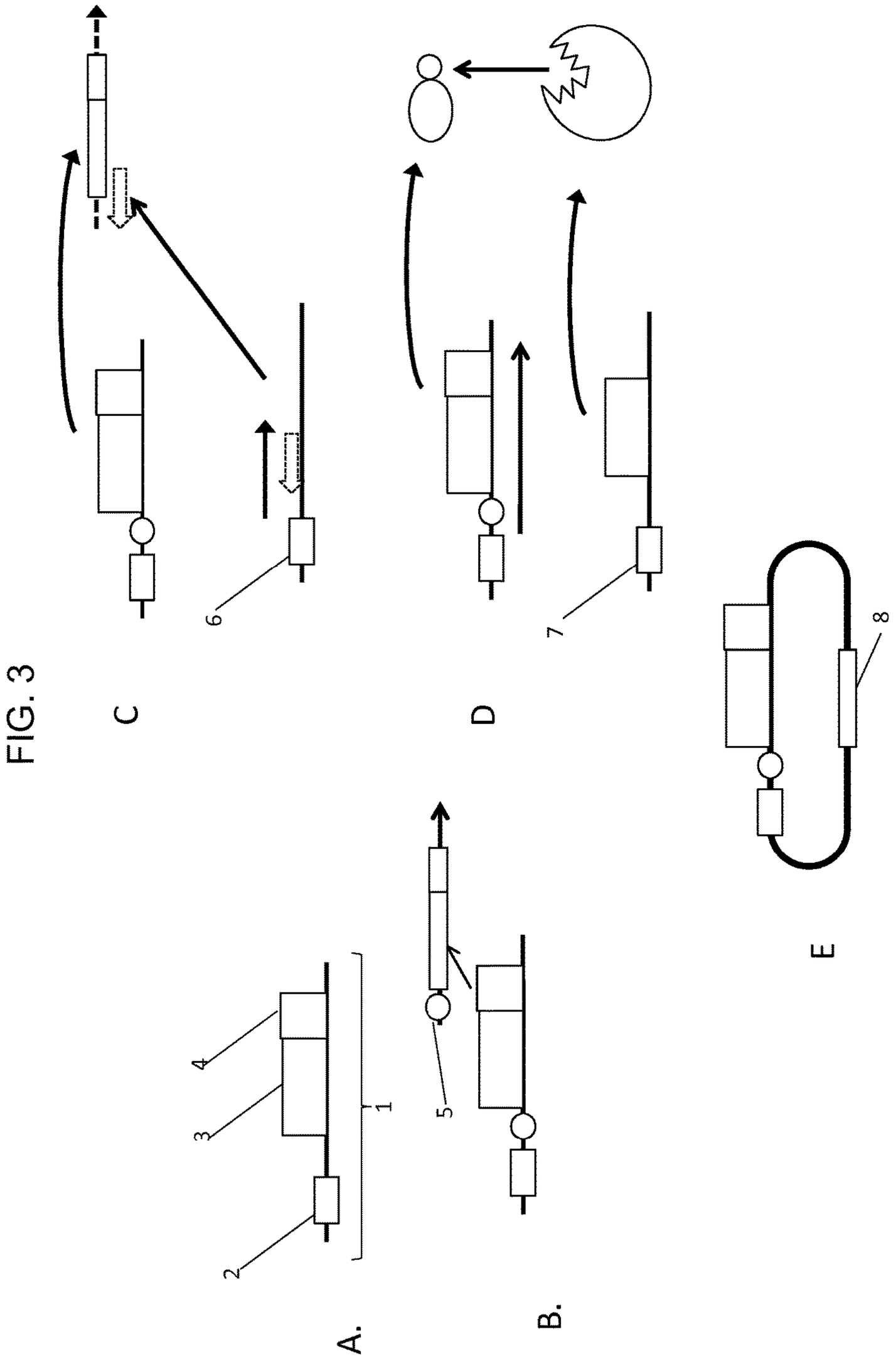
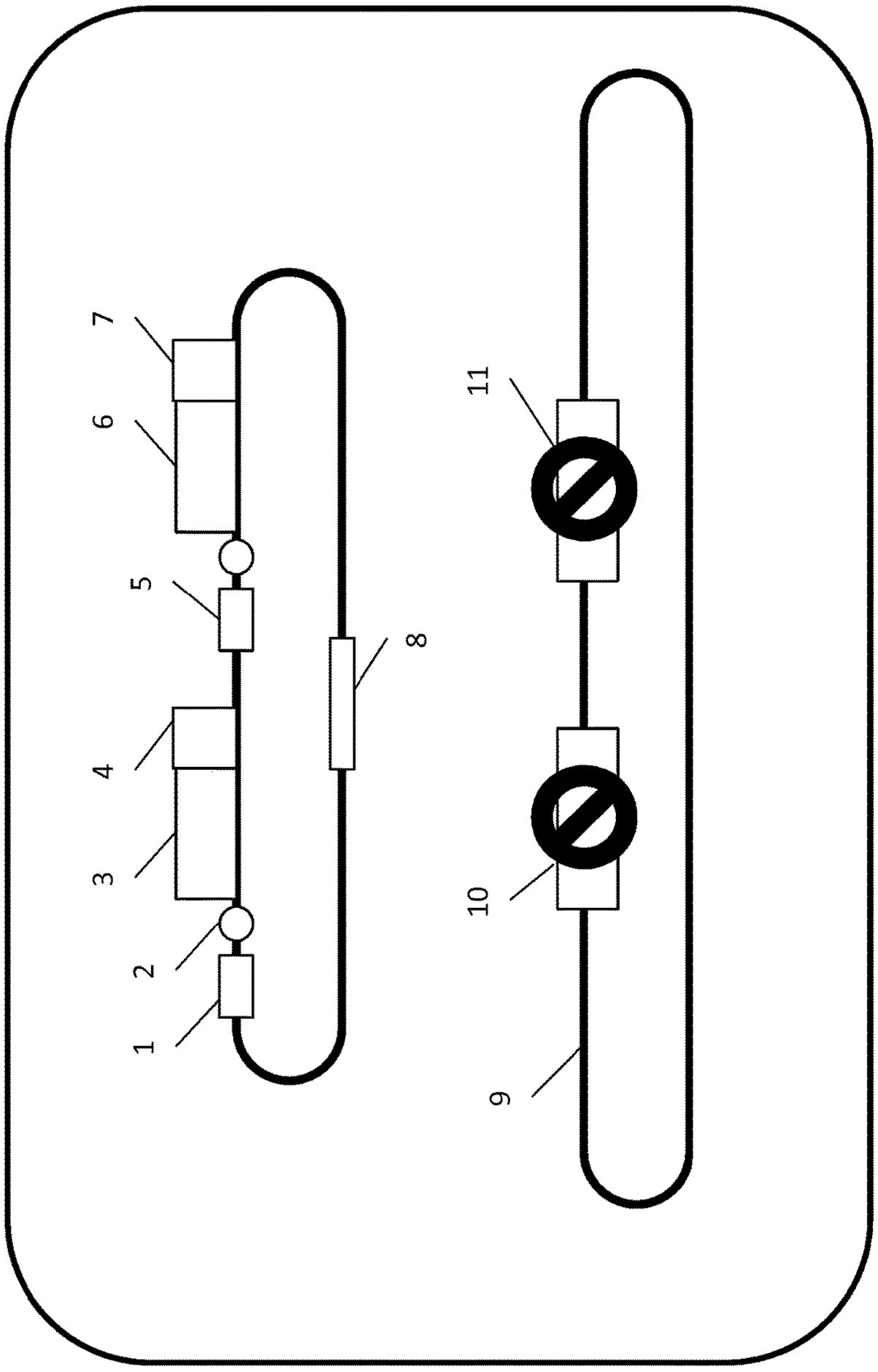


FIG. 3

FIG. 4



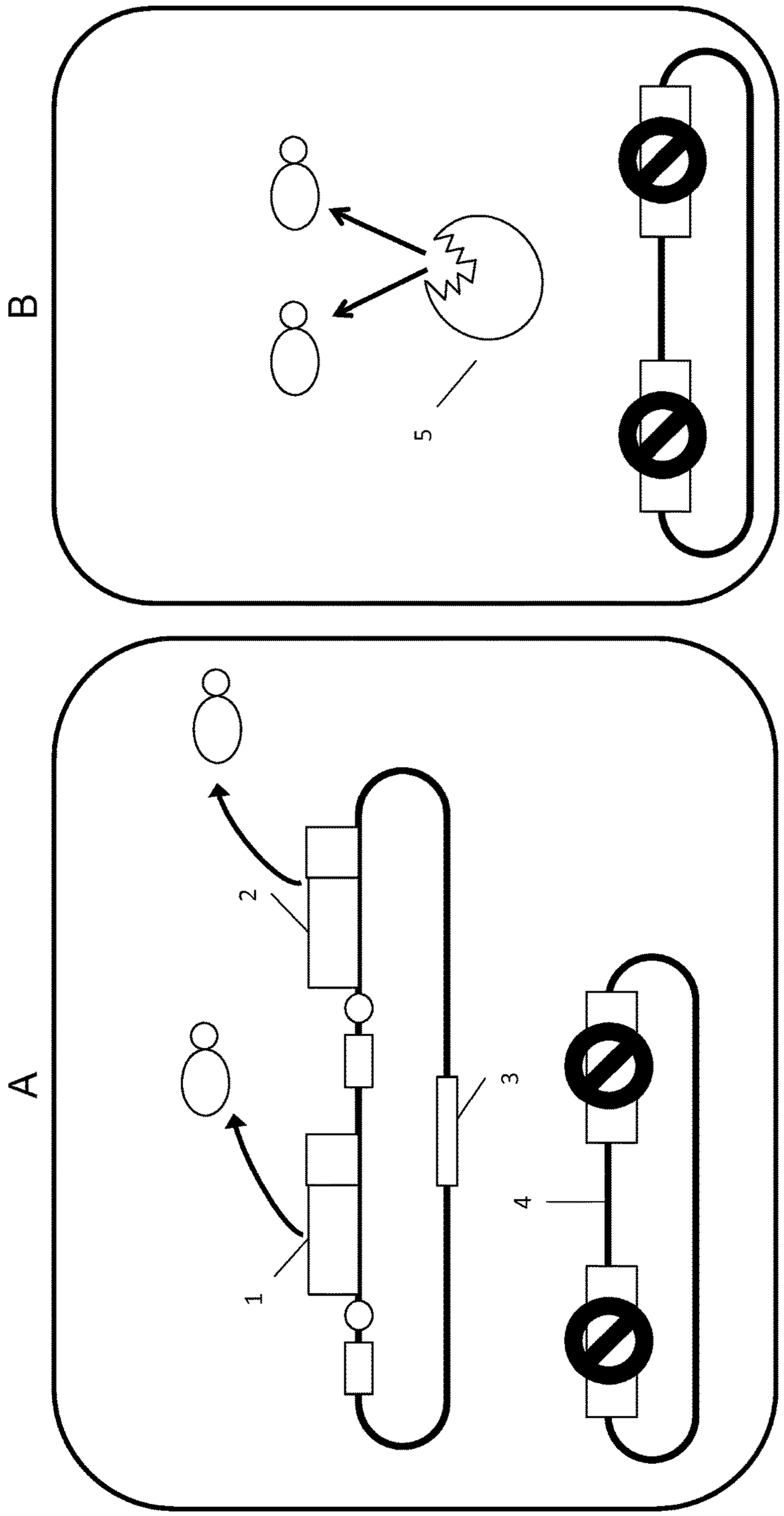


FIG. 5

METHODS AND MOLECULES FOR YIELD IMPROVEMENT INVOLVING METABOLIC ENGINEERING

RELATED APPLICATIONS

This application is a National Phase application of International Application No. PCT/US2010/036902 filed Jun. 1, 2010, which claims priority to and the benefit of U.S. Provisional Application No. 61/182,839, filed Jun. 1, 2009. The entire teachings of the above applications are incorporated herein by reference.

BACKGROUND OF THE INVENTION

In general, the invention relates to metabolic engineering of cells for the enhanced production of a cellular product.

Metabolic engineering involves the industrial production of chemicals from biological sources. Typically, a microbe such as a bacterium or a single-celled eukaryote is engineered to produce a compound in large amounts that is normally produced in small amounts or not at all. Examples of compounds produced by metabolic engineering include ethanol, butanol, lactic acid, various vitamins and amino acids, and artemisinin. Metabolic engineering generally involves genetic modification of a host organism, such as expression of foreign genes to make enzymes that synthesize compounds that may not be native to the host organism, overexpression of genes using strong promoters, introduction of mutations that alter allosteric regulation, and introduction of mutations that limit the production of alternative products.

It is generally desirable to produce compounds as cheaply and efficiently as possible. One major cost in metabolic engineering is the 'feedstock'—the mixture of nutrients used in the medium in which the microbe grows. The feedstock typically includes a carbohydrate source, a source of fixed nitrogen, sources of sulfur, phosphorus, and so on, as well as any specific nutritional requirements. One significant problem in metabolic engineering is that even under conditions of product production, much of the feedstock is channeled into other metabolic pathways that contribute to growth of the organism and production of its biomass. A second problem is the cost of the feedstock itself, especially when the feedstock includes, in addition to a carbohydrate, molecules that fulfill auxotrophic requirements. Therefore, there is a need in the art to limit production of biomass during metabolic engineering and also to reduce the cost of the feedstock.

SUMMARY OF THE INVENTION

The invention generally provides improved cells, molecules, and methods for synthesis of products by metabolic engineering. In a general embodiment, the invention provides an engineered cell that synthesizes a product more cost-effectively than current methods by making use of a cell with the following characteristics. The cell contains one or more proteins that include an enzymatic function with an engineered connection to a sequence that can promote degradation of the protein. The cell also includes a regulatory system such that upon addition or withdrawal of a regulatory factor, which may be a chemical, a protein, photons, temperature, or any other factor, the degradation of the protein is enhanced. As a result, the metabolism of the cell is altered so that the synthesis and/or secretion of a desired product is enhanced. In a further embodiment, the desired product is obtained from the cell or the medium. The enzymatic function may promote growth of the cell during an expansion phase or may allow the culturing and expansion of the cell with less or none of an expensive feedstock component.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the enzyme is a catabolic enzyme.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the enzyme is an anabolic enzyme.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the enzyme is an anabolic enzyme.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the cell is a bacterial cell.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the cell is a fungal cell.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the cell is an insect cell, a plant cell, a protozoan cell, or a mammalian cell.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the regulatory system controls synthesis of the protein.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degrada-

tion of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the regulatory system controls synthesis of a second factor that controls the degradation of the protein.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, synthesis and/or secretion of a desired product is consequently enhanced, and wherein the sequence that can promote degradation of the protein includes an amino acid sequence that differs from the sequence Ala-Ala-Asn-Asp-Glu-Asn-Tyr-Ala-Leu-Ala-Ala (SEQ ID NO: 1) by at most four amino acid substitutions or deletions.

In a distinct class of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, wherein the enzymatic function in an amino acid biosynthetic function.

In a preferred embodiment, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, wherein the enzymatic function is part of aromatic amino acid synthesis.

In a distinct set of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, wherein the enzymatic function is part of the tricarboxylic acid cycle.

In a distinct set of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, wherein the enzymatic function is part of fatty acid synthesis, the oxidative pentose phosphate pathway, or glycolysis.

In a distinct set of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, wherein the enzymatic function is a kinase, an acetyl-CoA-producing enzyme, an enzyme that joins two carbon-containing reactant molecules into a single, carbon-containing product molecule, and an allosterically regulated enzyme

In a distinct set of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, wherein enzymatic function is pyruvate kinase, shikimate kinase, pyruvate dehydrogenase, citrate synthase, and DAHP synthase.

In a distinct set of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, wherein the enzymatic function is hexokinase, glucokinase, glucose-6 phosphatase, glucose-6-phosphate dehydrogenase, glucose phosphate isomerase, phosphofructokinase, fructose bisphosphate aldolase, glyceraldehyde phosphate dehydrogenase, triose phosphate isomerase, phosphoglyceromutase, enolase, phosphoenolpyruvate carboxykinase, pyruvate kinase, pyruvate dehydrogenase, pyruvate decarboxylase, pyruvate-formate lyase, lactate dehydrogenase, pyruvate carboxylase, citrate synthase, aconitate hydratase, isocitrate dehydrogenase, 2-oxoglutarate dehydrogenase, dihydrolipoamide succinyltransferase, succinyl-CoA ligase, succinyl-CoA hydrolase, succinate dehydrogenase, fumarase, malate dehydrogenase, malate synthase, isocitrate lyase, 2-oxoglutarate synthase, glutamate synthase, glutamate dehydrogenase, acetate CoA-ligase, acetyl-CoA carboxylase, malonyl-CoA transferase, acyl-carrier protein acetyltransferase, glutamine synthase, pyrroline-5-carboxylase reductase, glutamate ammonia ligase, aspartate transaminase, ornithine carbamoyl-transferase, arginino-succinate synthetase, aspartate-carbamoyl-transferase, arginino-succinate lyase, arginase, a tRNA charging enzyme, tyrosine transaminase, anthranilate synthase, prephenate dehydratase, prephenate dehydrogenase, chorismate mutase, chorismate synthase, 3-phosphoshikimate carboxyvinyltransferase, shikimate kinase, shikimate dehydrogenase, 3-dehydroquinate dehydratase, 3-dehydroquinate synthase, DAHP synthase, D-phosphoglycerate dehydrogenase, phosphoserine transaminase, phosphoserine phosphatase, glycerol kinase, PRPP synthase, histidinol dehydrogenase, glucosamine acetyltransferase, glycogen synthase, 6-phosphoglucose lactonase, phosphogluconate dehydrogenase, ribose-5-phosphate isomerase, carbamoyl phosphate synthase, isopentenyl-diphosphate isomerase, dimethylallyl transferase, mevalonate kinase, HMG-CoA reductase, NADP/NAD oxidoreductase, formate dehydrogenase, hydrogenase, nitrate reductase, nitrite reductase, farnesyl-trans-transferase, geranyl-trans-transferase, ATP phosphoribosyl transferase, amido-P-ribosyl transferase, and arginine decarboxylase.

In a related embodiment, the invention also features nucleic acids encoding proteins, in which the nucleic acid comprises a sequence encoding a protein having any of the above enzymatic activities.

In a distinct class of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, wherein the regulatory system involves expression of an anti-sense RNA.

In a distinct class of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, wherein the regulatory system controls the expression of a protein that promotes degradation of the artificial protein.

In a distinct class of embodiments, the invention provides an engineered cell that contains a protein that includes an enzymatic function and a sequence that can promote degradation of the protein, a regulatory system such that upon addition or withdrawal of a regulatory factor, the degradation of the protein is enhanced, wherein the regulatory system controls replication or segregation of a plasmid.

The invention also provides nucleic acids encoding proteins, wherein the nucleic acid comprises a sequence encoding an enzyme fused to a sequence that can promote degradation of the protein, wherein the enzyme is an amino acid biosynthetic protein, a protein in the tricarboxylic acid cycle, a glycolytic enzyme, a fatty acid biosynthetic enzyme, or an enzyme of the oxidative pentose phosphate pathway, and wherein the nucleic acid further comprises an engineered operable linkage to a regulatory element.

The invention also provides nucleic acids encoding proteins, wherein the nucleic acid comprises a sequence encoding a shikimate kinase enzymatic activity fused to a sequence that can promote degradation of the protein, and wherein the nucleic acid optionally comprises an engineered operable linkage to a regulatory element.

The invention also provides methods of production, in which a cell containing a protein that includes an enzymatic function with an engineered connection to a sequence that can promote degradation of the protein is induced to undergo a regulatory switch that promotes degradation of the protein, enhanced synthesis of a desired product results, and the product is obtained from the culture of the cell.

In a preferred embodiment, the invention also provides methods of production, in which a cell containing a protein that includes an enzymatic function with an engineered connection to a sequence that can promote degradation of the protein is induced to undergo a regulatory switch that promotes degradation of the protein, enhanced synthesis of a desired product results, the product is obtained from the culture of the cell, and the product is purified.

In a more preferred embodiment, the invention provides methods of production of shikimic acid, in which a cell containing a protein that includes an shikimate kinase enzymatic activity with an engineered connection to a sequence that can promote degradation of the protein is induced to undergo a regulatory switch that promotes degradation of the protein, enhanced synthesis of a desired product results, the product is obtained from the culture of the cell, and the product is purified.

By "amino acid biosynthetic function" is meant an enzymatic activity corresponding to a point in metabolism at or after a point of feedback inhibition by an amino acid.

By "essential gene" of a cell (e.g., microbe) is meant a gene that is required for growth of the cell for the production of a given product.

Other features and advantages of the invention will be apparent from the detailed description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. **1**A and **1**B are schematic drawings showing the use of regulated degradation to enhance production by metabolic engineering. FIG. **1**A shows a genetic construction (**1**) that includes a transcriptional regulatory element (**2**), a translational element (**3**), a coding sequence for a protein of interest such as an enzyme (**4**), fused in-frame to a coding sequence for a peptide or protein element that promotes degradation (**5**), the fusion protein product (**6**) that includes an enzymatic element (large oval) and a degradation tag that can be recognized by a protein degradation system (small oval), a schematic metabolic pathway in which reactions are represented by arrows (**7**), with a particular reaction (**8**) catalyzed by the enzymatic element of the fusion protein, leading to production of an undesired product (diamond, **9**), as well as an alternative pathway leading to production of a desired product (triangle, **10**). FIG. **1**B shows the behavior of the system in response to a regulatory change, in which the levels of the protein (**6**) are reduced or eliminated; the reaction leading to the undesired product is also reduced or eliminated, leading to enhanced production of the desired product.

FIG. **2** is a schematic drawing showing an alternative metabolic pathway in which a desired product (triangle) is an intermediate in the production of an undesired product. In this configuration, the protein that is reduced upon a regulatory switch catalyzes a reaction that converts the desired product into another molecule. When the regulatory switch is activated, the protein is degraded and the desired product accumulates.

FIGS. **3**A-**3**E are schematic drawings showing genetic constructions for regulating the degradation of a protein. FIG. **3**A shows a DNA element (**1**) that includes a regulated promoter (**2**), a coding sequence for an enzyme of interest (**3**), and an in-frame coding sequence for a degradation tag (**4**). FIG. **3**B shows a DNA element similar to that in FIG. **3**A, except that it encodes an mRNA whose translation is regulated by a regulatory site within the mRNA (**5**). FIG. **3**C shows a cellular configuration that includes a gene encoding a protein with a degradation tag, wherein the gene is transcribed, and also includes a second element in which the transcription of an antisense RNA is controlled by a regulated promoter (**6**). When the promoter is induced, the antisense RNA is expressed and binds to the mRNA encoding the protein with the degradation tag, blocking its translation and/or inducing its degradation, for example, by nucleases recognizing double-stranded RNA. FIG. **3**D shows a cellular configuration that includes a gene encoding a protein with a degradation tag, wherein the gene is transcribed, and also includes a second element in which the transcription of a degradation factor is controlled by a regulated promoter (**7**). FIG. **3**E shows a plasmid containing a gene encoding a protein with a degradation tag, and also containing an origin of replication that functions in a conditional manner (**8**).

FIG. **4** is a schematic drawing showing a bacterial cell for production of L-Valine. The cell contains a plasmid encoding constitutive promoters (**1**, **5**) driving transcription of ilvE (**3**) and panB (**6**) fused to ssrA degradation tags variants (**4**, **7**). The protein product from each gene is translated using the encoded ribosome binding site (**2**). This plasmid contains a conditionally-replicated origin (**8**), allowing for facile curing of the plasmid (by a temperature shift, for example). The bacterial chromosome (**9**) contains mutations rendering the endogenous copies of ilvE (**10**) and panB (**11**) inactive.

FIGS. **5**A and **5**B are schematic drawings showing a bacterial cell for production of L-Valine. Under permissive conditions (FIG. **5**A), a conditionally-replicated plasmid (**3**) is maintained by a cell bearing loss-of-function chromosomal mutations (**4**) in specific metabolic enzymes. The plasmid encodes for the production of ssrA-tagged metabolic enzymes (**1**, **2**) which complement the chromosomal mutants. Under the permissive conditions, production of these enzymes outpaces degradation resulting in a steady state pool of the protein products. Upon shifting to the restrictive conditions (FIG. **5**B), the plasmid is lost from the cell, essentially terminating synthesis. Under these conditions, an energy-dependent protease (**5**) degrades the remaining ssrA-tagged protein products.

DETAILED DESCRIPTION

A central aspect of the invention is the insight that it is useful and feasible to essentially harness the power of directed proteolysis to eliminate essential proteins during the production phase of metabolic engineering. To illustrate this insight, the generalized principles are described and exemplary schemes provided.

Broadly speaking, the methods of the invention control either the production, using regulated promoters, or degradation, using fused peptide segments which promote proteolysis (termed 'degradation tags'), of one or more important or essential proteins. When a microbe carrying such a construction is to be grown to a large scale, conditions are

created in which the rate of production of the protein of interest exceeds the combined rates of degradation and dilution (via cell growth and division) of said protein. Such 'growth conditions' produce sufficient steady-state concentrations of the protein of interest to allow for growth and replication of the microbe. When synthesis of a particular product is desired, the fermentation conditions are perturbed such that production is slowed and/or degradation is hastened resulting in depletion of the protein of interest. In general, the protein of interest is an enzyme that controls a major competing metabolic flux that does not contribute to the particular product. Depletion of such an enzyme results in increased flux through the desired metabolic pathway thereby enhancing the production efficiency of the product of interest.

In one instantiation of this technique, the protein of interest is fused to a degradation tag and its production is placed under the control of a regulated promoter. Under 'growth conditions', the promoter is induced such that production outpaces the basal levels of degradation. Upon switching to 'production conditions', the regulated promoter is repressed, thereby largely or completely terminating synthesis. Targeted protein degradation continues unabated until the protein of interest is essentially completely removed from the cell.

In an alternative configuration, the gene of interest may reside on a conditionally-replicated plasmid vector (bearing a temperature-sensitive origin, for example). Under the permissive conditions, the plasmid is maintained by the cell, allowing for robust synthesis of the protein of interest. Upon moving to non-permissive conditions, the plasmid is lost from the cell, essentially terminating synthesis of the protein of interest and, through the aforementioned degradation pathways, resulting in removal of this protein from the cell.

Those skilled in the art of genetic engineering will recognize that the specific features of this approach can be varied and yet produce the same general results. For example, many microbial protein degradation systems, or components thereof (e.g., adaptors, unfoldases, or proteases), are not essential, so an alternative configuration is to express a component of a protein degradation system from a regulated promoter and to express the protein of interest, fused to a degradation tag, from its native promoter or a weak, foreign promoter. In this configuration, the production of the protease component is repressed during the growth phase and induced during the production phase. Thus, protein degradation of the protein of interest is minimal during the 'growth phase' but can be induced during the 'production phase.' This configuration has the advantage of allowing for the use of a native promoter to drive production of the targeted essential protein. Such an approach need not be limited to the endogenous degradation machinery. Foreign degradation components derived from other organisms may be introduced into the strain of interest and utilized as described above. Such approaches obviate the need to perturb the endogenous degradation system, extending the generality of the system to microbes such as *S. cerevisiae* in which such a degradation system (i.e., the 26S proteasome) is essential. Indeed, Grilly et. al. have demonstrated the efficacy of *E. coli*-derived degradation machinery expressed in *Saccharomyces cerevisiae* and generated a strain that allows for targeted, controlled degradation of suitably tagged proteins in *S. cerevisiae* (Grilly et al. *Mol Syst Biol* 3:127 [2007]). Additionally, degradation tags have been identified for multiple energy-dependent proteases including ClpAP, ClpXP, HslUV, and Lon (Gur et al. *PNAS* 106:44 18503-18508 [2008], Gur et al. *PNAS* 105:42 16113-16118 [2008], Burton et al. *Nat Struct Mol Bio* 12(3):245-251 [2005], Flynn et al. *Mol Cell* 11(3):671-683). As such, addition of the appropriate tag to the protein of interest allows for targeted degradation via each of these proteases in a variety of organisms.

When a cell is configured to express an inducible degradation factor with a protein of interest fused to a degradation tag and expressed from a distinctly regulated promoter, under some circumstances the degradation of the protein of interest is inadequate due to continued expression. In such circumstances, it is often useful to express an anti-sense RNA that can inhibit translation of the protein of interest, for example from the same inducible promoter that regulates the degradation factor.

Finally, the production of proteolysis inhibitors or activators may be regulated, either using inducible promoters or conditionally-replicated plasmids, such that targeted degradation is inhibited during the 'growth phase' and permitted during the 'production phase'. These alternative configurations illustrate that the general strategy of causing the disappearance of a protein during a 'production phase' may be implemented in various ways.

To allow for facile induction and repression of the genetic components (e.g., the degradation tagged gene of interest or a component of the degradation system), growth-phase-dependent promoters may be utilized. The *E. coli* promoter, osmY, is known to be strongly induced during stationary phase. The use of this, or a similarly regulated promoter, to drive production of a degradation component would allow for minimal degradation during culture growth (exponential phase) and efficient degradation once the culture had been saturated (stationary phase). As such, the gene of interest could be present during growth of the culture and later depleted allowing for efficient production of the small molecule of interest.

Alternatively, an exponential-phase promoter may be used to drive production of the protein of interest. During growth, production would outpace degradation, allowing for sufficient steady-state levels of this protein to support growth. Upon entering stationary phase, this promoter would be down-regulated, slowing production and allowing for degradation to remove the protein from the cell, thereby terminating growth and improving the production efficiency of the molecule of interest. The principles of the invention may also be applied in a eukaryotic system.

For example, yeasts are often used in the production of ethanol from a carbohydrate. In general, ethanol formation is promoted by pyruvate decarboxylase, while use of carbon for biomass production is promoted by the pyruvate dehydrogenase complex. Accordingly, to enhance the efficiency of ethanol production in yeast, pyruvate dehydrogenase is manipulated as follows. A chromosome gene encoding a subunit of the pyruvate dehydrogenase complex (PDH) is knocked out according to standard procedures. The corresponding gene is placed under control of a regulated promoter, such as a GAL1 promoter, GAL7 promoter or GAL10 promoter, which are inducible by galactose, or the CUP1 promoter, which is inducible by copper, zinc and other metal ions. The coding sequence for the subunit of the pyruvate dehydrogenase complex is also fused to a sequence encoding a protein segment that promotes ubiquitination. For example, an F box protein segment is used as a fusion partner to promote degradation of the subunit of the PDH. Zhou et al. (Molecular Cell [2000] 6:751-756, the entirety of which is incorporated by reference) describe how to construct an F box fusion to a second protein and express the protein in yeast and also in mammalian cells. In a specific

illustration, a CUP1(promoter)-Fbox-PDH subunit genetic construction is placed in a yeast cell with a knockout of the corresponding chromosomal gene encoding the PDH subunit, the yeast cell is grown in the presence of an inducing metal ion, the inducing metal ion is withdrawn, and enhanced ethanol production results.

Production of Lactic Acid

In scaled-up conditions for production of chemicals, it is typical to use low-cost carbohydrate sources such as glucose, sucrose, molasses, high-fructose corn syrup, depolymerized cellulosic biomass, or glycerol as a carbon source. To produce cellular constituents such as amino acids and fatty acids, much of the carbon flux from such carbon sources goes through pyruvate and acetyl-CoA. The latter molecule is the starting point for both the citric acid cycle (also known as the TCA cycle or the Krebs cycle), as well as fatty acid synthesis. Thus, when glucose or an equivalent molecule is used as a carbon source, the process for converting pyruvate to acetyl-CoA is an essential process for growth of typical organisms used in metabolic engineering such as yeast or *E. coli*.

According to the invention, for example, when the goal is to produce a lactic acid, it is useful to eliminate the competing reaction of the conversion of pyruvate to acetyl-CoA. It is generally not useful to simply mutate the gene or genes involved in this process, as they are often important or essential during the organism's growth phase. In the specific case of *E. coli*, two major systems exist for converting pyruvate to acetyl-CoA: pyruvate dehydrogenase and pyruvate-formate lyase. Mutational inactivation of both of these systems prevents growth on glucose as a sole carbon source. According to the invention, one of these systems, such as pyruvate-formate lyase (which functions under anaerobic conditions) is mutated, and pyruvate dehydrogenase is engineered to be active under conditions of growth, but is then post-translationally inactivated. Two specific methods of inactivation are provided by the invention, degradation by proteolysis and enzyme-mediated chemical modification such as phosphorylation. These forms of post-translational modification are optionally inducible and are preferably induced when switching from growth conditions to production conditions. It is also generally useful to turn off transcription of the relevant genes upon switching to production conditions.

In a specific embodiment of the invention, the proteolysis method may be employed as follows. Many bacteria, including *E. coli*, possess compartmentalized, energy-dependent proteases that recognize their substrates via short, fused peptide tags. Experiments in vitro and in vivo have shown that incorporation of such tags into foreign proteins is sufficient to direct efficient proteolysis of the targeted protein. The best characterized tag, ssrA, is derived from a system for degrading incorrectly translated proteins. Said system involves the ssrA tag sequence (Ala-Ala-Asn-Asp-Glu-Asn-Tyr-Ala-Leu-Ala-Ala in *E. coli*; SEQ ID NO: 1), an adaptor protein encoded by sspB that recognizes the ssrA-encoded peptide, and a series of downstream-functioning proteins (ClpX, ClpA, and ClpP) that unfold and degrade the tagged protein (Sauer et al., Cell 119:9-18 [2004]; Flynn et al., PNAS 98:10584-10589 [2001]). Normally, this ssrA tag sequence is incorporated into partially translated proteins where the ribosome has stalled due to a truncated or otherwise defective mRNA. According to the invention, this sequence or a variation thereof is incorporated into a protein of interest such as pyruvate dehydrogenase at the C-terminus. In one variation of the invention, the DNA sequence encoding the pyruvate dehydrogenase-ssrA fusion protein is expressed from an inducible/repressible promoter, and is repressed upon switching engineered bacteria from growth conditions to production conditions. Without wishing to be bound by theory, the pyruvate dehydrogenase-ssrA fusion protein is degraded at a constant rate, and when the transcription of the gene is halted, the mRNA naturally decays and the protein also decays due to the ssrA tag. According to the invention, the user may choose from a wild-type tag or various mutant tags, depending on the desired efficacy of binding between the protease and the substrate. Since the degradation rate of a protein-ssrA fusion will vary somewhat as a function of the protein sequence and the intracellular substrate concentration, some routine experimentation is required to identify an optimal ssrA degradation tag.

Interestingly, experiments have demonstrated that the adaptor protein, SspB is strictly required for efficient degradation of proteins bearing some mutant ssrA tags (for example, AANDENYADAS; SEQ ID NO: 2) (McGinness et al., *Mol. Cell* 22(5):701-707 [2006]). According to the invention, an alternative configuration is the regulated expression of SspB in a strain in which the chromosomal copy of pyruvate dehydrogenase has been fused to the mutated ssrA tag. In this way, the native control elements of pyruvate dehydrogenase remain unperturbed.

Extending this idea, adaptors from other bacteria (*C. crescentus* CC_2101, for example) have been identified which bind their cognate ssrA tags (AANDNFAEEFAVAA in *C. crescentus*; SEQ ID NO: 3) and are capable of delivering bound substrates to *E. coli* ClpXP for degradation (Chien et al., *Structure* 15(10):1296-1305; Griffith et al., *Mol Microbiol* 70(4):1012-1025; Chowdhury et al., *Protein Science* 19(2):242-254). Critically, variants of these foreign tags are not bound by the *E. coli* SspB variant allowing for control of suitably tagged substrates via the foreign adaptor. According to the invention, the chromosomal copy of pyruvate dehydrogenase is fused to such a degradation tag. The cognate adaptor is then introduced on a plasmid vector under the control of a regulated promoter. Pyruvate dehydrogenase is targeted for degradation only under conditions in which the foreign adaptor is produced. In this manner, both the endogenous protease system and control elements of pyruvate dehydrogenase remain unperturbed.

The aforementioned methods require fusion of the degradation tag to the C-terminus of the protein of interest. Experiments have shown that proteins can also be targeted for degradation by ClpXP via N-terminal degradation tags (Flynn et. al., *Mol Cell* 11(3):671-683). Thus, according to the invention, one may alternatively fuse N-terminal degradation tags to the protein of interest (for a representative example, see λO tag, below). Additionally, ClpAP is known to degrade proteins bearing an N-end rule residue (i.e., Leu, Tyr, Trp, or Phe) at their N-terminus. Fusion of endoprotease recognition sites which, when cleaved give rise to one of these N-end rule residues, may also be used to target proteins for degradation via the N-terminus (Wang et al., *Genes Dev* 21(4):403-408). For simplicity, the following discussion will focus on a single implementation in which the protein of interest is targeted for degradation via fusion to an unmodified *E. coli* ssrA tag. Any other tag or degradation system may also be utilized.

Sample degradation tags include those listed in Table 1.

TABLE 1

Wild-type *E. coli* ssrA tag:	AANDENYALAA (SEQ ID NO: 1)
Mutant 1:	AANDENYADAA (SEQ ID NO: 4)
Mutant 2:	AANDENYAAAA (SEQ ID NO: 5)
Mutant 3:	AANDENYAVAA (SEQ ID NO: 6)
Mutant 4:	AANDENYALDA (SEQ ID NO: 7)
Mutant 5:	AANDENYALVA (SEQ ID NO: 8)
Mutant 6:	AANDENYALAG (SEQ ID NO: 9)
Mutant 7:	AANDENYALGG (SEQ ID NO: 10)
Adaptor-dependent tag	AANDENYADAS (SEQ ID NO: 2)
Wild-type *C. crescentus* ssrA tag:	AANDNFAEEFAVAA (SEQ ID NO: 3)
ccSsra Specificity Mutant 1:	ADNDNFAEEFADAS (SEQ ID NO: 11)
λO tag (N-terminal tag)	MTNTAKILNFGRAS (SEQ ID NO: 12)

At low substrate concentrations, the mutant tags allow for a reduced rate of intracellular degradation relative to the wild-type tag.

For the case of lactic acid production, the result is that after switching to a medium that represses synthesis of the pyruvate dehydrogenase-ssrA protein, this protein is degraded over a period of 2-60 minutes depending on the needs of the user, and metabolic flux of carbon into acetyl-CoA from pyruvate essentially ceases. As a result, flux through lactate dehydrogenase is increased. The method of the invention may be employed in combination with other engineering steps that enhance production of lactic acid, such as overproduction of lactate dehydrogenase, mutation of the zwf gene, growth in anaerobic conditions, and so on.

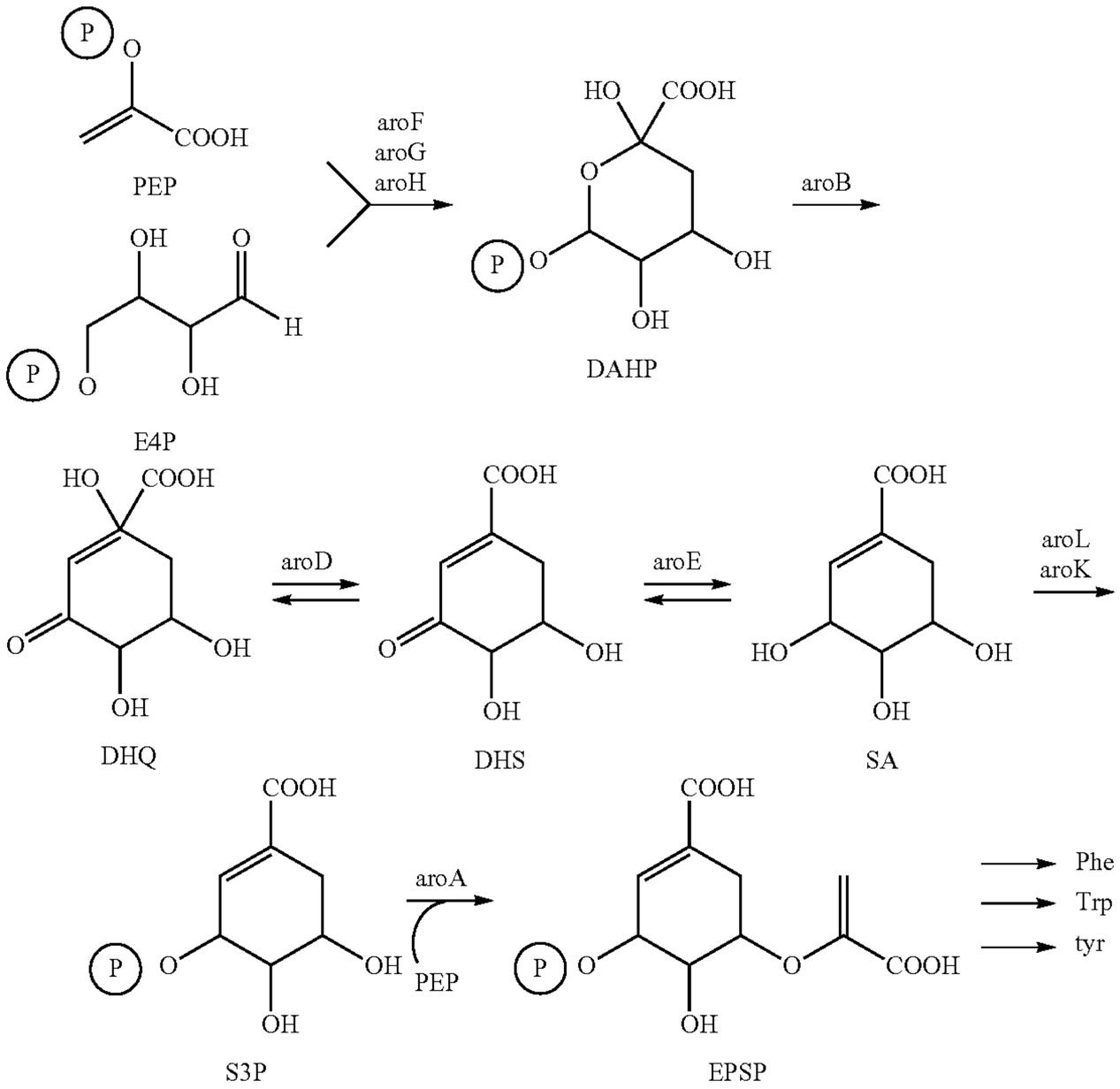
Metabolic engineering techniques to improve the biological production of amino acids have been applied with great success to the microbes *B. subtilis, C. glutamicum*, and *E. coli*. Using directed approaches, genes encoding enzymes that catalyze off-pathway reactions have been removed from the production strain allowing for increased metabolic flux through the pathway of interest. Additionally, random mutagenesis and selected breeding approaches have resulted in strains that overproduce the amino acid of interest (Park et al. PNAS [2007] 104(19):7797-7802). Mapping of said mutant strains often reveals that genes catalyzing off-target reactions have been inactivated confirming the efficacy of this approach. Oftentimes, the off-target pathways catalyze the production of alternative amino acids and thus inactivation of these genes results in strains auxotrophic for a variety of amino acids.

According to the invention, it is both useful and feasible to control the degradation of essential enzymes which catalyze these off-target reactions. Such controlled degradation approaches allow for growth of the strain under conditions in which these targeted enzymes are present and active, relieving the requirement for amino acid supplemented media. Upon changing to conditions of robust degradation or limited production, the targeted enzyme is depleted from the cell, resulting in increased metabolic flux through the pathway of interest and efficient production of the amino acid of interest.

In *E. coli* and the industrially relevant microbe *C. glutamicum*, production of the branched amino acids, L-Leucine, L-Valine and the coenzyme A precursor, pantothenate all utilize the metabolic intermediate, 2-ketoisovalerate. This intermediate is channeled to L-Leucine through the enzyme leuA, to L-Valine through ilvE and to panthonate through panB. According to the invention, when overproduction of L-Leucine is desired, ilvE and panB are targeted for degradation as follows. A plasmid bearing a temperature-sensitive origin as well as ssrA-tagged variants of ilvE and panB driven by a constitutive promoter is transformed into a host strain in which ilvE and panB have been knocked out of the chromosome. Under growth conditions, the plasmid is maintained and production outpaces degradation. Upon conversion to production conditions, the plasmid is cured from the cell, thereby effectively terminating synthesis and allowing for degradation to remove these enzymes from the cell. As such, metabolic flux is diverted toward the production of L-Leucine. Alternatively, when L-Valine production is desired, leuA and panB are targeted for degradation as described above. Critically, such approaches obviate the need to supplement the growth media with expensive amino acids (for example, ilvE-strains are auxotrophic for L-Valine and L-Isoleucine) while maintaining the ability to overproduce the small molecule of interest. A variety of other loss-of-function mutations are known to increase production of said amino acids (reviewed in Park, Lee Appl. Micribiol. Biotechnol. [2010] 85:491-596). According to the invention, such genes are targeted for degradation using the aforementioned approaches, allowing for efficient production of the desired amino acid under degradative conditions and robust cell growth on non-supplemented media under non-degradative conditions.

Shikimic Acid Production

Another example further illustrates the invention. Shikimic acid is an intermediate in aromatic amino acid synthesis, and is also used in the chemical synthesis of the drug Tamiflu® as well as in combinatorial chemical libraries. The pathway for aromatic amino acid synthesis is illustrated below.



In brief, phosphoenolpyruvate and erythrose-4-phosphate, both from central metabolism, are condensed to a single 7-carbon intermediate that is processed through a series of intermediates that ultimately diverge into separate pathways for phenylalanine, tryptophan, and tyrosine. Shikimic acid is produced by the aroE gene product, and is then converted to shikimate phosphate by shikimate kinase, which in *E. coli* is produced independently by two genes, aroL and aroK. Current methods for producing shikimic acid involve the null mutation of both aroL and aroK, blocking shikimate phosphate production and leading to accumulation of shikimic acid. The aroK aroL double mutant is auxotrophic for tryptophan, tyrosine, and phenylalanine, each of which is an expensive molecule that must be added to the feedstock when shikimic acid is produced by metabolic engineering.

According to the invention, a shikimic acid-producing strain may be engineered as follows. One of the shikimate kinase genes, e.g., aroL, is knocked out by standard procedures. The other, e.g., aroK, is expressed with an ssrA peptide fused to its C-terminus. This fusion protein is expressed from a regulated promoter, such as the lac promoter, a quorum-sensing promoter, a promoter that is repressed in low-fixed nitrogen, a promoter that is induced by growth on glucose and repressed by growth on glycerol, or any other promoter that works well in the chosen conditions for switching from a growth mode to a production mode. In this way, the use of tyrosine, tryptophan, and phenylalanine can be avoided.

This control of shikimate kinase levels can be coupled to other strategies to enhance shikimic acid production, some of which are known in the art of metabolic engineering. For example, in *E. coli*, transport of glucose or most other carbohydrates normally involves transfer of a phosphate from phosphoenolpyruvate onto glucose. It is often useful to employ an alternative system using a protein that mediates facilitated diffusion of glucose and related carbohydrates, instead of the PEP-dependent system; a gene such as the glf gene from *Zymomonas mobilis* is often used. One common method is to knock out the endogenous ptsI gene and instead express the glf gene. According to the invention, an alternative method is to express a ptsI-ssrA fusion protein from a regulated promoter, and to also constitutively express the glf gene.

It is also useful to mutate genes encoding proteins that produce alternative products such as quinic acid. Further, it is useful to inactivate the shikimate transporter gene shiA by mutation, thus preventing re-uptake of shikimate that has been secreted. These approaches are based on Kraemer et al. (Metabolic Engineering 5:277-283 [2003], incorporated by reference herein), which reviews these established techniques and strategies.

According to the invention, in addition to blocking function of shikimate kinase, it is often useful to block conversion of PEP to pyruvate, which is normally catalyzed by the enzyme pyruvate kinase. Accordingly, a pyruvate kinase-ssrA fusion protein is expressed from a regulated promoter and the wild-type pyruvate kinase gene is inactivated. The result is accumulation of PEP, which is then used by the engineered bacteria to produce shikimic acid.

More specifically, to produce shikimic acid in an economical manner, an *E. coli* strain that is otherwise wild-type, for example, MG1655 or W3110, may be engineered to have the following alterations:

1. The chromosomal copies of aroK and aroL genes are deleted or otherwise mutated.
2. The chromosomal copy of the ptsI gene is optionally deleted or otherwise mutated.

3. The glf gene of *Zymomonas mobilis* is constitutively expressed.
4. The chromosomal copy of the pyruvate kinase gene is optionally deleted.
5. The following gene fusions are constituted into an operon and expressed from a regulated promoter: aroK-ssrA, and optionally ptsI-ssrA, pyruvate kinase-ssrA. The operon is generated by total gene synthesis from a commercial supplier, such as DNA 2.0, Mr. Gene, Blue Heron Biotechnologies, or Genscript. The operon is integrated into the *E. coli* chromosome.
6. The following regulated promoter systems may be utilized:
   a. The bacteriophage lambda P_R promoter, in the presence of a single copy of the c1857 temperature-sensitive allele of the lambda repressor transcribed from a constitutive promoter.
   b. The lactose operon promoter, in the presence of a single copy of the lacI repressor gene transcribed from a constitutive promoter.
   c. A luxR-responsive promoter, in the presence of a gene encoding the LuxR protein.
7. The strain is optionally engineered to express a sucrose transport system and an invertase.

During the growth phase, the strain is grown in a minimal medium such as M9 medium with glucose, sucrose, or molasses as a carbon source, and in the absence of tryptophan, tyrosine, or phenylalanine. When the lambda P_R system is used, the strain is grown at 42° C. Upon switching to the production phase, the temperature is lowered to 30° C., whereupon shikimic acid is produced. Without wishing to be bound by theory, upon the shift to 30° C., the genes encoding shikimate kinase, pyruvate kinase, and the phosphotransferase I protein are repressed, and the corresponding proteins are degraded and not replaced, since mRNAs in *E. coli* are generally unstable and have a half-life of only a few minutes. The cessation of aromatic amino acid synthesis leads to an up-regulation of the initial steps of this pathway, such as the genes aroF, aroG, and aroH, which encode DAHP synthases. The loss of pyruvate kinase activity leads to an accumulation of phosphoenolpyruvate (PEP), one of the substrates of DAHP synthase. The loss of the phosphotransferase I protein leads to a cessation of glucose transport by the phosphotransferase system, further assisting in PEP accumulation. The loss of shikimate kinase activity results in accumulation of shikimic acid, which is collected by standard procedures.

The *E. coli* strain described above optionally includes other modifications described by Kraemer et al. (op. cit.), including but not limited to deletion of the shikimate transporter shiA, and use of an AroD/E-homologous protein from *N. tabacum* to reduce production of quinic acid.

It should be noted that the extent of repression of the various genes is determined by routine experimentation. For example, it is sometimes useful to separately regulate pyruvate kinase so that its activity is reduced but not completely abolished, so that the citric acid cycle may operate and some ATP may be produced by oxidative phosphorylation. Alternatively, pyruvate kinase may be left unmutated.

Production of Fatty Acids and Alcohols

Biofuels often derive from fatty acids that are derivatized into esters or reduced to fatty alcohols. The starting point for fatty acid synthesis is acetyl-CoA, which is also the starting point for the tricarboxylic acid cycle. According to the invention, it is useful to construct a gene encoding a fusion protein that includes citrate synthase and ssrA, expressed from a regulated promoter. Such a construction has the effect of preventing entry into the TCA cycle, with the result that acetyl-CoA is preferentially directed into fatty acid synthesis. Depending on which other metabolic engineering has been performed, production of ethanol may be enhanced.

As an alternative strategy to producing fatty acids, instead of amino acids, it is sometimes useful to block the synthesis of aromatic amino acids by blocking DAHP synthase. This has the effect of preventing new protein synthesis, leading to some accumulation of other amino acids and feedback inhibition of the enzymes that initiate pathways for their synthesis. Accordingly, a DAHP synthase-ssrA fusion protein is expressed from a regulated promoter, and the promoter is turned off when production of a fatty acid product or related product is desired. In the specific case of *E. coli*, three isotypes of DAHP synthase are encoded by the genes aroF, aroG, and aroH. To apply this method of the invention to *E. coli*, it is generally useful to inactivate the chromosomal copies of these genes by mutation, then construct a fusion of one of genes to DNA encoding the ssrA peptide, which is then placed under the control of a regulated promoter.

As a first illustration, consider the synthesis of dodecanoic acid (lauric acid; C12 fatty acid; $CH_3(CH_2)_{10}COOH$). Voelker and Davies (J. Bact. [1994] 176[23]7320-7327) described an engineered *E. coli* that expressed a plant C12 thioesterase and also carried a knockout of the fadD. The C12 thioesterase has the effect of releasing lauric acid from acyl carrier protein during fatty acid synthesis, and the fadD encodes a fatty acid degradation enzyme that recycles the carbon in fatty acids that cannot be incorporated into membranes. The C12 thioesterase-expressing fadD knockout strain synthesizes lauric acid at a high level. However, it is noteworthy that this strain grows and divides (FIG. 5 of Voelker and Davies), evidently converting much of the input carbon into biomass even though the C12 thioesterase is expressed constitutively at a high level. According to the invention, when a C12 thioesterase-expressing fadD knockout strain is also engineered to express a DAHP synthase-ssrA fusion protein from a regulated promoter, and the promoter is turned off, the DAHP synthase-ssrA fusion protein is degraded and not replaced, protein synthesis essentially ceases, and production of lauric acid is enhanced relative to the C12 thioesterase-expressing fadD knockout strain.

As a second illustration, consider the synthesis of isobutanol ($(CH_3)_2CHCH_2OH$). Atsumi et al. (Nature [3 Jan. 2008] 451:86-90) described an engineered *E. coli* that expressed an artificial operon that expressed high levels of isobutanol by a combination of valine biosynthesis genes, 2-ketoacid decarboxylase, and alcohol dehydrogenase. According to the invention, when a strain expressing valine synthesis genes, 2-ketoacid decarboxylase, and alcohol dehydrogenase is also engineered to express a DAHP synthase-ssrA fusion protein from a regulated promoter, and the promoter is turned off, the DAHP synthase-ssrA fusion protein is degraded and not replaced, protein synthesis essentially ceases, and production of isobutanol is enhanced relative to the parental isobutanol-secreting strain.

More broadly, Atsumi et al. described the production of a variety of alpha-keto carboxylic acids such as 2-ketobutyrate, 2-ketoisovalerate, 2-ketovalerate, 2-keto-3-methyl-valerate, 2-keto-4-methyl-valerate, and phenylpyruvate, which can be decarboxylated to create an aldehyde and then reduced by the serial actions of 2-ketoacid decarboxylase, and alcohol dehydrogenase, to create a series of useful alcohols. According to the invention, when such strains are also engineered to express a DAHP synthase-ssrA fusion

protein from a regulated promoter, and the promoter is turned off, the DAHP synthase-ssrA fusion protein is degraded and not replaced, protein synthesis essentially ceases, and production of the desired alcohols is enhanced relative to the parental alcohol-producing strains.

Sequences Provided by the Invention

The following protein and DNA sequences further illustrate the invention.

Shikimate kinase (AroK)-ssrA
(SEQ ID NO: 14)
MAEKRNIFLVGPMGAGKSTIGRQLAQQLNMEFYDSDQEIEKRTGADVG
WVFDLEGEEGFRDREEKVINELTEKQGIVLATGGGSVKSRETRNRLSA
RGVVVYLETTIEKQLARTQRDKKRPLLHVETPPREVLEALANERNPLY
EEIADVTIRTDDQSAKVVANQIIHMLESNAANDENYALAA

Shikimate kinase-linker-ssrA
(SEQ ID NO: 15)
MAEKRNIFLVGPMGAGKSTIGRQLAQQLNMEFYDSDQEIEKRTGADVG
WVFDLEGEEGFRDREEKVINELTEKQGIVLATGGGSVKSRETRNRLSA
RGVVVYLETTIEKQLARTQRDKKRPLLHVETPPREVLEALANERNPLY
EEIADVTIRTDDQSAKVVANQIIHMLESNGGSGGAANDENYALAA

λO-Shikimate kinase
(SEQ ID NO: 16)
MTNTAKILNFGRASMAEKRNIFLVGPMGAGKSTIGRQLAQQLNMEFYD
SDQEIEKRTGADVGWVFDLEGEEGFRDREEKVINELTEKQGIVLATGG
GSVKSRETRNRLSARGVVVYLETTIEKQLARTQRDKKRPLLHVETPPR
EVLEALANERNPLYEEIADVTIRTDDQSAKVVANQIIHMLESN

λO-linker-Shikimate kinase
(SEQ ID NO: 17)
MTNTAKILNFGRASGGSGGMAEKRNIFLVGPMGAGKSTIGRQLAQQLN
MEFYDSDQEIEKRTGADVGWVFDLEGEEGFRDREEKVINELTEKQGIV
LATGGGSVKSRETRNRLSARGVVVYLETTIEKQLARTQRDKKRPLLHV
ETPPREVLEALANERNPLYEEIADVTIRTDDQSAKVVANQIIHMLESN

PtsI-ssrA
(SEQ ID NO: 18)
MISGILASPGIAFGKALLLKEDEIVIDRKKISADQVDQEVERFLSGRA
KASAQLETIKTKAGETFGEEKEAIFEGHIMLLEDEELEQEIIALIKDK
HMTADAAAHEVIEGQASALEELDDEYLKERAADVRDIGKRLLRNILGL
KIIDLSAIQDEVILVAADLTPSETAQLNLKKVLGFITDAGGRTSHTSI
MARSLELPAIVGTGSVTSQVKNDDYLILDAVNNQVYVNPTNEVIDKMR
AVQEQVASEKAELAKLKDLPAITLDGHQVEVCANIGTVRDVEGAERNG
AEGVGLYRTEFLEMDRDALPTEEEQFAAYKAVAEACGSQAVIVRTMDI
GGDKELPYMNFPKEENPFLGWRAIRIAMDRREILRDQLRAILRASAFG
KLRIMFPMIISVEEVRALRKEIEIYKQELRDEGKAFDESIEIGVMVET
PAAATIARHLAKEVDFFSIGTNDLTQYTLAVDRGNDMISHLYQPMSPS
VLNLIKQVIDASHAEGKWTGMCGELAGDERATLLLLGMGLDEFSMSAI
SIPRIKKIIRNTNFEDAKVLAE
QALAQPTTDELMTLVNKFIEEKTICAANDENYALAA

-continued

Pyruvate kinase I-ssrA (pykF as opposed to pykA)
(SEQ ID NO: 19)
MKKTKIVCTIGPKTESEEMLAKMLDAGMNVMRLNFSHGDYAEHGQRIQ
NLRNVMSKTGKTAAILLDTKGPEIRTMKLEGGNDVSLKAGQTFTFTTD
KSVIGNSEMVAVTYEGFTTDLSVGNTVLVDDGLIGMEVTAIEGNKVIC
KVLNNGDLGENKGVNLPGVSIALPALAEKDKQDLIFGCEQGVDFVAAS
FIRKRSDVIEIREHLKAHGGENIHIISKIENQEGLNNFDEILEASDGI
MVARGDLGVEIPVEEVIFAQKMMIEKCIRARKVVITATQMLDSMIKNP
RPTRAEAGDVANAILDGTDAVMLSGESAKGKYPLEAVSIMATICERTD
RVMNSRLEFNNDNRKLRITEAVCRGAVETAEKLDAPLIVVATQGGKSA
RAVRKYFPDATILALTTNEKTAHQLVLSKGVVPQLVKEITSTDDFYRL
GKELALQSGLAHKGDVVVMVSGALVPSGTTNTASVHVLAANDENYALA
A

Citrate synthase-ssrA (gltA)
(SEQ ID NO: 20)
MADTKAKLTLNGDTAVELDVLKGTLGQDVIDIRTLGSKGVFTFDPGFT
STASCESKITFIDGDEGILLHRGFPIDQLATDSNYLEVCYILLNGEKP
TQEQYDEFKTTVTRHTMIHEQITRLFHAFRRDSHPMAVMCGITGALAA
FYHDSLDVNNPRHREIAAFRLLSKMPTMAAMCYKYSIGQPFVYPRNDL
SYAGNFLNMMFSTPCEPYEVNPILERAMDRILILHADHEQNASTSTVR
TAGSSGANPFACIAAGIASLWGPAHGGANEAALKMLEEISSVKHIPEF
VRRAKDKNDSFRLMGFGHRVYKNYDPRATVMRETCHEVLKELGTKDDL
LEVAMELENIALNDPYFIEKKLYPNVDFYSGIILKAMGIPSSMFTVIF
AMARTVGWIAHWSEMHSDGMKIARPRQLYTGYEKRDFKSDIKRAANDE
NYALAA

DAHP-ssrA (tyrosine-repressible)
(SEQ ID NO: 21)
MQKDALNNVHITDEQVLMTPEQLKAAFPLSLQQEAQIADSRKSISDII
AGRDPRLLVVCGPCSIHDPETALEYARRFKALAAEVSDSLYLVMRVYF
EKPRTTVGWKGLINDPHMDGSFDVEAGLQIARKLLLELVNMGLPLATE
ALDPNSPQYLGDLFSWSAIGARTTESQTHREMASGLSMPVGFKNGTDG
SLATAINAMRAAAQPHRFVGINQAGQVALLQTQGNPDGHVILRGGKAP
NYSPADVAQCEKEMEQAGLRPSLMVDCSHGNSNKDYRRQPAVAESVVA
QIKDGNRSIIGLMIESNIHEGNQSSEQPRSEMKYGVSVTDACISWEMT
DALLREIHQDLNGQLTARVAAANDENYALAA

ClpX (unfoldase from *E. coli*)
(SEQ ID NO: 22)
MTDKRKDGSGKLLYCSFCGKSQHEVRKLIAGPSVYICDECVDLCNDII
REEIKEVAPHRERSALPTPHEIRNHLDDYVIGQEQAKKVLAVAVYNHY
KRLRNGDTSNGVELGKSNILLIGPTGSGKTLLAETLARLLDVPFTMAD
ATTLTEAGYVGEDVENIIQKLLQKCDYDVQKAQRGIVYIDEIDKISRK
SDNPSITRDVSGEGVQQALLKLIEGTVAAVPPQGGRKHPQQEFLQVDT
SKILFICGGAFAGLDKVISHRVETGSGIGFGATVKAKSDKASEGELLA
QVEPEDLIKFGLIPEFIGRLPVVATLNELSEEALIQILKEPKNALTKQ

-continued

YQALFNLEGVDLEFRDEALDAIAKKAMARKTGARGLRSIVEAALLDTM

YDLPSMEDVEKVVIDESVIDGQSKPLLIYGKPEAQQASGE

ClpA (unfoldase from *E. coli*)

(SEQ ID NO: 23)

MLNQELELSLNMAFARAREHRHEFMTVEHLLLALLSNPSAREALEACS

VDLVALRQELEAFIEQTTPVLPASEEERDTQPTLSFQRVLQRAVFHVQ

SSGRNEVTGANVLVAIFSEQESQAAYLLRKHEVSRLDVVNFISHGTRK

DEPTQSSDPGSQPNSEEQAGGEERMENFTTNLNQLARVGGIDPLIGRE

KELERAIQVLCRRRKNNPLLVGESGVGKTAIAEGLAWRIVQGDVPEVM

ADCTIYSLDIGSLLAGTKYRGDFEKRFKALLKQLEQDTNSILFIDEIH

TIIGAGAASGGQVDAANLIKPLLSSGKIRVIGSTTYQEFSNIFEKDRA

LARRFQKIDITEPSIEETVQIINGLKPKYEAHHDVRYTAKAVRAAVEL

AVKYINDRHLPDKAIDVIDEAGARARLMPVSKRKKTVNVADIESVVAR

IARIPEKSVSQSDRDTLKNLGDRLKMLVFGQDKAIEALTEAIKMARAG

LGHEHKPVGSFLFAGPTGVGKTEVTVQLSKALGIELLRFDMSEYMERH

TVSRLIGAPPGYVGFDQGGLLTDAVIKHPHAVLLLDEIEKAHPDVFNI

LLQVMDNGTLTDNNGRKADFRNVVLVMTTNAGVRETERKSIGLIHQDN

STDAMEEIKKIFTPEFRNRLDNIIWFDHLSTDVIHQVVDKFIVELQVQ

LDQKGVSLEVSQEARNWLAEKGYDRAMGARPMARVIQDNLKKPLANEL

LFGSLVDGGQVTVALDKEKNELTYGFQSAQKHKAEAAH

ClpP (protease from *E. coli*)

(SEQ ID NO: 24)

MSYSGERDNFAPHMALVPMVIEQTSRGERSFDIYSRLLKERVIFLTGQ

VEDHMANLIVAQMLFLEAENPEKDIYLYINSPGGVITAGMSIYDTMQF

IKPDVSTICMGQAASMGAFLLTAGAKGKRFCLPNSRVMIHQPLGGYQG

QATDIEIHAREILKVKGRMNELMALHTGQSLEQIERDTERDRFLSAPE

AVEYGLVDSILTHRN

SspB (adaptor from *E. coli*)

(SEQ ID NO: 25)

MDLSQLTPRRPYLLRAFYEWLLDNQLTPHLVVDVTLPGVQVPMEYARD

GQIVLNIAPRAVGNLELANDEVRFNARFGGIPRQVSVPLAAVLAIYAR

ENGAGTMFEPEAAYDEDTSIMNDEEASADNETVMSVIDGDKPDHDDDT

HPDDEPPQPPRGGRPALRVVK

ClpS (N-end rule adaptor from *E. coli*)

(SEQ ID NO: 26)

MGKTNDWLDFDQLAEEKVRDALKPPSMYKVILVNDDYTPMEFVIDVLQ

KFFSYDVERATQLMLAVHYQGKAICGVFTAEVAETKVAMVNKYARENE

HPLLCTLEKA

ccSspB (adaptor from *C. crescentus*)

(SEQ ID NO: 27)

MSQTEPPEDLMQYEAMAQDALRGVVKAALKKAAAPGGLPEPHHLYITF

KTKAAGVSGPQDLLSKYPDEMTIVLQHQYWDLAPGETFFSVTLKFGGQ

PKRLSVPYAALTRFYDPSVQFALQFSAPEIIEDEPEPDPEPEDKANQG

ASGDEGPKIVSLDQFRKK

-continued

GBW168-aroK locus (insertion shown in lower-case font)

(SEQ ID NO: 28)

GAAGTTCTGGAAGCGTTGGCCAATGAACGCAATCCGCTGTATGAAGAG

ATTGCCGACGTGACCATTCGTACTGATGATCAAAGCGCTAAAGTGGTT

GCAAACCAGATTATTCACATGCTGGAAAGCAACgcagctaacgatgaa

aactacagcgaaaactatgctgacgctagctaatactagagctgatcc

ttcaactcagcaaaagttcgatttattcaacaaagccacgttgtgtct

caaaatctctgatgttacattgcacaagataaaaatatatcatcatga

acaataaaactgtctgcttacataaacagtaatacaaggggtgttatg

agccatattcaacgggaaacgtcttgctcccgtccgcgcttaaactcc

aacatggacgctgatttatatgggtataaatgggctcgcgataatgtc

gggcaatcaggtgcgacaatctatcgcttgtatgggaagcccgatgcg

ccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgtt

acagatgagatggtccgtctcaactggctgacggagtttatgcctacc

cgaccatcaagcattttatccgtactcctgatgatgcgtggttactca

ccaccgcgattcctgggaaaacagccttccaggtattagaagaatatc

ctgattcaggtgaaaatattgttgatgcgctggccgtgttcctgcgcc

ggttacattcgattcctgtttgtaattgtccttttaacagcgatcgtg

tatttcgtcttgctcaggcgcaatcacgcatgaataacggtttggttg

atgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaag

tctggaaagaaatgcacaagctcttgccattctcaccggattcagtcg

tcactcatggtgatttctcacttgataaccttatttttgacgagggga

aattaataggttgtattgatgttggacgggtcggaatcgcagaccgtt

accaggaccttgccattctttggaactgcctcggtgagttttctcctt

cattacagaaacggattttcaaaaatatggtattgataatcctgatat

gaataaattgcagtttcatttgatgctcgatgagtttttctaataaTT

CTGGCTTTATATACACTCGTCTGCGGGTACAGTAATTAAGGTGGATGT

CGCGTTATGGAGAGGATTGTCGTTACTCTCGGGGAACGTAGTTACCCA

AT

Xba-B0032-TACTAG-AroKfwd

(SEQ ID NO: 29)

ggccgcttctagagtcacacaggaaagtactagatggcagagaaacgc

aatatctttc

AroK-LAA-spe-pstrev

(SEQ ID NO: 30)

ggccctgcagcggccgctactagtattaagcagccagagcataatttt

catcgttagagcgttgctttccagcatgtgaataatc

pSB3C5

(SEQ ID NO: 31)

tactagtagcggccgctgcaggagtcactaagggttagttagttagat

tagcagaaagtcaaaagcctccgaccggaggcttttgactaaaacttc

ccttggggttatcattggggctcactcaaaggcggtaatcagataaaa

aaaatccttagctttcgctaaggatgatttctgctagagatggaatag

actggatggaggcggataaagttgcaggaccacttctgcgctcggccc

-continued

ttccggctggctggtttattgctgataaatctggagccggtgagcgtg

ggtctcgcggtatcattgcagcactggggccagatggtaagccctccc

gtatcgtagttatctacacgacggggagtcaggcaactatggatgaac

gaaatagacagatcgctgagataggtgcctcactgattaagcattggt

aactgtcagaccaagtttactcatatatactttagattgatttaaaac

ttcatttttaatttaaaaggatctaggtgaagatccttttgataatc

tcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcag

acccccttaataagatgatcttcttgagatcgttttggtctgcgcgtaa

tctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcga

aggttctctgagctaccaactctttgaaccgaggtaactggcttggag

gagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgc

atgacttcaagactaactcctctaaatcaattaccagtggctgctgcc

agtggtgcttttgcatgtctttccgggttggactcaagacgatagtta

ccggataaggcgcagcggtcggactgaacgggggttcgtgcatacag

tccagcttggagcgaactgcctacccggaactgagtgtcaggcgtgga

atgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaa

ggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctgg

tatctttatagtcctgtcgggtttcgccaccactgatttgagcgtcag

atttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttt

gccgcggccctctcacttccctgttaagtatcttcctggcatcttcca

ggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaa

cgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgt

atcacatattctgctgacgcaccggtgcagccttttttctcctgccac

atgaagcacttcactgacaccctcatcagtgccaacatagtaagccag

tatacactccgctagcgctgaggtctgcctcgtgaagaaggtgttgct

gactcataccaggcctgaatcgccccatcatccagccagaaagtgagg

gagccacggttgatgagagctttgttgtaggtggaccagttggtgatt

ttgaacttttgctttgccacggaacggtctgcgttgtcgggaagatgc

gtgatctgatccttcaactcagcaaaagttcgatttattcaacaaagc

cacgttgtgtctcaaaatctctgatgttacattgcacaagataaaaat

atatcatcatgaacaataaaactgtctgcttacataaacagtaataca

aggggtgtttactagaggttgatcgggcacgtaagaggttccaacttt

caccataatgaaataagatcactaccgggcgtatttttgagttatcg

agattttcaggagctaaggaagctaaaatggagaaaaaaatcacggga

tataccaccgttgatatatcccaatggcatcgtaaagaacattttgag

gcatttcagtcagttgctcaatgtacctataaccagaccgttcagctg

gatattacggccttttaaagaccgtaaagaaaaataagcacaagttt

tatccggcctttattcacattcttgcccgcctgatgaacgctcacccg

gagtttcgtatggccatgaaagacggtgagctggtgatctgggatagt

gttcacccttgttacaccgttttccatgagcaaactgaaacgttttcg

tccctctggagtgaataccacgacgatttccggcagtttctccacata

tattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccct

aaagggtttattgagaatatgtttttgtctcagccaatccctgggtg

agtttcaccagttttgatttaaacgtggccaatatggacaacttcttc

gcccccgttttcacgatgggcaaatattatacgcaaggcgacaaggtg

ctgatgccgctggcgatccaggttcatcatgccgtttgtgatggatcc

atgtcggccgcatgcttaatgaattacaacagtactgtgatgagtggc

agggcggggcgtaataatactagctccggcaaaaaaacgggcaaggtg

tcaccaccctgccctttttctttaaaaccgaaaagattacttcgcgtt

tgccacctgacgtctaagaaaaggaatattcagcaatttgcccgtgcc

gaagaaaggcccacccgtgaaggtgagccagtgagttgattgctacgt

aattagttagttagcccttagtgactcgaattcgcggccgcttctaga

g

Bba_F2620

(SEQ ID NO: 32)

tccctatcagtgatagagattgacatccctatcagtgatagagatact

gagcactactagagaaagaggagaaatactagatgaaaaacataaatg

ccgacgacacatacagaataattaataaaattaaagcttgtagaagca

ataatgatattaatcaatgcttatctgatatgactaaaatggtacatt

gtgaatattatttactcgcgatcatttatcctcattctatggttaaat

ctgatatttcaatcctagataattaccctaaaaaatggaggcaatatt

atgatgacgctaatttaataaaatatgatcctatagtagattattcta

actccaatcattcaccaattaattggaatatatttgaaaacaatgctg

taaataaaaaatctccaaatgtaattaaagaagcgaaaacatcaggtc

ttatcactgggtttagtttccctattcatacggctaacaatggcttcg

gaatgcttagttttgcacattcagaaaaagacaactatatagatagtt

tatttttacatgcgtgtatgaacataccattaattgttccttctctag

ttgataattatcgaaaaataaatatagcaaataataaatcaaacaacg

atttaaccaaaagagaaaaagaatgtttagcgtgggcatgcgaaggaa

aaagctcttgggatatttcaaaaatattaggttgcagtgagcgtactg

tcactttccatttaaccaatgcgcaaatgaaactcaatacaacaaacc

gctgccaaagtatttctaaagcaattttaacaggagcaattgattgcc

catactttaaaaattaataacactgatagtgctagtgtagatcactac

tagagccaggcatcaaataaaacgaaaggctcagtcgaaagactgggc

ctttcgttttatctgttgtttgtcggtgaacgctctctactagagtca

cactggctcaccttcgggtgggcctttctgcgtttatatactagagac

ctgtaggatcgtacaggtttacgcaagaaaatggtttgttatagtcga

ataaa

Bba_R0010

(SEQ ID NO: 33)

caatacgcaaaccgcctctccccgcgcgttggccgattcattaatgca

gctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacg

caattaatgtgagttagctcactcattaggcaccccaggctttacact

-continued

ttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaat

ttcacaca

Ribosome binding site (Bba_B0032)
(SEQ ID NO: 34)

tcacacaggaaag

aroK (open reading frame)
(SEQ ID NO: 35)

atggcagagaaacgcaatatctttctggttgggcctatgggtgccgga

aaaagcactattgggcgccagttagctcaacaactcaatatggaattt

tacgattccgatcaagagattgagaaacgaaccggagctgatgtgggc

tgggttttcgatttagaaggcgaagaaggcttccgcgatcgcgaagaa

aaggtcatcaatgagttgaccgagaaacagggtattgtgctggctact

ggcggcggctctgtgaaatCccgtgaaacgcgtaaccgtCTttccgct

cgtggcgttgtcgtttatcttgaaacgaccatcgaaaagcaacttgca

cgcacgcagcgtgataaaaaacgcccgttgctgcacgttgaaacaccg

ccgcgtgaagttctggaagcgttggccaatgaacgcaatccgctgtat

gaagagattgccgacgtgaccattcgtactgatgatcaaagcgctaaa

gtggttgcaaaccagattattcacatgctggaaagcaac

sspB (open reading frame)
(SEQ ID NO: 36)

atggatttgtcacagctaacaccacgtcgtccctatctgctgcgtgca

ttctatgagtggttgctggataaccagctcacgccgcacctggtggtg

gatgtgacgctccctggcgtgcaggttcctatggaatatgcgcgtgac

gggcaaatcgtactcaacattgcgccgcgtgctgtcggcaatctggaa

ctggcgaatgatgaggtgcgctttaacgcgcgctttggtggcattccg

cgtcaggtttctgtgccgctggctgccgtgctggctatctacgcccgt

gaaaatggcgcaggcacgatgtttgagcctgaagctgcctacgatgaa

gataccagcatcatgaatgatgaagaggcatcggcagacaacgaaacc

gttatgtcggttattgatggcgacaagccagatcacgatgatgacact

catcctgacgatgaacctccgcagccaccacgcggtggtcgaccggca

ttacgcgttgtgaagtaa

Nucleic acid sequence for AANDENYALAA
(SEQ ID NO: 37)

gcagctaacgatgaaaattatgctctggctgcttaa

Nucleic acid sequence for AANDENYALVA
(SEQ ID NO: 38)

gcagctaacgatgaaaattatgctctggttgcttaa

Nucleic acid sequence for AANDENYADAS
(SEQ ID NO: 39)

gcagctaacgatgaaaattatgctgacgctagctaa

Nucleic acid sequence for AANDENYALDD
(SEQ ID NO: 40)

gcagctaacgatgaaaattatgactggacgactaa

Representative assembled construct
F2620-B0032-AroK-LVA-pSB3C5 (circular)
(SEQ ID NO: 41)

gaattcgcggccgcttctagtccctatcagtgatagagattgacatcc

ctatcagtgatagagatactgagcactactagagaaagaggagaaata

ctagatgaaaaacataaatgccgacgacacatacagaataattaataa

-continued

aattaaagcttgtagaagcaataatgatattaatcaatgcttatctga

tatgactaaaatggtacattgtgaatattatttactcgcgatcattta

tcctcattctatggttaaatctgatatttcaatcctagataattaccc

taaaaaatggaggcaatattatgatgacgctaatttaataaaatatga

tcctatagtagattattctaactccaatcattcaccaattaattggaa

tatatttgaaaacaatgctgtaaataaaaaatctccaaatgtaattaa

agaagcgaaaacatcaggtcttatcactgggtttagtttccctattca

tacggctaacaatggcttcggaatgcttagttttgcacattcagaaaa

agacaactatatagatagtttatttttacatgcgtgtatgaacatacc

attaattgttccttctctagttgataattatcgaaaaataaatatagc

aaataataaatcaaacaacgatttaaccaaaagagaaaaagaatgttt

agcgtgggcatgcgaaggaaaaagctcttgggatatttcaaaaatatt

aggttgcagtgagcgtactgtcactttccatttaaccaatgcgcaaat

gaaactcaatacaacaaaccgctgccaaagtatttctaaagcaatttt

aacaggagcaattgattgcccatactttaaaaattaataacactgata

gtgctagtgtagatcactactagagccaggcatcaaataaaacgaaag

gctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtg

aacgctctctactagagtcacactggctcaccttcgggtgggcctttc

tgcgtttatatactagagacctgtaggatcgtacaggtttacgcaaga

aaatggtttgttatagtcgaataaatactagagtcacacaggaaagta

ctagatggcagagaaacgcaatatctttctggttgggcctatgggtgc

cggaaaaagcactattgggcgccagttagctcaacaactcaatatgga

attttacgattccgatcaagagattgagaaacgaaccggagctgatgt

gggctgggttttcgatttagaaggcgaagaaggcttccgcgatcgcga

agaaaaggtcatcaatgagttgaccgagaaacagggtattgtgctggc

tactggcggcggctctgtgaaatcccgtgaaacgcgtaaccgtctttc

cgctcgtggcgttgtcgtttatcttgaaacgaccatcgaaaagcaact

tgcacgcacgcagcgtgataaaaaacgcccgttgctgcacgttgaaac

accgccgcgtgaagttctggaagcgttggccaatgaacgcaatccgct

gtatgaagagattgccgacgtgaccattcgtactgatgatcaaagcgc

taaagtggttgcaaaccagattattcacatgctggaaagcaacgcagc

taacgatgaaaattatgactggttgcttaatactagtagcggccgctg

caggagtcactaagggttagttagttagattagcagaaagtcaaaagc

ctccgaccggaggcttttgactaaaacttcccttggggttatcattgg

ggctcactcaaaggcggtaatcagataaaaaaaatccttagctttcgc

taaggatgatttctgctagagatggaatagactggatggaggcggata

aagttgcaggaccacttctgcgctcggcccttccggctggctggttta

ttgctgataaatctggagccggtgagcgtgggtctcgcggtatcattg

cagcactggggccagatggtaagccctcccgtatcgtagttatctaca

cgacggggagtcaggcaactatggatgaacgaaatagacagatcgctg

agataggtgcctcactgattaagcattggtaactgtcagaccaagttt

-continued

actcatatatactttagattgatttaaaacttcatttttaatttaaaa

ggatctaggtgaagatccttttttgataatctcatgaccaaaatccctt

aacgtgagttttcgttccactgagcgtcagacccccttaataagatgat

cttcttgagatcgttttggtctgcgcgtaatctcttgctctgaaaacg

aaaaaaccgccttgcagggcggtttttcgaaggttctctgagctacca

actctttgaaccgaggtaactggcttggaggagcgcagtcaccaaaac

ttgtcctttcagtttagccttaaccggcgcatgacttcaagactaact

cctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgt

ctttccgggttggactcaagacgatagttaccggataaggcgcagcgg

teggactgaacgggggttcgtgcatacagtccagcttggagcgaact

gcctacccggaactgagtgtcaggcgtggaatgagacaaacgcggcca

taacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcg

cacgagggagccgccaggggaaacgcctggtatctttatagtcctgtc

gggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtca

ggggggcggagcctatggaaaaacggctttgccgcggccctctcactt

ccctgttaagtatcttcctggcatcttccaggaaatctccgccccgtt

cgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagt

cagtgagcgaggaagcggaatatatcctgtatcacatattctgctgac

gcaccggtgcagccttttttctcctgccacatgaagcacttcactgac

accctcatcagtgccaacatagtaagccagtatacactccgctagcgc

tgaggtctgcctcgtgaagaaggtgttgctgactcataccaggcctga

atcgccccatcatccagccagaaagtgagggagccacggttgatgaga

gctttgttgtaggtggaccagttggtgattttgaacttttgctttgcc

acggaacggtctgcgttgtcgggaagatgcgtgatctgatccttcaac

tcagcaaaagttcgatttattcaacaaagccacgttgtgtctcaaaat

ctctgatgttacattgcacaagataaaaatatatcatcatgaacaata

aaactgtctgcttacataaacagtaatacaaggggtgtttactagagg

ttgatcgggcacgtaagaggttccaactttcaccataatgaaataaga

tcactaccgggcgtatttttgagttatcgagattttcaggagctaagg

aagctaaaatggagaaaaaaatcacgggatataccaccgttgatatat

cccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctc

aatgtacctataaccagaccgttcagctggatattacggccttttaa

agaccgtaaagaaaaataagcacaagttttatccggcctttattcaca

ttcttgcccgcctgatgaacgctcacccggagtttcgtatggccatga

aagacggtgagctggtgatctgggatagtgttcacccttgttacaccg

ttttccatgagcaaactgaaacgttttcgtccctctggagtgaatacc

acgacgatttccggcagtttctccacatatattcgcaagatgtggcgt

gttacggtgaaaacctggcctatttccctaaagggtttattgagaata

tgtttttttgtctcagccaatccctgggtgagtttcaccagttttgatt

taaacgtggccaatatggacaacttatcgcccccgttttcacgatggg

caaatattatacgcaaggcgacaaggtgctgatgccgctggcgatcca

ggttcatcatgccgtttgtgatggcttccatgtcggccgcatgcttaa

tgaattacaacagtactgtgatgagtggcagggcggggcgtaataata

ctagctccggcaaaaaaacgggcaaggtgtcaccaccctgccctttt

ctttaaaaccgaaaagattacttcgcgtttgccacctgacgtctaaga

aaaggaatattcagcaatttgcccgtgccgaagaaaggcccacccgtg

aaggtgagccagtgagttgattgctacgtaattagttagttagcccct

agtgactc

Branched amino acid production (*C. glutamicum*)
cg-ilvE-AANDENYALVA

(SEQ ID NO: 42)

atgacgtcattagagttcacagtaacccgtaccgaaaatccgacgtca

cccgatcgtctgaaggaaattcttgccgcaccgaagttcggtaagttc

ttcaccgaccacatggtgaccattgactggaacgagtcggaaggctgg

cacaacgcccaattagtgccatacgcgccgattcctatggatcctgcc

accaccgtattccactacggacaggcaatttttgagggaattaaggcc

taccgccattcggacgaaaccatcaagactttccgtcctgatgaaaac

gccgagcgtatgcagcgttcagcagctcgaatggcaatgccacagttg

ccaaccgaggactttattaaagcacttgaactgctggtagacgcggat

caggattgggttcctgagtacggcggagaagcttccctctacctgcgc

ccattcatgatctccaccgaaattggcttgggtgtcagcccagctgat

gcctacaagttcctggtcatcgcatccccagtcggcgcttacttcacc

ggtggaatcaagcctgtttccgtctggctgagcgaagattacgtccgc

gctgcacccggcggaactggtgacgccaaatttgctggcaactacgcg

gcttctttgcttgcccagtcccaggctgcggaaaagggctgtgaccag

gtcgtatggttggatgccatcgagcacaagtacatcgaagaaatgggt

ggcatgaacttgggttcatctaccgcaacggcgaccaagtcaagctag

tcacccctgaactttccggctcactacttccaggcatcacccgcaagt

cacttctacaagtagcacgcgacttgggatacgaagtagaagagcgaa

agatcaccaccaccgagtgggaagaagacgcaaagtctggcgccatga

ccgaggcatttgcttgcggtactgcagctgttatcacccctgttggca

ccgtgaaatcagctcacggcaccttcgaagtgaacaacaatgaagtcg

gagaaatcacgatgaagcttcgtgaaaccctcaccggaattcagcaag

gaaacgttgaagaccaaaacggatggctttacccactggttggcGCAG

CTAACGATGAAAATTATGCTCTGGTGGCTtaa

Branched amino acid production (*C. glutamicum*)
cg-panB-AANDENYALGG

(SEQ ID NO: 43)

atgtcaggcattgatgcaaagaaaatccgcacccgtcatttccgcgaa

gctaaagtaaacggccagaaagtttcggttctcaccagctatgatgcg

ctttcggcgcgcatttttgatgaggctggcgtcgatatgctccttgtt

ggtgattccgctgccaacgttgtgctgggtcgcgataccaccttgtcg

atcaccttggatgagatgattgtgctggccaaggcggtgacgatcgct

acgaagcgtgcgcttgtggtggttgatctgccgtttggtacctatgag

-continued

gtgagcccaaatcaggcggtggagtccgcgatccgggtcatgcgtgaa

acgggtgcggctgcggtgaagatcgagggtggcgtggagatcgcgcag

acgattcgacgcattgttgatgctggaattccggttgtcggccacatc

gggtacaccccgcagtccgagcattccttgggcggccacgtggttcag

ggtcgtggcgcgagttctggaaagctcatcgccgatgcccgcgcgttg

gagcaggcgggtgcgtttgcggttgtgttggagatggttccagcagag

gcagcgcgcgaggttaccgaggatctttccatcaccactatcggaatc

ggtgccggcaatggcacagatgggcaggttttggtgtggcaggatgcc

ttcggcctcaaccgcggcaagaagccacgcttcgtccgcgagtacgcc

accttgggcgattccttgcacgacgccgcgcaggcctacatcgccgat

atccacgcgggtaccttcccaggcgaagcggagtcctttGCAGCTAAC

GATGAAAATTATGCTCTGGGCGGCtaa

Branched amino acid production (*C. glutamicum*)
cg-leuA-AANDENYALAG

(SEQ ID NO: 44)

atgcttcaccacatgacttcgcgtgcgaatctacttcttcttcgccgc

ggcgggtcccagaggtctatgtctcctaacgatgcattcatctccgca

cctgccaagatcgaaaccccagttgggcctcgcaacgaaggccagcca

gcatggaataagcagcgtggctcctcaatgccagttaaccgctacatg

cctttcgaggttgaggtagaagatatttctctgccggaccgcacttgg

ccagataaaaaaatcaccgttgcacctcagtggtgtgctgttgacctg

cgtgacggcaaccaggctctgattgatccgatgtctcctgagcgtaag

cgccgcatgtttgagctgctggttcagatgggcttcaaagaaatcgag

gtcggtttcccttcagcttcccagactgattttgatttcgttcgtgag

atcatcgaaaagggcatgatccctgacgatgtcaccattcaggttctg

gttcaggctcgtgagcacctgattcgccgtacttttgaagcttgcgaa

ggcgcaaaaaacgttatcgtgcacttctacaactccacctccatcctg

cagcgcaacgtggtgttccgcatggacaaggtgcaggtgaagaagctg

gctaccgatgccgctgaactaatcaagaccatcgctcaggattaccca

gacaccaactggcgctggcagtactcccctgagtccttcaccggcact

gaggttgagtacgccaaggaagttgtggacgcagttgttgaggtcatg

gatccaactcctgagaaccaatgatcatcaacctgccttccaccgttg

agatgatcacccctaacgtttacgcagactccattgaatggatgcacc

gcaatctaaaccgtcgtgattccattatcctgtccctgcacccgcaca

atgaccgtggcaccggcgttggcgcagctgagctgggctacatggctg

gcgctgaccgcatcgaaggctgcctgttcggcaacggcgagcgcaccg

gcaacgtctgcctggtcaccctggcactgaacatgctgacccagggcg

ttgaccctcagctggacttcaccgatatacgccagatccgcagcaccg

ttgaatactgcaaccagctgcgcgttcctgagcgccacccatacggcg

gtgacctggtcttcaccgctttctccggttcccaccaggacgctgtga

acaagggtctggacgccatggctgccaaggttcagccaggtgctagct

ccactgaagtttcttgggagcagctgcgcgacaccgaatgggaggttc

-continued

cttacctgcctatgatccaaaggatgtcggtcgcgactacgaggctgt

tatccgcgtgaactcccagtccggcaagggcggcgttgcttacatcat

gaagaccgatcacggtctgcagatccctcgctccatgcaggttgagtt

ctccaccgttgtccagaacgtcaccgacgctgagggcggcgaggtcaa

ctccaaggcaatgtgggatatcttcgccaccgagtacctggagcgcac

cgcaccagttgagcagatcgcgctgcgcgtcgagaacgctcagaccga

aaacgaggatgcatccatcaccgccgagctcatccacaacggcaagga

cgtcaccgtcgatggccgcggcaacggcccactggccgcttacgccaa

cgcgctggagaagctgggcatcgacgttgagatccaggaatacaacca

gcacgcccgcacctcgggcgacgatgcagaagcagccgcctacgtgct

ggctgaggtcaacggccgcaaggtctggggcgtcggcatcgctggctc

catcacctacgcttcgctgaaggcagtgacctccgccgtaaaccgcgc

gctggacgtcaaccacgaggcagtcctggctggcggcgttGCAGCTAA

CGATGAAAATTATGCTCTGGCTGGCtaa

Branched amino acid production (*C. glutamicum*)
cg-p-ilvE-AANDENYALVA

(SEQ ID NO: 45)

mtsleftvtrtenptspdrlkeilaapkfgkfftdhmvtidwnesegw

hnaqlvpyapipmdpattvfhygqaifegikayrhsdetiktfrpden

aermqrsaarmampqlptedfikalellvdadqdwypeyggeaslylr

pfmisteiglgvspadaykflviaspvgayftggikpvsvwlsedyvr

aapggtgdakfagnyaasllaqsqaaekgcdqvvvvldaiehkyieem

ggmnlgfiyrngdqvklvtpelsgsllpgitrksllqvardlgyevee

rkitteweedaksgamteafacgtaavitpvgtvksahgtfevnnne

vgeitmklretltgiqqgnvedqngwlyplvgAANDENYALVA

Branched amino acid production (*C. glutamicum*)
cg-p-panB-AANDENYALGG

(SEQ ID NO: 46)

msgidakkirtrhfreakvngqkvsvltsydalsarifdcagvdmllv

gdsaanvvlgrdttlsitldemivlakavtiatkralvvvdlpfgtye

vspnqavesairvmretgaaavkieggveiaqtirrivdagipvvghi

gytpqsehslgghvvqgrgassgkliadaraleqagafavvlemvpae

aarevtedlsittigigagngtdgqvlvwqdafglnrgkkprfvreya

tlgdslhdaaqayiadihagtfpgeaesfAANDENYALGG

Branched amino acid production (*C. glutamicum*)
cg-p-leuA-AANDENYALAG

(SEQ ID NO: 47)

mlhhmtsranllllrrggsqrsmspndafisapakietpvgprnegqp

awnkqrgssmpvnrympfevevedislpdrtwpdkkitvapqwcavdl

rdgnqalidpmsperkrrmfellvqmgfkeievgfpsasqtdfdfvre

iiekgmipddvtiqvlvqarehlirrtfeacegaknvivhfynstsil

qrnvvfrmdkvqvkklatdaaeliktiaqdypdtnwrwqyspesftgt

eveyakevvdavvevmdptpenpmiinlpstvemitpnvyadsiewmh

rnlnrrdsiilslhphndrgtgvgaaelgymagadriegclfgngert

gnvclvtlalnmltqgvdpqldftdirqirstveycnqlrvperhpyg

-continued

gdlvftafsgshqdavnkgldamaakvqpgasstevsweqlrdtewev

pylpidpkdvgrdyeavirvnsqsgkggvayimktdhglqiprsmqve

fstvvqnvtdaeggevnskamwdifateylertapvegialrvenaqt

enedasitaelihngkdvtvdgrgngplaayanaleklgidveiqeyn

qhartsgddaeaaayvlaevngrkvwgvgiagsityaslkavtsavnr

aldvnheavlaggvAANDENYALAG

Branched amino acid production (*E. coli*)
ec-ilvE-AANDENYALVA

(SEQ ID NO: 48)

atgaccacgaagaaagctgattacatttggttcaatggggagatggtt

cgctgggaagacgcgaaggtgcatgtgatgtcgcacgcgctgcactat

ggcacttcggtttttgaaggcatccgttgctacgactcgcacaaagga

ccggttgtattccgccatcgtgagcatatgcagcgtctgcatgactcc

gccaaaatctatcgcttcccggtttcgcagagcattgatgagctgatg

gaagcttgtcgtgacgtgatccgcaaaaacaatctcaccagcgcctat

atccgtccgctgatcttcgtcggtgatgttggcatgggagtaaacccg

ccagcgggatactcaaccgacgtgattatcgctgctttcccgtgggga

gcgtatctgggcgcagaagcgctggagcaggggatcgatgcgatggtt

tcctcctggaaccgcgcagcaccaaacaccatcccgacggcggcaaaa

gccggtggtaactacctctcttccctgctggtgggtagcgaagcgcgc

cgccacggttatcaggaaggtatcgcgctggatgtgaacggttatatc

tctgaaggcgcaggcgaaaacctgtttgaagtgaaagatggtgtgctg

ttcaccccaccgttcacctcctccgcgctgccgggtattacccgtgat

gccatcatcaaactggcgaaagagctgggaattgaagtacgtgagcag

gtgctgtcgcgcgaatccctgtacctggcggatgaagtgtttatgtcc

ggtacggcggcagaaatcacgccagtgcgcagcgtagacggtattcag

gttggcgaaggccgttgtggcccggttaccaaacgcattcagcaagcc

ttcttcggcctcttcactggcgaaaccgaagataaatggggctggtta

gatcaagttaatcaaGCAGCTAACGATGAAAATTATGCTCTGGTGGCT

taa

Branched amino acid production (*E. coli*)
ec-panB-AANDENYALGG

(SEQ ID NO: 49)

atgaaaccgaccaccatctccttactgcagaagtacaaacaggaaaaa

aaacgtttcgcgaccatcaccgcttatgactatagcttcgccaaactc

tttgctgatgaagggcttaacgtcatgctggtgggcgattcgctgggc

atgacggttcaggggcacgactccaccctgccagttaccgttgccgat

atcgcctaccacactgccgccgtacgtcgcggcgcaccaaactgcctg

ctgctggctgacctgccgtttatggcgtatgccacgccggaacaagcc

ttcgaaaacgccgcaacggttatgcgtgccggtgctaacatggtcaaa

attgaaggcggtgagtggctggtagaaaccgtacaaatgctgaccgaa

cgtgccgttcctgtatgtggtcacttaggtttaacaccacagtcagtg

aatattttcggtggctacaaagttcaggggcgcggcgatgaagcgggc

-continued

gatcaactgctcagcgatgcattagccttagaagagctggggcacagc

tgctggtgctggaatgcgtgccggttgaactggcaaaacgtattaccg

aagcactggcgatcccggttattggcattggcgcaggcaacgtcactg

acgggcagatcctcgtgatgcacgacgcctttggtattaccggcggtc

acattcctaaattcgctaaaaatttcctcgccgaaacgggcgacatcc

gcgcggctgtgcggcagtatatggctgaagtggagtccggcgtttatc

cgggcgaagaacacagtttccatGCAGCTAACGATGAAAATTATGCTC

TGGGCGGCtaa

Branched amino acid production (*E. coli*)
ec-leuA-AANDENYALAG

(SEQ ID NO: 50)

atgagccagcaagtcattattttcgataccacattgcgcgacggtgaa

caggcgttacaggcaagcttgagtgtgaaagaaaaactgcaaattgcg

ctggcccttgagcgtatgggtgttgacgtgatggaagtcggtttcccc

gtctcttcgccgggcgattttgaatcggtgcaaaccatcgcccgccag

gttaaaaacagccgcgtatgtgcgttagctcgctgcgtggaaaaagat

atcgacgtggcggccgaatccctgaaagtcgccgaagccttccgtatt

catacctttattgccacttcgccaatgcacatcgccaccaagctgcgc

agcacgctggacgaggtgatcgaacgcgctatctatatggtgaaacgc

gcccgtaattacaccgatgatgttgaattttcttgcgaagatgccggg

cgtacacccattgccgatctggcgcgagtggtcgaagcggcgattaat

gccggtgccaccaccatcaacattccggacaccgtgggctacaccatg

ccgtttgagttcgccggaatcatcagcggcctgtatgaacgcgtgcct

aacatcgacaaagccattataccgtacatacccacgacgatttgggcc

tggcggtcggaaactcactggcggcggtacatgccggtgcacgccagg

tggaaggcgcaatgaacgggatcggcgagcgtgccggaaactgttccc

tggaagaagtcatcatggcgatcaaagttcgtaaggatattctcaacg

tccacaccgccattaatcaccaggagatatggcgcaccagccagttag

ttagccagatttgtaatatgccgatcccggcaaacaaagccattgttg

gcagcggcgcattcgcacactcctccggtatacaccaggatggcgtgc

tgaaaaaccgcgaaaactacgaaatcatgacaccagaatctattggtc

tgaaccaaatccagctgaatctgacctctcgttcggggcgtgcggcgg

tgaaacatcgcatggatgagatggggtataaagaaagtgaatataatt

tagacaatttgtacgatgcttcctgaagctggcggacaaaaaaggtca

ggtgtttgattacgatctggaggcgctggccttcatcggtaagcagca

agaagagccggagcatttccgtctggattacttcagcgtgcagtctgg

ctctaacgatatcgccaccgccgccgtcaaactggcctgtggcgaaga

agtcaaagcagaagccgccaacggtaacggtccggtcgatgccgtcta

tcaggcaattaaccgcatcactgaatataacgtcgaactggtgaaata

cagcctgaccgccaaaggccacggtaaagatgcgctgggtcaggtgga

tatcgtcgctaactacaacggtcgccgcttccacggcgtcggcctggc

taccgatattgtcgagtcatctgccaaagccatggtgcacgttctgaa

-continued

caatatctggcgtgccgcagaagtcgaaaaagagttgcaacgcaaagc

tcaacacaacgaaaacaacaaggaaaccgtgGCAGCTAACGATGAAAA

TTATGCTCTGGCTGGCtga

Branched amino acid production (*E. coli*)
ec-p-ilvE-AANDENYALVA

(SEQ ID NO: 51)

mttkkadyiwfngemvrwedakvhvmshalhygtsvfegircydshkg

pvvfrhrehmqrlhdsakiyrfpvsqsidelmeacrdvirknnltsay

irplifvgdygmgvnppagystdviiaafpwgaylgaealeqgidamv

sswnraapntiptaakaggnylssllvgsearrhgyqegialdvngyi

segagenlfevkdgvlftppftssalpgitrdaiiklakelgievreq

vlsreslyladevfmsgtaaeitpvrsvdgiqvgegrcgpvtkriqqa

ffglftgetedkwgwldqvnqAANDENYALVA

Branched amino acid production (*E. coli*)
ec-p-panB-AANDENYALGG

(SEQ ID NO: 52)

mkpttisllqkykqekkrfatitaydysfaklfadeglnvmlvgdslg

mtvqghdstlpvtvadiayhtaavrrgapnclllladlpfmayatpeqa

fenaatvmraganrnvkieggewlvetvqmlteravpvcghlgltpqs

vnifggykvqgrgdeagdqllsdalaleaagaqllvlecvpvelakri

tealaipvigigagnvtdgqilvmhdafgitgghipkfaknflaetgd

iraavrqymaevesgvypgeehsfhAANDENYALGG

Branched amino acid production (*E. coli*)
ec-p-leuA-AANDENYALAG

(SEQ ID NO: 53)

msqqviifdttlidgeqalqaslsvkeklqialalermgvdvmevgfp

vsspgdfesvqtiarqvknsrvcalarcvekdidvaaeslkvaeafri

htfiatspmhiatklrstldevieraiymvkrarnytddvefscedag

rtpiadlarvveaainagattinipdtvgytmpfefagiisglyervp

nidkaiisvhthddlglavgnslaavhagarqvegamngigeragncs

leevimaikvrkdilnyhtainhqeiwrtsqlvsqicnmpipankaiv

gsgafahssgihqdgvlknrenyeimtpesiglnqiqlnltsrsgraa

vkhrmdemgykeseynldnlydaflkladkkgqvfdydlealafigkq

qeepehfrldyfsvqsgsndiataavklacgeevkaeaangngpvdav

yqainriteynvelvkysltakghgkdalgqvdivanyngrrfhgvgl

atdivessakamvhvlnniwraaevekelqrkaqhnennketvAANDE

NYALAG

EXAMPLES

The present invention is illustrated by the following examples, which are in no way intended to be limiting of the invention.

Example 1. Synthesis of Shikimic Acid from a Microbe Containing an Engineered Shikimate Kinase Gene

An *E. coli* strain capable of being grown in the absence of aromatic amino acids and producing shikimic acid was engineered as follows. The strain was engineered to express a shikimate kinase isoform, the product of the aroK gene, from a plasmid, while the chromosomal genes encoding shikimate kinase were non-functional. The plasmid-borne shikimate kinase isoform was engineered to have a degradation tag at its C-terminus. In this case and throughout the invention, it was and is useful to inspect the three-dimensional structure of a protein to verify that a chosen terminus is compatible with addition of a degradation tag. The solved structure of the aroK product, PDB file 1KAG, was inspected and the steric availability of the C-terminus was verified.

Plasmid vectors were generated which allow for conditional expression of *E. coli* shikimate kinase I, aroK. Using standard plasmid construction techniques, the coding sequence for aroK was fused to each of the four degradation tags, AANDENYALAA (SEQ ID NO: 1), AANDENYALVA (SEQ ID NO: 8), AANDENYADAS (SEQ ID NO: 2), and AANDENYALDD (SEQ ID NO: 13). This fusion construct was inserted downstream of either the IPTG-inducible lac promoter (SEQ ID NO: 33) or the HSL-inducible LuxR-derived promoter, F2620 (SEQ ID NO: 32). Each construct contained the ribosome binding site (SEQ ID NO: 34) and resided on the plasmid backbone, pSB3C5 (SEQ ID NO: 31), a chloramphenicol-resistant low-copy plasmid bearing a p15a origin of replication. Nucleotide sequences for each component are listed below, as well as a sample assembled sequence for the construct F2620-B0032-AroK-LVA (SEQ ID NO: 41) as present in pSB3C5.

The complete cloning process for the generation of plasmid F2620-B0032-AroK-LAA (pSB3C5) is described here and the general principles were applied to the generation of the other plasmids. The open reading frame of aroK was PCR amplified from *E. coli* DH5α chromosomal DNA using primers Xba-B0032-TACTAG-AroKfwd (SEQ ID NO: 29) and AroK-LAA-spe-pstrev (SEQ ID NO: 30) resulting in product PCR1-LAA. F2620 (SEQ ID NO: 32) was generated by PCR resulting in product PCR2-F2620. PCR1-LAA was then incubated with restriction enzymes XbaI and PstI in NEB Buffer #2 supplemented with BSA for 2 hours at 37° C.; PCR2-F2620 was incubated with restriction enzymes EcoRI and SpeI under identical conditions. Successful PCR amplification and restriction digestion was analyzed by gel electrophoresis. After removing heat-denatured restriction enzymes using a Qiagen PCR purification kit, digested PCR1-LAA and PCR2-F2620 were mixed in a stoichiometric ratio with plasmid backbone pSB3C5 which had been treated with EcoRI and PstI. The 3-component mixture was incubated with T4 DNA ligase for 2 hours at room temperature. Chemically competent *E. coli* NEB 10β cells were then transformed with this ligation product and plated on LB/chloramphenicol. Individual colonies were picked and grown in liquid culture overnight.

Strains of *E. coli* termed GBW181, GBW182, and GBW183 were engineered as follows. The relevant features were that GBW181, GWB182, and GWB183 contained a version of aroK with a C-terminal "AANDENYADAS" (SEQ ID NO: 2), "AANDENYALVA" (SEQ ID NO: 8), and "AANDENYALDD" (SEQ ID NO: 13), variants of the AANDENYALAA (SEQ ID NO: 1) degradation tag (see table above). Of these, the AANDENYALVA (SEQ ID NO: 8) tag triggered the greatest degradation, while the AANDENYALDD (SEQ ID NO: 13) did not cause degradation and served as a negative control.

In these constructions, the aroK-tag genes were regulated by a strong promoter that was induced by homoserine lactone. Specifically, the aroK gene was expressed from the element F2620 (SEQ ID NO: 32), which encodes a luxR

transcriptional regulatory protein that is activated by homoserine lactone (HSL), a LuxR-regulated promoter directing transcription of the *E. coli* aroK gene fused to a DNA segment encoding AANDENYALVA (SEQ ID NO: 8), and a p15a origin of replication. The chromosomal copies of aroK and aroL were mutated by conventional procedures.

In the following experiments, cells were grown in M9 medium that included 0.4% glucose, 1 μg/ml thiamin, and "tryptophan dropout medium" (Sigma-Aldrich, St. Louis, Mo.), which contains most amino acids but lacks the expensive amino acid tryptophan. This assay system had the advantage that cells would grow more quickly than in a minimal medium without amino acids, while faithfully representing the behavior of cells grown in a minimal medium supplemented only with a carbohydrate source.

The relative degradation-promoting activities of the three different tags were confirmed in a preliminary experiment. Strains 181 and 183 were found to grow in selective medium in the absence of the inducer HSL, while strain 182 only grew in the presence of about 10 nM HSL. These results indicated that low-level expression of the non-induced promoter produced sufficient aroK protein in strains 181 and 183 for tryptophan production, while the aroK protein from strain 182 was too rapidly degraded to allow sufficient tryptophan synthesis for growth.

Cells were inoculated from a single colony and grown with aeration at 37° C. for about 16 hours with 10 nM homoserine lactone to induce the aroK-AANDENYALVA (SEQ ID NO: 8) protein. The culture reached an OD_{700} of about 0.5. At this point, the culture was spun down, resuspended in twice the prior volume, washed in M9 medium without additions, and split into cultures with 10 nM homoserine lactone or with no homoserine lactone, in M9 medium, glucose, thiamin, and tryptophan dropout medium. After about 4 hours, the cultures were spun down and the supernatants were filter-sterilized.

The supernatants were tested for levels of shikimic acid by a bioassay as follows, based on the ability of shikimic acid to support growth of an aroE mutant of *E. coli*. Each supernatant was diluted 2-fold into fresh medium containing about 10^4 of an aroK mutant strain of *E. coli*, JW3242-1 (Coli Genetic Stock Center, New Haven, Conn.). In addition, serial dilutions of shikimic acid were added to similar cultures. The cultures were grown for 24 hours and optical densities compared. Based on this analysis, the shikimic acid level in the culture lacking homoserine lactone was about 10 μg/ml. The culture with 10 nM homoserine lactone produced no detectable shikimic acid.

These results indicated that shikimic acid can be produced from a culture grown in the absence of an aromatic amino acid.

Production of shikimic acid was also observed in a culture of strain 182 grown in the absence of amino acid supplements. A culture is grown in the presence of homoserine lactone in, for example, M9 medium containing glucose, sucrose, glycerol, molasses, or treated cellulosic biomass, is grown to a late logarithmic stage, the homoserine lactone is removed, and shikimic acid is produced by the cells as the aroK product is degraded and not replaced. The resulting shikimic acid is purified from the supernatant. To further improve shikimic acid yields, strain 182 is engineered to express the glf gene from *Zymomonas mobilis*.

Example 2. Production of Shikimic Acid from a Microbial Strain in Which Shikimate Kinase is Fused to a Degradation Tag and Expressed from an Episome with Conditional Replication

In an alternative method of the invention, an *E. coli* strain that could be grown in the absence of aromatic amino acids and produce shikimic acid was engineered as follows. Four variants were constructed from a plasmid derivative of the low-copy vector pSC101, in which the origin of the plasmid was temperature-sensitive for replication. The plasmid encoded the *E. coli* aroK gene expressed from its endogenous promoter. The four plasmid variant coding sequences for the degradation tags AANDENYALAA (SEQ ID NO: 1), AANDENYALVA (SEQ ID NO: 8), AANDENYADAS (SEQ ID NO: 2) and the non-degrading control variant AANDENYALDD (SEQ ID NO: 13) were fused to the 3' end of the aroK coding sequence. These vectors also encoded a chloramphenicol-resistance marker. Expression of shikimate kinase from the *E. coli* chromosome was defective.

The four strains were inoculated into the M9 glucose thiamin tryptophan-dropout medium described in Example 1 and incubated with aeration at 30° C. for 16 hours. The strains encoding shikimate kinase with the AANDENYADAS (SEQ ID NO: 2) and AANDENYALDD (SEQ ID NO: 13) tags reached near-saturation while the strains encoding shikimate kinase with the AANDENYALAA (SEQ ID NO: 1) and AANDENYALVA (SEQ ID NO: 8) tags showed no detectable growth. The strain encoding the shikimate kinase-AANDENYADAS (SEQ ID NO: 2) fusion protein was pelleted in a centrifuge and resuspended in fresh medium for a net 2-fold dilution, and then incubated at 37° C. for about 5.5 hours with aeration. The cells were pelleted in a centrifuge, and the supernatant was withdrawn, filter-sterilized, and tested for shikimic acid levels in the bioassay essentially as described in Example 1. Based on the results of this bioassay, the shikimic acid in the filter-sterilized supernatant of the culture was about 0.05 micrograms/ml.

Without wishing to be bound by theory, shikimic acid was produced by the following mechanism. When the culture bearing plasmid with the shikimate kinase-AANDENYADAS (SEQ ID NO: 2) expression construction and the temperature-sensitive origin of replication was transferred to 37° C., replication of the plasmid largely or completely stopped, and the plasmid was lost from many cells during cell division. Once the plasmid was lost from a given cell, the remaining shikimate kinase-AANDENYADAS (SEQ ID NO: 2) protein was degraded and not replaced, leaving the cell without shikimate kinase enzyme activity. Such cells produced shikimic acid and secreted this molecule into the medium.

OTHER EMBODIMENTS

From the foregoing description, it is apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

All publications, patent applications, and patents mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication, patent application, or patent was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 53

<210> SEQ ID NO 1
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 1

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Ala Ala
1 5 10

<210> SEQ ID NO 2
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 2

Ala Ala Asn Asp Glu Asn Tyr Ala Asp Ala Ser
1 5 10

<210> SEQ ID NO 3
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Caulobacter crescentus

<400> SEQUENCE: 3

Ala Ala Asn Asp Asn Phe Ala Glu Glu Phe Ala Val Ala Ala
1 5 10

<210> SEQ ID NO 4
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 4

Ala Ala Asn Asp Glu Asn Tyr Ala Asp Ala Ala
1 5 10

<210> SEQ ID NO 5
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 5

Ala Ala Asn Asp Glu Asn Tyr Ala Ala Ala Ala
1 5 10

<210> SEQ ID NO 6
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 6

Ala Ala Asn Asp Glu Asn Tyr Ala Val Ala Ala
1 5 10

<210> SEQ ID NO 7
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 7

-continued

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Asp Ala
1 5 10

<210> SEQ ID NO 8
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 8

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Val Ala
1 5 10

<210> SEQ ID NO 9
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 9

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Ala Gly
1 5 10

<210> SEQ ID NO 10
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 10

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Gly Gly
1 5 10

<210> SEQ ID NO 11
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Caulobacter crescentus

<400> SEQUENCE: 11

Ala Asp Asn Asp Asn Phe Ala Glu Glu Phe Ala Asp Ala Ser
1 5 10

<210> SEQ ID NO 12
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 12

Met Thr Asn Thr Ala Lys Ile Leu Asn Phe Gly Arg Ala Ser
1 5 10

<210> SEQ ID NO 13
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 13

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Asp Asp
1 5 10

<210> SEQ ID NO 14
<211> LENGTH: 184
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 14

-continued

```
Met Ala Glu Lys Arg Asn Ile Phe Leu Val Gly Pro Met Gly Ala Gly
1               5                   10                  15

Lys Ser Thr Ile Gly Arg Gln Leu Ala Gln Gln Leu Asn Met Glu Phe
            20                  25                  30

Tyr Asp Ser Asp Gln Glu Ile Glu Lys Arg Thr Gly Ala Asp Val Gly
        35                  40                  45

Trp Val Phe Asp Leu Glu Gly Glu Glu Gly Phe Arg Asp Arg Glu Glu
    50                  55                  60

Lys Val Ile Asn Glu Leu Thr Glu Lys Gln Gly Ile Val Leu Ala Thr
65                  70                  75                  80

Gly Gly Gly Ser Val Lys Ser Arg Glu Thr Arg Asn Arg Leu Ser Ala
                85                  90                  95

Arg Gly Val Val Val Tyr Leu Glu Thr Thr Ile Glu Lys Gln Leu Ala
            100                 105                 110

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

Pro Arg Glu Val Leu Glu Ala Leu Ala Asn Glu Arg Asn Pro Leu Tyr
    130                 135                 140

Glu Glu Ile Ala Asp Val Thr Ile Arg Thr Asp Asp Gln Ser Ala Lys
145                 150                 155                 160

Val Val Ala Asn Gln Ile Ile His Met Leu Glu Ser Asn Ala Ala Asn
                165                 170                 175

Asp Glu Asn Tyr Ala Leu Ala Ala
            180
```

<210> SEQ ID NO 15
<211> LENGTH: 189
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 15

```
Met Ala Glu Lys Arg Asn Ile Phe Leu Val Gly Pro Met Gly Ala Gly
1               5                   10                  15

Lys Ser Thr Ile Gly Arg Gln Leu Ala Gln Gln Leu Asn Met Glu Phe
            20                  25                  30

Tyr Asp Ser Asp Gln Glu Ile Glu Lys Arg Thr Gly Ala Asp Val Gly
        35                  40                  45

Trp Val Phe Asp Leu Glu Gly Glu Glu Gly Phe Arg Asp Arg Glu Glu
    50                  55                  60

Lys Val Ile Asn Glu Leu Thr Glu Lys Gln Gly Ile Val Leu Ala Thr
65                  70                  75                  80

Gly Gly Gly Ser Val Lys Ser Arg Glu Thr Arg Asn Arg Leu Ser Ala
                85                  90                  95

Arg Gly Val Val Val Tyr Leu Glu Thr Thr Ile Glu Lys Gln Leu Ala
            100                 105                 110

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

Pro Arg Glu Val Leu Glu Ala Leu Ala Asn Glu Arg Asn Pro Leu Tyr
    130                 135                 140

Glu Glu Ile Ala Asp Val Thr Ile Arg Thr Asp Asp Gln Ser Ala Lys
145                 150                 155                 160

Val Val Ala Asn Gln Ile Ile His Met Leu Glu Ser Asn Gly Gly Ser
```

-continued

```
                165                 170                 175

Gly Gly Ala Ala Asn Asp Glu Asn Tyr Ala Leu Ala Ala
            180                 185


<210> SEQ ID NO 16
<211> LENGTH: 187
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 16

Met Thr Asn Thr Ala Lys Ile Leu Asn Phe Gly Arg Ala Ser Met Ala
1               5                   10                  15

Glu Lys Arg Asn Ile Phe Leu Val Gly Pro Met Gly Ala Gly Lys Ser
            20                  25                  30

Thr Ile Gly Arg Gln Leu Ala Gln Gln Leu Asn Met Glu Phe Tyr Asp
        35                  40                  45

Ser Asp Gln Glu Ile Glu Lys Arg Thr Gly Ala Asp Val Gly Trp Val
    50                  55                  60

Phe Asp Leu Glu Gly Glu Glu Gly Phe Arg Asp Arg Glu Glu Lys Val
65                  70                  75                  80

Ile Asn Glu Leu Thr Glu Lys Gln Gly Ile Val Leu Ala Thr Gly Gly
                85                  90                  95

Gly Ser Val Lys Ser Arg Glu Thr Arg Asn Arg Leu Ser Ala Arg Gly
            100                 105                 110

Val Val Val Tyr Leu Glu Thr Thr Ile Glu Lys Gln Leu Ala Arg Thr
        115                 120                 125

Gln Arg Asp Lys Lys Arg Pro Leu Leu His Val Glu Thr Pro Pro Arg
    130                 135                 140

Glu Val Leu Glu Ala Leu Ala Asn Glu Arg Asn Pro Leu Tyr Glu Glu
145                 150                 155                 160

Ile Ala Asp Val Thr Ile Arg Thr Asp Asp Gln Ser Ala Lys Val Val
                165                 170                 175

Ala Asn Gln Ile Ile His Met Leu Glu Ser Asn
            180                 185


<210> SEQ ID NO 17
<211> LENGTH: 192
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide

<400> SEQUENCE: 17

Met Thr Asn Thr Ala Lys Ile Leu Asn Phe Gly Arg Ala Ser Gly Gly
1               5                   10                  15

Ser Gly Gly Met Ala Glu Lys Arg Asn Ile Phe Leu Val Gly Pro Met
            20                  25                  30

Gly Ala Gly Lys Ser Thr Ile Gly Arg Gln Leu Ala Gln Gln Leu Asn
        35                  40                  45

Met Glu Phe Tyr Asp Ser Asp Gln Glu Ile Glu Lys Arg Thr Gly Ala
    50                  55                  60

Asp Val Gly Trp Val Phe Asp Leu Glu Gly Glu Glu Gly Phe Arg Asp
65                  70                  75                  80

Arg Glu Glu Lys Val Ile Asn Glu Leu Thr Glu Lys Gln Gly Ile Val
                85                  90                  95

Leu Ala Thr Gly Gly Gly Ser Val Lys Ser Arg Glu Thr Arg Asn Arg
```

-continued

```
            100                 105                 110

Leu Ser Ala Arg Gly Val Val Val Tyr Leu Glu Thr Thr Ile Glu Lys
        115                 120                 125

Gln Leu Ala Arg Thr Gln Arg Asp Lys Lys Arg Pro Leu Leu His Val
    130                 135                 140

Glu Thr Pro Pro Arg Glu Val Leu Glu Ala Leu Ala Asn Glu Arg Asn
145                 150                 155                 160

Pro Leu Tyr Glu Glu Ile Ala Asp Val Thr Ile Arg Thr Asp Asp Gln
                165                 170                 175

Ser Ala Lys Val Val Ala Asn Gln Ile Ile His Met Leu Glu Ser Asn
            180                 185                 190


<210> SEQ ID NO 18
<211> LENGTH: 586
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 18

Met Ile Ser Gly Ile Leu Ala Ser Pro Gly Ile Ala Phe Gly Lys Ala
1               5                   10                  15

Leu Leu Leu Lys Glu Asp Glu Ile Val Ile Asp Arg Lys Lys Ile Ser
            20                  25                  30

Ala Asp Gln Val Asp Gln Glu Val Glu Arg Phe Leu Ser Gly Arg Ala
        35                  40                  45

Lys Ala Ser Ala Gln Leu Glu Thr Ile Lys Thr Lys Ala Gly Glu Thr
    50                  55                  60

Phe Gly Glu Glu Lys Glu Ala Ile Phe Glu Gly His Ile Met Leu Leu
65                  70                  75                  80

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

His Met Thr Ala Asp Ala Ala Ala His Glu Val Ile Glu Gly Gln Ala
            100                 105                 110

Ser Ala Leu Glu Glu Leu Asp Asp Glu Tyr Leu Lys Glu Arg Ala Ala
        115                 120                 125

Asp Val Arg Asp Ile Gly Lys Arg Leu Leu Arg Asn Ile Leu Gly Leu
    130                 135                 140

Lys Ile Ile Asp Leu Ser Ala Ile Gln Asp Glu Val Ile Leu Val Ala
145                 150                 155                 160

Ala Asp Leu Thr Pro Ser Glu Thr Ala Gln Leu Asn Leu Lys Lys Val
                165                 170                 175

Leu Gly Phe Ile Thr Asp Ala Gly Gly Arg Thr Ser His Thr Ser Ile
            180                 185                 190

Met Ala Arg Ser Leu Glu Leu Pro Ala Ile Val Gly Thr Gly Ser Val
        195                 200                 205

Thr Ser Gln Val Lys Asn Asp Asp Tyr Leu Ile Leu Asp Ala Val Asn
    210                 215                 220

Asn Gln Val Tyr Val Asn Pro Thr Asn Glu Val Ile Asp Lys Met Arg
225                 230                 235                 240

Ala Val Gln Glu Gln Val Ala Ser Glu Lys Ala Glu Leu Ala Lys Leu
                245                 250                 255

Lys Asp Leu Pro Ala Ile Thr Leu Asp Gly His Gln Val Glu Val Cys
            260                 265                 270

Ala Asn Ile Gly Thr Val Arg Asp Val Glu Gly Ala Glu Arg Asn Gly
        275                 280                 285
```

-continued

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

Asp Ala Leu Pro Thr Glu Glu Glu Gln Phe Ala Ala Tyr Lys Ala Val
305                 310                 315                 320

Ala Glu Ala Cys Gly Ser Gln Ala Val Ile Val Arg Thr Met Asp Ile
                325                 330                 335

Gly Gly Asp Lys Glu Leu Pro Tyr Met Asn Phe Pro Lys Glu Glu Asn
            340                 345                 350

Pro Phe Leu Gly Trp Arg Ala Ile Arg Ile Ala Met Asp Arg Arg Glu
        355                 360                 365

Ile Leu Arg Asp Gln Leu Arg Ala Ile Leu Arg Ala Ser Ala Phe Gly
    370                 375                 380

Lys Leu Arg Ile Met Phe Pro Met Ile Ile Ser Val Glu Glu Val Arg
385                 390                 395                 400

Ala Leu Arg Lys Glu Ile Glu Ile Tyr Lys Gln Glu Leu Arg Asp Glu
                405                 410                 415

Gly Lys Ala Phe Asp Glu Ser Ile Glu Ile Gly Val Met Val Glu Thr
            420                 425                 430

Pro Ala Ala Ala Thr Ile Ala Arg His Leu Ala Lys Glu Val Asp Phe
        435                 440                 445

Phe Ser Ile Gly Thr Asn Asp Leu Thr Gln Tyr Thr Leu Ala Val Asp
    450                 455                 460

Arg Gly Asn Asp Met Ile Ser His Leu Tyr Gln Pro Met Ser Pro Ser
465                 470                 475                 480

Val Leu Asn Leu Ile Lys Gln Val Ile Asp Ala Ser His Ala Glu Gly
                485                 490                 495

Lys Trp Thr Gly Met Cys Gly Glu Leu Ala Gly Asp Glu Arg Ala Thr
            500                 505                 510

Leu Leu Leu Leu Gly Met Gly Leu Asp Glu Phe Ser Met Ser Ala Ile
        515                 520                 525

Ser Ile Pro Arg Ile Lys Lys Ile Ile Arg Asn Thr Asn Phe Glu Asp
    530                 535                 540

Ala Lys Val Leu Ala Glu Gln Ala Leu Ala Gln Pro Thr Thr Asp Glu
545                 550                 555                 560

Leu Met Thr Leu Val Asn Lys Phe Ile Glu Glu Lys Thr Ile Cys Ala
                565                 570                 575

Ala Asn Asp Glu Asn Tyr Ala Leu Ala Ala
            580                 585


<210> SEQ ID NO 19
<211> LENGTH: 481
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 19

Met Lys Lys Thr Lys Ile Val Cys Thr Ile Gly Pro Lys Thr Glu Ser
1               5                   10                  15

Glu Glu Met Leu Ala Lys Met Leu Asp Ala Gly Met Asn Val Met Arg
            20                  25                  30

Leu Asn Phe Ser His Gly Asp Tyr Ala Glu His Gly Gln Arg Ile Gln
        35                  40                  45

Asn Leu Arg Asn Val Met Ser Lys Thr Gly Lys Thr Ala Ala Ile Leu
    50                  55                  60

Leu Asp Thr Lys Gly Pro Glu Ile Arg Thr Met Lys Leu Glu Gly Gly
65                  70                  75                  80
```

-continued

```
Asn Asp Val Ser Leu Lys Ala Gly Gln Thr Phe Thr Phe Thr Thr Asp
                85                  90                  95

Lys Ser Val Ile Gly Asn Ser Glu Met Val Ala Val Thr Tyr Glu Gly
            100                 105                 110

Phe Thr Thr Asp Leu Ser Val Gly Asn Thr Val Leu Val Asp Asp Gly
        115                 120                 125

Leu Ile Gly Met Glu Val Thr Ala Ile Glu Gly Asn Lys Val Ile Cys
    130                 135                 140

Lys Val Leu Asn Asn Gly Asp Leu Gly Glu Asn Lys Gly Val Asn Leu
145                 150                 155                 160

Pro Gly Val Ser Ile Ala Leu Pro Ala Leu Ala Glu Lys Asp Lys Gln
                165                 170                 175

Asp Leu Ile Phe Gly Cys Glu Gln Gly Val Asp Phe Val Ala Ala Ser
            180                 185                 190

Phe Ile Arg Lys Arg Ser Asp Val Ile Glu Ile Arg Glu His Leu Lys
        195                 200                 205

Ala His Gly Gly Glu Asn Ile His Ile Ile Ser Lys Ile Glu Asn Gln
    210                 215                 220

Glu Gly Leu Asn Asn Phe Asp Glu Ile Leu Glu Ala Ser Asp Gly Ile
225                 230                 235                 240

Met Val Ala Arg Gly Asp Leu Gly Val Glu Ile Pro Val Glu Glu Val
                245                 250                 255

Ile Phe Ala Gln Lys Met Met Ile Glu Lys Cys Ile Arg Ala Arg Lys
            260                 265                 270

Val Val Ile Thr Ala Thr Gln Met Leu Asp Ser Met Ile Lys Asn Pro
        275                 280                 285

Arg Pro Thr Arg Ala Glu Ala Gly Asp Val Ala Asn Ala Ile Leu Asp
    290                 295                 300

Gly Thr Asp Ala Val Met Leu Ser Gly Glu Ser Ala Lys Gly Lys Tyr
305                 310                 315                 320

Pro Leu Glu Ala Val Ser Ile Met Ala Thr Ile Cys Glu Arg Thr Asp
                325                 330                 335

Arg Val Met Asn Ser Arg Leu Glu Phe Asn Asn Asp Asn Arg Lys Leu
            340                 345                 350

Arg Ile Thr Glu Ala Val Cys Arg Gly Ala Val Glu Thr Ala Glu Lys
        355                 360                 365

Leu Asp Ala Pro Leu Ile Val Val Ala Thr Gln Gly Gly Lys Ser Ala
    370                 375                 380

Arg Ala Val Arg Lys Tyr Phe Pro Asp Ala Thr Ile Leu Ala Leu Thr
385                 390                 395                 400

Thr Asn Glu Lys Thr Ala His Gln Leu Val Leu Ser Lys Gly Val Val
                405                 410                 415

Pro Gln Leu Val Lys Glu Ile Thr Ser Thr Asp Asp Phe Tyr Arg Leu
            420                 425                 430

Gly Lys Glu Leu Ala Leu Gln Ser Gly Leu Ala His Lys Gly Asp Val
        435                 440                 445

Val Val Met Val Ser Gly Ala Leu Val Pro Ser Gly Thr Thr Asn Thr
    450                 455                 460

Ala Ser Val His Val Leu Ala Ala Asn Asp Glu Asn Tyr Ala Leu Ala
465                 470                 475                 480

Ala
```

-continued

<210> SEQ ID NO 20
<211> LENGTH: 438
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 20

Met Ala Asp Thr Lys Ala Lys Leu Thr Leu Asn Gly Asp Thr Ala Val
1 5 10 15

Glu Leu Asp Val Leu Lys Gly Thr Leu Gly Gln Asp Val Ile Asp Ile
20 25 30

Arg Thr Leu Gly Ser Lys Gly Val Phe Thr Phe Asp Pro Gly Phe Thr
35 40 45

Ser Thr Ala Ser Cys Glu Ser Lys Ile Thr Phe Ile Asp Gly Asp Glu
50 55 60

Gly Ile Leu Leu His Arg Gly Phe Pro Ile Asp Gln Leu Ala Thr Asp
65 70 75 80

Ser Asn Tyr Leu Glu Val Cys Tyr Ile Leu Leu Asn Gly Glu Lys Pro
85 90 95

Thr Gln Glu Gln Tyr Asp Glu Phe Lys Thr Thr Val Thr Arg His Thr
100 105 110

Met Ile His Glu Gln Ile Thr Arg Leu Phe His Ala Phe Arg Arg Asp
115 120 125

Ser His Pro Met Ala Val Met Cys Gly Ile Thr Gly Ala Leu Ala Ala
130 135 140

Phe Tyr His Asp Ser Leu Asp Val Asn Asn Pro Arg His Arg Glu Ile
145 150 155 160

Ala Ala Phe Arg Leu Leu Ser Lys Met Pro Thr Met Ala Ala Met Cys
165 170 175

Tyr Lys Tyr Ser Ile Gly Gln Pro Phe Val Tyr Pro Arg Asn Asp Leu
180 185 190

Ser Tyr Ala Gly Asn Phe Leu Asn Met Met Phe Ser Thr Pro Cys Glu
195 200 205

Pro Tyr Glu Val Asn Pro Ile Leu Glu Arg Ala Met Asp Arg Ile Leu
210 215 220

Ile Leu His Ala Asp His Glu Gln Asn Ala Ser Thr Ser Thr Val Arg
225 230 235 240

Thr Ala Gly Ser Ser Gly Ala Asn Pro Phe Ala Cys Ile Ala Ala Gly
245 250 255

Ile Ala Ser Leu Trp Gly Pro Ala His Gly Gly Ala Asn Glu Ala Ala
260 265 270

Leu Lys Met Leu Glu Glu Ile Ser Ser Val Lys His Ile Pro Glu Phe
275 280 285

Val Arg Arg Ala Lys Asp Lys Asn Asp Ser Phe Arg Leu Met Gly Phe
290 295 300

Gly His Arg Val Tyr Lys Asn Tyr Asp Pro Arg Ala Thr Val Met Arg
305 310 315 320

Glu Thr Cys His Glu Val Leu Lys Glu Leu Gly Thr Lys Asp Asp Leu
325 330 335

Leu Glu Val Ala Met Glu Leu Glu Asn Ile Ala Leu Asn Asp Pro Tyr
340 345 350

Phe Ile Glu Lys Lys Leu Tyr Pro Asn Val Asp Phe Tyr Ser Gly Ile
355 360 365

Ile Leu Lys Ala Met Gly Ile Pro Ser Ser Met Phe Thr Val Ile Phe
370 375 380

-continued

```
Ala Met Ala Arg Thr Val Gly Trp Ile Ala His Trp Ser Glu Met His
385                 390                 395                 400

Ser Asp Gly Met Lys Ile Ala Arg Pro Arg Gln Leu Tyr Thr Gly Tyr
                405                 410                 415

Glu Lys Arg Asp Phe Lys Ser Asp Ile Lys Arg Ala Ala Asn Asp Glu
            420                 425                 430

Asn Tyr Ala Leu Ala Ala
        435


<210> SEQ ID NO 21
<211> LENGTH: 367
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 21

Met Gln Lys Asp Ala Leu Asn Asn Val His Ile Thr Asp Glu Gln Val
1               5                   10                  15

Leu Met Thr Pro Glu Gln Leu Lys Ala Ala Phe Pro Leu Ser Leu Gln
            20                  25                  30

Gln Glu Ala Gln Ile Ala Asp Ser Arg Lys Ser Ile Ser Asp Ile Ile
        35                  40                  45

Ala Gly Arg Asp Pro Arg Leu Leu Val Val Cys Gly Pro Cys Ser Ile
    50                  55                  60

His Asp Pro Glu Thr Ala Leu Glu Tyr Ala Arg Arg Phe Lys Ala Leu
65                  70                  75                  80

Ala Ala Glu Val Ser Asp Ser Leu Tyr Leu Val Met Arg Val Tyr Phe
                85                  90                  95

Glu Lys Pro Arg Thr Thr Val Gly Trp Lys Gly Leu Ile Asn Asp Pro
            100                 105                 110

His Met Asp Gly Ser Phe Asp Val Glu Ala Gly Leu Gln Ile Ala Arg
        115                 120                 125

Lys Leu Leu Leu Glu Leu Val Asn Met Gly Leu Pro Leu Ala Thr Glu
    130                 135                 140

Ala Leu Asp Pro Asn Ser Pro Gln Tyr Leu Gly Asp Leu Phe Ser Trp
145                 150                 155                 160

Ser Ala Ile Gly Ala Arg Thr Thr Glu Ser Gln Thr His Arg Glu Met
                165                 170                 175

Ala Ser Gly Leu Ser Met Pro Val Gly Phe Lys Asn Gly Thr Asp Gly
            180                 185                 190

Ser Leu Ala Thr Ala Ile Asn Ala Met Arg Ala Ala Ala Gln Pro His
        195                 200                 205

Arg Phe Val Gly Ile Asn Gln Ala Gly Gln Val Ala Leu Leu Gln Thr
    210                 215                 220

Gln Gly Asn Pro Asp Gly His Val Ile Leu Arg Gly Gly Lys Ala Pro
225                 230                 235                 240

Asn Tyr Ser Pro Ala Asp Val Ala Gln Cys Glu Lys Glu Met Glu Gln
                245                 250                 255

Ala Gly Leu Arg Pro Ser Leu Met Val Asp Cys Ser His Gly Asn Ser
            260                 265                 270

Asn Lys Asp Tyr Arg Arg Gln Pro Ala Val Ala Glu Ser Val Val Ala
        275                 280                 285

Gln Ile Lys Asp Gly Asn Arg Ser Ile Ile Gly Leu Met Ile Glu Ser
    290                 295                 300

Asn Ile His Glu Gly Asn Gln Ser Ser Glu Gln Pro Arg Ser Glu Met
305                 310                 315                 320
```

-continued

```
Lys Tyr Gly Val Ser Val Thr Asp Ala Cys Ile Ser Trp Glu Met Thr
                325                 330                 335

Asp Ala Leu Leu Arg Glu Ile His Gln Asp Leu Asn Gly Gln Leu Thr
            340                 345                 350

Ala Arg Val Ala Ala Ala Asn Asp Glu Asn Tyr Ala Leu Ala Ala
        355                 360                 365


<210> SEQ ID NO 22
<211> LENGTH: 424
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 22

Met Thr Asp Lys Arg Lys Asp Gly Ser Gly Lys Leu Leu Tyr Cys Ser
1               5                   10                  15

Phe Cys Gly Lys Ser Gln His Glu Val Arg Lys Leu Ile Ala Gly Pro
            20                  25                  30

Ser Val Tyr Ile Cys Asp Glu Cys Val Asp Leu Cys Asn Asp Ile Ile
        35                  40                  45

Arg Glu Glu Ile Lys Glu Val Ala Pro His Arg Glu Arg Ser Ala Leu
    50                  55                  60

Pro Thr Pro His Glu Ile Arg Asn His Leu Asp Asp Tyr Val Ile Gly
65                  70                  75                  80

Gln Glu Gln Ala Lys Lys Val Leu Ala Val Ala Val Tyr Asn His Tyr
                85                  90                  95

Lys Arg Leu Arg Asn Gly Asp Thr Ser Asn Gly Val Glu Leu Gly Lys
            100                 105                 110

Ser Asn Ile Leu Leu Ile Gly Pro Thr Gly Ser Gly Lys Thr Leu Leu
        115                 120                 125

Ala Glu Thr Leu Ala Arg Leu Leu Asp Val Pro Phe Thr Met Ala Asp
    130                 135                 140

Ala Thr Thr Leu Thr Glu Ala Gly Tyr Val Gly Glu Asp Val Glu Asn
145                 150                 155                 160

Ile Ile Gln Lys Leu Leu Gln Lys Cys Asp Tyr Asp Val Gln Lys Ala
                165                 170                 175

Gln Arg Gly Ile Val Tyr Ile Asp Glu Ile Asp Lys Ile Ser Arg Lys
            180                 185                 190

Ser Asp Asn Pro Ser Ile Thr Arg Asp Val Ser Gly Glu Gly Val Gln
        195                 200                 205

Gln Ala Leu Leu Lys Leu Ile Glu Gly Thr Val Ala Ala Val Pro Pro
    210                 215                 220

Gln Gly Gly Arg Lys His Pro Gln Gln Glu Phe Leu Gln Val Asp Thr
225                 230                 235                 240

Ser Lys Ile Leu Phe Ile Cys Gly Gly Ala Phe Ala Gly Leu Asp Lys
                245                 250                 255

Val Ile Ser His Arg Val Glu Thr Gly Ser Gly Ile Gly Phe Gly Ala
            260                 265                 270

Thr Val Lys Ala Lys Ser Asp Lys Ala Ser Glu Gly Glu Leu Leu Ala
        275                 280                 285

Gln Val Glu Pro Glu Asp Leu Ile Lys Phe Gly Leu Ile Pro Glu Phe
    290                 295                 300

Ile Gly Arg Leu Pro Val Val Ala Thr Leu Asn Glu Leu Ser Glu Glu
305                 310                 315                 320

Ala Leu Ile Gln Ile Leu Lys Glu Pro Lys Asn Ala Leu Thr Lys Gln
```

-continued

```
                325                 330                 335

Tyr Gln Ala Leu Phe Asn Leu Glu Gly Val Asp Leu Glu Phe Arg Asp
            340                 345                 350

Glu Ala Leu Asp Ala Ile Ala Lys Lys Ala Met Ala Arg Lys Thr Gly
        355                 360                 365

Ala Arg Gly Leu Arg Ser Ile Val Glu Ala Ala Leu Leu Asp Thr Met
    370                 375                 380

Tyr Asp Leu Pro Ser Met Glu Asp Val Glu Lys Val Val Ile Asp Glu
385                 390                 395                 400

Ser Val Ile Asp Gly Gln Ser Lys Pro Leu Leu Ile Tyr Gly Lys Pro
                405                 410                 415

Glu Ala Gln Gln Ala Ser Gly Glu
            420
```

<210> SEQ ID NO 23
<211> LENGTH: 758
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 23

```
Met Leu Asn Gln Glu Leu Glu Leu Ser Leu Asn Met Ala Phe Ala Arg
1               5                   10                  15

Ala Arg Glu His Arg His Glu Phe Met Thr Val Glu His Leu Leu Leu
            20                  25                  30

Ala Leu Leu Ser Asn Pro Ser Ala Arg Glu Ala Leu Glu Ala Cys Ser
        35                  40                  45

Val Asp Leu Val Ala Leu Arg Gln Glu Leu Glu Ala Phe Ile Glu Gln
    50                  55                  60

Thr Thr Pro Val Leu Pro Ala Ser Glu Glu Glu Arg Asp Thr Gln Pro
65                  70                  75                  80

Thr Leu Ser Phe Gln Arg Val Leu Gln Arg Ala Val Phe His Val Gln
                85                  90                  95

Ser Ser Gly Arg Asn Glu Val Thr Gly Ala Asn Val Leu Val Ala Ile
            100                 105                 110

Phe Ser Glu Gln Glu Ser Gln Ala Ala Tyr Leu Leu Arg Lys His Glu
        115                 120                 125

Val Ser Arg Leu Asp Val Val Asn Phe Ile Ser His Gly Thr Arg Lys
    130                 135                 140

Asp Glu Pro Thr Gln Ser Ser Asp Pro Gly Ser Gln Pro Asn Ser Glu
145                 150                 155                 160

Glu Gln Ala Gly Gly Glu Glu Arg Met Glu Asn Phe Thr Thr Asn Leu
                165                 170                 175

Asn Gln Leu Ala Arg Val Gly Gly Ile Asp Pro Leu Ile Gly Arg Glu
            180                 185                 190

Lys Glu Leu Glu Arg Ala Ile Gln Val Leu Cys Arg Arg Arg Lys Asn
        195                 200                 205

Asn Pro Leu Leu Val Gly Glu Ser Gly Val Gly Lys Thr Ala Ile Ala
    210                 215                 220

Glu Gly Leu Ala Trp Arg Ile Val Gln Gly Asp Val Pro Glu Val Met
225                 230                 235                 240

Ala Asp Cys Thr Ile Tyr Ser Leu Asp Ile Gly Ser Leu Leu Ala Gly
                245                 250                 255

Thr Lys Tyr Arg Gly Asp Phe Glu Lys Arg Phe Lys Ala Leu Leu Lys
            260                 265                 270
```

-continued

```
Gln Leu Glu Gln Asp Thr Asn Ser Ile Leu Phe Ile Asp Glu Ile His
        275                 280                 285

Thr Ile Ile Gly Ala Gly Ala Ala Ser Gly Gly Gln Val Asp Ala Ala
    290                 295                 300

Asn Leu Ile Lys Pro Leu Leu Ser Ser Gly Lys Ile Arg Val Ile Gly
305                 310                 315                 320

Ser Thr Thr Tyr Gln Glu Phe Ser Asn Ile Phe Glu Lys Asp Arg Ala
                325                 330                 335

Leu Ala Arg Arg Phe Gln Lys Ile Asp Ile Thr Glu Pro Ser Ile Glu
            340                 345                 350

Glu Thr Val Gln Ile Ile Asn Gly Leu Lys Pro Lys Tyr Glu Ala His
        355                 360                 365

His Asp Val Arg Tyr Thr Ala Lys Ala Val Arg Ala Ala Val Glu Leu
    370                 375                 380

Ala Val Lys Tyr Ile Asn Asp Arg His Leu Pro Asp Lys Ala Ile Asp
385                 390                 395                 400

Val Ile Asp Glu Ala Gly Ala Arg Ala Arg Leu Met Pro Val Ser Lys
                405                 410                 415

Arg Lys Lys Thr Val Asn Val Ala Asp Ile Glu Ser Val Val Ala Arg
            420                 425                 430

Ile Ala Arg Ile Pro Glu Lys Ser Val Ser Gln Ser Asp Arg Asp Thr
        435                 440                 445

Leu Lys Asn Leu Gly Asp Arg Leu Lys Met Leu Val Phe Gly Gln Asp
    450                 455                 460

Lys Ala Ile Glu Ala Leu Thr Glu Ala Ile Lys Met Ala Arg Ala Gly
465                 470                 475                 480

Leu Gly His Glu His Lys Pro Val Gly Ser Phe Leu Phe Ala Gly Pro
                485                 490                 495

Thr Gly Val Gly Lys Thr Glu Val Thr Val Gln Leu Ser Lys Ala Leu
            500                 505                 510

Gly Ile Glu Leu Leu Arg Phe Asp Met Ser Glu Tyr Met Glu Arg His
        515                 520                 525

Thr Val Ser Arg Leu Ile Gly Ala Pro Pro Gly Tyr Val Gly Phe Asp
    530                 535                 540

Gln Gly Gly Leu Leu Thr Asp Ala Val Ile Lys His Pro His Ala Val
545                 550                 555                 560

Leu Leu Leu Asp Glu Ile Glu Lys Ala His Pro Asp Val Phe Asn Ile
                565                 570                 575

Leu Leu Gln Val Met Asp Asn Gly Thr Leu Thr Asp Asn Asn Gly Arg
            580                 585                 590

Lys Ala Asp Phe Arg Asn Val Val Leu Val Met Thr Thr Asn Ala Gly
        595                 600                 605

Val Arg Glu Thr Glu Arg Lys Ser Ile Gly Leu Ile His Gln Asp Asn
    610                 615                 620

Ser Thr Asp Ala Met Glu Glu Ile Lys Lys Ile Phe Thr Pro Glu Phe
625                 630                 635                 640

Arg Asn Arg Leu Asp Asn Ile Ile Trp Phe Asp His Leu Ser Thr Asp
                645                 650                 655

Val Ile His Gln Val Val Asp Lys Phe Ile Val Glu Leu Gln Val Gln
            660                 665                 670

Leu Asp Gln Lys Gly Val Ser Leu Glu Val Ser Gln Glu Ala Arg Asn
        675                 680                 685

Trp Leu Ala Glu Lys Gly Tyr Asp Arg Ala Met Gly Ala Arg Pro Met
```

-continued

690 695 700

Ala Arg Val Ile Gln Asp Asn Leu Lys Lys Pro Leu Ala Asn Glu Leu
705 710 715 720

Leu Phe Gly Ser Leu Val Asp Gly Gly Gln Val Thr Val Ala Leu Asp
725 730 735

Lys Glu Lys Asn Glu Leu Thr Tyr Gly Phe Gln Ser Ala Gln Lys His
740 745 750

Lys Ala Glu Ala Ala His
755

<210> SEQ ID NO 24
<211> LENGTH: 207
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 24

Met Ser Tyr Ser Gly Glu Arg Asp Asn Phe Ala Pro His Met Ala Leu
1 5 10 15

Val Pro Met Val Ile Glu Gln Thr Ser Arg Gly Glu Arg Ser Phe Asp
20 25 30

Ile Tyr Ser Arg Leu Leu Lys Glu Arg Val Ile Phe Leu Thr Gly Gln
35 40 45

Val Glu Asp His Met Ala Asn Leu Ile Val Ala Gln Met Leu Phe Leu
50 55 60

Glu Ala Glu Asn Pro Glu Lys Asp Ile Tyr Leu Tyr Ile Asn Ser Pro
65 70 75 80

Gly Gly Val Ile Thr Ala Gly Met Ser Ile Tyr Asp Thr Met Gln Phe
85 90 95

Ile Lys Pro Asp Val Ser Thr Ile Cys Met Gly Gln Ala Ala Ser Met
100 105 110

Gly Ala Phe Leu Leu Thr Ala Gly Ala Lys Gly Lys Arg Phe Cys Leu
115 120 125

Pro Asn Ser Arg Val Met Ile His Gln Pro Leu Gly Gly Tyr Gln Gly
130 135 140

Gln Ala Thr Asp Ile Glu Ile His Ala Arg Glu Ile Leu Lys Val Lys
145 150 155 160

Gly Arg Met Asn Glu Leu Met Ala Leu His Thr Gly Gln Ser Leu Glu
165 170 175

Gln Ile Glu Arg Asp Thr Glu Arg Asp Arg Phe Leu Ser Ala Pro Glu
180 185 190

Ala Val Glu Tyr Gly Leu Val Asp Ser Ile Leu Thr His Arg Asn
195 200 205

<210> SEQ ID NO 25
<211> LENGTH: 165
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 25

Met Asp Leu Ser Gln Leu Thr Pro Arg Arg Pro Tyr Leu Leu Arg Ala
1 5 10 15

Phe Tyr Glu Trp Leu Leu Asp Asn Gln Leu Thr Pro His Leu Val Val
20 25 30

Asp Val Thr Leu Pro Gly Val Gln Val Pro Met Glu Tyr Ala Arg Asp
35 40 45

Gly Gln Ile Val Leu Asn Ile Ala Pro Arg Ala Val Gly Asn Leu Glu

-continued

```
    50                  55                  60

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

Arg Gln Val Ser Val Pro Leu Ala Ala Val Leu Ala Ile Tyr Ala Arg
                85                  90                  95

Glu Asn Gly Ala Gly Thr Met Phe Glu Pro Glu Ala Ala Tyr Asp Glu
            100                 105                 110

Asp Thr Ser Ile Met Asn Asp Glu Glu Ala Ser Ala Asp Asn Glu Thr
        115                 120                 125

Val Met Ser Val Ile Asp Gly Asp Lys Pro Asp His Asp Asp Asp Thr
    130                 135                 140

His Pro Asp Asp Glu Pro Pro Gln Pro Pro Arg Gly Gly Arg Pro Ala
145                 150                 155                 160

Leu Arg Val Val Lys
                165


<210> SEQ ID NO 26
<211> LENGTH: 106
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 26

Met Gly Lys Thr Asn Asp Trp Leu Asp Phe Asp Gln Leu Ala Glu Glu
1               5                   10                  15

Lys Val Arg Asp Ala Leu Lys Pro Pro Ser Met Tyr Lys Val Ile Leu
            20                  25                  30

Val Asn Asp Asp Tyr Thr Pro Met Glu Phe Val Ile Asp Val Leu Gln
        35                  40                  45

Lys Phe Phe Ser Tyr Asp Val Glu Arg Ala Thr Gln Leu Met Leu Ala
    50                  55                  60

Val His Tyr Gln Gly Lys Ala Ile Cys Gly Val Phe Thr Ala Glu Val
65                  70                  75                  80

Ala Glu Thr Lys Val Ala Met Val Asn Lys Tyr Ala Arg Glu Asn Glu
                85                  90                  95

His Pro Leu Leu Cys Thr Leu Glu Lys Ala
            100                 105


<210> SEQ ID NO 27
<211> LENGTH: 162
<212> TYPE: PRT
<213> ORGANISM: Caulobacter crescentus

<400> SEQUENCE: 27

Met Ser Gln Thr Glu Pro Pro Glu Asp Leu Met Gln Tyr Glu Ala Met
1               5                   10                  15

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

Ala Ala Pro Gly Gly Leu Pro Glu Pro His His Leu Tyr Ile Thr Phe
        35                  40                  45

Lys Thr Lys Ala Ala Gly Val Ser Gly Pro Gln Asp Leu Leu Ser Lys
    50                  55                  60

Tyr Pro Asp Glu Met Thr Ile Val Leu Gln His Gln Tyr Trp Asp Leu
65                  70                  75                  80

Ala Pro Gly Glu Thr Phe Phe Ser Val Thr Leu Lys Phe Gly Gly Gln
                85                  90                  95

Pro Lys Arg Leu Ser Val Pro Tyr Ala Ala Leu Thr Arg Phe Tyr Asp
```

-continued

```
            100                 105                 110

Pro Ser Val Gln Phe Ala Leu Gln Phe Ser Ala Pro Glu Ile Ile Glu
        115                 120                 125

Asp Glu Pro Glu Pro Asp Pro Glu Pro Glu Asp Lys Ala Asn Gln Gly
    130                 135                 140

Ala Ser Gly Asp Glu Gly Pro Lys Ile Val Ser Leu Asp Gln Phe Arg
145                 150                 155                 160

Lys Lys
```

<210> SEQ ID NO 28
<211> LENGTH: 1252
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 28

```
gaagttctgg aagcgttggc caatgaacgc aatccgctgt atgaagagat tgccgacgtg     60

accattcgta ctgatgatca aagcgctaaa gtggttgcaa accagattat tcacatgctg    120

gaaagcaacg cagctaacga tgaaaactac agcgaaaact atgctgacgc tagctaatac    180

tagagctgat ccttcaactc agcaaaagtt cgatttattc aacaaagcca cgttgtgtct    240

caaaatctct gatgttacat tgcacaagat aaaaatatat catcatgaac aataaaactg    300

tctgcttaca taaacagtaa tacaaggggt gttatgagcc atattcaacg ggaaacgtct    360

tgctcccgtc cgcgcttaaa ctccaacatg gacgctgatt tatatgggta taaatgggct    420

cgcgataatg tcgggcaatc aggtgcgaca atctatcgct tgtatgggaa gcccgatgcg    480

ccagagttgt ttctgaaaca tggcaaaggt agcgttgcca atgatgttac agatgagatg    540

gtccgtctca actggctgac ggagtttatg cctctcccga ccatcaagca ttttatccgt    600

actcctgatg atgcgtggtt actcaccacc gcgattcctg ggaaaacagc cttccaggta    660

ttagaagaat atcctgattc aggtgaaaat attgttgatg cgctggccgt gttcctgcgc    720

cggttacatt cgattcctgt ttgtaattgt ccttttaaca gcgatcgtgt atttcgtctt    780

gctcaggcgc aatcacgcat gaataacggt ttggttgatg cgagtgattt tgatgacgag    840

cgtaatggct ggcctgttga acaagtctgg aaagaaatgc acaagctctt gccattctca    900

ccggattcag tcgtcactca tggtgatttc tcacttgata accttatttt tgacgagggg    960

aaattaatag gttgtattga tgttggacgg gtcggaatcg cagaccgtta ccaggacctt   1020

gccattcttt ggaactgcct cggtgagttt tctccttcat tacagaaacg gctttttcaa   1080

aaatatggta ttgataatcc tgatatgaat aaattgcagt ttcatttgat gctcgatgag   1140

tttttctaat aattctggct ttatatacac tcgtctgcgg gtacagtaat taaggtggat   1200

gtcgcgttat ggagaggatt gtcgttactc tcggggaacg tagttaccca at           1252
```

<210> SEQ ID NO 29
<211> LENGTH: 58
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 29

```
ggccgcttct agagtcacac aggaaagtac tagatggcag agaaacgcaa tatctttc       58
```

-continued

<210> SEQ ID NO 30
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 30

ggccctgcag cggccgctac tagtattaag cagccagagc ataattttca tcgttagctg 60

cgttgctttc cagcatgtga ataatc 86

<210> SEQ ID NO 31
<211> LENGTH: 2738
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 31

tactagtagc ggccgctgca ggagtcacta agggttagtt agttagatta gcagaaagtc 60

aaaagcctcc gaccggaggc ttttgactaa aacttccctt ggggttatca ttggggctca 120

ctcaaaggcg gtaatcagat aaaaaaaatc cttagctttc gctaaggatg atttctgcta 180

gagatggaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 240

ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta 300

tcattgcagc actggggcca gatggtaagc cctcccgtat cgtagttatc tacacgacgg 360

ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 420

ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt gatttaaaac 480

ttcattttta atttaaaagg atctaggtga agatcctttt tgataatctc atgaccaaaa 540

tcccttaacg tgagttttcg ttccactgag cgtcagaccc cttaataaga tgatcttctt 600

gagatcgttt tggtctgcgc gtaatctctt gctctgaaaa cgaaaaaacc gccttgcagg 660

gcggtttttc gaaggttctc tgagctacca actctttgaa ccgaggtaac tggcttggag 720

gagcgcagtc accaaaactt gtcctttcag tttagcctta accggcgcat gacttcaaga 780

ctaactcctc taaatcaatt accagtggct gctgccagtg gtgcttttgc atgtctttcc 840

gggttggact caagacgata gttaccggat aaggcgcagc ggtcggactg aacgggggggt 900

tcgtgcatac agtccagctt ggagcgaact gcctacccgg aactgagtgt caggcgtgga 960

atgagacaaa cgcggccata acagcggaat gacaccggta aaccgaaagg caggaacagg 1020

agagcgcacg agggagccgc caggggaaac gcctggtatc tttatagtcc tgtcgggttt 1080

cgccaccact gatttgagcg tcagatttcg tgatgcttgt caggggggcg gagcctatgg 1140

aaaaacggct ttgccgcggc cctctcactt ccctgttaag tatcttcctg gcatcttcca 1200

ggaaatctcc gccccgttcg taagccattt ccgctcgccg cagtcgaacg accgagcgta 1260

gcgagtcagt gagcgaggaa gcggaatata tcctgtatca catattctgc tgacgcaccg 1320

gtgcagcctt ttttctcctg ccacatgaag cacttcactg acaccctcat cagtgccaac 1380

atagtaagcc agtatacact ccgctagcgc tgaggtctgc ctcgtgaaga aggtgttgct 1440

gactcatacc aggcctgaat cgccccatca tccagccaga aagtgaggga gccacggttg 1500

atgagagctt tgttgtaggt ggaccagttg gtgattttga acttttgctt tgccacggaa 1560

cggtctgcgt tgtcgggaag atgcgtgatc tgatccttca actcagcaaa agttcgattt 1620

-continued

attcaacaaa gccacgttgt gtctcaaaat ctctgatgtt acattgcaca agataaaaat 1680

atatcatcat gaacaataaa actgtctgct tacataaaca gtaatacaag gggtgtttac 1740

tagaggttga tcgggcacgt aagaggttcc aactttcacc ataatgaaat aagatcacta 1800

ccgggcgtat tttttgagtt atcgagattt tcaggagcta aggaagctaa aatggagaaa 1860

aaaatcacgg gatataccac cgttgatata tcccaatggc atcgtaaaga acattttgag 1920

gcatttcagt cagttgctca atgtacctat aaccagaccg ttcagctgga tattacggcc 1980

tttttaaaga ccgtaaagaa aaataagcac aagttttatc cggcctttat tcacattctt 2040

gcccgcctga tgaacgctca cccggagttt cgtatggcca tgaaagacgg tgagctggtg 2100

atctgggata gtgttcaccc ttgttacacc gttttccatg agcaaactga aacgttttcg 2160

tccctctgga gtgaatacca cgacgatttc cggcagtttc tccacatata ttcgcaagat 2220

gtggcgtgtt acggtgaaaa cctggcctat ttccctaaag ggtttattga gaatatgttt 2280

tttgtctcag ccaatccctg ggtgagtttc accagttttg atttaaacgt ggccaatatg 2340

gacaacttct tcgcccccgt tttcacgatg ggcaaatatt atacgcaagg cgacaaggtg 2400

ctgatgccgc tggcgatcca ggttcatcat gccgtttgtg atggcttcca tgtcggccgc 2460

atgcttaatg aattacaaca gtactgtgat gagtggcagg gcggggcgta ataatactag 2520

ctccggcaaa aaaacgggca aggtgtcacc accctgccct ttttctttaa aaccgaaaag 2580

attacttcgc gtttgccacc tgacgtctaa gaaaaggaat attcagcaat ttgcccgtgc 2640

cgaagaaagg cccacccgtg aaggtgagcc agtgagttga ttgctacgta attagttagt 2700

tagcccttag tgactcgaat tcgcggccgc ttctagag 2738

<210> SEQ ID NO 32
<211> LENGTH: 1061
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 32

tccctatcag tgatagagat tgacatccct atcagtgata gagatactga gcactactag 60

agaaagagga gaaatactag atgaaaaaca taaatgccga cgacacatac agaataatta 120

ataaaattaa agcttgtaga agcaataatg atattaatca atgcttatct gatatgacta 180

aaatggtaca ttgtgaatat tatttactcg cgatcattta tcctcattct atggttaaat 240

ctgatatttc aatcctagat aattacccta aaaaatggag gcaatattat gatgacgcta 300

atttaataaa atatgatcct atagtagatt attctaactc caatcattca ccaattaatt 360

ggaatatatt tgaaaacaat gctgtaaata aaaaatctcc aaatgtaatt aaagaagcga 420

aaacatcagg tcttatcact gggtttagtt tccctattca tacggctaac aatggcttcg 480

gaatgcttag ttttgcacat tcagaaaaag acaactatat agatagttta tttttacatg 540

cgtgtatgaa cataccatta attgttcctt ctctagttga taattatcga aaaataaata 600

tagcaaataa taaatcaaac aacgatttaa ccaaaagaga aaaagaatgt ttagcgtggg 660

catgcgaagg aaaaagctct tgggatattt caaaaatatt aggttgcagt gagcgtactg 720

tcactttcca tttaaccaat gcgcaaatga aactcaatac aacaaaccgc tgccaaagta 780

tttctaaagc aattttaaca ggagcaattg attgcccata ctttaaaaat taataacact 840

gatagtgcta gtgtagatca ctactagagc caggcatcaa ataaaacgaa aggctcagtc 900

-continued

gaaagactgg gcctttcgtt ttatctgttg tttgtcggtg aacgctctct actagagtca 960

cactggctca ccttcgggtg ggcctttctg cgtttatata ctagagacct gtaggatcgt 1020

acaggtttac gcaagaaaat ggtttgttat agtcgaataa a 1061

<210> SEQ ID NO 33
<211> LENGTH: 200
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 33

caatacgcaa accgcctctc cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca 60

ggtttcccga ctggaaagcg ggcagtgagc gcaacgcaat taatgtgagt tagctcactc 120

attaggcacc ccaggcttta cactttatgc ttccggctcg tatgttgtgt ggaattgtga 180

gcggataaca atttcacaca 200

<210> SEQ ID NO 34
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 34

tcacacagga aag 13

<210> SEQ ID NO 35
<211> LENGTH: 519
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 35

atggcagaga aacgcaatat ctttctggtt gggcctatgg gtgccggaaa aagcactatt 60

gggcgccagt tagctcaaca actcaatatg gaattttacg attccgatca agagattgag 120

aaacgaaccg gagctgatgt gggctgggtt ttcgatttag aaggcgaaga aggcttccgc 180

gatcgcgaag aaaaggtcat caatgagttg accgagaaac agggtattgt gctggctact 240

ggcggcggct ctgtgaaatc ccgtgaaacg cgtaaccgtc tttccgctcg tggcgttgtc 300

gtttatcttg aaacgaccat cgaaaagcaa cttgcacgca cgcagcgtga taaaaaacgc 360

ccgttgctgc acgttgaaac accgccgcgt gaagttctgg aagcgttggc caatgaacgc 420

aatccgctgt atgaagagat tgccgacgtg accattcgta ctgatgatca aagcgctaaa 480

gtggttgcaa accagattat tcacatgctg gaaagcaac 519

<210> SEQ ID NO 36
<211> LENGTH: 498
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 36

-continued

atggatttgt cacagctaac accacgtcgt ccctatctgc tgcgtgcatt ctatgagtgg 60

ttgctggata accagctcac gccgcacctg gtggtggatg tgacgctccc tggcgtgcag 120

gttcctatgg aatatgcgcg tgacgggcaa atcgtactca acattgcgcc gcgtgctgtc 180

ggcaatctgg aactggcgaa tgatgaggtg cgctttaacg cgcgctttgg tggcattccg 240

cgtcaggttt ctgtgccgct ggctgccgtg ctggctatct acgcccgtga aaatggcgca 300

ggcacgatgt ttgagcctga agctgcctac gatgaagata ccagcatcat gaatgatgaa 360

gaggcatcgg cagacaacga aaccgttatg tcggttattg atggcgacaa gccagatcac 420

gatgatgaca ctcatcctga cgatgaacct ccgcagccac cacgcggtgg tcgaccggca 480

ttacgcgttg tgaagtaa 498

<210> SEQ ID NO 37
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 37

gcagctaacg atgaaaatta tgctctggct gcttaa 36

<210> SEQ ID NO 38
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 38

gcagctaacg atgaaaatta tgctctggtt gcttaa 36

<210> SEQ ID NO 39
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 39

gcagctaacg atgaaaatta tgctgacgct agctaa 36

<210> SEQ ID NO 40
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 40

gcagctaacg atgaaaatta tgctctggac gactaa 36

<210> SEQ ID NO 41
<211> LENGTH: 4379
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

-continued

<400> SEQUENCE: 41

gaattcgcgg ccgcttctag tccctatcag tgatagagat tgacatccct atcagtgata 60

gagatactga gcactactag agaaagagga gaaatactag atgaaaaaca taaatgccga 120

cgacacatac agaataatta ataaaattaa agcttgtaga agcaataatg atattaatca 180

atgcttatct gatatgacta aaatggtaca ttgtgaatat tatttactcg cgatcattta 240

tcctcattct atggttaaat ctgatatttc aatcctagat aattacccta aaaaatggag 300

gcaatattat gatgacgcta atttaataaa atatgatcct atagtagatt attctaactc 360

caatcattca ccaattaatt ggaatatatt tgaaaacaat gctgtaaata aaaaatctcc 420

aaatgtaatt aaagaagcga aaacatcagg tcttatcact gggtttagtt tccctattca 480

tacggctaac aatggcttcg gaatgcttag ttttgcacat tcagaaaaag acaactatat 540

agatagttta tttttacatg cgtgtatgaa cataccatta attgttcctt ctctagttga 600

taattatcga aaaataaata tagcaaataa taaatcaaac aacgatttaa ccaaaagaga 660

aaaagaatgt ttagcgtggg catgcgaagg aaaaagctct tgggatattt caaaaatatt 720

aggttgcagt gagcgtactg tcactttcca tttaaccaat gcgcaaatga aactcaatac 780

aacaaaccgc tgccaaagta tttctaaagc aattttaaca ggagcaattg attgcccata 840

ctttaaaaat taataacact gatagtgcta gtgtagatca ctactagagc caggcatcaa 900

ataaaacgaa aggctcagtc gaaagactgg gcctttcgtt ttatctgttg tttgtcggtg 960

aacgctctct actagagtca cactggctca ccttcgggtg ggcctttctg cgtttatata 1020

ctagagacct gtaggatcgt acaggtttac gcaagaaaat ggtttgttat agtcgaataa 1080

atactagagt cacacaggaa agtactagat ggcagagaaa cgcaatatct ttctggttgg 1140

gcctatgggt gccggaaaaa gcactattgg gcgccagtta gctcaacaac tcaatatgga 1200

attttacgat tccgatcaag agattgagaa acgaaccgga gctgatgtgg gctgggtttt 1260

cgatttagaa ggcgaagaag gcttccgcga tcgcgaagaa aaggtcatca atgagttgac 1320

cgagaaacag ggtattgtgc tggctactgg cggcggctct gtgaaatccc gtgaaacgcg 1380

taaccgtctt tccgctcgtg gcgttgtcgt ttatcttgaa acgaccatcg aaaagcaact 1440

tgcacgcacg cagcgtgata aaaaacgccc gttgctgcac gttgaaacac cgccgcgtga 1500

agttctggaa gcgttggcca atgaacgcaa tccgctgtat gaagagattg ccgacgtgac 1560

cattcgtact gatgatcaaa gcgctaaagt ggttgcaaac cagattattc acatgctgga 1620

aagcaacgca gctaacgatg aaaattatgc tctggttgct taatactagt agcggccgct 1680

gcaggagtca ctaagggtta gttagttaga ttagcagaaa gtcaaaagcc tccgaccgga 1740

ggcttttgac taaaacttcc cttggggtta tcattggggc tcactcaaag gcggtaatca 1800

gataaaaaaa atccttagct ttcgctaagg atgatttctg ctagagatgg aatagactgg 1860

atggaggcgg ataaagttgc aggaccactt ctgcgctcgg cccttccggc tggctggttt 1920

attgctgata aatctggagc cggtgagcgt gggtctcgcg gtatcattgc agcactgggg 1980

ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtca ggcaactatg 2040

gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca ttggtaactg 2100

tcagaccaag tttactcata tatactttag attgatttaa aacttcattt ttaatttaaa 2160

aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt 2220

tcgttccact gagcgtcaga ccccttaata agatgatctt cttgagatcg ttttggtctg 2280

cgcgtaatct cttgctctga aaacgaaaaa accgccttgc agggcggttt ttcgaaggtt 2340

-continued

ctctgagcta ccaactcttt gaaccgaggt aactggcttg gaggagcgca gtcaccaaaa 2400

cttgtccttt cagtttagcc ttaaccggcg catgacttca agactaactc ctctaaatca 2460

attaccagtg gctgctgcca gtggtgcttt tgcatgtctt tccgggttgg actcaagacg 2520

atagttaccg gataaggcgc agcggtcgga ctgaacgggg ggttcgtgca tacagtccag 2580

cttggagcga actgcctacc cggaactgag tgtcaggcgt ggaatgagac aaacgcggcc 2640

ataacagcgg aatgacaccg gtaaaccgaa aggcaggaac aggagagcgc acgagggagc 2700

cgccagggga aacgcctggt atctttatag tcctgtcggg tttcgccacc actgatttga 2760

gcgtcagatt tcgtgatgct tgtcaggggg gcggagccta tggaaaaacg gctttgccgc 2820

ggccctctca cttccctgtt aagtatcttc ctggcatctt ccaggaaatc tccgccccgt 2880

tcgtaagcca tttccgctcg ccgcagtcga acgaccgagc gtagcgagtc agtgagcgag 2940

gaagcggaat atatcctgta tcacatattc tgctgacgca ccggtgcagc ctttttttctc 3000

ctgccacatg aagcacttca ctgacaccct catcagtgcc aacatagtaa gccagtatac 3060

actccgctag cgctgaggtc tgcctcgtga agaaggtgtt gctgactcat accaggcctg 3120

aatcgcccca tcatccagcc agaaagtgag ggagccacgg ttgatgagag ctttgttgta 3180

ggtggaccag ttggtgattt tgaacttttg ctttgccacg gaacggtctg cgttgtcggg 3240

aagatgcgtg atctgatcct tcaactcagc aaaagttcga tttattcaac aaagccacgt 3300

tgtgtctcaa aatctctgat gttacattgc acaagataaa aatatatcat catgaacaat 3360

aaaactgtct gcttacataa acagtaatac aaggggtgtt tactagaggt tgatcgggca 3420

cgtaagaggt tccaactttc accataatga aataagatca ctaccgggcg tatttttga 3480

gttatcgaga ttttcaggag ctaaggaagc taaaatggag aaaaaaatca cgggatatac 3540

caccgttgat atatcccaat ggcatcgtaa agaacatttt gaggcatttc agtcagttgc 3600

tcaatgtacc tataaccaga ccgttcagct ggatattacg gccttttaa agaccgtaaa 3660

gaaaaataag cacaagtttt atccggcctt tattcacatt cttgcccgcc tgatgaacgc 3720

tcacccggag tttcgtatgg ccatgaaaga cggtgagctg gtgatctggg atagtgttca 3780

cccttgttac accgttttcc atgagcaaac tgaaacgttt tcgtccctct ggagtgaata 3840

ccacgacgat ttccggcagt ttctccacat atattcgcaa gatgtggcgt gttacggtga 3900

aaacctggcc tatttcccta aagggtttat tgagaatatg ttttttgtct cagccaatcc 3960

ctgggtgagt ttcaccagtt ttgatttaaa cgtggccaat atggacaact tcttcgcccc 4020

cgttttcacg atgggcaaat attatacgca aggcgacaag gtgctgatgc cgctggcgat 4080

ccaggttcat catgccgttt gtgatggctt ccatgtcggc cgcatgctta atgaattaca 4140

acagtactgt gatgagtggc agggcggggc gtaataatac tagctccggc aaaaaaacgg 4200

gcaaggtgtc accaccctgc ccttttttctt taaaaccgaa aagattactt cgcgtttgcc 4260

acctgacgtc taagaaaagg aatattcagc aatttgcccg tgccgaagaa aggcccaccc 4320

gtgaaggtga gccagtgagt tgattgctac gtaattagtt agttagccct tagtgactc 4379

<210> SEQ ID NO 42
<211> LENGTH: 1137
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium glutamicum

<400> SEQUENCE: 42

atgacgtcat tagagttcac agtaacccgt accgaaaatc cgacgtcacc cgatcgtctg 60

-continued

aaggaaattc ttgccgcacc gaagttcggt aagttcttca ccgaccacat ggtgaccatt 120

gactggaacg agtcggaagg ctggcacaac gcccaattag tgccatacgc gccgattcct 180

atggatcctg ccaccaccgt attccactac ggacaggcaa tttttgaggg aattaaggcc 240

taccgccatt cggacgaaac catcaagact ttccgtcctg atgaaaacgc cgagcgtatg 300

cagcgttcag cagctcgaat ggcaatgcca cagttgccaa ccgaggactt tattaaagca 360

cttgaactgc tggtagacgc ggatcaggat tgggttcctg agtacggcgg agaagcttcc 420

ctctacctgc gcccattcat gatctccacc gaaattggct tgggtgtcag cccagctgat 480

gcctacaagt tcctggtcat cgcatcccca gtcggcgctt acttcaccgg tggaatcaag 540

cctgtttccg tctggctgag cgaagattac gtccgcgctg cacccggcgg aactggtgac 600

gccaaatttg ctggcaacta cgcggcttct ttgcttgccc agtcccaggc tgcggaaaag 660

ggctgtgacc aggtcgtatg gttggatgcc atcgagcaca agtacatcga agaaatgggt 720

ggcatgaacc ttgggttcat ctaccgcaac ggcgaccaag tcaagctagt caccctgaa 780

ctttccggct cactacttcc aggcatcacc cgcaagtcac ttctacaagt agcacgcgac 840

ttgggatacg aagtagaaga gcgaaagatc accaccaccg agtgggaaga agacgcaaag 900

tctggcgcca tgaccgaggc atttgcttgc ggtactgcag ctgttatcac ccctgttggc 960

accgtgaaat cagctcacgg caccttcgaa gtgaacaaca atgaagtcgg agaaatcacg 1020

atgaagcttc gtgaaaccct caccggaatt cagcaaggaa acgttgaaga ccaaaacgga 1080

tggctttacc cactggttgg cgcagctaac gatgaaaatt atgctctggt ggcttaa 1137

<210> SEQ ID NO 43
<211> LENGTH: 843
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium glutamicum

<400> SEQUENCE: 43

atgtcaggca ttgatgcaaa gaaaatccgc acccgtcatt tccgcgaagc taaagtaaac 60

ggccagaaag tttcggttct caccagctat gatgcgcttt cggcgcgcat ttttgatgag 120

gctggcgtcg atatgctcct tgttggtgat tccgctgcca acgttgtgct gggtcgcgat 180

accaccttgt cgatcacctt ggatgagatg attgtgctgg ccaaggcggt gacgatcgct 240

acgaagcgtg cgcttgtggt ggttgatctg ccgtttggta cctatgaggt gagcccaaat 300

caggcggtgg agtccgcgat ccgggtcatg cgtgaaacgg gtgcggctgc ggtgaagatc 360

gagggtggcg tggagatcgc gcagacgatt cgacgcattg ttgatgctgg aattccggtt 420

gtcggccaca tcgggtacac cccgcagtcc gagcattcct tgggcggcca cgtggttcag 480

ggtcgtggcg cgagttctgg aaagctcatc gccgatgccc gcgcgttgga gcaggcgggt 540

gcgtttgcgg ttgtgttgga gatggttcca gcagaggcag cgcgcgaggt taccgaggat 600

ctttccatca ccactatcgg aatcggtgcc ggcaatggca cagatgggca ggttttggtg 660

tggcaggatg ccttcggcct caaccgcggc aagaagccac gcttcgtccg cgagtacgcc 720

accttgggcg attccttgca cgacgccgcg caggcctaca tcgccgatat ccacgcgggt 780

accttcccag gcgaagcgga gtcctttgca gctaacgatg aaaattatgc tctgggcggc 840

taa 843

<210> SEQ ID NO 44
<211> LENGTH: 1950
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium glutamicum

-continued

<400> SEQUENCE: 44

atgcttcacc acatgacttc gcgtgcgaat ctacttcttc ttcgccgcgg cgggtcccag 60

aggtctatgt ctcctaacga tgcattcatc tccgcacctg ccaagatcga aaccccagtt 120

gggcctcgca acgaaggcca gccagcatgg aataagcagc gtggctcctc aatgccagtt 180

aaccgctaca tgcctttcga ggttgaggta gaagatattt ctctgccgga ccgcacttgg 240

ccagataaaa aaatcaccgt tgcacctcag tggtgtgctg ttgacctgcg tgacggcaac 300

caggctctga ttgatccgat gtctcctgag cgtaagcgcc gcatgtttga gctgctggtt 360

cagatgggct tcaaagaaat cgaggtcggt ttcccttcag cttcccagac tgattttgat 420

ttcgttcgtg agatcatcga aaagggcatg atccctgacg atgtcaccat tcaggttctg 480

gttcaggctc gtgagcacct gattcgccgt acttttgaag cttgcgaagg cgcaaaaaac 540

gttatcgtgc acttctacaa ctccacctcc atcctgcagc gcaacgtggt gttccgcatg 600

gacaaggtgc aggtgaagaa gctggctacc gatgccgctg aactaatcaa gaccatcgct 660

caggattacc cagacaccaa ctggcgctgg cagtactccc ctgagtcctt caccggcact 720

gaggttgagt acgccaagga agttgtggac gcagttgttg aggtcatgga tccaactcct 780

gagaacccaa tgatcatcaa cctgccttcc accgttgaga tgatcacccc taacgtttac 840

gcagactcca ttgaatggat gcaccgcaat ctaaaccgtc gtgattccat tatcctgtcc 900

ctgcacccgc acaatgaccg tggcaccggc gttggcgcag ctgagctggg ctacatggct 960

ggcgctgacc gcatcgaagg ctgcctgttc ggcaacggcg agcgcaccgg caacgtctgc 1020

ctggtcaccc tggcactgaa catgctgacc cagggcgttg accctcagct ggacttcacc 1080

gatatacgcc agatccgcag caccgttgaa tactgcaacc agctgcgcgt tcctgagcgc 1140

cacccatacg gcggtgacct ggtcttcacc gctttctccg gttcccacca ggacgctgtg 1200

aacaagggtc tggacgccat ggctgccaag gttcagccag gtgctagctc cactgaagtt 1260

tcttgggagc agctgcgcga caccgaatgg gaggttcctt acctgcctat cgatccaaag 1320

gatgtcggtc gcgactacga ggctgttatc cgcgtgaact cccagtccgg caagggcggc 1380

gttgcttaca tcatgaagac cgatcacggt ctgcagatcc ctcgctccat gcaggttgag 1440

ttctccaccg ttgtccagaa cgtcaccgac gctgagggcg gcgaggtcaa ctccaaggca 1500

atgtgggata tcttcgccac cgagtacctg gagcgcaccg caccagttga gcagatcgcg 1560

ctgcgcgtcg agaacgctca gaccgaaaac gaggatgcat ccatcaccgc cgagctcatc 1620

cacaacggca aggacgtcac cgtcgatggc cgcggcaacg gcccactggc cgcttacgcc 1680

aacgcgctgg agaagctggg catcgacgtt gagatccagg aatacaacca gcacgcccgc 1740

acctcgggcg acgatgcaga agcagccgcc tacgtgctgg ctgaggtcaa cggccgcaag 1800

gtctggggcg tcggcatcgc tggctccatc acctacgctt cgctgaaggc agtgacctcc 1860

gccgtaaacc gcgcgctgga cgtcaaccac gaggcagtcc tggctggcgg cgttgcagct 1920

aacgatgaaa attatgctct ggctggctaa 1950

<210> SEQ ID NO 45
<211> LENGTH: 378
<212> TYPE: PRT
<213> ORGANISM: Corynebacterium glutamicum

<400> SEQUENCE: 45

Met Thr Ser Leu Glu Phe Thr Val Thr Arg Thr Glu Asn Pro Thr Ser
1 5 10 15

-continued

```
Pro Asp Arg Leu Lys Glu Ile Leu Ala Ala Pro Lys Phe Gly Lys Phe
            20                  25                  30

Phe Thr Asp His Met Val Thr Ile Asp Trp Asn Glu Ser Glu Gly Trp
        35                  40                  45

His Asn Ala Gln Leu Val Pro Tyr Ala Pro Ile Pro Met Asp Pro Ala
    50                  55                  60

Thr Thr Val Phe His Tyr Gly Gln Ala Ile Phe Glu Gly Ile Lys Ala
65                  70                  75                  80

Tyr Arg His Ser Asp Glu Thr Ile Lys Thr Phe Arg Pro Asp Glu Asn
                85                  90                  95

Ala Glu Arg Met Gln Arg Ser Ala Ala Arg Met Ala Met Pro Gln Leu
            100                 105                 110

Pro Thr Glu Asp Phe Ile Lys Ala Leu Glu Leu Leu Val Asp Ala Asp
        115                 120                 125

Gln Asp Trp Val Pro Glu Tyr Gly Gly Glu Ala Ser Leu Tyr Leu Arg
    130                 135                 140

Pro Phe Met Ile Ser Thr Glu Ile Gly Leu Gly Val Ser Pro Ala Asp
145                 150                 155                 160

Ala Tyr Lys Phe Leu Val Ile Ala Ser Pro Val Gly Ala Tyr Phe Thr
                165                 170                 175

Gly Gly Ile Lys Pro Val Ser Val Trp Leu Ser Glu Asp Tyr Val Arg
            180                 185                 190

Ala Ala Pro Gly Gly Thr Gly Asp Ala Lys Phe Ala Gly Asn Tyr Ala
        195                 200                 205

Ala Ser Leu Leu Ala Gln Ser Gln Ala Ala Glu Lys Gly Cys Asp Gln
    210                 215                 220

Val Val Trp Leu Asp Ala Ile Glu His Lys Tyr Ile Glu Glu Met Gly
225                 230                 235                 240

Gly Met Asn Leu Gly Phe Ile Tyr Arg Asn Gly Asp Gln Val Lys Leu
                245                 250                 255

Val Thr Pro Glu Leu Ser Gly Ser Leu Leu Pro Gly Ile Thr Arg Lys
            260                 265                 270

Ser Leu Leu Gln Val Ala Arg Asp Leu Gly Tyr Glu Val Glu Glu Arg
        275                 280                 285

Lys Ile Thr Thr Thr Glu Trp Glu Glu Asp Ala Lys Ser Gly Ala Met
    290                 295                 300

Thr Glu Ala Phe Ala Cys Gly Thr Ala Ala Val Ile Thr Pro Val Gly
305                 310                 315                 320

Thr Val Lys Ser Ala His Gly Thr Phe Glu Val Asn Asn Asn Glu Val
                325                 330                 335

Gly Glu Ile Thr Met Lys Leu Arg Glu Thr Leu Thr Gly Ile Gln Gln
            340                 345                 350

Gly Asn Val Glu Asp Gln Asn Gly Trp Leu Tyr Pro Leu Val Gly Ala
        355                 360                 365

Ala Asn Asp Glu Asn Tyr Ala Leu Val Ala
    370                 375
```

<210> SEQ ID NO 46
<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Corynebacterium glutamicum

<400> SEQUENCE: 46

```
Met Ser Gly Ile Asp Ala Lys Lys Ile Arg Thr Arg His Phe Arg Glu
```

-continued

```
1               5                   10                  15

Ala Lys Val Asn Gly Gln Lys Val Ser Val Leu Thr Ser Tyr Asp Ala
            20                  25                  30

Leu Ser Ala Arg Ile Phe Asp Glu Ala Gly Val Asp Met Leu Leu Val
        35                  40                  45

Gly Asp Ser Ala Ala Asn Val Val Leu Gly Arg Asp Thr Thr Leu Ser
    50                  55                  60

Ile Thr Leu Asp Glu Met Ile Val Leu Ala Lys Ala Val Thr Ile Ala
65                  70                  75                  80

Thr Lys Arg Ala Leu Val Val Val Asp Leu Pro Phe Gly Thr Tyr Glu
                85                  90                  95

Val Ser Pro Asn Gln Ala Val Glu Ser Ala Ile Arg Val Met Arg Glu
            100                 105                 110

Thr Gly Ala Ala Ala Val Lys Ile Glu Gly Gly Val Glu Ile Ala Gln
        115                 120                 125

Thr Ile Arg Arg Ile Val Asp Ala Gly Ile Pro Val Val Gly His Ile
    130                 135                 140

Gly Tyr Thr Pro Gln Ser Glu His Ser Leu Gly Gly His Val Val Gln
145                 150                 155                 160

Gly Arg Gly Ala Ser Ser Gly Lys Leu Ile Ala Asp Ala Arg Ala Leu
                165                 170                 175

Glu Gln Ala Gly Ala Phe Ala Val Val Leu Glu Met Val Pro Ala Glu
            180                 185                 190

Ala Ala Arg Glu Val Thr Glu Asp Leu Ser Ile Thr Thr Ile Gly Ile
        195                 200                 205

Gly Ala Gly Asn Gly Thr Asp Gly Gln Val Leu Val Trp Gln Asp Ala
    210                 215                 220

Phe Gly Leu Asn Arg Gly Lys Lys Pro Arg Phe Val Arg Glu Tyr Ala
225                 230                 235                 240

Thr Leu Gly Asp Ser Leu His Asp Ala Ala Gln Ala Tyr Ile Ala Asp
                245                 250                 255

Ile His Ala Gly Thr Phe Pro Gly Glu Ala Glu Ser Phe Ala Ala Asn
            260                 265                 270

Asp Glu Asn Tyr Ala Leu Gly Gly
        275                 280
```

<210> SEQ ID NO 47
<211> LENGTH: 649
<212> TYPE: PRT
<213> ORGANISM: Corynebacterium glutamicum

<400> SEQUENCE: 47

```
Met Leu His His Met Thr Ser Arg Ala Asn Leu Leu Leu Leu Arg Arg
1               5                   10                  15

Gly Gly Ser Gln Arg Ser Met Ser Pro Asn Asp Ala Phe Ile Ser Ala
            20                  25                  30

Pro Ala Lys Ile Glu Thr Pro Val Gly Pro Arg Asn Glu Gly Gln Pro
        35                  40                  45

Ala Trp Asn Lys Gln Arg Gly Ser Ser Met Pro Val Asn Arg Tyr Met
    50                  55                  60

Pro Phe Glu Val Glu Val Glu Asp Ile Ser Leu Pro Asp Arg Thr Trp
65                  70                  75                  80

Pro Asp Lys Lys Ile Thr Val Ala Pro Gln Trp Cys Ala Val Asp Leu
                85                  90                  95
```

-continued

```
Arg Asp Gly Asn Gln Ala Leu Ile Asp Pro Met Ser Pro Glu Arg Lys
            100                 105                 110

Arg Arg Met Phe Glu Leu Leu Val Gln Met Gly Phe Lys Glu Ile Glu
        115                 120                 125

Val Gly Phe Pro Ser Ala Ser Gln Thr Asp Phe Asp Phe Val Arg Glu
    130                 135                 140

Ile Ile Glu Lys Gly Met Ile Pro Asp Asp Val Thr Ile Gln Val Leu
145                 150                 155                 160

Val Gln Ala Arg Glu His Leu Ile Arg Arg Thr Phe Glu Ala Cys Glu
                165                 170                 175

Gly Ala Lys Asn Val Ile Val His Phe Tyr Asn Ser Thr Ser Ile Leu
            180                 185                 190

Gln Arg Asn Val Val Phe Arg Met Asp Lys Val Gln Val Lys Lys Leu
        195                 200                 205

Ala Thr Asp Ala Ala Glu Leu Ile Lys Thr Ile Ala Gln Asp Tyr Pro
    210                 215                 220

Asp Thr Asn Trp Arg Trp Gln Tyr Ser Pro Glu Ser Phe Thr Gly Thr
225                 230                 235                 240

Glu Val Glu Tyr Ala Lys Glu Val Val Asp Ala Val Val Glu Val Met
                245                 250                 255

Asp Pro Thr Pro Glu Asn Pro Met Ile Ile Asn Leu Pro Ser Thr Val
            260                 265                 270

Glu Met Ile Thr Pro Asn Val Tyr Ala Asp Ser Ile Glu Trp Met His
        275                 280                 285

Arg Asn Leu Asn Arg Arg Asp Ser Ile Ile Leu Ser Leu His Pro His
    290                 295                 300

Asn Asp Arg Gly Thr Gly Val Gly Ala Ala Glu Leu Gly Tyr Met Ala
305                 310                 315                 320

Gly Ala Asp Arg Ile Glu Gly Cys Leu Phe Gly Asn Gly Glu Arg Thr
                325                 330                 335

Gly Asn Val Cys Leu Val Thr Leu Ala Leu Asn Met Leu Thr Gln Gly
            340                 345                 350

Val Asp Pro Gln Leu Asp Phe Thr Asp Ile Arg Gln Ile Arg Ser Thr
        355                 360                 365

Val Glu Tyr Cys Asn Gln Leu Arg Val Pro Glu Arg His Pro Tyr Gly
    370                 375                 380

Gly Asp Leu Val Phe Thr Ala Phe Ser Gly Ser His Gln Asp Ala Val
385                 390                 395                 400

Asn Lys Gly Leu Asp Ala Met Ala Ala Lys Val Gln Pro Gly Ala Ser
                405                 410                 415

Ser Thr Glu Val Ser Trp Glu Gln Leu Arg Asp Thr Glu Trp Glu Val
            420                 425                 430

Pro Tyr Leu Pro Ile Asp Pro Lys Asp Val Gly Arg Asp Tyr Glu Ala
        435                 440                 445

Val Ile Arg Val Asn Ser Gln Ser Gly Lys Gly Gly Val Ala Tyr Ile
    450                 455                 460

Met Lys Thr Asp His Gly Leu Gln Ile Pro Arg Ser Met Gln Val Glu
465                 470                 475                 480

Phe Ser Thr Val Val Gln Asn Val Thr Asp Ala Glu Gly Gly Glu Val
                485                 490                 495

Asn Ser Lys Ala Met Trp Asp Ile Phe Ala Thr Glu Tyr Leu Glu Arg
            500                 505                 510

Thr Ala Pro Val Glu Gln Ile Ala Leu Arg Val Glu Asn Ala Gln Thr
```

-continued

```
        515                 520                 525

Glu Asn Glu Asp Ala Ser Ile Thr Ala Glu Leu Ile His Asn Gly Lys
    530                 535                 540

Asp Val Thr Val Asp Gly Arg Gly Asn Gly Pro Leu Ala Ala Tyr Ala
545                 550                 555                 560

Asn Ala Leu Glu Lys Leu Gly Ile Asp Val Glu Ile Gln Glu Tyr Asn
                565                 570                 575

Gln His Ala Arg Thr Ser Gly Asp Asp Ala Glu Ala Ala Ala Tyr Val
            580                 585                 590

Leu Ala Glu Val Asn Gly Arg Lys Val Trp Gly Val Gly Ile Ala Gly
        595                 600                 605

Ser Ile Thr Tyr Ala Ser Leu Lys Ala Val Thr Ser Ala Val Asn Arg
    610                 615                 620

Ala Leu Asp Val Asn His Glu Ala Val Leu Ala Gly Gly Val Ala Ala
625                 630                 635                 640

Asn Asp Glu Asn Tyr Ala Leu Ala Gly
                645
```

<210> SEQ ID NO 48
<211> LENGTH: 963
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 48

```
atgaccacga agaaagctga ttacatttgg ttcaatgggg agatggttcg ctgggaagac     60

gcgaaggtgc atgtgatgtc gcacgcgctg cactatggca cttcggtttt tgaaggcatc    120

cgttgctacg actcgcacaa aggaccggtt gtattccgcc atcgtgagca tatgcagcgt    180

ctgcatgact ccgccaaaat ctatcgcttc ccggtttcgc agagcattga tgagctgatg    240

gaagcttgtc gtgacgtgat ccgcaaaaac aatctcacca gcgcctatat ccgtccgctg    300

atcttcgtcg gtgatgttgg catgggagta aacccgccag cgggatactc aaccgacgtg    360

attatcgctg ctttcccgtg gggagcgtat ctgggcgcag aagcgctgga gcaggggatc    420

gatgcgatgg tttcctcctg gaaccgcgca gcaccaaaca ccatcccgac ggcggcaaaa    480

gccggtggta actacctctc ttccctgctg gtgggtagcg aagcgcgccg ccacggttat    540

caggaaggta tcgcgctgga tgtgaacggt tatatctctg aaggcgcagg cgaaaacctg    600

tttgaagtga aagatggtgt gctgttcacc ccaccgttca cctcctccgc gctgccgggt    660

attacccgtg atgccatcat caaactggcg aaagagctgg gaattgaagt acgtgagcag    720

gtgctgtcgc gcgaatccct gtacctggcg gatgaagtgt ttatgtccgg tacggcggca    780

gaaatcacgc cagtgcgcag cgtagacggt attcaggttg gcgaaggccg ttgtggcccg    840

gttaccaaac gcattcagca agccttcttc ggcctcttca ctggcgaaac cgaagataaa    900

tggggctggt tagatcaagt taatcaagca gctaacgatg aaaattatgc tctggtggct    960

taa                                                                  963
```

<210> SEQ ID NO 49
<211> LENGTH: 828
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 49

```
atgaaaccga ccaccatctc cttactgcag aagtacaaac aggaaaaaaa acgtttcgcg     60

accatcaccg cttatgacta tagcttcgcc aaactctttg ctgatgaagg gcttaacgtc    120
```

-continued

atgctggtgg gcgattcgct gggcatgacg gttcaggggc acgactccac cctgccagtt 180

accgttgccg atatcgccta ccacactgcc gccgtacgtc gcggcgcacc aaactgcctg 240

ctgctggctg acctgccgtt tatggcgtat gccacgccgg aacaagcctt cgaaaacgcc 300

gcaacggtta tgcgtgccgg tgctaacatg gtcaaaattg aaggcggtga gtggctggta 360

gaaaccgtac aaatgctgac cgaacgtgcc gttcctgtat gtggtcactt aggtttaaca 420

ccacagtcag tgaatatttt cggtggctac aaagttcagg ggcgcggcga tgaagcgggc 480

gatcaactgc tcagcgatgc attagcctta gaagctgctg gggcacagct gctggtgctg 540

gaatgcgtgc cggttgaact ggcaaaacgt attaccgaag cactggcgat cccggttatt 600

ggcattggcg caggcaacgt cactgacggg cagatcctcg tgatgcacga cgcctttggt 660

attaccggcg gtcacattcc taaattcgct aaaaatttcc tcgccgaaac gggcgacatc 720

cgcgcggctg tgcggcagta tatggctgaa gtggagtccg gcgtttatcc gggcgaagaa 780

cacagtttcc atgcagctaa cgatgaaaat tatgctctgg gcggctaa 828

<210> SEQ ID NO 50
<211> LENGTH: 1605
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 50

atgagccagc aagtcattat ttcgatacc acattgcgcg acggtgaaca ggcgttacag 60

gcaagcttga gtgtgaaaga aaaactgcaa attgcgctgg cccttgagcg tatgggtgtt 120

gacgtgatgg aagtcggttt ccccgtctct tcgccgggcg attttgaatc ggtgcaaacc 180

atcgcccgcc aggttaaaaa cagccgcgta tgtgcgttag ctcgctgcgt ggaaaaagat 240

atcgacgtgg cggccgaatc cctgaaagtc gccgaagcct tccgtattca tacctttatt 300

gccacttcgc caatgcacat cgccaccaag ctgcgcagca cgctggacga ggtgatcgaa 360

cgcgctatct atatggtgaa acgcgcccgt aattacaccg atgatgttga attttcttgc 420

gaagatgccg ggcgtacacc cattgccgat ctggcgcgag tggtcgaagc ggcgattaat 480

gccggtgcca ccaccatcaa cattccggac accgtgggct acaccatgcc gtttgagttc 540

gccggaatca tcagcggcct gtatgaacgc gtgcctaaca tcgacaaagc cattatctcc 600

gtacataccc acgacgattt gggcctggcg gtcggaaact cactggcggc ggtacatgcc 660

ggtgcacgcc aggtggaagg cgcaatgaac gggatcggcg agcgtgccgg aaactgttcc 720

ctggaagaag tcatcatggc gatcaaagtt cgtaaggata ttctcaacgt ccacaccgcc 780

attaatcacc aggagatatg gcgcaccagc cagttagtta gccagatttg taatatgccg 840

atcccggcaa acaaagccat tgttggcagc ggcgcattcg cacactcctc cggtatacac 900

caggatggcg tgctgaaaaa ccgcgaaaac tacgaaatca tgacaccaga atctattggt 960

ctgaaccaaa tccagctgaa tctgacctct cgttcggggc gtgcggcggt gaaacatcgc 1020

atggatgaga tggggtataa agaaagtgaa tataatttag acaatttgta cgatgctttc 1080

ctgaagctgg cggacaaaaa aggtcaggtg tttgattacg atctggaggc gctggccttc 1140

atcggtaagc agcaagaaga gccggagcat ttccgtctgg attacttcag cgtgcagtct 1200

ggctctaacg atatcgccac cgccgccgtc aaactggcct gtggcgaaga agtcaaagca 1260

gaagccgcca acggtaacgg tccggtcgat gccgtctatc aggcaattaa ccgcatcact 1320

gaatataacg tcgaactggt gaaatacagc ctgaccgcca aaggccacgg taaagatgcg 1380

-continued

ctgggtcagg tggatatcgt cgctaactac aacggtcgcc gcttccacgg cgtcggcctg 1440

gctaccgata ttgtcgagtc atctgccaaa gccatggtgc acgttctgaa caatatctgg 1500

cgtgccgcag aagtcgaaaa agagttgcaa cgcaaagctc aacacaacga aaacaacaag 1560

gaaaccgtgg cagctaacga tgaaaattat gctctggctg gctga 1605

<210> SEQ ID NO 51
<211> LENGTH: 320
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 51

Met Thr Thr Lys Lys Ala Asp Tyr Ile Trp Phe Asn Gly Glu Met Val
1 5 10 15

Arg Trp Glu Asp Ala Lys Val His Val Met Ser His Ala Leu His Tyr
20 25 30

Gly Thr Ser Val Phe Glu Gly Ile Arg Cys Tyr Asp Ser His Lys Gly
35 40 45

Pro Val Val Phe Arg His Arg Glu His Met Gln Arg Leu His Asp Ser
50 55 60

Ala Lys Ile Tyr Arg Phe Pro Val Ser Gln Ser Ile Asp Glu Leu Met
65 70 75 80

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

Ile Arg Pro Leu Ile Phe Val Gly Asp Val Gly Met Gly Val Asn Pro
100 105 110

Pro Ala Gly Tyr Ser Thr Asp Val Ile Ile Ala Ala Phe Pro Trp Gly
115 120 125

Ala Tyr Leu Gly Ala Glu Ala Leu Glu Gln Gly Ile Asp Ala Met Val
130 135 140

Ser Ser Trp Asn Arg Ala Ala Pro Asn Thr Ile Pro Thr Ala Ala Lys
145 150 155 160

Ala Gly Gly Asn Tyr Leu Ser Ser Leu Leu Val Gly Ser Glu Ala Arg
165 170 175

Arg His Gly Tyr Gln Glu Gly Ile Ala Leu Asp Val Asn Gly Tyr Ile
180 185 190

Ser Glu Gly Ala Gly Glu Asn Leu Phe Glu Val Lys Asp Gly Val Leu
195 200 205

Phe Thr Pro Pro Phe Thr Ser Ser Ala Leu Pro Gly Ile Thr Arg Asp
210 215 220

Ala Ile Ile Lys Leu Ala Lys Glu Leu Gly Ile Glu Val Arg Glu Gln
225 230 235 240

Val Leu Ser Arg Glu Ser Leu Tyr Leu Ala Asp Glu Val Phe Met Ser
245 250 255

Gly Thr Ala Ala Glu Ile Thr Pro Val Arg Ser Val Asp Gly Ile Gln
260 265 270

Val Gly Glu Gly Arg Cys Gly Pro Val Thr Lys Arg Ile Gln Gln Ala
275 280 285

Phe Phe Gly Leu Phe Thr Gly Glu Thr Glu Asp Lys Trp Gly Trp Leu
290 295 300

Asp Gln Val Asn Gln Ala Ala Asn Asp Glu Asn Tyr Ala Leu Val Ala
305 310 315 320

<210> SEQ ID NO 52
<211> LENGTH: 275

-continued

<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 52

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

Lys Arg Phe Ala Thr Ile Thr Ala Tyr Asp Tyr Ser Phe Ala Lys Leu
20 25 30

Phe Ala Asp Glu Gly Leu Asn Val Met Leu Val Gly Asp Ser Leu Gly
35 40 45

Met Thr Val Gln Gly His Asp Ser Thr Leu Pro Val Thr Val Ala Asp
50 55 60

Ile Ala Tyr His Thr Ala Ala Val Arg Arg Gly Ala Pro Asn Cys Leu
65 70 75 80

Leu Leu Ala Asp Leu Pro Phe Met Ala Tyr Ala Thr Pro Glu Gln Ala
85 90 95

Phe Glu Asn Ala Ala Thr Val Met Arg Ala Gly Ala Asn Met Val Lys
100 105 110

Ile Glu Gly Gly Glu Trp Leu Val Glu Thr Val Gln Met Leu Thr Glu
115 120 125

Arg Ala Val Pro Val Cys Gly His Leu Gly Leu Thr Pro Gln Ser Val
130 135 140

Asn Ile Phe Gly Gly Tyr Lys Val Gln Gly Arg Gly Asp Glu Ala Gly
145 150 155 160

Asp Gln Leu Leu Ser Asp Ala Leu Ala Leu Glu Ala Ala Gly Ala Gln
165 170 175

Leu Leu Val Leu Glu Cys Val Pro Val Glu Leu Ala Lys Arg Ile Thr
180 185 190

Glu Ala Leu Ala Ile Pro Val Ile Gly Ile Gly Ala Gly Asn Val Thr
195 200 205

Asp Gly Gln Ile Leu Val Met His Asp Ala Phe Gly Ile Thr Gly Gly
210 215 220

His Ile Pro Lys Phe Ala Lys Asn Phe Leu Ala Glu Thr Gly Asp Ile
225 230 235 240

Arg Ala Ala Val Arg Gln Tyr Met Ala Glu Val Glu Ser Gly Val Tyr
245 250 255

Pro Gly Glu Glu His Ser Phe His Ala Ala Asn Asp Glu Asn Tyr Ala
260 265 270

Leu Gly Gly
275

<210> SEQ ID NO 53
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 53

Met Ser Gln Gln Val Ile Ile Phe Asp Thr Thr Leu Arg Asp Gly Glu
1 5 10 15

Gln Ala Leu Gln Ala Ser Leu Ser Val Lys Glu Lys Leu Gln Ile Ala
20 25 30

Leu Ala Leu Glu Arg Met Gly Val Asp Val Met Glu Val Gly Phe Pro
35 40 45

Val Ser Ser Pro Gly Asp Phe Glu Ser Val Gln Thr Ile Ala Arg Gln
50 55 60

-continued

```
Val Lys Asn Ser Arg Val Cys Ala Leu Ala Arg Cys Val Glu Lys Asp
65                  70                  75                  80

Ile Asp Val Ala Ala Glu Ser Leu Lys Val Ala Glu Ala Phe Arg Ile
                85                  90                  95

His Thr Phe Ile Ala Thr Ser Pro Met His Ile Ala Thr Lys Leu Arg
            100                 105                 110

Ser Thr Leu Asp Glu Val Ile Glu Arg Ala Ile Tyr Met Val Lys Arg
        115                 120                 125

Ala Arg Asn Tyr Thr Asp Asp Val Glu Phe Ser Cys Glu Asp Ala Gly
    130                 135                 140

Arg Thr Pro Ile Ala Asp Leu Ala Arg Val Val Glu Ala Ala Ile Asn
145                 150                 155                 160

Ala Gly Ala Thr Thr Ile Asn Ile Pro Asp Thr Val Gly Tyr Thr Met
                165                 170                 175

Pro Phe Glu Phe Ala Gly Ile Ile Ser Gly Leu Tyr Glu Arg Val Pro
            180                 185                 190

Asn Ile Asp Lys Ala Ile Ile Ser Val His Thr His Asp Asp Leu Gly
        195                 200                 205

Leu Ala Val Gly Asn Ser Leu Ala Ala Val His Ala Gly Ala Arg Gln
    210                 215                 220

Val Glu Gly Ala Met Asn Gly Ile Gly Glu Arg Ala Gly Asn Cys Ser
225                 230                 235                 240

Leu Glu Glu Val Ile Met Ala Ile Lys Val Arg Lys Asp Ile Leu Asn
                245                 250                 255

Val His Thr Ala Ile Asn His Gln Glu Ile Trp Arg Thr Ser Gln Leu
            260                 265                 270

Val Ser Gln Ile Cys Asn Met Pro Ile Pro Ala Asn Lys Ala Ile Val
        275                 280                 285

Gly Ser Gly Ala Phe Ala His Ser Ser Gly Ile His Gln Asp Gly Val
    290                 295                 300

Leu Lys Asn Arg Glu Asn Tyr Glu Ile Met Thr Pro Glu Ser Ile Gly
305                 310                 315                 320

Leu Asn Gln Ile Gln Leu Asn Leu Thr Ser Arg Ser Gly Arg Ala Ala
                325                 330                 335

Val Lys His Arg Met Asp Glu Met Gly Tyr Lys Glu Ser Glu Tyr Asn
            340                 345                 350

Leu Asp Asn Leu Tyr Asp Ala Phe Leu Lys Leu Ala Asp Lys Lys Gly
        355                 360                 365

Gln Val Phe Asp Tyr Asp Leu Glu Ala Leu Ala Phe Ile Gly Lys Gln
    370                 375                 380

Gln Glu Glu Pro Glu His Phe Arg Leu Asp Tyr Phe Ser Val Gln Ser
385                 390                 395                 400

Gly Ser Asn Asp Ile Ala Thr Ala Ala Val Lys Leu Ala Cys Gly Glu
                405                 410                 415

Glu Val Lys Ala Glu Ala Ala Asn Gly Asn Gly Pro Val Asp Ala Val
            420                 425                 430

Tyr Gln Ala Ile Asn Arg Ile Thr Glu Tyr Asn Val Glu Leu Val Lys
        435                 440                 445

Tyr Ser Leu Thr Ala Lys Gly His Gly Lys Asp Ala Leu Gly Gln Val
    450                 455                 460

Asp Ile Val Ala Asn Tyr Asn Gly Arg Arg Phe His Gly Val Gly Leu
465                 470                 475                 480

Ala Thr Asp Ile Val Glu Ser Ser Ala Lys Ala Met Val His Val Leu
```

-continued

```
                485                 490                 495

Asn Asn Ile Trp Arg Ala Ala Glu Val Glu Lys Glu Leu Gln Arg Lys
            500                 505                 510

Ala Gln His Asn Glu Asn Asn Lys Glu Thr Val Ala Ala Asn Asp Glu
        515                 520                 525

Asn Tyr Ala Leu Ala Gly
    530
```

What is claimed is:

1. A cell comprising:

(a) an engineered protein expressed from a plasmid, wherein the protein comprises a first moiety having a metabolic enzymatic activity engineered to fuse with a second moiety having a degradation tag,

wherein the cell does not contain a functional chromosomal copy of the first moiety,

wherein the cell comprises a mutation or deletion in the chromosomal copy of the first moiety,

wherein expression of the engineered protein is under the control of a regulatory system comprising a regulated promoter, and

(b) a non-native degradation protein selected from an adaptor, unfoldase, or protease.

2. The cell of claim **1**, wherein said second moiety differs from the amino acid sequence of SEQ ID NO: 1 by at most four amino acid substitutions or deletions.

3. The cell of claim **2**, wherein said second moiety comprises the amino acid sequence of any one of SEQ ID NOs: 1-2 and 4-10.

4. The cell of claim **1**, wherein said regulated promoter is selected from the group consisting of a lac operon promoter, a nitrogen-regulated promoter, a quorum sensing promoter, and a temperature-sensitive promoter.

5. The cell of claim **1**, wherein said cell is a microbial cell.

6. The cell of claim **5**, wherein said cell is a bacterial cell.

7. The cell of claim **5**, wherein said cell is a fungal cell.

8. A method for producing a desired product, said method comprising:

(a) culturing in a suitable medium the cell of claim **1** under conditions that allow expression of the engineered protein; and

(b) recovering said desired product from said cell or said medium.

9. The cell of claim **1**, wherein said degradation protein includes one or more of: SspB adaptor protein, ClpX unfoldase, ClpA unfoldase, ClpP protease, and ClpS adaptor.

10. The cell of claim **9**, wherein

the SspB adaptor protein comprises the amino acid sequence of SEQ ID NO: 25 or 27; or

the ClpX unfoldase comprises the amino acid sequence of SEQ ID NO: 22; or the ClpA unfoldase comprises the amino acid sequence of SED ID NO: 23; or the ClpP protease comprises the amino acid sequence of SEQ ID NO: 24; or the ClpS adaptor comprises the amino acid sequence of SEQ ID NO: 26.

11. The cell of claim **1**, wherein said second moiety comprises the sequence of SEQ ID NO: 3 or 11.

12. The cell of claim **1**, wherein expression of the engineered protein under the control of the regulatory system is increased by a regulatory factor added to medium.

13. The cell of claim **1**, wherein expression of the engineered protein under the control of a regulatory system is reduced by reduction of a regulatory factor in medium.

14. The cell of claim **13**, wherein reduced expression of the engineered protein results in a reduction in the amount of the engineered protein in the cell.

15. The cell of claim **1**, wherein expression of the engineered protein is under the control of a regulatory system comprising a conditionally-replicated plasmid.

16. A cell comprising:

an engineered protein expressed from a plasmid, wherein the protein comprises a first moiety having a metabolic enzymatic activity engineered to fuse with a second moiety having a degradation tag,

wherein the cell does not contain a functional chromosomal copy of the first moiety,

wherein the cell comprises a mutation or deletion in the chromosomal copy of the first moiety,

wherein expression of the engineered protein is under the control of a regulatory system comprising a regulated promoter,

wherein said regulated promoter is selected from the group consisting of a lac operon promoter, a nitrogen-regulated promoter, a quorum sensing promoter, and a temperature-sensitive promoter, and

a heterologous nucleic acid encoding a degradation protein.

17. The cell of claim **16**, wherein said second moiety differs from the amino acid sequence of SEQ ID NO: 1 by at most four amino acid substitutions or deletions.

18. The cell of claim **17**, wherein said second moiety comprises the amino acid sequence of any one of SEQ ID NOs: 1-2 and 4-10.

19. The cell of claim **16**, wherein said cell is a microbial cell.

20. The cell of claim **19**, wherein said cell is a bacterial cell.

21. The cell of claim **19**, wherein said cell is a fungal cell.

22. The cell of claim **16**, wherein said wherein said degradation protein includes one or more of: SspB adaptor protein, ClpX unfoldase, ClpP protease, and ClpS adaptor.

23. The cell of claim **22**, wherein

the SspB adaptor protein comprises the amino acid sequence of SEQ ID NO: 25 or 27;

or the ClpX unfoldase comprises the amino acid sequence of SEQ ID NO: 22; or

the ClpA unfoldase comprises the amino acid sequence of SED ID NO: 23; or

the ClpP protease comprises the amino acid sequence of SEQ ID NO: 24; or

the ClpS adaptor comprises the amino acid sequence of SEQ ID NO: 26.

24. The cell of claim **16**, wherein expression of the engineered protein under control of the regulated promoter is increased by a regulatory factor added to medium.

25. The cell of claim **16**, wherein expression of the engineered protein under control of the regulated promoter is reduced by reduction of a regulatory factor in medium.

26. The cell of claim **25**, wherein reduced expression of the engineered protein results in a reduction in the amount of the engineered protein in the cell.

27. The cell of claim **16**, wherein expression of the engineered protein is under the control of a regulatory system comprising a conditionally-replicated plasmid.

* * * * *