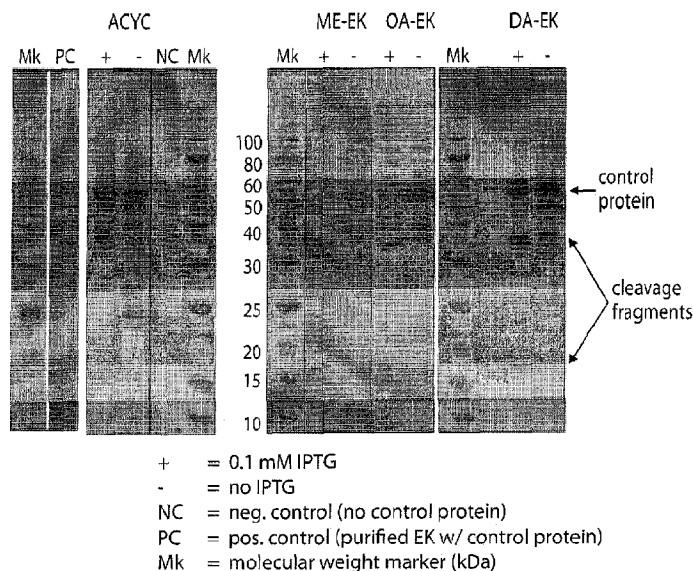




(86) Date de dépôt PCT/PCT Filing Date: 2011/08/31  
(87) Date publication PCT/PCT Publication Date: 2012/03/08  
(45) Date de délivrance/Issue Date: 2021/02/23  
(85) Entrée phase nationale/National Entry: 2013/02/22  
(86) N° demande PCT/PCT Application No.: US 2011/049997  
(87) N° publication PCT/PCT Publication No.: 2012/030980  
(30) Priorité/Priority: 2010/08/31 (US61/378,828)

(51) Cl.Int./Int.Cl. *C12N 15/63* (2006.01)  
(72) Inventeur/Inventor:  
SWARTZ, JAMES R., US  
(73) Propriétaires/Owners:  
GREENLIGHT BIOSCIENCES, INC., US;  
THE BOARD OF TRUSTEES OF THE LELAND  
STANFORD JUNIOR UNIVERSITY, US  
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : PROCÉDES POUR LA COMMANDE DE FLUX DANS DES VOIES METABOLIQUES PAR MANIPULATION DE PROTEASE  
(54) Title: METHODS FOR CONTROL OF FLUX IN METABOLIC PATHWAYS THROUGH PROTEASE MANIPULATION



(57) **Abrégé/Abstract:**

The embodiments described herein pertain to cells, and methods for preparing cells, that can be used as biocatalysts by altering enzymes that compete for a substrate or product of a pathway of interest such that the targeted enzyme is sensitive to a site-specific protease, which protease is expressed but relocated in the cell to a site where it is not in contact with the targeted enzyme in the intact cell. Upon cell lysis, the protease contacts the target enzyme, which is then inactivated by protease cleavage.

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
8 March 2012 (08.03.2012)

PCT

(10) International Publication Number  
**WO 2012/030980 A1**(51) International Patent Classification:  
C12N 15/63 (2006.01)(74) Agent: BAKER, Hunter, C.; Wolf, Greenfield & Sacks,  
P.C., 600 Atlantic Avenue, Boston, MA 02210-2206  
(US).(21) International Application Number:  
PCT/US2011/049997(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,  
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,  
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,  
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,  
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,  
NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU,  
RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ,  
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA,  
ZM, ZW.(22) International Filing Date:  
31 August 2011 (31.08.2011)

(25) Filing Language: English

(26) Publication Language: English

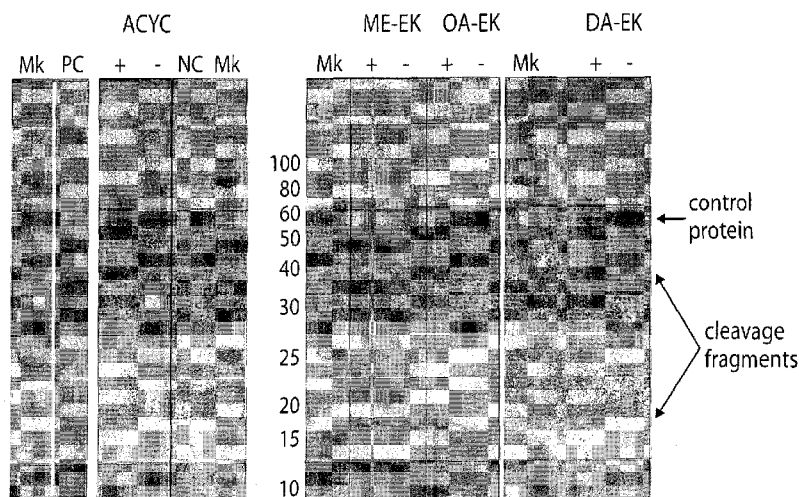
(30) Priority Data:  
61/378,828 31 August 2010 (31.08.2010) US(71) Applicants (for all designated States except US):  
GREENLIGHT BIOSCIENCES, INC. [US/US]; 200  
Boston Avenue, Suite 4100, Medford, MA 02155 (US).  
THE BOARD OF TRUSTEES OF THE LELAND  
STANFORD JUNIOR UNIVERSITY [US/US]; 1705  
El Camino Real, Palo Alto, CA 94306-1106 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): SWARTZ, James, R.  
[US/US]; 1860 White Oak Drive, Menlo Park, CA 94025  
(US).(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG,  
ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,  
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,  
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,  
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,  
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHODS FOR CONTROL OF FLUX IN METABOLIC PATHWAYS THROUGH PROTEASE MANIPULATION



+ = 0.1 mM IPTG  
 - = no IPTG  
 NC = neg. control (no control protein)  
 PC = pos. control (purified EK w/ control protein)  
 Mk = molecular weight marker (kDa)

Fig. 1

(57) Abstract: The embodiments described herein pertain to cells, and methods for preparing cells, that can be used as biocatalysts by altering enzymes that compete for a substrate or product of a pathway of interest such that the targeted enzyme is sensitive to a site-specific protease, which protease is expressed but re-located in the cell to a site where it is not in contact with the targeted enzyme in the intact cell. Upon cell lysis, the protease contacts the target enzyme, which is then inactivated by protease cleavage.

**WO 2012/030980 A1** 

---

**Published:**

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

# **METHODS FOR CONTROL OF FLUX IN METABOLIC PATHWAYS THROUGH PROTEASE MANIPULATION**

## **RELATED APPLICATIONS**

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional application, U.S.S.N. 61/378,828, filed August 31, 2010.

## **BACKGROUND OF THE INVENTION**

[0002] Production of chemicals via synthetic enzymatic pathways in microbial hosts has proven useful for many important classes of molecules, including isoprenoids, polyketides, nonribosomal peptides, bioplastics, and chemical building blocks. Due to the inherent modularity of biological information, synthetic biology holds great potential for expanding this list of microbially produced compounds even further. Yet embedding a novel biochemical pathway in the metabolic network of a host cell can disrupt the subtle regulatory mechanisms that the cell has evolved over the millennia. Indeed, the final yield of a compound is often limited by deleterious effects on the engineered cell's metabolism that are difficult to predict due to the limited understanding of the complex interactions that occur within the cell. The unregulated consumption of cellular resources, metabolic burden of heterologous protein production, and accumulation of pathway intermediates/products that are inhibitory or toxic to the host are all significant issues that may limit overall yield.

[0003] The concept of metabolic engineering has emerged to fulfill this purpose, which can be defined as purposeful modification of metabolic and cellular networks by employing various experimental techniques to achieve desired goals. What distinguishes metabolic engineering from genetic engineering and strain improvement is that it considers metabolic and other cellular network as a whole to identify targets to be engineered. In this sense, metabolic flux is an essential concept in the practice of metabolic engineering. Although gene expression levels and the concentrations of proteins and metabolites in the cell can provide clues to the status of the metabolic network, they have inherent limitations in fully describing the cellular phenotype due to the lack of information on the correlations among these cellular components. Metabolic fluxes represent the reaction rates in metabolic pathways, and serve to integrate these factors through a mathematical framework. Thus, metabolic fluxes can be considered as one way of representing the phenotype of the cell as a result of interplays among various cell components; the observed metabolic flux profiles

reflect the consequences of interconnected transcription, translation, and enzyme reactions incorporating complex regulations.

**[0004]** Cell-free synthesis offers advantages over *in vivo* production methods. Importantly, the lack of a cell wall *in vitro* allows for direct control of the synthesis environment. The redox potential, pH, or ionic strength can be altered directly, and over more broad, non-physiological ranges, due to lack of cell growth or need for cell viability. Furthermore, direct recovery of products can more easily be achieved. Cell-free systems can direct most, if not all, of the metabolic resources of the cell towards the exclusive production from one pathway. The present compositions and methods address these issues.

### SUMMARY OF THE INVENTION

**[0005]** Systems, compositions, and methods are provided for controlling metabolic pathway flux through modification of targeted enzymes that compete with enzymes involved in a pathway of interest, including modification to maintain or alter cellular concentrations of the targeted enzymes during a cell growth phase, followed by manipulation to decrease concentrations of the targeted enzymes during a production phase, when the product of the pathway of interest is produced. The cell growth phase necessarily involves intact cells, while the production phase may be performed with lysates of such cells.

**[0006]** In the systems, compositions, and methods provided herein, target enzymes (*e.g.*, enzymes that compete for substrates of enzymes in the pathway of interest) are targeted for cleavage by a site-specific protease. Such targeted enzymes are undesirable during production phase of the metabolic pathway. The targeted enzyme(s) may be genetically modified to comprise at least one cleavage site (also referred to herein as a site-specific protease recognition site) for a site-specific protease or may have a naturally-occurring cleavage site. The cleavage site is typically located such that native enzyme activity is minimally affected and that when the targeted enzyme is cleaved, it is inactivated.

**[0007]** The corresponding site-specific protease, *i.e.*, the protease that recognizes the cleavage site on the targeted enzyme, is genetically modified to incorporate a peptide sequence that provides for relocation of the site-specific protease to a cellular or extra-cellular compartment other than the naturally-occurring compartment. When the site-specific protease is relocated during the cell growth phase of the cell cycle, it is not in contact with, and thus is not able to cleave, the targeted enzyme. When the cell is lysed prior to the metabolite production phase, the site-specific protease is brought into contact with and cleaves the targeted enzyme, thereby inactivating the targeted enzyme.

**[0008]** In some embodiments, the site-specific protease is sequestered in the periplasm. In other embodiments, the site-specific protease is secreted into the growth medium.

**[0009]** The site-specific protease may be endogenous to the cell, or the coding sequence of the site-specific protease may be introduced into the cell for expression.

**[0010]** In one aspect, provided here are cells genetically modified to comprise (i) a site-specific protease coding sequence comprising a relocation sequence; and (ii) a coding sequence for at least one targeted enzyme, which competes with an enzyme involved in a pathway of interest, wherein the targeted enzyme coding sequence comprises a site-specific protease recognition site. In some embodiments, the site-specific protease coding sequence is operably linked to an inducible promoter.

**[0011]** In some embodiments, the relocation sequence is a periplasmic targeting sequence. In some embodiments, the periplasmic targeting sequence is selected from the group consisting of:

MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1),  
 MKKTAIAIAVALAGFATVAQA (SEQ ID NO:2)  
 MKQSTIALALLPLLFTPVTKA (SEQ ID NO:3),  
 MMITLRKLPLAVAVAAGVMSAQAMA (SEQ ID NO:4),  
 MNKKVLTLSAVMASMLFGAAHA (SEQ ID NO:5) ,  
 MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:6),  
 MKKNIAFLASMFVFSIATNAYA (SEQ ID NO:7)  
 MKKIWLALAGLVLAFSASA (SEQ ID NO:8),  
 MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:9),  
 MKQALRVAFGFLILWASVLHA (SEQ ID NO:10),and  
 MRVLLFLLLSLFMLPAFS (SEQ ID NO:11).

**[0012]** In some embodiments, the site-specific protease coding sequence encodes a site-specific protease selected from the group consisting of TEV NIa, HRV 3C, Enterokinase, Factor Xa, and Thrombin.

**[0013]** In certain embodiments, the targeted enzyme coding sequence is genetically modified to comprise the site-specific protease recognition site.

**[0014]** In some embodiments, the targeted enzyme competes with an enzyme that increases the rate of precursor supplied to the pathway of interest. In some embodiments, the targeted enzyme competes with a key pathway entry enzyme.

**[0015]** In certain embodiments, the gene encoding the native counterpart enzyme is replaced with the DNA sequence coding for the targeted enzyme. In certain embodiments,

the gene encoding the native counterpart enzyme is knocked out (*e.g.*, inactivated or replaced by an inactive DNA molecule). In some embodiments, the targeted enzyme is over-expressed in the cell.

**[0016]** In particular embodiments, the site-specific protease or the targeted enzyme is present on either an episomal vector or a chromosome.

**[0017]** In certain embodiments, the pathway of interest is the synthesis of: a) an antibiotic; b) a biosurfactant; c) an amino acid; d) an organic acid; e) a fatty acid; f) an alcohol or polyol; g) a flavor or fragrance; h) a nucleotide; i) a vitamin; j) a pigment; k) a sugar or polysaccharide; l) a biopolymer or plastic; m) an *E. coli* metabolite; n) shikimic acid and/or shikimate; o) an isoprenoid or terpene; p) poly-3-hydroxybutyrate; or q) isobutanol and/or 1-butanol.

**[0018]** In some embodiments, the pathway of interest is the synthesis of shikimic acid and/or shikimate, an isoprenoid or terpene, poly-3-hydroxybutyrate, isobutanol, and/or 1-butanol.

**[0019]** In some embodiments, the cell growth medium has been modified by the addition or enhancement of a factor that increases or preserves the activity of the targeted enzyme.

**[0020]** In one aspect, provided herein are cell lysates of any of the cells described herein.

**[0021]** In another aspect, provided herein are systems for producing a product of a pathway of interest, the systems comprising one or more lysates of any of the cells described herein, and optionally one or more substrates, enzymes, nutrients, co-factors, buffers, reducing agents, and ATP generating systems.

**[0022]** In yet another aspect, provided herein are methods of producing a product of a pathway of interest, the method comprising growing any one of the cells described herein to a desired cell density, lysing at least a portion of the cells, and optionally combining the lysate with one or more substrates, enzymes, nutrients, co-factors, buffers, reducing agents, and/or ATP generating systems, wherein the enzymes in the pathway of interest catalyze production of the product and the site-specific protease cleaves the targeted enzyme. In some embodiments, the method further comprises combining the lysate with one or more additional cell lysates.

**[0023]** In still another aspect, provided herein are vectors that comprise a site-specific protease coding sequence comprising a relocation sequence, wherein, optionally, the coding sequence is operably linked to an inducible promoter.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0024]** *Figure 1* depicts images of Coomassie stained poly-acrylamide gel electrophoresis (PAGE) gels of samples taken after cleavage incubations using cell extracts obtained from control cultures and from site-specific protease expression cultures with and without induction of site-specific protease expression. The cultures being ACYC, ME-EK, OA-EK and DA-EK, are, Abbreviations: pACYC-Duet vector control (ACYC); MalE-enterokinase, SEQ ID NO:19 (ME-EK); OmpA-enterokinase, SEQ ID NO:18 (OA-EK); DsbA-enterokinase, SEQ ID NO:17 (DA-EK); and isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG).

**[0025]** *Figure 2* depicts images of an immunoblot of samples taken after cleavage incubations using cell extracts obtained from control cultures, from periplasmically-targeted site-specific enterokinase cultures, as well as from purified recombinant enterokinase.

## DETAILED DESCRIPTION OF THE EMBODIMENTS

**[0026]** The systems, compositions, and methods described herein are based on the idea that genetically modified cells can be engineered for controlling metabolic pathway flux through genetic modification of targeted enzymes that compete with enzymes involved in a pathway of interest, including genetic modification to maintain or alter cellular concentrations of the targeted enzymes during a cell growth phase, followed by manipulation to decrease concentrations of the active targeted enzymes during the production phase, when the product of the pathway of interest is produced. During the production phase, targeted enzyme inactivation results from cleavage of the enzyme by a site-specific protease. The site-specific protease may be an endogenous or exogenous protease. In certain embodiments, the targeted enzyme is genetically modified to comprise a motif (*e.g.*, a site-specific protease recognition site) that is cleaved by a site-specific protease.

**[0027]** In some embodiments, expression of the site-specific protease is inducible. In certain embodiments, the site-specific protease is genetically modified to be relocated to a cellular or extra-cellular compartment (*e.g.*, a cellular or extra-cellular compartment) where the site-specific protease does not naturally occur, such that the site-specific protease does not inactivate the targeted enzyme when thus relocated. In certain embodiments, the site-specific protease is genetically modified to include a peptide sequence that provides for periplasmic targeting of the protease, that is where the protease is sequestered in the periplasm of the cell.

[0028] In certain embodiments, the targeted enzyme is one that competes with an enzyme involved in a pathway of interest. In some embodiments, the targeted enzyme competes with a pathway entry enzyme. In certain embodiments, the targeted enzyme competes with is a rate-limiting enzyme. In certain embodiments, the targeted enzyme competes with an enzyme that increases the rate of precursor supplied to the pathway of interest or that supplies another required substrate or cofactor.

**[0029]** In some embodiments, the gene encoding the native counterpart enzyme is replaced with the DNA sequence coding for the targeted enzyme. In certain embodiments, the gene encoding the native counterpart enzyme is knocked out (*e.g.*, inactivated or replaced by artificial DNA). In certain embodiments, the targeted enzyme is over-expressed (*e.g.*, enzyme expression is increased above physiologically-normal enzyme expression levels) in the cell.

**[0030]** In certain embodiments, the targeted enzyme and/or the site-specific protease is present on either an episomal vector or a chromosome.

**[0031]** In certain embodiments, the cell growth medium has been modified by the addition or enhancement of a factor (*e.g.*, a nutrient, co-factor, reducing agent) that increases or preserves the activity of the targeted enzyme.

**[0032]** In one aspect, provided is a genetically modified cell, which comprises a site-specific protease modified to comprise a relocation sequence, *e.g.*, a periplasmic targeting sequence. The cell may further comprise a targeted enzyme that has been genetically modified to introduce a recognition sequence for the site-specific protease, which inactivates the targeted enzyme upon cleavage.

**[0033]** In another aspect, lysates of such a genetically modified cell are provided. The lysate may be combined with one or more of substrate, nutrients, cofactors, buffers, reducing agents, ATP generating systems, and/or other agents to generate a cell-free system for producing a product of interest.

**[0034]** In another aspect, provided are methods for producing a product of interest, the method comprising growing cells that express a site-specific protease genetically modified to incorporate a peptide sequence that provides for its relocation, wherein a targeted enzyme comprising a recognition sequence for the site-specific protease is expressed in the cytoplasm of the cell, lysing the cells to bring the site-specific protease into contact with the targeted enzyme, and producing the product of the pathway in a cell-free system comprising the lysate. Additional substrate, nutrients, cofactors, buffers, reducing agents, ATP generating systems, and/or other agents, may be added to the cell-free system. In one embodiment the product of interest is produced at a high flux rate.

**[0035]** In yet another aspect, provided are methods for producing a product of interest, the method comprising growing cells that are genetically modified to over-express at least one periplasmic targeted site-specific protease, lysing the cells, and producing the product of the pathway of interest in a cell-free system comprising the lysate. Additional substrate,

nutrients, cofactors, buffers, reducing agents, ATP generating systems, and/or other agents may be added to the cell-free system.

#### *Production Methods*

**[0036]** High yield production of a product of interest is accomplished by providing a cell in which cytoplasmic enzymes comprising a pathway of interest are expressed, *e.g.*, at physiologically normal levels, or at greater than physiologically normal levels; and where at least one targeted enzyme involved in the pathway is (a) expressed at normal levels and (b) comprises an inactivating recognition for a site-specific protease. In some embodiments, the site-specific protease is expressed in the cell but relocated to a site where it does not contact the target enzyme.

**[0037]** During cell culture it may be desirable to control the components of the growth medium of the cells in order to avoid exposure of the periplasmic sequestered protease to conditions that affect activity, *e.g.*, exposure to metals and the like. For example, it has been found that DAHP synthase in the shikimic acid pathway can be inactivated through copper-catalyzed oxidation, and thus it is desirable to modify the culture conditions by increasing the concentration of co-factors, such as manganese and magnesium metals, in the growth medium to outcompete available copper (see, *e.g.*, Bauerle *et al.*, *J. Bacteriol.* (1999) 181:1636-1642; and Stadtman *et al.*, *J. Biol. Chem.* (1991) 266:2005-2008.).

Thus, in certain embodiments, cofactor(s) are provided or concentrations of co-factor(s) are altered in the growth medium to enhance enzyme activation in the periplasm or other relocated enzyme site.

**[0038]** For production purposes, a lysate of the cell is utilized, wherein the periplasmically sequestered site-specific protease is brought into operable contact with the enzymes of the pathway of interest expressed in the cytoplasm. Cells are lysed by any convenient method that substantially maintains enzyme activity, *e.g.*, sonication, French press, and the like, as known in the art. The lysate may be fractionated, particulate matter spun out, or may be used in the absence of additional processing steps. The cell lysate may be further combined with one or more substrates, enzymes, nutrients, co-factors, buffers, reducing agents, and/or ATP generating systems, as required for enzyme activity. Such a system, in certain embodiments, may be referred to herein as a "cell-free system," *e.g.*, an isolated system containing a cell lysate or extract expressly engineered to synthesize an enzyme or cascade of enzymes that, when acting in a given sequence (*e.g.*, in an enzymatic pathway) and proportion over a determined substrate, results in the preferential generation of a product, the compound of

interest. A compound of interest may be a chemical entity (*e.g.*, a small organic molecule), which can be used as an active pharmaceutical ingredient (API), chemical precursor, or intermediate.

**[0039]** As used herein, a “substrate” is a compound or mixture of compounds capable of providing the required elements needed to synthesize a compound of interest.

**[0040]** As used herein, a “small organic molecule” or “small molecule” refers to an organic molecule with a molecular weight of less than 800 g/mol (*e.g.*, less than 700 g/mol, less than 600 g/mol, less than 500 g/mol, less than 400 g/mol, less than 300 g/mol, less than 200 g/mol, less than 100 g/mol, between 50 to 800 g/mol, inclusive, between 100 to 800 g/mol, inclusive, or between 100 to 500 g/mol, inclusive). In certain embodiments, the small organic molecule is a therapeutically active agent such as a drug (*e.g.*, a small organic molecule approved by the U.S. Food and Drug Administration as provided in the Code of Federal Regulations (CFR)). The small organic molecule may also comprise a metal. In this instance, the small organic molecule is also referred to as an “small organometallic molecule.”

**[0041]** As used herein, a “reducing equivalent” or “reducing agent” is a chemical species which transfers the equivalent of one electron in a redox reaction. Examples of reducing equivalents are a lone electron (for example in reactions involving metal ions), a hydrogen atom (consisting of a proton and an electron), and a hydride ion (:H<sup>-</sup>) which carries two electrons (for example in reactions involving nicotinamide adenine dinucleotide (NAD<sup>+</sup>)). A “reducing equivalent acceptor” is a chemical species that accepts the equivalent of one electron in a redox reaction.

**[0042]** As used herein, an “adenosine triphosphate regeneration system” or “ATP regeneration system” is a chemical or biochemical system that converts adenosine, AMP, and ADP into ATP. Examples of ATP regeneration systems include those involving glucose metabolism, glutamate metabolism, and photosynthesis.

**[0043]** Lysates of cells of different genetic backgrounds (*e.g.* previously altered or genetically engineered) or species, or that are prepared by different strategies can be mixed and simultaneously or sequentially used in a process with the cell lysate. The lysate can be free or immobilized, and can be reused or disposed of at each stage of the process. For example, in certain embodiments, the cell lysate is a lysate of an *E. coli* organism engineered to overexpress one or more enzymes in the pathway of interest. In certain embodiments, the cell lysate is a combination of different cell lysates, *e.g.*, a combination of two, three, four, five, six, seven, eight, nine, or ten different cell lysates, obtained from two, three, four, five,

six, seven, eight, nine, or ten different *E. coli* organisms each engineered to overexpress one or more enzymes in the pathway of interest.

**[0044]** The methods described herein provide for high yields of the desired product, which yield is greater than the yield that can be achieved with a native microbial host. Productivity (*i.e.*, rate of production per unit of volume or biomass) may also be increased. In one embodiment, the yield of product is at least about 2-fold above the basal rate, at least about 5-fold above the basal rate, at least about 10-fold above the basal rate, at least about 25-fold above the basal rate, at least about 50-fold above the basal rate, or more, *e.g.*, between about 2-fold to about 100-fold above the basal rate. In certain embodiments, the rate of yield of the product using the inventive methods is between about 0.1 to 20 grams of product/L/h.

**[0045]** Different inocula can be adapted to different conditions (*e.g.*, two batches grown on two different carbon sources) or can have different genotypes and then mixed to carry out the process (*e.g.*, to get simultaneous consumption of a mix of carbon sources or sequential processing of a metabolite through a pathway divided into two separate batches of cells). A process can also take place sequentially by allowing one set of reactions to proceed in one vessel and then transferring the supernatant to a second vessel.

**[0046]** The reactions may utilize a large scale reactor, small scale, or may be multiplexed to perform a plurality of simultaneous syntheses. Continuous reactions will use a feed mechanism to introduce a flow of reagents, and may isolate the end-product as part of the process. Batch systems are also of interest, where additional reagents may be introduced over time to prolong the period of time for active synthesis. A reactor may be run in any mode such as batch, extended batch, semi-batch, semi-continuous, fed-batch, and continuous, and which will be selected in accordance with the application purpose.

**[0047]** The reactions may be of any volume, either in a small scale (*e.g.*, usually at least about 1 ml and not more than about 15 ml) or in a scaled up reaction (*e.g.*, where the reaction volume is at least about 15 ml, usually at least about 50 ml, more usually at least about 100 ml, and may be 500 ml, 1000 ml, or greater up to many thousands of liters of volume). Reactions may be conducted at any scale.

**[0048]** Various salts and buffers may be included, where ionic species may be optimized with regard to product production. When changing the concentration of a particular component of the reaction medium another component may be changed accordingly. Also, the concentration levels of components in the reactor may be varied over time. The adjuster of the thiol/disulfide oxidation/reduction potential may be dithiothreitol, ascorbic acid,

glutathione, and/or their oxidized forms. Other adjusters of the general redox potential may also be used.

**[0049]** In a semi-continuous operation mode, the reactor may be operated in dialysis, diafiltration batch or fed-batch mode. A feed solution may be supplied to the reactor through the same membrane or a separate injection unit. Synthesized product is accumulated in the reactor, and then is isolated and purified according to the usual method for purification after completion of the system operation. Alternatively, product can be removed during the process either in a continuous or discontinuous mode with the option of returning part or all of the remaining compounds to the reactor.

**[0050]** Where there is a flow of reagents, the direction of liquid flow can be perpendicular and/or tangential to a membrane. Tangential flow is effective for recycling ATP and for preventing membrane plugging and may be superimposed on perpendicular flow. Flow perpendicular to the membrane may be caused or effected by a positive pressure pump or a vacuum suction pump or by applying trans-membrane pressure using other methods known in the art. The solution in contact with the outside surface of the membrane may be cyclically changed, and may be in a steady tangential flow with respect to the membrane. The reactor may be stirred internally or externally.

**[0051]** The amount of product produced in a reaction can be measured in various ways, for example, by enzymatic assays which produce a colored or fluorometric product or by high-performance liquid chromatography (HPLC) methods. In certain embodiments, the product is measured utilizing an assay which measures the activity or concentration of the particular product being produced.

#### *Pathways of Interest*

**[0052]** As used herein, the term “enzyme pathway” or “pathway of interest” refers to a cellular or acellular (cell-free) system for converting a substrate to a product of interest, where the system comprises a plurality of enzymes and may additionally comprise substrates acted upon by one or more of the enzymes, products of the enzyme-catalyzed reactions, co-factors utilized by the enzymes, and the like. The system may be present in an intact cell, or in a lysate of a cell. Many metabolic pathways are known and have been described in microbial systems, and are accessible in public databases; see, *e.g.*, Smolke, Ed., *The Metabolic Pathway Engineering Handbook: Tools and Applications*, CRC Press, New York (2009); Stephanopoulos, Nielsen, and Aristidou, Eds., *Metabolic Engineering: Principles and Methodology*, Academic Press, New York (1998); Greenberg, *Metabolic Pathways:*

*Energetics, Tricarboxylic Acid Cycle, and Carbohydrates*, Academic Press, New York (1967); and D.M. Greenberg's multi-volume series entitled *Metabolic pathways*, Volumes 1-7.

[0053] Pathways of interest include, for example, pathways involved in carbohydrate, amino acid, nucleic acid, steroid, fatty acid, and natural product biosynthesis, and encompass the synthesis of various chemical compounds and materials, including, but not limited to:

- a) antibiotics; *e.g.*, actinomycin, bleomycin, rifamycin, chloramphenicol, carbapenems, tetracycline, lincomycin, erythromycin, streptomycin, cyclohexamide, puromycin, cycloserine, bacitracin, penicillin, cephalosporin, vancomycin, polymyxin, and gramicidin;
- b) biosurfactants; *e.g.*, rhamnolipids, sophorolipids, glycolipids, and lipopeptides;
- c) biological fuels; *e.g.*, bioethanol, biodiesel, and biobutanol;
- d) amino acids; *e.g.*, L-glutamate, L-lysine, L-phenylalanine, L-aspartic acid, L-isoleucine, L-valine, L-tryptophan, L-proline (hydroxyproline), L-threonine, L-methionine, L-tyrosine, and D-p-hydroxyphenylglycine;
- e) organic acids; *e.g.*, citric acid, lactic acid, gluconic acid, acetic acid, propionic acid, succinic acid, fumaric acid, and itaconic acid;
- f) fatty acids; *e.g.*, arachidonic acid, polyunsaturated fatty acid (PUBA), and  $\alpha$ -linoleic acid;
- g) alcohols and polyols; *e.g.*, glycerol, mannitol, erythritol, xylitol, poly-3-hydroxybutyrate, isobutanol, and 1-butanol;
- h) flavors and fragrances; *e.g.*, vanillin, benzaldehyde, dihydroxyacetone, 4-(R)-decanolide, and 2-actyl-1-pyrroline;
- i) nucleotides; *e.g.*, 5'-guanylic acid and 5'-inosinic acid;
- j) vitamins; *e.g.*, vitamin C, vitamin F, vitamin B2, provitamin D2, vitamin B12, folic acid, nicotinamide, biotin, 2-keto-L-gulonic acid, and provitamin Q10;
- k) pigments; *e.g.*, astaxanthin,  $\beta$ -carotene, leucopene, monascorubrin, and rubropunctatin;
- l) sugars and polysaccharides; *e.g.*, ribose, sorbose, xanthan, gellan, and dextran; and
- m) biopolymers and plastics; *e.g.*, polyhydroxyalkanoates (PHA), poly- $\gamma$ -glutamic acid, and 1,3-propanediol.

[0054] Other examples of pathways of interest include the synthesis of various *E. coli* metabolites. A "metabolite" is any substance used or produced during metabolism (*e.g.*, an enzyme, substrate, or product). Herein, a metabolite is often, although not always, the product of an enzyme in the pathway of interest. Exemplary *E. coli* metabolites include, but

are not limited to, 2,3-dihydroxybenzoic acid, 2-ketoglutarate, 3-phosphoglycerate, 4-hydroxybenzoate, 6-phosphogluconate, acetoacetyl-CoA, acetyl-CoA, acetylphosphate, adenine, adenosine, adenosine phosphosulfate, ADP, ADP-glucose, Alanine, AMP, anthranilate, arginine, Asparagine, Aspartate, ATP, carbamylaspartate, cis-aconitate, citrate, citrulline, CMP, coenzyme A, CTP, cyclic AMP, cytidine, cytosine, dAMP, dATP, dCTP, deoxyadenosine, deoxyguanosine, deoxyribose-5-P, dGMP, dihydroorotate, dihydroxyacetone phosphate, dTDP, dTTP, erythrose-4-phosphate, FAD, flavin mononucleotide, fructose-1,6-bisphosphate, fructose-6-phosphate, fumarate, GDP, gluconate, gluconolactone, glucosamine-6-phosphate, glucose-6-phosphate, glucose-1-phosphate, glutamate, glutamine, glutathione, glutathione disulfide, glyceraldehyde-3-phosphate, glycerate, glycerol-3-phosphate, GMP, GTP, guanine, guanosine, histidine, histidinol, homocysteine, inosine diphosphate, inosine monophosphate, inosine triphosphate, isoleucine, lysine, malate, malonyl-CoA, methionine, myo-inositol, N-Acetyl-glucosamine-1P, N-acetyl-ornithine, NAD +, NADH, NADP +, NADPH, ornithine, oxaloacetate, phenylalanine, phenylpyruvate, phosphoenolpyruvate, proline, propionyl-CoA, PRPP, pyruvate, quinolate, riboflavin, ribose-5-phosphate, ribulose-5-phosphate, S-adenosyl-L-methionine, serine, shikimic acid, shikimate, succinate, succinyl-CoA, threonine, tryptophan, tyrosine, UDP, UDP-glucose, UDP-glucuronate, UDP-N-acetylglucosamine, uridine, UTP, valine, and xylulose-5-phosphate.

**[0055]** In certain embodiments, the pathway of interest provides for the synthesis of shikimic acid and/or shikimate (shikimate is the anionic form of shikimic acid) and synthetic intermediates thereto, an isoprenoid or terpene (*e.g.*, amorphadiene, farnesene, lycopene, astaxanthin, vitamin A, menthol, beta-carotene), poly-3-hydroxybutyrate, isobutanol, and 1-butanol.

**[0056]** A number of reactions may be catalyzed by enzymes in a pathway of interest. Broad classes of enzymes, which can be identified by enzyme classification number, provided in parentheses, include, but are not limited to:

(EC 1) oxidoreductases; *e.g.*, dehydrogenases, oxidases, reductases, oxidoreductases, synthases, oxygenases, monooxygenases, dioxygenases, lipooxygenases, hydrogenases, transhydrogenases, peroxidases, catalases, epoxidases, hydroxylases, demethylases, desaturases, dismutases, hydroxyltransferases, dehalogenases, and deiodinases;

(EC2) transferases; *e.g.*, transaminases, kinases, dikinases, methyltransferases, hydroxymethyltransferases, formyltransferases, formiminotransferases, carboxytransferases, carbamoyltransferases, amidinotransferases, transaldolases, transketolases, acetyltransferases,

acyltransferases palmitoyltransferases, succinyltransferases, malonyltransferases, galloyltransferases, sinapoyltransferases, tigloyltransferases, tetradecanoyltransferases, hydroxycinnamoyltransferases, feruloyltransferases, mycolyltransferases, benzoyltransferases, piperoyltransferases, trimethyltridecanoyltransferase, myristoyltransferases, coumaroyltransferases, thiolases, aminoacyltransferases, phosphorylases, hexosyltransferases, pentosyltransferases, sialyltransferases, pyridinylases, diphosphorylases, cyclotransferases, sulfurylases, adenosyltransferases, carboxyvinyltransferases, isopentenyltransferases, aminocarboxypropyltransferases, dimethylallyltransferases, farnesyltransferases, hexaprenyltransferases, decaprenylcistransferases, pentaprenyltransferases, nonaprenyltransferases, geranylgeranyltransferases, aminocarboxypropyltransferases, oximinotransferases, purinetransferases, phosphodismutases, phosphotransferases, nucleotidyltransferases, polymerases, cholinephosphotransferases, phosphorylmutases, sulfurtransferases, sulfotransferases, and CoA-transferases;

(EC3) hydrolases; *e.g.*, lipases, esterases, amylases, peptidases, hydrolases, lactonases, deacylases, deacetylases, pheophorbidases, depolymerases, thiolesterases, phosphatases, diphosphatases, triphosphatases, nucleotidases, phytases, phosphodiesterases, phospholipases, sulfatases, cyclases, oligonucleotidases, ribonucleases, exonucleases, endonucleases, glycosidases, nucleosidases, glycosylases, aminopeptidases, dipeptidases, carboxypeptidases, metallocarboxypeptidases, omega-peptidases, serine endopeptidases, cystein endopeptidases, aspartic endopeptidases, metalloendopeptidases, threonine endopeptidases, aminases, amidases, desuccinylases, deformylases, acylases, deiminases, deaminases, dihydrolases, cyclohydrolases, nitrilases, ATPases, GTPases, halidases, dehalogenases, and sulfohydrolases;

(EC 4) lyases; *e.g.*, decarboxylases, carboxylases, carboxykinases, aldolases, epoxylyases, oxoacid-lyases, carbon-carbon lyases, dehydratases, hydratases, synthases, endolyases, exolyases, ammonia-lyases, amidine-lyases, amine-lyases, carbon-sulfur lyases, carbon-halide lyases, phosphorus-oxygen lyases, and dehydrochlorinases;

(EC 5) isomerases; *e.g.*, isomerases, racemases, mutases, tautomerases, phosphomutases, phosphoglucomutases, aminomutases, cycloisomerase, cyclases, topoisomerases; and

(EC 6) ligases; *e.g.*, synthetases, tRNA-ligases, acid-thiol ligases, amide synthases, peptide synthases, cycloligases, carboxylases, DNA-ligases, RNA-ligases, and cyclases.

**[0057]** More specific classes of enzymes include, without limitation, sub-classes of oxidoreductases, transferases, lyases, isomerases, and ligases, as provided below.

**[0058]** Exemplary oxidoreductases include, but are not limited to:

(EC 1.1) oxidoreductases acting on the CH-OH group of donors, and an acceptor;

(EC 1.2) oxidoreductases acting on the aldehyde or oxo group of donors, and an acceptor;

(EC 1.3) oxidoreductases acting on the CH-CH group of donors, and an acceptor;

(EC 1.4) oxidoreductases acting on the CH-NH<sub>2</sub> group of donors, and an acceptor;

(EC 1.5) oxidoreductases acting on the CH-NH group of donors, and an acceptor;

(EC 1.6) oxidoreductases acting on NADH or NADPH, and an acceptor;

(EC 1.7) oxidoreductases acting on other nitrogenous compounds as donors, and an acceptor;

(EC 1.8) oxidoreductases acting on a sulfur group of donors, and an acceptor;

(EC 1.9) oxidoreductases acting on a heme group of donors, and an acceptor;

(EC 1.10) oxidoreductases acting on diphenols and related substances as donors, and an acceptor;

(EC 1.11) oxidoreductases acting on a peroxide as acceptor;

(EC 1.12) oxidoreductases acting on hydrogen as donor, and an acceptor;

(EC 1.13) oxidoreductases acting on single donors with incorporation of molecular oxygen, incorporating one or two oxygen atoms;

(EC 1.14) oxidoreductases acting on paired donors, with incorporation or reduction of molecular oxygen, with the donor being 2-oxoglutarate, NADH, NADPH, reduced flavin, flavoprotein, pteridine, iron-sulfur protein, ascorbate;

(EC 1.15) oxidoreductases acting on superoxide radicals as acceptor;

(EC 1.16) oxidoreductases oxidizing metal ions, and an acceptor;

(EC 1.17) oxidoreductases acting on CH or CH<sub>2</sub> groups, and an acceptor;

(EC 1.18) oxidoreductases acting on iron-sulfur proteins as donors, and an acceptor;

(EC 1.19) oxidoreductases acting on reduced flavodoxin as donor, and an acceptor;

(EC 1.20) oxidoreductases acting on phosphorus or arsenic in donors, and an acceptor;

and

(EC 1.21) oxidoreductases acting on X-H and Y-H to form an X-Y bond, and an acceptor; where acceptors for each donor category may include, without limitation: NAD, NADP, heme protein, oxygen, disulfide, quinone, an iron-sulfur protein, a flavin, a nitrogenous group, a cytochrome, dinitrogen, and H<sup>+</sup>.

- [0059]** Exemplary transferases include, but are not limited to:
- (EC 2.1) transferases transferring one-carbon groups;
  - (EC 2.2) transferases transferring aldehyde or ketonic groups;
  - (EC 2.3) Acyltransferases;
  - (EC 2.4) Glycosyltransferases;
  - (EC 2.5) transferases transferring alkyl or aryl groups, other than methyl groups;
  - (EC 2.6) transferases transferring nitrogenous groups;
  - (EC 2.7) transferases transferring phosphorus-containing groups;
  - (EC 2.8) transferases transferring sulfur-containing groups; and
  - (EC 2.9) transferases transferring selenium-containing groups.
- [0060]** Exemplary hydrolases include, but are not limited to:
- (EC 3.1) hydrolases acting on ester bonds;
  - (EC 3.2) Glycosylases;
  - (EC 3.3) hydrolases acting on ether bonds;
  - (EC 3.4) hydrolases acting on peptide bonds (peptidases);
  - (EC 3.5) hydrolases acting on carbon-nitrogen bonds, other than peptide bonds;
  - (EC 3.6) hydrolases acting on acid anhydrides;
  - (EC 3.7) hydrolases acting on carbon-carbon bonds;
  - (EC 3.8) hydrolases acting on halide bonds;
  - (EC 3.9) hydrolases acting on phosphorus-nitrogen bonds;
  - (EC 3.10) hydrolases acting on sulfur-nitrogen bonds;
  - (EC 3.11) hydrolases acting on carbon-phosphorus bonds;
  - (EC 3.12) hydrolases acting on sulfur-sulfur bonds; and
  - (EC 3.13) hydrolases acting on carbon-sulfur bonds.
- [0061]** Exemplary lyases include, but are not limited to:
- (EC 4.1) Carbon-carbon lyases;
  - (EC 4.2) Carbon-oxygen lyases;
  - (EC 4.3) Carbon-nitrogen lyases;
  - (EC 4.4) Carbon-sulfur lyases;
  - (EC 4.5) Carbon-halide lyases; and
  - (EC 4.6) Phosphorus-oxygen lyases.
- [0062]** Exemplary isomerases include, but are not limited to:
- (EC 5.1) Racemases and epimerases;
  - (EC 5.2) cis-trans-Isomerases;

(EC 5.3) Intramolecular isomerases;  
(EC 5.4) Intramolecular transferases (mutases); and  
(EC 5.5) Intramolecular lyases.

**[0063]** Exemplary ligases include, but are not limited to:

(EC 6.1) ligases forming carbon-oxygen bonds;  
(EC 6.2) ligases forming carbon-sulfur bonds;  
(EC 6.3) ligases forming carbon-nitrogen bonds;  
(EC 6.4) ligases forming carbon-carbon bonds;  
(EC 6.5) ligases forming phosphoric ester bonds; and  
(EC 6.6) ligases forming nitrogen-metal bonds.

**[0064]** Isozymes (also known as isoenzymes) are enzymes that differ in amino acid sequence but catalyze the same chemical reaction. At some points in a pathway of interest, two or more isozymes may be present. Isozymes may display different kinetic parameters, and/or different regulatory properties.

**[0065]** Enzymes involved in a pathway of interest or associated pathway may also be classified according to the role of the enzyme. Direct involvement enzymes (class 1) in a cell or cell lysate catalyze a reaction in the pathway. It is typical of pathways that such direct enzymes are one of a chain, where a product of a first enzyme is the substrate of a second enzyme, the product of the second enzyme is the substrate of a third enzyme, and so forth, which eventually results in the product of interest. Indirect involvement enzymes (class 2) in a cell or cell lysate react in an associated pathway, usually in the production of a substrate used in the pathway of interest.

**[0066]** Within a pathway, enzymes will vary in turnover rate and the effectiveness with which a product is produced. As a result, certain enzymes in a pathway become rate-limiting. Increasing the concentration of rate-limiting enzymes in a pathway (relative to non-rate limiting enzymes) allows increased flux through the pathway of interest (see, *e.g.*, Zamboni *et al. Nature Protocols* (2009) 4:878-892). Often rate-limiting enzymes are associated with toxicity when over-produced, and thus the available concentrations of such enzymes is desirably modulated. See, *e.g.*, PCT Publication No. WO 2010/077806.

**[0067]** A third class of enzymes in a cell or cell lysate are competing enzymes (class 3), which utilize a substrate or product of the pathway of interest. A characteristic of a competing enzyme is that the kinetics of the substrate conversion are sufficiently high that the presence of the enzyme decreases the overall yield and/or the rate of production of the

desired final product catalyzed by the pathway of interest. A normal cell may require the expression of competing enzymes, and therefore rather than knocking out expression of competing enzymes completely, it is desirable to selectively decrease the concentration of the active enzyme by the methods described herein.

**[0068]** For convenience of naming, an enzyme in the pathway may be categorized as a first, pathway entry enzyme, or a subsequent downstream enzyme or enzymes. For convenience, the pathway entry enzyme may be referred to herein as  $E_1$ , and the downstream enzymes may be consecutively numbered,  $E_2, E_3, \dots E_n$ . Pathways of interest for use in the methods of the described herein will usually comprise at least one enzyme, at least two enzymes, at least three enzymes, at least four enzymes, or more, *e.g.*, between 1 to 50 enzymes, between 1 to 40 enzymes, between 1 to 30 enzymes, between 1 to 20 enzymes, between 1 to 10 enzymes, between 1 to 5 enzymes, between 1 to 2 enzymes, between 2 to 50 enzymes, between 2 to 40 enzymes, between 2 to 30 enzymes, between 2 to 20 enzymes, between 2 to 10 enzymes, between 2 to 5 enzymes, between 2 to 4 enzymes, between 5 to 50 enzymes, between 5 to 40 enzymes, between 5 to 30 enzymes, between 5 to 20 enzymes, between 5 to 10 enzymes, between 5 to 8 enzymes, between 10 to 50 enzymes, between 10 to 40 enzymes, between 10 to 30 enzymes, or between 10 to 20 enzymes, inclusive.

**[0069]** Enzymes in a pathway may be naturally occurring, or modified to optimize a particular characteristic of interest, *e.g.*, substrate specificity, reaction kinetics, solubility, and/or insensitivity to feedback inhibition. In addition, in some cases, the gene expressing the enzyme will be optimized for codon usage within the host cell. In some embodiments, the complete pathway comprises enzymes from a single organism, however such is not required, and combining enzymes from multiple organisms is also contemplated. For some purposes, a pathway may be endogenous to the host cell, but such is also not required, and a complete pathway or components of a pathway may be introduced into a host cell. Where the system is provided in an intact cell, the complete set of enzymes of the pathway of interest can be present in the cell. For purposes of cell-free production, one or more enzymes may be added to the lysate, or alternatively may be produced by the lysate, so as to complete the pathway. The enzyme may be isolated and/or purified.

**[0070]** In the pathway system, a first substrate ( $S_1$ ) is acted upon by the pathway entry enzyme, and is converted to a first product, although it will be understood by one of skill in the art that an enzyme may act upon more than one substrate simultaneously, and may produce more than one product, such that two or more pathways may be interconnected at a single enzyme. The first product is a substrate ( $S_2$ ) for downstream enzyme  $E_2$ , and is

converted to a second product by  $E_2$ . Depending on the complexity of the pathway, the second product may be the final product (PF), or may be a substrate ( $S_3$ ) for a third downstream enzyme ( $E_3$ ), and is converted to a third product by  $E_3$ , which may be a substrate ( $S_4$ ) for a fourth enzyme. The final enzyme in the pathway, which may be, for example,  $E_2$ ,  $E_3$ , or  $E_4$ , produces the product of interest (PF). It is a characteristic of enzymatic pathways that the product of one enzyme is the substrate for the next enzyme.

**[0071]** Products may be stable or relatively labile, but in some instances the final product is sufficiently stable that it can be isolated from the cell, cell lysate, or reaction mixture.

Competing enzymes utilize a substrate or product of the pathway of interest, which may include any one of PF,  $S_1$ ,  $S_2$ ,  $S_3$ , and/or  $S_4$ , and may be referred to as competing enzymes ( $E_C$ ).

**[0072]** In some embodiments, the initial substrate,  $S_1$ , is a central metabolite, or cellular “commodity”. The central pathways of metabolism include glycolysis and the citric acid cycle. Such  $S_1$  compounds may not be specific to the pathway of interest, but are compounds widely found in various cells and are substrates for multiple enzymes and pathways.

Examples of commodity substrates include, without limitation, glucose, ATP, pyruvate, phosphoenol pyruvate, and the like. A pathway entry enzyme,  $E_1$ , may convert a commodity substrate to a product that is a selective substrate for one or a relatively small number of enzymes. In some embodiments,  $S_1$  compounds or cellular commodities are added to the cell lysate.

**[0073]** In some embodiments, a key entry enzyme is defined as one that performs the first committed step in a pathway to a product of interest. This step may involve the biochemical commitment of a compound to the pathway of a product of interest. Examples of key entry enzymes include, but are not limited to, those set forth in Table 1.

<b>Key Entry Enzyme(s)</b>	<b>Biosynthetic Pathway</b>	<b>Example Products</b>	<b><i>E. coli</i> enzyme</b>
amidophosphoribosyl transferase	purine biosynthesis	GMP, GDP, GTP, dGDP, dGTP, AMP, ADP, ATP, dADP, dATP, inosine monophosphate	PurF
orotate phosphoribosyltransferase	pyrimidine biosynthesis	UMP, UDP, UTP, CDP, CTP	PyrE
2-dehydro-3-deoxyphosphoheptonate aldolase	chorismate biosynthesis	Shikimate, Tyrosine, Phenylalanine, Tryptophan	AroE,F,G

Table 1: Exemplary list of key pathway entry enzymes			
Key Entry Enzyme(s)	Biosynthetic Pathway	Example Products	<i>E. coli</i> enzyme
phosphoribosyltransferase HisG	histidine biosynthesis	Histidine	HisG
acetolactate / acetohydroxybutanoate synthase	isoleucine, leucine, valine biosynthesis	Isoleucine, Leucine, Valine	IlvH,M,N
UDP-N-acetylglucosamine acyltransferase	lipopolysaccharide biosynthesis	Lipid A disaccharide	LpxA
aspartate aminotransferase	lysine, threonine and methionine biosynthesis	Lysine, Threonine, Methionine	AspC
arginine decarboxylase	putrescine biosynthesis	Putrescine	SpeA
GTP cyclohydrolase I	tetrahydrofolate biosynthesis	Tetrahydrofolate	FolE
acetyl-CoA carboxylase	fatty acid biosynthesis	Malonyl-CoA	AccA,B,C,D

[0074] A specific non-limiting example of a pathway, provided for illustrative purposes, is the pathway for the synthesis of shikimic acid. In this pathway, for example, a reaction between the cellular commodity compounds phosphoenolpyruvate ( $S_{1A}$ ) and erythrose-4-phosphate ( $S_{1B}$ ) is catalyzed by the enzyme DAHP synthase ( $E_1$ ) to form 3-deoxy-D-arabinoheptulose-7-phosphate (DAHP). DAHP ( $S_2$ ) is transformed to 3-dehydroquininate (3-DHQ) by the second enzyme in the pathway, DHQ synthase ( $E_2$ ). 3-DHQ ( $S_3$ ) is dehydrated to 3-dehydroshikimate by the third enzyme in the pathway, 3-DHQ dehydratase ( $E_3$ ). 3-dehydroshikimate ( $S_4$ ) is reduced to shikimic acid (PF) by the fourth enzyme in the pathway, shikimate dehydrogenase ( $E_4$ ), using NADPH as a cofactor. The enzymes of the pathway are known in the art and have been characterized in a number of organisms, including, for example, *E. coli*, in which the enzymes are encoded by the genetic loci as follows: DAHP synthase (aroG, aroF, aroH); DHQ synthase (aroB); 3-DHQ dehydratase (aroD); shikimate dehydrogenase (aroE); see, e.g., PCT Publication No. WO2010/074760.

#### *Metabolic Flux*

[0075] “Flux” or “metabolic flux” refers to the rate that molecules pass through a pathway or reaction of interest. Among the factors that control flux are rate of catalysis of enzymes in

the pathway, the availability of substrate, the concentration of enzymes in a cell, and/or the proximity of enzymes in a pathway.

[0076] While a high rate of flux through a pathway of interest is desirable, at the same time it can create toxicity issues if a product not normally accumulated at high levels in the cell is produced at a high rate relative to that occurring under normal conditions. It is understood that a high rate of flux is pathway specific, and refers to the concentration of pathway product over time, such as, for example, production of a product at a rate of about 0.1 grams (g) to about 20 g of product/L/h. In some embodiments, the rate of product production is about 0.5, about 1, about 2, about 5, about 10, or about 15 grams of product/L/h.

[0077] Methods of determining flux rates are known and used in the art; see, e.g., Wiechert *et al.*, *Metab. Eng.* (2001) 3:265-283; Wiechert *et al.*, *Metab. Eng.* (2001)3:195-206; and metabolic engineering texts such as Lee and Papoutsakis, Eds., *Metabolic Engineering*, Marcel Dekker, Inc. New York (1999); Stephanopoulos, Nielsen, and Aristidou, Eds., *Metabolic Engineering: Principles and Methodology*, Academic Press, New York (1998); and Nielsen and Eggeling, Eds., *Metabolic Engineering*, Springer, London (2001).

Flux may be calculated from measurable quantities using techniques such as metabolic flux analysis (MFA), for example, by direct measurement of the conversion rate of isotopically labeled substrate.

#### *Site-specific protease*

[0078] As used herein, the term refers to a protease, for example, an endoprotease, which cleaves selectively at a specific amino acid motif. In some embodiments, the amino acid motif is a motif of at least 4 amino acid residues to reduce background protein cleavage, and may be a motif of at least 5 amino acids, at least 6 amino acids, or more. Such proteases are known to those of skill in the art, and include, without limitation, tobacco etch virus protease (ENLYFQ<sup>G/S</sup>); yellow fever virus protease (GARR<sup>G/S</sup>); thrombin (LVPRGS); and Factor Xa (I<sup>E</sup><sub>D</sub>GR).

[0079] In certain embodiments, described herein are methods of generating site-specific proteases having a periplasmic relocation sequence. Synthesis of genes coding for each site-specific protease enables codon optimization for expression in *E. coli* cells. Polymerase chain reaction (PCR), or other methods familiar to those skilled in the art, can be used for the addition of the periplasmic relocation sequence. Genes are sub-cloned into

vectors enabling controlled expression (*e.g.*, vectors enabling controlled expression using the T7 induction system, such as a vector from the pET series).

<b>Table 2: Examples of specific proteases</b>			
<b>Name</b>	<b>MEROPS</b>	<b>UniProt</b>	<b>Specific Site</b>
TEV NIa	C04.004	P04517 (2036-2279)	E-N-L-Y-F-Q↓G (SEQ ID NO:12)
HRV 3C	C03.007	P04936 (1508-1690)	L-E-V-L-F-Q↓G-P (SEQ ID NO:13)
Enterokinase	S01.156	P98072 (801-1035) bovine	D-D-D-D-K↓ (SEQ ID NO:14)
Factor Xa	S01.216	P00743 (234-468) bovine	I-E-G-R↓ (SEQ ID NO:15)
Thrombin (fIIa)	S01.217	P00735 (367-623) bovine	L-V-P-R↓G (SEQ ID NO:16)

**[0080]** In certain embodiments, a site-specific protease coding sequence is introduced into the microbial cell. The site-specific protease coding sequence may be operably linked to a constitutive or regulatable promoter, *e.g.*, an inducible promoter, which may be referred to herein as a site-specific protease expression construct. The site-specific protease expression construct may be provided on an episomal vector, *e.g.*, a plasmid, YAC, BAC, or viral vector, as known in the art. Alternatively, the site-specific protease expression construct may be integrated into the chromosome of the microbial host cell.

**[0081]** In some embodiments, the expression rate of the site-specific protease is experimentally adjusted to optimize efficiency of export of the protease out of the cell or into the periplasm of the cell. Poor translocation can result from insufficient capacity of export machinery. Methods for adjustment of expression rate include, without limitation, modification of copy number of the plasmid carrying the gene coding for the site-specific protease to be exported to the periplasm. Replicons known and used in the art include P15A (10 copies/cell), ColA (30 copies/cell), ColE1 (40 copies/cell), and RSF1030 (>100 copies/cell). The ribosome binding site in the 5' UTR of the gene coding for the site-specific protease to be exported to the periplasm may be modified, where a library of ribosome binding sites with varying strengths can be created and tested for use in the embodiments described herein, see *Nature Biotechnology* 27: 946-950, 2009; *Nature Biotechnology*

14:629, 1996. The promoter region

upstream of the gene coding for the protein, *e.g.*, a site-specific protease, to be exported may be modified to adjust the rate of transcription, where a library of promoter regions with varying strengths can be created and tested for use in the embodiments described herein, see PNAS 102:12678-12683, 2005; *BMC Biotechnology* 7:34, 2007.

#### *Protease cleavage site*

**[0082]** As described above, a site-specific protease cleaves targeted enzymes in a sequence-specific manner, *e.g.*, at ENLYFQ<sup>G</sup>/<sub>S</sub>, GARR<sup>G</sup>/<sub>S</sub>, LVPRGS, or F<sup>E</sup>/<sub>D</sub>GR. The methods described herein may utilize, in part, a modification of targeted genetic sequences to make a targeted enzyme labile to site-specific protease digestion and consequent inactivation. In addition, the site-specific protease may be selected such that the pathway enzymes of interest are not cleaved by the protease.

**[0083]** The enzyme targeted for site-specific protease cleavage may be an enzyme that competes with an enzyme in the pathway of interest, *e.g.*, an enzyme that utilizes a substrate or product of the pathway and by so doing reduces flux through the pathway or reduces production of the desired product. The targeted enzyme may be altered by amino acid insertion or substitution to generate a site-specific protease cleavage site, for example, by a neutral substitution with conservative or no amino acid changes. By “neutral substitution” it is intended that the amino acid at a position within the motif (*e.g.*, site-specific protease recognition site) is substituted with an amino acid that maintains the targeted enzyme’s activity. Such an amino acid sequence may be referred to as “protease labile sequence.” In certain embodiments, the site for cleavage is selected such that, following cleavage, the targeted enzyme’s activity is destroyed. In some embodiments, the enzyme’s activity is not completely destroyed (*e.g.*, 100% inactive, or reduced by 100%), but rather is reduced by 99%, about 90%, about 80%, about 70%, about 60%, or about 50%. In some embodiments, a targeted enzyme is considered to be inactive if the level of enzymatic activity is such that the targeted enzyme is not capable of competing with enzymes in the pathway of interest.

**[0084]** A targeted enzyme may be selected by pathway analysis to identify enzymes that divert/decelerate flux away from the product of interest or that catalyze a side reaction that siphons flux from a product of interest. It may first be determined whether these enzyme activities can be removed from the host strain (*e.g.*, through genomic knockout) without adversely affecting cell growth. In cases where complete removal of enzyme activity would

substantially harm or otherwise negatively affect the metabolic health of the host, the enzyme(s) performing this activity are targeted for degradation by a relocated site-specific protease. In this way, flux to the product of interest may be increased.

**[0085]** In some embodiments, a library of targeted enzyme sequences with protease cleavage sites inserted at varying positions is initially created to determine the optimal site-specific recognition site and/or placement of the site. Positions may be chosen based on the structure (*e.g.*, tertiary structure) of the targeted enzyme, for example, by inserting the site-specific recognition site in surface-exposed loop regions. Where structural information for the targeted enzyme is not available, the positioning of the recognition site may be determined by comparing amino acid sequences of targeted enzyme homologs and identifying hydrophilic regions with high sequence and length variability. Hydrophilic residues are commonly exposed on the enzyme surface, and high sequence and length variability can indicate that the region is not critical for enzyme function and can tolerate sequence modification. For example, see *Journal of Virology* 78:5079, 2004; US patent 7223390, which teach that limited proteolytic digestion followed by chromatographic co-fractionation and N-terminal polypeptide sequencing/mass spectroscopy can be used to determine such candidate sites which are accessible to protease digestion and nonessential for forming an integral enzyme/protein structure.

**[0086]** Modification of an amino acid sequence of the targeted enzyme to introduce a site-specific recognition site may be accomplished by methods of gene engineering known in the art, for example using PCR-based and ligation-based methods, or by methods of genomic modification known to those skilled in the art. See, for example Phusion site-directed mutagenesis kit (New England Biolabs); *Journal of Virological Methods* 149:85, 2008; *Nucleic Acids Research* 32:e174, 2004; *Journal of Bacteriology* 180: 2063, 1998; *Proceedings of the National Academy of Sciences of the United States of America* 97:6640, 2000.

#### *Periplasmic sequestration*

**[0087]** In some embodiments, site-specific protease relocation is to the periplasmic space of the cell. In such aspects, described herein are methods of generating cells, lysates, and uses thereof, whereby a protease is genetically modified to incorporate a peptide relocation sequence that provides for periplasmic targeting of the protease. Periplasmic relocation sequences (also referred to herein as targeting peptide sequences, targeting signals, signal peptides, or signal sequences), in some embodiments, are found at the N-terminus of bacterial

secretory proteins. They vary in length from about 15 to about 70 amino acids. The primary sequences of amino acids also vary, but may have a common overall structure, including the following parts: (i) an N-terminal part having a variable length and carrying, for example, a net positive charge; (ii) a central hydrophobic core of about 6 to about 15 amino acids; and (iii) about 4 to about 6 amino acids defining the cleavage site for signal peptidases.

**[0088]** Periplasmic relocation sequences suitable for use herein may be derived from a protein that is secreted in a Gram negative bacterium. The secreted protein may be encoded by the bacterium or by a bacteriophage that infects the bacterium. Examples of suitable Gram negative bacterial sources of secreted proteins include but are not limited to, members of the genera *Escherichia*, *Pseudomonas*, *Klebsiella*, *Salmonella*, *Caulobacter*, *Methylomonas*, *Acetobacter*, *Achromobacter*, *Acinetobacter*, *Aeromonas*, *Agrobacterium*, *Alcaligenes*, *Azotobacter*, *Burkholderia*, *Citrobacter*, *Comamonas*, *Enterobacter*, *Erwinia*, *Rhizobium*, *Vibrio*, and *Xanthomonas*.

**[0089]** There are three pathways for translocation of proteins across cytoplasmic membranes: (i) SecB-dependent, (ii) signal recognition particle (SRP), and (iii) twin arginine translocation (TAT) pathways. SecB-dependent and signal recognition particle pathways both use the SecYEG translocon. The twin arginine translocation pathway uses the TatABCE complex. SecB-dependent translocation is most commonly used, but this pathway is not able to transport folded proteins. Rapid cytoplasmic folding may necessitate use of TAT pathway. Examples of bacterial secreted proteins having periplasmic relocation sequences include, but are not limited to, proteins encoded by the following genes: *ompA*, *geneIII*, *E. coli* alkaline phosphatase, *lamB*, *malE*, *secE*, *secY*, and *prlA-4*. One skilled in the art can identify the periplasmic relocation sequence located at the N-terminus of each of these proteins and other bacterial secretory proteins. It is also known by one skilled in the art that some amino acid substitutions, additions, and/or deletions may be made in a periplasmic relocation sequences while retaining the periplasmic targeting function. Thus, a functional periplasmic relocation sequence as described herein may be a fully natural or a modified sequence.

**[0090]** In some embodiments, the site-specific protease is relocated to an extracytoplasmic location. Signals for secretion in bacterial cells are known in the art, for example, sec and Tat translocons in *B. subtilis* (see for review Buist *et al.* (2006) *Microbiology* 152(Pt 10):2867-74; and Tjalsma *et al.* (2004) *Microbiol Mol Biol Rev.* 68(2):207-33).

ClyA export from bacterial cells is also known in the art, for example see Ludwig *et al.* (2010) *J Bacteriol.* 192(15):4001-11.

**[0091]** In other embodiments, the site-specific protease is relocated to a nuclear compartment. Peptide signals for nuclear relocation are known in the art, for example the SV40 large T-cell antigen peptide: PKKKRKV (see Kalderon *et al.* (1984) *Cell* 39 (3 Pt 2): 499–509); and the nucleoplasmin peptide: KR[PAATKKAGQA]KKKK (see Dingwall *et al.* (1988) *J Cell Biol.* 107 (3): 841–9). In other embodiments, localization may be made to the thylakoid membrane, see, for example Schnell (1998) *Annual Review of Plant Physiology and Plant Molecular Biology* Vol. 49: 97-126; Ko and Cashmore (1989) *EMBO J.* 8(11): 3187–3194; de Boer *et al.* (1991) *EMBO J.* 10(10): 2765–2772; and Mackle and Zilinskas (1994) *J. Bacteriol.* 176(7):1857-1864.

**[0092]** Fusion proteins described herein may comprise a periplasmic relocation sequence and/or a site-specific protease. In certain embodiments, the optimal periplasmic relocation sequence for the site-specific protease is empirically determined from a selection of such sequences. The efficiency of secretion of the site-specific protease may depend on various parameters, including the relocation sequence, the protease being relocated, host cell strain used, protease expression level, and other parameters. For example, a library of modified genes with varying 5' regions coding for different periplasmic relocation sequences can be created using PCR or other methods known to those skilled in the art. This library is subcloned in a vector enabling controlled expression (*e.g.*, a vector enabling controlled expression using the T7 induction system such as a vector from the pET series), and tested for export efficiency as well as protein activity (see *J. Biol. Chem.* 267:4882, 1992; *J. Biol. Chem.* 267:12375, 1992; US2007/0111283 A1; and *Biotechnology Advances* 23:177, 2005).

**[0093]** Examples of periplasmic relocation sequences include, without limitation:

<b>Table 3. Examples of periplasmic relocation sequences</b>				
<b>Name</b>	<b>Signal peptide</b>	<b>kDa</b>	<b>Pathway</b>	<b>Source</b>
MalE	MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1)	2.7	Sec	<i>E. coli</i>
OmpA	MKKTAIAlAVALAGFATVAQA (SEQ ID NO:2)	2.0		<i>E. coli</i>
PhoA	MKQSTIALALLPLLFTPVTKA (SEQ ID NO:3)	2.3	Sec	<i>E. coli</i>
LamB	MMITLRKLPLAVAVAAGVMSAQAMA (SEQ ID NO:4)	2.6	Sec	<i>E. coli</i>
MglB	MNKKVLTLSAVMASMLFGAAHA	2.4	Sec	<i>E. coli</i>

Table 3. Examples of periplasmic relocation sequences				
Name	Signal peptide	kDa	Pathway	Source
	(SEQ ID NO:5)			
PeIB	MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:6)	2.2	Sec	<i>E. caratovora</i>
STII	MKKNIAFLASMFVFSIATNAYA (SEQ ID NO:7)	2.6		<i>E. coli</i>
DsbA	MKKIWLALAGLVLAFSASA (SEQ ID NO:8)	2.0	SRP	<i>E. coli</i>
SfmC	MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:9)	2.6	SRP	<i>E. coli</i>
ToIB	MKQALRVAFGLILWASVLHA (SEQ ID NO:10)	2.4	SRP	<i>E. coli</i>
TorT	MRVLI.FI.II.SI.FMI.PAFS (SEQ ID NO:11)	2.1	SRP	<i>E. coli</i>

**[0094]** A cleavage site is optionally located between the relocation sequence and the site-specific protease to allow separation. The cleavage site may be any site that can be used in separating the relocation sequence and the site-specific protease coding sequence. Any cleavage site making use of any method for protein cleavage may be used. Periplasmic relocation sequences from *E. coli* may contain within their signal sequence a motif recognized by leader peptidase, Lep, for signal sequence processing and cleavage. Other methods that have been successfully used and are well known to one skilled in the art include protease cleavage methods, *e.g.* thrombin, factor Xa protease and other endo peptidases, such as trypsin. The genes encoding the fusion protein can be synthesized to include a cleavage site for one of these proteases between the relocation sequence and the protease sequence. Another system for fusion and cleavage is the intein/chitin binding domain system (S. Chong, *et al.*, *Gene*, Vol. 192, 271-281 (1997)). This system makes use of the self-cleaving properties of intein proteins.

#### *Nucleic Acids, Polypeptides, and Cells*

**[0095]** The nucleic acids used as described herein, whether RNA, iRNA (also known as RNAi), antisense nucleic acid, cDNA, genomic DNA, vectors, artificial chromosomes, viruses, or hybrids thereof may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed/generated recombinantly. A nucleic acid molecule or nucleic

acid molecules that encode any of the targeted enzymes or site-specific proteases associated with the embodiments described herein can be introduced into a cell or cells using methods and techniques that are known in the art. For example, nucleic acid molecules can be introduced into cells by standard protocols such as transformation, including chemical transformation and electroporation, transduction, and particle bombardment. Expressing a nucleic acid molecule(s) encoding a targeted enzyme or site-specific protease may also be accomplished by integrating the nucleic acid molecule into the genome of the cell. Nucleic acid molecule(s) can be integrated into a cell's genomic DNA using standard techniques known in the art. Recombinant polypeptides generated from these nucleic acids can be individually isolated or cloned and tested for a desired activity. "Polypeptide" and "protein," as used herein, encompasses targeted enzymes and site-specific proteases. Any recombinant expression system can be used, including, but not limited to, bacterial, mammalian, yeast, insect, or plant cell expression systems. The nucleic acids used as described herein can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, *e.g.*, Adams *et al.*, *J. Am. Chem. Soc.* (1983) 105:661; Belousov *et al.*, *Nucleic Acids Res.* (1997) 25:3440-3444; Frenkel *et al.*, *Free Radic. Biol. Med.* (1995) 19:373-380; Blommers *et al.*, *Biochemistry* (1994) 33:7886-7896; Narang *et al.*, *Meth. Enzymol.* (1979) 68:90; Brown *et al.*, *Meth. Enzymol.* (1979) 68:109; Beaucage *et al.*, *Tetrahedron Letters* (1981) 22:1859; and U.S. Patent 4,458,066.

**[0096]** Host cells for pathway engineering include a wide variety of heterotrophic and autotrophic microorganisms, including, but not limited to, bacteria, fungi and protozoans. In certain embodiments, host cells include those for which means by which a polypeptide can be directed to a cellular compartment or extracellular compartments are known. In some embodiments, the cell is any type of cell that recombinantly expresses any one or more of the nucleic acids (*e.g.*, site-specific protease or targeted enzyme coding sequences) described herein. Such cells include prokaryotic and eukaryotic cells. In some embodiments, the cell is a bacterial cell, such as *Escherichia spp.*, *Streptomyces spp.*, *Zyomonas spp.*, *Acetobacter spp.*, *Citrobacter spp.*, *Synechocystis spp.*, *Rhizobium spp.*, *Clostridium spp.*, *Corynebacterium spp.*, *Streptococcus spp.*, *Xanthomonas spp.*, *Lactobacillus spp.*, *Lactococcus spp.*, *Bacillus spp.*, *Alcaligenes spp.*, *Pseudomonas spp.*, *Aeromonas spp.*, *Azotobacter spp.*, *Comamonas spp.*, *Mycobacterium spp.*, *Rhodococcus spp.*, *Gluconobacter spp.*, *Ralstonia spp.*, *Acidithiobacillus spp.*, *Microlunatus spp.*, *Geobacter spp.*, *Geobacillus spp.*, *Arthrobacter spp.*, *Flavobacterium spp.*, *Serratia spp.*, *Saccharopolyspora spp.*, *Thermus spp.*, *Stenotrophomonas spp.*, *Chromobacterium spp.*, *Sinorhizobium spp.*, *Saccharopolyspora*

*spp.*, *Agrobacterium spp.* and *Pantoea spp.* The bacterial cell can be a Gram-negative cell such as an *Escherichia coli* (*E. coli*) cell, or a Gram-positive cell such as a species of *Bacillus*. In other embodiments the cell is a fungal cell such as yeast cells, *e.g.*, *Saccharomyces spp.*, *Schizosaccharomyces spp.*, *Pichia spp.*, *Paffia spp.*, *Kluyveromyces spp.*, *Candida spp.*, *Talaromyces spp.*, *Brettanomyces spp.*, *Pachysolen spp.*, *Debaryomyces spp.*, *Yarrowia spp.* and industrial polyploid yeast strains. Other non-limiting examples of fungi include *Aspergillus spp.*, *Penicillium spp.*, *Fusarium spp.*, *Rhizopus spp.*, *Acremonium spp.*, *Neurospora spp.*, *Sordaria spp.*, *Magnaporthe spp.*, *Allomyces spp.*, *Ustilago spp.*, *Botrytis spp.*, and *Trichoderma spp.* In other embodiments, the cell is an algal cell, a plant cell, or a mammalian cell. It should be appreciated that some cells compatible with the embodiments described herein may express an endogenous copy of one or more of the genes described herein as well as a recombinant copy. Species of interest include, without limitation, *S. cerevisiae*, *E. coli*, *Pseudomonas* species, *Klebsiella* species, and *Synechocystis* species. To avoid unwanted degradation of the relocated site-specific protease, the host strain can be modified to remove various compartmental proteases (*e.g.*, periplasmic proteases) and/or to augment with proteins such as chaperones and maturases to assist with protein folding. Such modifications and augmentations employ methods known to those skilled in the art, see, *e.g.*, U.S. Patents 4,946,783 and 6,921,659, and Chen *et al.*, *Biotechnology and Bioengineering* (2004) 85: 463-474.

[0097] In some embodiments, one or more genes described herein are expressed recombinantly in a bacterial cell. Bacterial cells described herein can be cultured in media of any type (rich or minimal) and any composition. In some embodiments, the cells are culture in minimal medium. As would be understood by one of ordinary skill in the art, routine optimization would allow for use of a variety of types of media. The selected medium can be supplemented with various additional components. Some non-limiting examples of supplemental components include glucose, antibiotics, isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), tetracycline or anhydro-tetracycline (aTc) for gene induction and ATCC Trace Mineral Supplement. Similarly, other aspects of the medium, and growth conditions of the cells of the embodiments described herein may be optimized through routine experimentation. pH and temperature are non-limiting examples of factors which can be optimized. In some embodiments, the concentration and amount of a supplemental component may be optimized. In some embodiments, how often the media is supplemented with one or more supplemental components, and the amount of time that the media is cultured is optimized.

**[0098]** Techniques for the manipulation of nucleic acids, *e.g.*, subcloning, labeling probes (*e.g.*, random-primer labeling using Klenow polymerase, nick translation, amplification), sequencing, hybridization, and the like are well described in the scientific and patent literature, see, *e.g.*, Sambrook, Ed., *Molecular Cloning: A Laboratory Manual* (2nd Ed.) Vols 1-3, Cold Spring Harbor Laboratory (1989); Ausubel, Ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., New York (1997); and Tijssen, Ed., *Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization with Nucleic Acid Probes, Part I. Theory and Nucleic Acid Preparation*, Elsevier, N.Y. (1993).

**[0099]** It should be appreciated that the genes encoding targeted enzymes and site-specific proteases associated with the embodiments described herein can be obtained from a variety of sources. As one of ordinary skill in the art would be aware, homologous genes for these targeted enzymes and site-specific proteases exist in many species and can be identified by homology searches, for example through a protein BLAST search, available at the NCBI internet site .

Genes encoding these targeted enzymes and site-specific proteases can be PCR-amplified from DNA from any source which contains the given targeted enzyme and site-specific protease, for example using degenerate primers, as would be understood by one of ordinary skill in the art. In some embodiments, the gene encoding a given targeted enzyme and site-specific protease can be synthetic (artificial), for example, DNA synthesized from sugars, nitrogen-based compounds, phosphates, and other compound/reagents required for DNA synthesis. Any means of obtaining the genes encoding for the targeted enzymes and site-specific proteases discussed here are compatible with aspects of the embodiments described herein.

**[00100]** The embodiments described herein also provide isolated polypeptides and proteins (*e.g.*, targeted enzymes and site-specific proteases) encoded by the nucleic acids. Such polypeptides are useful, for example, alone or as fusion proteins. Polypeptides associated with the cells, cell lysates, and methods described herein can be isolated from biological samples including tissue or cell homogenates, and can also be expressed recombinantly in a variety of prokaryotic and eukaryotic expression systems by constructing an expression vector appropriate to the expression system, introducing the expression vector into the expression system, and isolating the recombinantly expressed polypeptide or protein, *e.g.*, a targeted enzyme or site-specific protease. Polypeptides can also be synthesized chemically using well-established methods of peptide synthesis. In some embodiments, isolated

enzymes and/or protease may be added directly to cell lysates during the production phase of metabolite production.

**[00101]** A variety of methodologies well known to the skilled practitioner can be utilized to obtain isolated polypeptides associated with the embodiments described herein. The polypeptide may be purified from cells by chromatographic means or immunological recognition. Alternatively, an expression vector may be introduced into cells to cause production of the polypeptide. In another method, mRNA transcripts may be microinjected or otherwise introduced into cells to cause production of the encoded polypeptide.

Translation of mRNA in cell free extracts such as the reticulocyte lysate system also may be used to produce polypeptides. Those skilled in the art also can readily follow known methods for isolating polypeptides. These include, but are not limited to, immunochromatography, HPLC, size exclusion chromatography, ion exchange chromatography and immune affinity chromatography.

**[00102]** The expression of the molecules, such as the targeted enzymes and site-specific proteases, described herein may be determined using routine methods known to those of ordinary skill in the art. These methods include, but are not limited to: direct RNA amplification, reverse transcription of RNA to cDNA, real-time RT-PCR, amplification of cDNA, hybridization, and immunologically based assay methods, which include, but are not limited to, immunohistochemistry, antibody sandwich capture assay, ELISA, and enzyme-linked immunospot assay (EliSpot assay). For example, the determination of the presence of level of nucleic acid molecules in a cell can be carried out via any standard nucleic acid determination assay, including the polymerase chain reaction (PCR), or assaying with labeled hybridization probes. Such hybridization methods include, but are not limited to microarray techniques.

**[00103]** The embodiments described herein thus involve, in one aspect, targeted enzymes and site specific proteases, genes encoding those enzymes and proteases, functional modifications and variants of the foregoing, as well as uses relating thereto.

**[00104]** The embodiments described herein also includes degenerate coding sequences which include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT and AGC. Each of the six codons is equivalent for the purposes of encoding a serine residue. Thus, it will be apparent to one of ordinary skill in the art that any of the serine-encoding nucleotide triplets may be employed to direct the protein synthesis apparatus, *in vitro* or *in vivo*, to incorporate a serine residue into an elongating polypeptide. Similarly, nucleotide sequence triplets which encode

other amino acid residues include, but are not limited to: CCA, CCC, CCG and CCT (proline codons); CGA, CGC, CGG, CGT, AGA and AGG (arginine codons); ACA, ACC, ACG and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the embodiments described herein embrace degenerate nucleic acids that differ from the biologically isolated nucleic acids in codon sequence due to the degeneracy of the genetic code. The embodiments also embraces codon optimization to suit optimal codon usage of a host cell, *e.g.*, an *E. coli* or other bacterial cell.

**[00105]** The embodiments described herein also provides modified nucleic acid molecules which include additions, substitutions and deletions of one or more nucleotides. In some embodiments, these modified nucleic acid molecules and/or the polypeptides they encode retain at least one activity or function of the unmodified nucleic acid molecule and/or the polypeptides, such as enzymatic (*e.g.*, protease) activity. In certain embodiments, the modified nucleic acid molecules encode modified polypeptides, for example, polypeptides having conservative amino acid substitutions as are described elsewhere herein. The modified nucleic acid molecules are structurally related to the unmodified nucleic acid molecules, and in some embodiments are sufficiently structurally related to the unmodified nucleic acid molecules so that the modified and unmodified nucleic acid molecules hybridize under stringent conditions known to one of skill in the art.

**[00106]** For example, modified nucleic acid molecules which encode polypeptides having single amino acid changes can be prepared. Each of these nucleic acid molecules can have one, two or three nucleotide substitutions exclusive of nucleotide changes corresponding to the degeneracy of the genetic code as described herein. Likewise, modified nucleic acid molecules that encode polypeptides having two amino acid changes can be prepared which have, *e.g.*, 2-6 nucleotide changes. Numerous modified nucleic acid molecules like these will be readily envisioned by one of skill in the art, including for example, substitutions of nucleotides in codons encoding amino acids 2 and 3, 2 and 4, 2 and 5, 2 and 6, and so on. In the foregoing example, each combination of two amino acids is included in the set of modified nucleic acid molecules, as well as all nucleotide substitutions which code for the amino acid substitutions. Additional nucleic acid molecules that encode polypeptides having additional substitutions (*e.g.*, 3 or more), additions or deletions (*e.g.*, by introduction of a stop codon or a splice site(s)) also can be prepared as readily envisioned by one of ordinary skill in the art. Any of the foregoing nucleic acids or polypeptides can be tested by routine

experimentation for retention of structural relation or activity, *e.g.*, enzymatic activity, to the nucleic acids and/or polypeptides disclosed herein.

**[00107]** In certain embodiments, variants of the targeted enzymes and site-specific proteases described herein are contemplated. As used herein, a “variant” of an targeted enzyme or site-specific protease is a protein which contains one or more modifications to the primary amino acid sequence of the targeted enzyme or site-specific protease. Modifications which create a variant can be made to an targeted enzyme or site-specific protease, for example, to reduce or eliminate an activity of the targeted enzyme in the presence of the site-specific protease or to alter the cellular distribution of the site-specific protease.

**[00108]** Modifications to a targeted enzyme or a site-specific protease may be made to the nucleic acid which encodes the protein, and can include deletions, point mutations, truncations, amino acid substitutions and additions of amino acids or non-amino acid moieties. Alternatively, modifications can be made directly to the polypeptide(s) or protein, such as by cleavage, addition of a linker molecule, addition of a detectable moiety, such as biotin, addition of a fatty acid, and the like. Modifications also embrace fusion proteins. One of skill in the art will know methods for predicting the effect on protein conformation of a change in amino acid sequence, and can thus “design” a variant targeted enzyme or site-specific protease according to those known methods. One example of such a method is described by Dahiyat and Mayo in *Science* 278:82 87, 1997, whereby proteins can be designed *de novo*. The method can be applied to a known protein to vary only a portion of the protein sequence. By applying the computational methods of Dahiyat and Mayo, specific variants of a protein can be proposed and tested to determine whether the variant retains a desired conformation.

**[00109]** In some embodiments, variants include targeted enzymes and site-specific proteases that are modified specifically to alter a feature of the protein unrelated to its desired physiological activity. For example, cysteine residues can be substituted or deleted to prevent unwanted disulfide linkages. Similarly, certain amino acids can be changed to enhance expression of a protein by eliminating proteolysis by proteases in an expression system (*e.g.*, dibasic amino acid residues in yeast expression systems in which KEX2 protease activity is present).

**[00110]** Mutations of a nucleic acid which encode a targeted enzyme or site-specific protease may preserve the amino acid reading frame of the coding sequence, and in some instances, do not create regions in the nucleic acid which are likely to hybridize to form

secondary structures, such a hairpins or loops, which can be deleterious to expression of the variant protein.

**[00111]** Mutations can be made by selecting an amino acid substitution, or by random mutagenesis of a selected site in a nucleic acid which encodes the targeted enzyme or site-specific protease. Variant proteins are then expressed and tested for one or more activities to determine which mutation provides a variant enzyme or protease with the desired properties, *e.g.*, site-specific protease cleavage. Further mutations can be made to variants (or to non-variant proteins) which are silent as to the amino acid sequence of the protein, but which provide codons for efficient translation in a particular host. Examples of codons for efficient translation of a nucleic acid in, *e.g.*, *E. coli*, are well known to those of ordinary skill in the art. Still, other mutations can be made to the noncoding sequences of a gene or cDNA clone to enhance expression of the protein. The activity of variants of protein can be tested by cloning the gene encoding the variant protein into a bacterial or mammalian expression vector, introducing the vector into an appropriate host cell, expressing the variant protein, and testing for a functional capability of the protein as disclosed herein.

**[00112]** The skilled artisan will also know that conservative amino acid substitutions may be made in proteins to provide functionally equivalent variants of the foregoing targeted enzymes and site-specific proteases, *i.e.*, the variants retain the functional capabilities of the proteins. As used herein, a “conservative amino acid substitution” refers to an amino acid substitution which does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Variants can be prepared according to methods for protein sequence known to one of ordinary skill in the art such as are found in references which compile such methods, *e.g.*, *Molecular Cloning: A Laboratory Manual*, J. Sambrook, *et al.*, eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, *et al.*, eds., John Wiley & Sons, Inc., New York. Exemplary functionally equivalent variants of the protein include conservative amino acid substitutions in the amino acid sequences of the enzymes and proteases disclosed herein. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D.

**[00113]** In certain embodiments, fewer than all of the amino acids can be changed when preparing variant enzymes and proteases. Where particular amino acid residues are known to confer function and function is required, such amino acids will not be replaced, or alternatively, will be replaced by conservative amino acid substitutions. In some

embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 residues are changed when preparing variant proteins. In some embodiments, the fewest number of substitutions is made. Thus, one method for generating variant enzymes or proteases is to substitute all other amino acids for a particular single amino acid, then determine the activity of the variant, then repeat the process with one or more of the enzymes or proteins having the best activity.

**[00114]** Conservative amino-acid substitutions in the amino acid sequence of protein to produce functionally equivalent variants of protein may be made by alteration of a nucleic acid encoding a targeted enzyme or site-specific protease. Such substitutions can be made by a variety of methods known to one of ordinary skill in the art. For example, amino acid substitutions may be made by PCR-directed mutation, site-directed mutagenesis according to the method of Kunkel (Kunkel, Proc. Nat. Acad. Sci. U.S.A. (1985) 82: 488-492), or by chemical synthesis of a gene encoding a protein. For example, in some embodiments, a nucleic acid encoding a targeted enzyme is modified to comprise a site-specific protease cleavage site, and the expressed enzyme variant still retains the functional activity of the unmodified enzyme.

#### *Vectors and Expression Constructs*

**[00115]** Vectors useful for the transformation of an isolated DNA fragment encoding a targeted enzyme and/or a site-specific protease into suitable host cells (*e.g.*, *E. coli* cells) are well known in the art. As used herein, a “vector” may be any of a number of nucleic acids into which a desired sequence or sequences (*e.g.*, sequences encoding targeted enzymes or site-specific proteases) may be inserted by restriction and ligation for transport between different genetic environments or for expression in a host cell. Vectors are typically composed of DNA, although RNA vectors are also available. Vectors include, but are not limited to, plasmids, fosmids, phagemids, virus genomes, and artificial chromosomes. The vector can contain sequences directing transcription and translation of the relevant gene, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment that controls transcriptional termination. Vectors may also be used which promote the integration of the chimeric gene encoding a fusion protein into the host cell genome. Such vectors may be for random integration, site-directed integration, or for homologous recombination. A vector may have features allowing single cross-over or double-crossover types of homologous recombination.

One or multiple copies may be integrated into a host cell (*e.g.*, bacterial or yeast cell) genome.

**[00116]** A cloning vector is able to replicate autonomously or integrated in the genome in a host cell. A cloning vector is further characterized by one or more endonuclease restriction sites at which the vector may be cut (*e.g.*, via enzymatic digestion) in a determinable fashion and into which a desired DNA sequence may be ligated. In this way, the new recombinant vector retains its ability to replicate in the host cell. When plasmid vectors are used, replication of the desired sequence may occur many times as the plasmid increases in copy number within the host or just a single time per host before the host reproduces by mitosis. When phage vectors are used, replication may occur actively during a lytic phase or passively during a lysogenic phase.

**[00117]** An expression vector is one into which a desired DNA sequence may be inserted by restriction and ligation such that it is operably joined to regulatory sequences and may be expressed as an RNA transcript. Vectors may further contain one or more marker sequences suitable for use in the identification of cells which have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (*e.g.*,  $\beta$ -galactosidase, luciferase, or alkaline phosphatase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies or plaques (*e.g.*, green fluorescent protein). Vectors include those capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably joined.

**[00118]** As used herein, a coding sequence and regulatory sequences are said to be “operably” linked or joined when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences. If it is desired that the coding sequences be translated into a functional protein, two DNA sequences are said to be operably joined if induction of a promoter in the 5' regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably joined to a coding sequence if the promoter region were capable of effecting transcription of

that DNA sequence such that the resulting transcript can be translated, for example, into the desired site-specific protease.

**[00119]** When the nucleic acid molecule that encodes any of the targeted enzymes or site-specific proteases described herein is expressed in a cell, a variety of transcription control sequences (*e.g.*, promoter/enhancer sequences) can be used to direct its expression. The promoter can be a native promoter, *i.e.*, the promoter of the gene in its endogenous context, which provides normal regulation of expression of the gene. In some embodiments the promoter can be constitutive, *i.e.*, the promoter is unregulated allowing for continual transcription of its associated gene. A variety of conditional promoters also can be used, such as promoters controlled by the presence or absence of a molecule. For example, some inducible promoters are activated by chemicals such as isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) or Tetracycline (Tet).

**[00120]** The precise nature of the regulatory sequences needed for gene expression may vary between species or cell types, and may include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. In particular, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined gene. Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired. Vectors may optionally include 5' leader or signal sequences. The choice and design of an appropriate vector is within the ability and discretion of one of ordinary skill in the art.

**[00121]** Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, *e.g.*, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by introducing into the cells heterologous DNA (or RNA). That heterologous DNA (or RNA) is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell. In some embodiments two or more nucleic acids may be cloned into the same expression vector or plasmid.

**[00122]** The methods described herein may make use of constitutive or regulated expression of various nucleic sequences encoding the targeted enzymes and/or site-specific proteases. Expression may be regulated by various cues, for example, induction by chemicals (*e.g.*, IPTG), change of growth phase, depletion of a nutrient, temperature shifts (*e.g.*, heat

shock promoter), and/or light. In some embodiments, inducible promoters are regulated by the presence of an inducing agent, for example, a chemical such as lactose, arabinose, or tetracycline, as known in the art. Where “high level” expression is indicated, the concentration of the expressed protein in the cell can be at least about 2-fold above basal levels, at least 10-fold above basal levels, at least 25-fold above basal levels, at least 50-fold above basal levels, or more, *e.g.*, between about 2-fold to about 100-fold above basal levels.

**[00123]** It may be desirable to experimentally adjust expression rate to optimize efficiency of cell or cell compartment export of the site-specific proteases described herein. Poor translocation, for example, to the periplasm of a cell can result from insufficient capacity of export machinery. Methods for adjustment of expression rate of the site-specific protease include, without limitation, modification of copy number of the plasmid carrying the gene coding for the protease to be exported to the periplasm. Replicons known and used in the art include P15A (10 copies/cell), ColA (30 copies/cell), ColE1 (40 copies/cell), and RSF1030 (>100 copies/cell). The ribosome binding site in the 5' UTR of the gene coding for the protease to be exported to the periplasm may be modified, where a library of ribosome binding sites with varying strengths can be created and tested; see, *e.g.*, Salis *et al.*, *Nature Biotechnology* (2009) 27: 946-950; and Simmons *et al.*, *Nature Biotechnology* (1996) 14:629-634 .

The promoter region upstream of the gene coding for the protease to be exported may be modified to adjust the rate of transcription, where a library of promoter regions with varying strengths can be created and tested; see, *e.g.*, Alper *et al.*, *PNAS* (2005) 102:12678-12683; and De Mey *et al.*, *BMC Biotechnology* (2007) 7:34 .

#### *Inducible expression*

**[00124]** The methods described herein may make use of regulated expression of various coding sequences. Expression may be regulated by various cues, for example, induction by chemicals, change of growth phase, depletion of a nutrient, temperature shifts, and light. In some embodiments inducible promoters regulated by the presence of an inducing agent, *e.g.* a chemical such as lactose, arabinose, and tetracycline, known in the art.

**[00125]** Expression and cloning vectors usually contain a promoter that is recognized by the host organism and is operably linked to the coding sequence of interest. Promoters are untranslated sequences located upstream (5') to the start codon of a structural gene that control the transcription and translation of particular nucleic acid sequence to which they are operably linked. Such promoters may fall into two classes, inducible and constitutive.

Inducible promoters are promoters that initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, *e.g.*, the presence or absence of a nutrient or a change in temperature. At this time a large number of promoters recognized by a variety of potential host cells are well known. While the native promoter may be used, in some embodiments, heterologous promoters are used, as they may permit greater transcription of the targeted enzyme or site-specific protease and higher yields.

**[00126]** Promoters suitable for use with prokaryotic hosts include the  $\beta$ -lactamase and lactose promoter systems, alkaline phosphatase, a tryptophan (*trp*) promoter system, and numerous hybrid promoters such as the *tac* promoter. However, other known bacterial promoters are also suitable, *e.g.*, the *lacI* promoter, the T3 promoter, the T7 promoter, the arabinose promoter, the *gpt* promoter, the lambda PR promoter, the lambda PL promoter, promoters from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), and the acid phosphatase promoter. Their nucleotide sequences have been published, thereby enabling a skilled worker to operably ligate them to a sequence of interest using linkers or adapters. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the coding sequence. In certain cases, also, the host cell may be modified genetically to adjust concentrations of metabolite or inducer transporter proteins so that all cells in a culture will be induced equivalently.

**[00127]** Promoters suitable for eukaryotic cells, *e.g.* yeast cells, are also known in the art. Virtually all eukaryotic genes have an AT-rich region located approximately 25 to 30 bases upstream from the site where transcription is initiated. Another sequence found 70 to 80 bases upstream from the start of transcription of many genes is a CXCAAT region where X may be any nucleotide. At the 3' end of most eukaryotic genes is an AATAAA sequence that may be the signal for addition of the poly-A tail to the 3' end of the coding sequence. All of these sequences are suitably inserted into eukaryotic expression vectors. Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglyceratekinase or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

**[00128]** Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate

dehydrogenase, and enzymes responsible for maltose and galactose utilization. Yeast enhancers also are advantageously used with yeast promoters.

### EXAMPLES

[00129] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the embodiments described herein, and are not intended to limit the scope of the invention or to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.*, amounts, temperature, and the like), but some experimental errors and deviations may be present. Unless indicated otherwise, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric pressure.

[00130] *Segregating site-specific protease.* The steps in the process of protease segregation and target enzyme engineering include (i) analyzing a pathway to identify an enzyme or enzymes to be targeted for inactivation, (ii) engineering a protease cleavage site by modifying the DNA sequence that encodes the targeted enzyme(s), (iii) verifying activity of the targeted enzyme, (iv) engineering periplasmically segregated proteases, (v) verifying activity of periplasmically-expressed proteases, (vi) demonstrating metabolically healthy cell growth, and (vii) demonstrating target enzyme(s) inactivation and increased flux to the product of interest post-lysis in an active, cell-free reaction.

#### *Example 1. Protease segregation to periplasm*

[00131] A library of plasmids is constructed containing several protease genes with various periplasmic signal sequences targeting the protease enzymes to the periplasm. For example, five specific proteases and six periplasmic signal sequences could be combined to make a library of thirty members.

[00132] Nucleotide sequences coding for periplasmic signal sequences are added to each protease gene through polymerase chain reaction (PCR) amplification. Alternatively, the protease genes are synthesized to contain the periplasmic leader signal sequence. Complete synthesis enables codon optimization for efficient expression in the host strain. Suitable proteases and periplasmic targeting/leader sequences are described above in Tables 2 and 3, respectively.

[00133] The library of periplasmically-targeted protease genes is subcloned into a pDuet vector, or other suitable vector, for inducible expression. *E.coli* strain BL21(DE3), or similar

strain expressing the T7 polymerase, is transformed with the periplasmically-targeted protease library. Expression modification is achieved through use of varying levels of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (each vector expresses LacI, and the T7 promoter contains a *lac* operator site), through use of various promoters and ribosome binding sites, as well as through use of copy number variation among different pDuet vectors. Other plasmids, expression systems, or strains familiar to those skilled in the art may also be used.

**[00134]** Separate cultures harboring individual library members are grown to intermediate optical density (OD) in rich defined media prior to expression induction with 0.02 – 1 mM IPTG. Expression is induced for several hours to enable buildup of protease in the periplasm. Post-induction growth is monitored to determine which library members, if any, hinder cell growth presumably due to off-target cleavage of native *E. coli* periplasmic proteins. Severe growth inhibition may indicate that the expressed, active protease targets an important or essential protein(s). This may be alleviated by identifying the off-target protein(s) and modifying their amino acid sequence(s) to eliminate the specific cleavage site(s).

Periplasmically-targeted protease is extracted using osmotic shock, or other methods known to those skilled in the art. Verification of expression of full-length protein is determined by denaturing protein gel electrophoresis with appropriate size standards. More efficient periplasmic export can be achieved by optimizing expression through use of varying IPTG levels, differing plasmid origins of replication on the plasmid being used, and/or modification of ribosomal binding site (RBS).

**[00135]** The specific activity of periplasmically-targeted protease is determined in periplasmic extract by designing a target protein that contains a cleavage site and that, upon cleavage by the site-specific protease, will give a measurable signal. One example is outlined by Sarath and Schwartzbach in *Current Protocols in Protein Science* (2001) 21.9.1 – 21.9.10. This method includes the use of a fluorescent reporter protein containing a purification tag on either the N- or C-terminus. A cleavage site is inserted between the purification tag and the open reading frame of the fluorescent reporter, creating a target for a specific protease. Purified reporter target can be added to periplasmic extract containing the site-specific protease, and this mixture can be separated using an affinity resin that binds the purification tag. Cleaved, unbound fluorescent reporter is separated, and fluorescence is measured to determine the extent of protease cleavage. Alternatively, the engineered fluorescent reporter can be expressed in the cytoplasm of the strain expressing a periplasmically-targeted

protease. Uncleaved reporter can be separated from the lysate, and the level of fluorescence in the lysate can be used to determine the extent of protease target cleavage.

[00136] A specific non-limiting example of a pathway is the pathway for the synthesis of shikimic acid. In this pathway, for example, a reaction between the cellular commodity compounds phosphoenolpyruvate ( $S_{1A}$ ) and erythrose-4-phosphate ( $S_{1B}$ ) is catalyzed by the isoenzymes of DAHP synthase, AroF, G, or H ( $E_1$ ) to form 3-deoxy-D-arabinoseheptulose-7-phosphate (DAHP). DAHP ( $S_2$ ) is transformed to 3-dehydroquinate (3-DHQ) by the second enzyme in the pathway, DHQ synthase, AroB ( $E_2$ ). 3-DHQ ( $S_3$ ) is dehydrated to 3-dehydroshikimate by the third enzyme in the pathway, 3-DHQ dehydratase, AroD ( $E_3$ ). 3-dehydroshikimate ( $S_4$ ) is reduced to shikimic acid ( $P_F$ ) by the fourth enzyme in the pathway, shikimate dehydrogenase, AroE ( $E_4$ ), using NADPH as a cofactor. The enzymes of the pathway are known in the art and have been characterized in a number of organisms, including, for example, *E. coli*, in which the enzymes are encoded by the genetic loci as follows: DAHP synthase (*aroG*, *F*, or *H*); DHQ synthase (*aroB*); 3-DHQ dehydratase (*aroD*); shikimate dehydrogenase (*aroE*).

[00137] In the shikimic acid pathway, the overexpression of *aroE*, *aroF*, and *tktA* has been shown to increase the production of shikimic acid either directly with *aroE* and *aroF*, or indirectly with *tktA*, which increases the supply of the substrate molecule erythrose-4-phosphate.

[00138] There are also competing enzymes relevant to this pathway, which utilize the desired final product as a substrate, which enzymes are shikimate kinase (AroK) and shikimate kinase II (AroL). The presence of active competing enzymes may be undesirable, as they reduce the amount of desired product in the reaction mixture.

[00139] The sequence encoding the competing enzyme, shikimate kinase II (Genbank accession number NP\_414922.1) is made susceptible to proteolysis by a site-specific protease, by introducing a protease cleavage site in the target protein through introduction of specific mutations, by recombination, or multiplex genome engineering. The site for protease inactivation is introduced at a suitable site in the enzyme. Alternatively, shikimate kinase II is

inactivated by the methods set forth in Datsenko, K.A., and B.L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.* 97:6640-6645, and shikimate kinase I is made susceptible to proteolysis by a site-specific protease.

**[00140]** Included in the production of shikimic acid may be expression of DAHP synthase and/or additional enzymes in the biochemical pathway to shikimic acid:

<b>Table 4. Examples of enzymes in the shikimic acid biochemical pathway.</b>		
<b>EC#</b>	<b>Enzyme</b>	<b>Gene</b>
2.2.1.1	transketolase	<i>tktA</i>
4.2.3.4	dehydroquinase synthase	<i>aroB</i>
4.2.1.10	dehydroquinase dehydratase	<i>aroD</i>
1.1.1.25	shikimate dehydrogenase	<i>aroE</i>

**[00141]** The sequence of the enzyme is analyzed to identify the amino acid sequence closest to the protease cleavage motif, so that the least number of modifications to the amino acid sequence is needed to result in a protease cleavage site. The ideal site is on the surface of the protein, to minimize the effects on enzyme function. A site near the surface also ensures that it is accessible to the protease when folded. Alternatively, an enzyme from a different organism that carries out the same function and contains a protease cleavage site can be used instead of a native enzyme that is not protease labile.

**[00142]** A suitable site for cleavage (a slash indicates alternative amino acids) is the sequence ENLYFQ<sup>G</sup>/<sub>S</sub> (cleaved by the tobacco etch virus protease). An alternative site for cleavage is GARR<sup>G</sup>/<sub>S</sub> (cleaved by the yellow fever virus protease). An alternative site for cleavage is LVPRGS (cleaved by thrombin). An alternative site for cleavage is I<sup>E</sup>/<sub>D</sub>GR (cleaved by thrombin). The genetically modified sequence, which may be referred to the protease-labile form of the sequence, is expressed from a plasmid with concomitant inactivation of the chromosomal copy, or used to alter the chromosomal sequence of the host organism. In an alternative embodiment, the gene that encodes AroL is inactivated, and AroK is made susceptible to proteolysis.

**[00143]** Enzymatic assays are conducted, after a brief scanning of the sequences, to ensure that the chosen protease cleaves the target protein but not any other protein needed for shikimic acid biosynthesis. To accomplish this, the cells modified with the protease-labile competing enzyme are further modified to express the cognate protease under the control of an inducible promoter. The cells are grown in medium containing glucose as a carbon source. The production of shikimic acid in the cells is measured using nuclear magnetic resonance (<sup>1</sup>HNMR) spectroscopy or high-performance liquid chromatography (HPLC), as described by Knop *et al.* (2001) *J. Am. Chem. Soc.* 123:10173-10182 or van Hess *et al.* (1999) *Talanta*, 5 January 1999, Pages 173-177], or any other analytical chemistry technique, such as GC-MS.

[00144] Alternatively individual enzymatic assays, *e.g.*, in which the depletion of an enzymatic substrate or accumulation of its product is measured in a solution with the purified enzyme protein, are performed.

[00145] If any protein in the shikimic acid pathway is found to be susceptible to proteolysis, other than the competing enzyme(s), cleavage sites are removed by making conservative amino acid changes by site-directed mutagenesis. Alternatively, an enzyme from a different organism that carries out the same function, but does not contain a protease cleavage site, can be used instead of a native enzyme that is protease labile. Such analysis is repeated if needed until an appropriate protease is found.

***Example 2. Expression and activity of periplasmically-targeted enterokinase***

[00146] *Plasmids, cell growth and expression.* The gene coding for bovine enterokinase (Genbank Accession L19663.1) was chemically synthesized with codon content optimized for expression in *E. coli* and with the following periplasmic signal sequences:

<b>Table 5. Periplasmic signal sequences</b>	
DsbA (DA)	atgaaaaagatttgctggcggctggctggttagtttagcgttagcgcacgcggcg (SEQ ID NO:17)
OmpA (OA)	atgaaaaaacggcaattgcgatagcgggtgcgctagctggtttgccacggtgccgcaggct (SEQ ID NO:18)
MalE (ME)	atgaaaataaaaacaggtgcacgcacatcctcgattatccgcattaacgcacgatgatgtttccgcctcggtctcgcc (SEQ ID NO:19)

[00147] These constructs were subcloned into pACYC-Duet (EMD Chemicals, Gibbstown, NJ), and *E. coli* strain BL21(DE3) was transformed with the resulting plasmids. 250 ml cultures were grown from glycerol stocks for 6 h at 37 °C and 300 rpm in EZ rich medium with 34 µg/ml chloramphenicol (cam) to an optical density (OD600) of 0.6 – 0.8. A strain carrying the empty pACYC-Duet vector was used as a control. Expression was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cultures were grown for 16 h at 25 °C. After induction, cells were harvested and the periplasmic fraction was extracted.

[00148] *Periplasmic extraction and activity assay.* Cells were harvested from 12 ml of culture by centrifugation at 3000xg for 30 min at 4 °C. The supernatant was discarded, and the cell pellet was gently re-suspended in 2 ml ice-cold 1 mM Tris-HCl. This suspension was then centrifuged at 3000xg for 30 min at 4 °C. The supernatant was collected as the

periplasmic fraction. Activity assays were conducted by incubating 3.6 µg of the periplasmic fraction with 4 µg of control protein (an enterokinase cleavable protein provided by EMD) in reaction buffer containing 50 mM NaCl, 20 mM Tris-HCl, 2 mM CaCl<sub>2</sub> pH 7.4. The reactions were incubated at 30 °C for 24 h. The reactions were quenched by adding 19.5 µL of sodium dodecyl sulfate (SDS)-PAGE running buffer containing 0.04 M dithiothreitol (DTT), and boiling at 99 °C for 5 min. Reactions were run on a NuPAGE® 12% Bis-Tris Gel at 200 V for 50 minutes, and gels were stained with Coomassie dye. Cleavage of the 48 kDa control protein by enterokinase (EK) gives two proteolytic fragments of 32 and 16 kDa. The gels shown in Figure 1 were used to visualize the protease cleavage products and indicate that the periplasmically localized enterokinase is active in cleaving the target protein.

### ***Example 3. Expression of periplasmically-targeted enterokinase***

**[00149]** The gene sequence for bovine enterokinase was codon-optimized for expression in *E. coli* and subcloned into the expression vector pACYCDuet-1 (Novagen) with the addition of the OmpA periplasmic expression signal sequence. The construct was transformed into *E. coli* strain BL21(DE3). Cells were grown to an optical density (OD<sub>600</sub>) of 0.6, expression was induced by addition of 0.8 mM IPTG and continued growth at 37 °C for 3 hours. The culture underwent 2.4 doublings during induction, indicating that enterokinase expression did not inhibit growth. Cells were harvested by centrifugation, resuspended in assay buffer (20 mM Tris HCl pH 7.4; 50 mM NaCl; 2 mM CaCl<sub>2</sub>) and lysed using a homogenizer.

### ***Example 4. Enterokinase activity in whole-cell lysate***

**[00150]** Whole cell lysate from enterokinase-expressing cells was tested for protease activity by incubation with a cleavage control protein containing an enterokinase cleavage site (Asp-Asp-Asp-Asp-Lys) and compared to the activity of recombinant enterokinase. Recombinant enterokinase and cleavage control protein were obtained from Novagen. 40 µl of lysate or 2 units of recombinant enterokinase in 40 µl of assay buffer (20 mM Tris HCl pH 7.4; 50 mM NaCl; 2 mM CaCl<sub>2</sub>) were incubated with 4 µg of cleavage control protein at room temperature. A similarly prepared lysate from a strain not expressing enterokinase was used as a negative control. Zero- and 2-hour samples were analyzed by immunoblot with S protein horseradish peroxidase (HRP) conjugate that recognizes the amino terminal S tag on the cleavage control protein. Figure 2 shows detection of the 16 kDa protease cleavage fragment. The target protein was specifically cleaved by recombinant enterokinase (lane 2)

and by lysate from a strain expressing periplasmically-targeted enterokinase (lane 4), but not by a lysate lacking enterokinase expression (lane 6).

***Example 5. Target enzyme inactivation by protease release from the periplasm***

[00151] The gene sequence for bovine enterokinase is codon optimized for expression in *E. coli* cells and subcloned into an inducible plasmid expression vector with the addition of a periplasmic expression signal sequence. A gene that encodes the AroL enzyme is modified to include an enterokinase-specific cleavage site. *E. coli* cells are modified to express the site-specific protease in the periplasm and the modified target enzyme in the cytoplasm. The cells are grown to an optical density (OD<sub>600</sub>) of 0.6, expression of the enterokinase is induced, and the cells are allowed to continue to growth at 37 °C for 3 hours.

[00152] Cells are lysed using a homogenizer, thereby bringing into contact the site-specific protease and the modified enzyme. The cell lysate is combined with one or more substrates, enzymes, nutrients, co-factors, buffers, reducing agents, and/or ATP generating systems, as required for activity of the pathway enzymes, thereby forming a cell-free system. The redox potential, pH, and ionic strength of the cell-free system is optimized. The production of shikimic acid in the cell-free system is measured using <sup>1</sup>HNMR spectroscopy or HPLC.

***Example 6. PurF inactivation by protease release from the periplasm***

[00153] A library is generated, each member of which encodes an amidophosphoribosyl transferase enzyme (*E. coli* PurF) that contains an enterokinase-specific cleavage site at different locations within the polypeptide sequence of the enzyme. Choice of cleavage site location is aided by the published PurF tertiary structure as reported in Muchmore *et al.*, *Protein Science* (1998) 7:39-51. For example, Muchmore *et al.* describes a region comprising a flexible solvent exposed loop that is suitable for cleavage site incorporation. The library is generated by PCR, gene synthesis or other standard molecular biology methods and constructed in vectors for controlled expression in *E. coli* cells. *E. coli* cells are transformed with the vectors expressing the modified enzyme library. The cells are grown to an optical density (OD<sub>600</sub>) of 0.6, expression of the library is induced, and the cells are allowed to continue growth at 37 °C for 3 hours.

[00154] Cells are lysed using a homogenizer, and cell lysate is treated with purified recombinant enterokinase (Novagen) to verify enterokinase-mediate cleavage of library members. Cleavage of PurF is monitored by SDS-PAGE. Library members that are cleaved by recombinant enterokinase are then tested for activity. Specifically, lysate from cells

overexpressing these library members is combined with PRPP and glutamine (10 mM each), as required for enzyme activity, thereby forming a cell-free system. Samples of the reaction are analyzed at a series of time points. The activity of modified PurF is monitored by the appearance of glutamate and the concomitant disappearance of glutamine in the cell-free reaction as measured by HPLC using pre-column derivatization with orthophthalaldehyde. Inactivation of modified PurF is tested by addition of recombinant enterokinase to the assay described above.

[00155] The most active members of the PurF library that show activity in the absence of enterokinase, and are cleaved and inactivated by enterokinase in the course of the reaction, are chosen for integration into a cell containing a pathway of interest that is inhibited by PurF activity (*e.g.*, a pathway requiring PRPP, such as a pathway producing a pyrimidine-derived compound). The cell additionally contains periplasmically expressed enterokinase and any enzymes required for the pathway of interest. The library member is knocked into the chromosomal locus of PurF, thus replacing the endogenous gene. It is verified that growth is not adversely affected by gene replacement.

[00156] Cells are lysed using a homogenizer, thereby bringing into contact the site-specific protease and the modified enzyme. The cell lysate is combined with one or more substrates, enzymes, nutrients, co-factors, buffers, reducing agents, and/or ATP generating systems, as required for activity of the pathway enzymes, thereby forming a cell-free system. The redox potential, pH, and ionic strength of the cell-free system is optimized. The production of the product of interest in the cell-free system is measured using appropriate methods.

#### OTHER EMBODIMENTS

[00157] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art.

If a definition set forth in this application is contrary to or otherwise inconsistent with a definition set forth in the patents, applications, published applications and other publications, the definition set forth in this application prevails.

[00158] Citation of publications or documents is not intended as an admission that any of such publications or documents are pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

**[00159]** As used herein, “a” or “an” means “at least one” or “one or more” unless otherwise indicated.

**[00160]** As used herein, the term “about” and “approximately” are used interchangeably, and the meaning of the terms can be obtained by considering the context of the terms as they are used in the specification and claims of the application. In some embodiments, “about” a particular numerical value encompasses a range of 30% of the value. In other embodiments, “about” a particular numerical value encompasses a range of 20% of the value. In yet other embodiments, “about” a particular numerical value encompasses a range of 10% of the value.

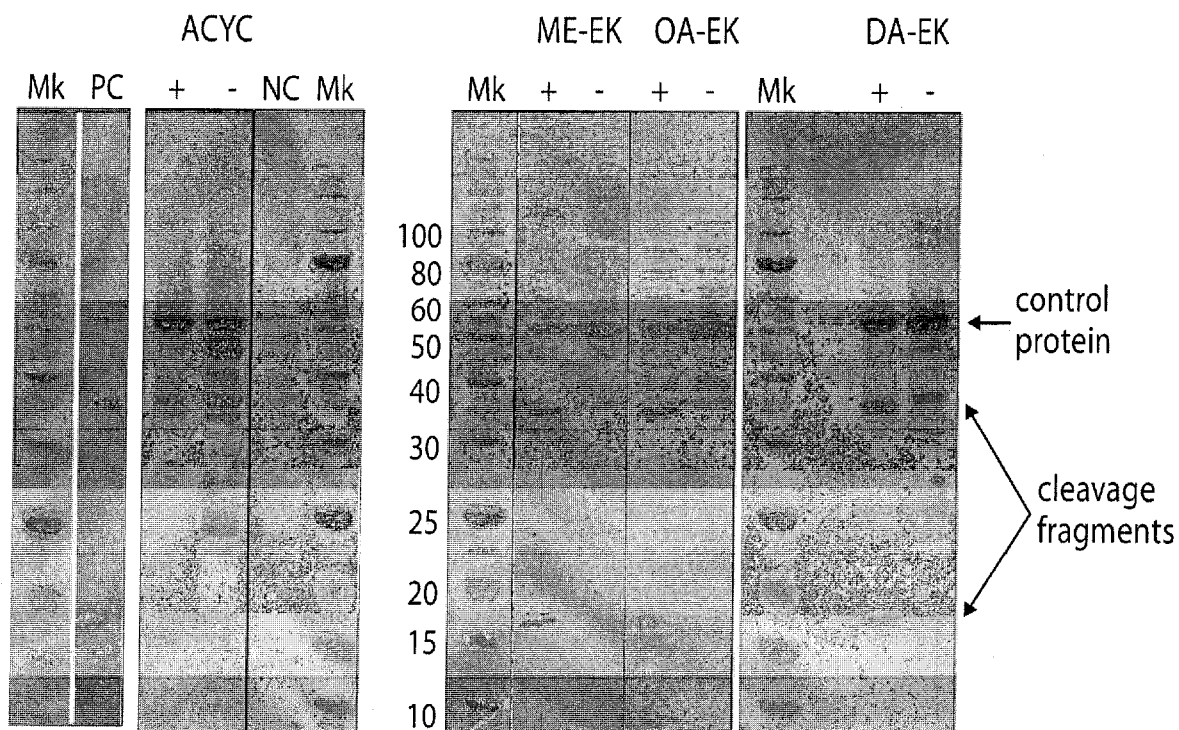
**What is claimed is:**

1. A method of producing a product of a biosynthetic pathway of interest, the method comprising:
  - (a) growing to a desired cell density genetically modified bacterial cells that encode:
    - (i) enzymes of the biosynthetic pathway of interest that catalyze production of the product,
    - (ii) a targeted enzyme that competes for substrates or cofactors with the enzymes of the biosynthetic pathway of interest, decreases the overall yield, and/or the rate of production of the product, and includes a HRV 3C protease recognition sequence, and
    - (iii) a HRV 3C protease that cleaves the HRV 3C protease recognition sequence of the targeted enzyme and is genetically modified to include a periplasmic targeting sequence;
  - (b) lysing at least a portion of the cells to produce a cell lysate, thereby bringing the HRV 3C protease into contact with the targeted enzyme to cleave and inactive or reduce the activity of the targeted enzyme;
  - (c) combining the cell lysate with one or more substrates, enzymes, nutrients, cofactors, buffers, reducing agents, or ATP generating systems; and
  - (d) incubating the cell lysate of step (c) to produce the product of the biosynthetic pathway of interest.
2. The method of claim 1, wherein the method further comprises combining the cell lysate of step (b) with one or more additional cell lysates.
3. The method of claim 1, wherein the periplasmic targeting sequence is selected from the group consisting of:
  - MKIKTGARILALSALTTMMFSASALA (SEQ ID NO:1),
  - MKKTAIAlAVALAGFATVAQA (SEQ ID NO:2),
  - MKQSTIALALLPLLFTPVTKA (SEQ ID NO:3),
  - MMITLRKLPLAVAVAAGVMSAQAMA (SEQ ID NO:4),

MNKKVLTLSAVMASMLFGAAHA (SEQ ID NO:5),  
MKYLLPTAAAGLLLLAAQPAMA (SEQ ID NO:6),  
MKKNIAFLASMFVFSIATNAYA (SEQ ID NO:7),  
MKKIWLALAGLVLAFSASA (SEQ ID NO:8),  
MMTKIKLLMLIIFYLIISASAHA (SEQ ID NO:9),  
MKQALRVAFGFLILWASVLHA (SEQ ID NO:10), and  
MRVLLFLLLSLFMLPAFS (SEQ ID NO:11).

4. The method of claim 1, wherein the HRV 3C protease is encoded by a nucleotide sequence that is operably linked to an inducible promoter.
5. The method of claim 1, wherein the targeted enzyme competes with an enzyme that increases the rate of precursor supplied to the pathway of interest.
6. The method of claim 5, wherein the targeted enzyme competes with a key pathway entry enzyme of the biosynthetic pathway of interest.
7. The method of claim 1, wherein a gene encoding a native counterpart of the targeted enzyme is replaced with a gene encoding the targeted enzyme.
8. The method of claim 1, wherein a gene encoding a native counterpart of the targeted enzyme is knocked out.
9. The method of claim 1, wherein the targeted enzyme is over-expressed in the bacterial cells.
10. The method of claim 1, wherein the targeted enzyme is present on either an episomal vector or a chromosome.

11. The method of claim 1, wherein the product is selected from the group consisting of antibiotics, biosurfactants, biological fuels, amino acids, organic acids, fatty acids, alcohols, polyols, flavors, fragrances, nucleotides, vitamins, pigments, sugars, polysaccharides, plastics, isoprenoids, and terpenes.
12. The method of claim 1, wherein the bacterial cells are *Escherichia coli* cells.
13. The method of claim 1, wherein the activity of the targeted enzyme is reduced by about 50% upon cleavage by the HRV 3C protease.
14. The method of claim 13, wherein the activity of the targeted enzyme is reduced by about 90% upon cleavage by the HRV 3C protease.
15. The method of claim 1, wherein the targeted enzyme is essential for growth of the bacterial cells.



+ = 0.1 mM IPTG  
 - = no IPTG  
 NC = neg. control (no control protein)  
 PC = pos. control (purified EK w/ control protein)  
 Mk = molecular weight marker (kDa)

Fig. 1

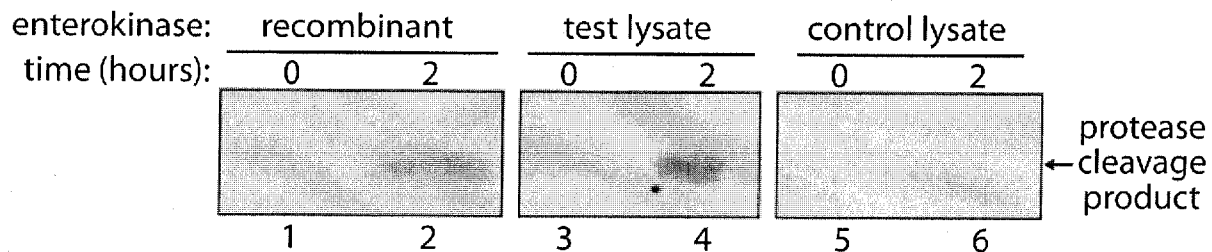
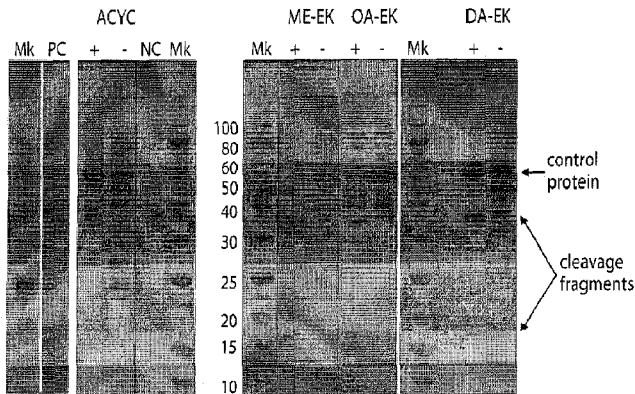


Fig. 2



+ = 0.1 mM IPTG  
 - = no IPTG  
 NC = neg. control (no control protein)  
 PC = pos. control (purified EK w/ control protein)  
 Mk = molecular weight marker (kDa)