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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 September 2014 (18.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/144649 A1

(51) International Patent Classification:
C12N 9/10 (2006.01)

(21) International Application Number:
PCT/US2014/029150

(22) International Filing Date:
14 March 2014 (14.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/788,346 15 March 2013 (15.03.2013) US

(71) Applicant: **MODULAR GENETICS, INC.** [US/US]; 12-T Cabot Road, Woburn, Massachusetts 01801 (US).

(72) Inventors: **JARRELL, Kevin A.**; 19 Moccasin Hill Road, Lincoln, Massachusetts 01773 (US). **REZNIK, Gabriel O.**; 124 Bishops Forest Drive, Waltham, Massachusetts 02452 (US). **VISHWANATH, Prashanth**; 186 Gardner Street, Apt. 1-3, Arlington, Massachusetts 02474 (US). **PYNN, Michelle**; 2A Florence Street, Andover, Massachusetts 01810 (US).

(74) Agents: **JARRELL, Brenda Herschbach** et al.; Choate, Hall & Stewart LLP, Two International Place, Boston, Massachusetts 02110 (US).

(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, BN, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

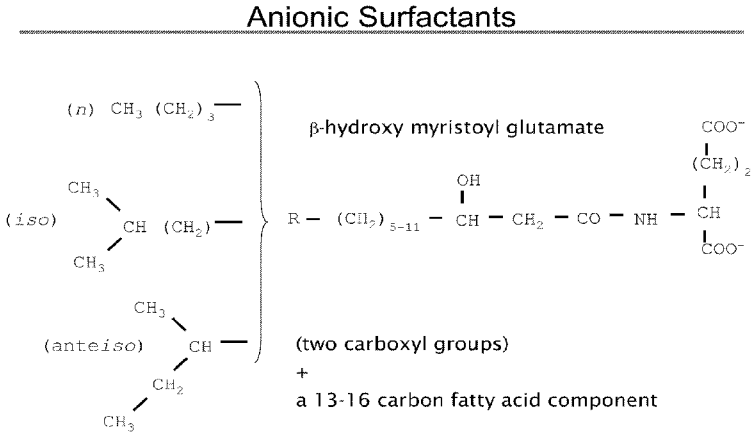
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: GENERATION OF ACYL AMINO ACIDS



(57) Abstract: In certain embodiments, the present invention comprises compositions and methods useful in the generation of acyl amino acids. In certain embodiments, the present invention provides an engineered polypeptide comprising a peptide synthetase domain; in some such embodiments, the engineered polypeptide comprises only a single peptide synthetase domain. In some embodiments, the present invention provides an engineered peptide synthetase that is substantially free of a thioesterase domain, and/or a reductase domain. In certain embodiments, the present invention provides an acyl amino acid composition comprising a plurality of different forms of an acyl amino acid. In some such compositions, substantially all of the acyl amino acids within the composition contain the same amino acid moiety and differ with respect to acyl moiety. We also described populations where the fatty acid si for example 95% one length (C14, myristic).

WO 2014/144649 A1

GENERATION OF ACYL AMINO ACIDS

Cross-Reference to Related Applications

[0001] This application claims the benefit of U.S. Provisional Application No. , filed March 15, 2013, the contents of which are hereby incorporated herein in its entirety.

Background

[0002] Acyl amino acids are commercially important compounds. Many have advantageous characteristics and are sold as surfactants, antibiotics, anti-insect agents and as a variety of other important agents.

[0003] Traditionally, acyl amino acids have been manufactured chemically. Such chemical manufacturing methods are hampered by a variety of shortcomings including the ease of obtaining and storing the starting materials, the necessity of using harsh and sometimes dangerous chemical reagents in the manufacturing process, the difficulty and efficiency of the synthesis itself, the fiscal and environmental cost of disposing of chemical by-products, etc. Thus, new compositions and methods for the efficient and cost-effective synthesis of acyl amino acids and manufacture on a commercial scale would be beneficial.

[0004] Recently, important technologies have been developed that permit production of acyl amino acids by engineered peptide synthetase polypeptides (See US Patent Number 7981685, issued July 19, 2011 and incorporated herein by reference in its entirety). Improvements and/or supplements to such technologies would be desirable and beneficial.

Summary of the Invention

[0005] In certain embodiments, the present invention comprises compositions and methods useful in the generation of acyl amino acids. In certain embodiments, the present invention provides an engineered polypeptide comprising a peptide synthetase domain; in some such embodiments, the engineered polypeptide comprises only a single peptide synthetase domain. In some embodiments, the present invention provides an engineered peptide synthetase that is substantially free of a thioesterase domain, and/or a reductase domain.

[0006] In certain embodiments, the present invention provides an acyl amino acid composition comprising a plurality of different forms of an acyl amino acid. In some such compositions, substantially all of the acyl amino acids within the composition contain the same amino acid moiety and differ with respect to acyl moiety. We also described populations where the fatty acid si for example 95% one length (C14, myristic)

[0007] In some embodiments, the present invention provides a method of making an acyl amino acid composition by contacting an engineered peptide synthetase with an amino acid substrate and an acyl entity substrate for the engineered peptide synthetase, under conditions and for a time sufficient for an acyl amino acid composition to be made. In some embodiments, the method comprises providing a cell engineered to express the engineered peptide synthetase. In some embodiments, the engineered peptide synthetase does not include a thioesterase domain; in some embodiments, the engineered peptide synthestase does not include a reductase domain; in some embodiments, the engineered peptide synthetase includes neither a thioesterase domain not a reductase domain.

[0008] In some embodiments, an amino acid substrate is or comprises an amino acid as set forth herein.

[0009] In some embodiments, an acyl entity substrate is or comprises a fatty acid moiety. In some embodiments, an acyl entity substrate is or comprises a fatty acid.

[0010] The present invention provides cells engineered to express at least one engineered peptide synthetase that synthesizes an acyl amino acid.

[0011] In some embodiments, the present invention comprises an an acyl amino acid composition produced by an engineered peptide synthetase.

[0012] The present invention provides methods of preparing a product comprising: providing or obtaining an acyl amino acid composition prepared in an engineered host (e.g., microbial) cell; optionally enriching the acyl amino acid composition for a particular acyl amino acid; and, in some embodiments, combining the enriched acyl amino acid composition with at least one other component to produce a product.

[0013] In some embodiments, the invention provides a method comprising steps of: contacting an engineered peptide synthetase polypeptide that comprises a single peptide synthetase domain and lacks either of a thioesterase domain, and/or a reductase domain with (i)

an amino acid substate of the peptide synthetase polypeptide; and (ii) an acyl moiety substrate of the peptide synthetase polypeptide, the contacting being performed under conditions and for a time sufficient that the engineered peptide synthetase polypeptide covalently links the acyl moiety from the acyl moiety substrate to the amino acid so that an acyl amino acid is generated.

Description of Certain Embodiments

Definitions

[0014] *Acyl amino acid*: The term "acyl amino acid" as used herein refers to an amino acid that is covalently linked to a fatty acid moiety. In some embodiments, the amino acid and fatty acid are covalently linked via an amide bond formed between a carboxylic acid group of a fatty acid and an amino group of an amino acid. In some embodiments, a fatty acid moiety or entity utilized or included in an acyl amino acid includes a β-hydroxyl group; in some embodiments, a fatty acid moiety or entity utilized or included in an acyl amino acid does not include a β-hydroxyl group. In some embodiments, a fatty acid moiety utilized or included in an acyl amino acid includes a β-amino group; in some embodiments, a fatty acid moiety or entity utilized or included in an acyl amino acid does not include a β-aminno group. In some embodiments, a fatty acid moiety utilized or included in an acyl amino acid is unmodified at the β-position.

[0015] *Amino acid*: As used herein, the term "amino acid," in its broadest sense, refers to any compound and/or substance that can be utilized in peptide synthesis (e.g., ribosomal or non-ribosomal synthesis). In some embodiments, an amino acid is any compound and/or substance that can be incorporated into a polypeptide chain, e.g., through formation of one or more peptide bonds. In some embodiments, an amino acid is any compound and/or substance that is a substrate for a peptide synthetase; in some such embodiments, an amino acid is any compound and/or substance onto which a peptide synthetase can link an acyl entity, for example through formation of an amide bond. In some embodiments, an amino acid has the general structure H_2N–C(H)(R)–COOH. In some embodiments, an amino acid is a naturally-occurring amino acid. In some embodiments, an amino acid is a synthetic amino acid; in some embodiments, an amino acid is a D-amino acid; in some embodiments, an amino acid is an L-amino acid. "Standard amino acid" refers to any of the twenty standard L-amino acids commonly found in

naturally occurring peptides. "Nonstandard amino acid" refers to any amino acid, other than the standard amino acids, regardless of whether it is prepared synthetically or obtained from a natural source. In some embodiments, an amino acid, including a carboxy- and/or amino-terminal amino acid in a polypeptide, can contain a structural modification as compared with the general structure above. For example, in some embodiments, an amino acid may be modified by methylation, amidation, acetylation, and/or substitution as compared with the general structure. In some embodiments, such modification may, for example, alter the circulating half life of a polypeptide containing the modified amino acid as compared with one containing an otherwise identical unmodified amino acid. In some embodiments, such modification does not significantly alter a relevant activity of a polypeptide containing the modified amino acid, as compared with one containing an otherwise identical unmodified amino acid. As will be clear from context, in some embodiments, the term "amino acid" is used to refer to a free amino acid; in some embodiments it is used to refer to an amino acid residue of a polypeptide. In some embodiments, a "naturally occurring" amino acid is one of the standard group of twenty amino acids that are the building blocks of polypeptides of most organisms, including alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine. In certain embodiments a "naturally occurring" amino acid may be one of those amino acids that are used less frequently and are typically not included in this standard group of twenty but are nevertheless still used by one or more organisms and incorporated into certain polypeptides. For example, the codons UAG and UGA normally encode stop codons in most organisms. However, in some organisms the codons UAG and UGA encode the amino acids selenocysteine and pyrrolysine. Thus, in certain embodiments, selenocysteine and pyrrolysine are naturally occurring amino acids.

[0016] *Associated with:* Two events or entities are "associated" with one another, as that term is used herein, if the presence, level and/or form of one is correlated with that of the other. For example, a particular entity (e.g., polypeptide) is considered to be associated with a particular disease, disorder, or condition, if its presence, level and/or form correlates with incidence of and/or susceptibility of the disease, disorder, or condition (e.g., across a relevant

population). In some embodiments, two or more entities are physically "associated" with one another if they interact, directly or indirectly, so that they are and remain in physical proximity with one another. In some embodiments, two or more entities that are physically associated with one another are covalently linked to one another; in some embodiments, two or more entities that are physically associated with one another are not covalently linked to one another but are non-covalently associated, for example by means of hydrogen bonds, van der Waals interaction, hydrophobic interactions, magnetism, and combinations thereof.

[0017] *Beta-hydroxy fatty acid linkage domain*: The term "beta-hydroxy fatty acid linkage domain" as used herein refers to a polypeptide domain that covalently links a beta-hydroxy fatty acid to an amino acid to form an acyl amino acid. A variety of beta-hydroxy fatty acid linkage domains are known to those skilled in the art. However, different beta-hydroxy fatty acid linkage domains often exhibit specificity for one or more beta-hydroxy fatty acids. As one non-limiting example, the beta-hydroxy fatty acid linkage domain from surfactin synthetase is specific for the beta-hydroxy myristic acid, which contains 13 to 15 carbons in the fatty acid chain. Thus, the beta-hydroxy fatty acid linkage domain from surfactin synthetase can be used in accordance with the present invention to construct an engineered polypeptide useful in the generation of an acyl amino acid that comprises the fatty acid beta-hydroxy myristic acid.

[0018] *Beta-hydroxy fatty acid*: The term "beta-hydroxy fatty acid" as used herein refers to a fatty acid chain comprising a hydroxy group at the beta position of the fatty acid chain. As is understood by those skilled in the art, the beta position corresponds to the third carbon of the fatty acid chain, the first carbon being the carbon of the carboxylate group. Thus, when used in reference to an acyl amino acid of the present invention, where the carboxylate moiety of the fatty acid has been covalently attached to the nitrogen of the amino acid, the beta position corresponds to the carbon two carbons removed from the carbon having the ester group. A beta-hydroxy fatty acid to be used in accordance with the present invention may contain any number of carbon atoms in the fatty acid chain. As non-limiting examples, a beta-hydroxy fatty acid may contain 3, 4, 5, 6, 7, 8,9, 10, 11, 12, 3, 14, 15, 15, 16, 17, 18, 19, 20 or more carbon atoms. Beta-hydroxy fatty acids to be used in accordance with the present invention may contain linear carbon chains, in which each carbon of the chain, with the exception of the terminal carbon atom and the carbon attached to the nitrogen of the amino acid, is directly covalently linked to two

other carbon atoms. Additionally or alternatively, beta-hydroxy fatty acids to be used in accordance with the present invention may contain branched carbon chains, in which at least one carbon of the chain is directly covalently linked to three or more other carbon atoms. Beta-hydroxy fatty acids to be used in accordance with the present invention may contain one or more double bonds between adjacent carbon atoms. Alternatively, beta-hydroxy fatty acids to be used in accordance with the present invention may contain only single-bonds between adjacent carbon atoms. A non-limiting exemplary beta-hydroxy fatty acid that may be used in accordance with the present invention is or comprises a beta-hydroxy, acid which contains 13 to 15 carbons in the fatty acid chain; in some embodiments, an exemplary beta-hydroxy fatty acid that may be used in accordance with the present invention is or comprises myristic acid myrisitc is usually used to mean 14 carbons Those of ordinary skill in the art will be aware of various beta-hydroxy fatty acids that can be used in accordance with the present invention. Different beta-hydroxy fatty acid linkage domains that exhibit specificity for other beta-hydroxy fatty acids (e.g., naturally or non-naturally occurring beta-hydroxy fatty acids) may be used in accordance with the present invention to generate any acyl amino acid of the practitioner's choosing.

[0019] *Characteristic sequence element:* As used herein, the phrase "characteristic sequence element" refers to a sequence element found in a polymer (e.g., in a polypeptide or nucleic acid) that represents a characteristic portion of that polymer. In some embodiments, presence of a characteristic sequence element correlates with presence or level of a particular activity or property of the polymer. In some embodiments, presence (or absence) of a characteristic sequence element defines a particular polymer as a member (or not a member) of a particular family or group of such polymers. A characteristic sequence element typically comprises at least two monomers (e.g., amino acids or nucleotides). In some embodiments, a characteristic sequence element includes at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, or more monomers (e.g., contiguously linked monomers). In some embodiments, a characteristic sequence element includes at least first and second stretches of continguous monomers spaced apart by one or more spacer regions whose length may or may not vary across polymers that share the sequence element.

[0020] *Combination therapy:* As used herein, the term "combination therapy" refers to those

situations in which a subject is simultaneously exposed to two or more therapeutic agents. In some embodiments, such agents are administered simultaneously; in some embodiments, such agents are administered sequentially; in some embodiments, such agents are administered in overlapping regimens.

[0021] *Comparable:* The term "comparable", as used herein, refers to two or more agents, entities, situations, sets of conditions, etc that may not be identical to one another but that are sufficiently similar to permit comparison therebetween so that conclusions may reasonably be drawn based on differences or similarities observed. Those of ordinary skill in the art will understand, in context, what degree of identity is required in any given circumstance for two or more such agents, entities, situations, sets of conditions, etc to be considered comparable.

[0022] *Corresponding to*: As used herein, the term "corresponding to" is often used to designate the position/identity of a residue in a polymer, such as an amino acid residue in a polypeptide or a nucleotide residue in a nucleic acid. Those of ordinary skill will appreciate that, for purposes of simplicity, residues in such a polymer are often designated using a canonical numbering system based on a reference related polymer, so that a residue in a first polymer "corresponding to" a residue at position 190 in the reference polymer, for example, need not actually be the 190^{th} residue in the first polymer but rather corresponds to the residue found at the 190^{th} position in the reference polymer; those of ordinary skill in the art readily appreciate how to identify "corresponding" amino acids, including through use of one or more commercially-available algorithms specifically designed for polymer sequence comparisons.

[0023] *Domain, Polypeptide domain*: The terms "domain" and "polypeptide domain" as used herein generally refer to polypeptide moieties that display a particular activity, even when isolated (e.g., cleaved) from other polypeptides or polypeptide domains. In some embodiments, a polypeptide domain folds into a particular discrete structure in three-dimensional space. In some embodiments, a polypeptide domain within a longer polypeptide is separated from one or more other polypeptide domains within the longer polypeptide by virture of a linker element, for example, that may comprise a substantially unstructured stretch of amino acids. In some embodiments, the terms refer to domains that naturally occur in longer polypeptides; in some embodiments, the term refers to engineered polypeptide moieties that correspond and/or show

significant homology and/or identity to such naturally occurring polypeptide moieties, or to other reference polypeptide moieties (e.g., historical engineered moieties). In some embodiments, an engineered domain that corresponds and/or shows significant homology and/or identity to a naturally occurring or other reference moiety shares a characteristic structure (e.g., primary structure such as the amino acid sequence of the domain, and/or secondary, tertiary, quaternary, etc. structures); alternatively or additionally, such an engineered domain may exhibit one or more distinct functions that it shares with its reference polypeptide moieties. As will be understood by those skilled in the art, in many cases polypeptides are modular and are comprised of one or more polypeptide domains; in some such embodiments, each domain exhibits one or more distinct functions that contribute to the overall function of the polypeptide. In some embodiments, the structure and/or function of many such domains are known to those skilled in the art.

[0024] *Engineered*: The term "engineered" as used herein refers to a non-naturally occurring moiety that has been created by the hand of man. For example, in reference to a polypeptide, an "engineered polypeptide" refers to a polypeptide that has been designed and/or produced by the hand of man. In some embodiments, an engineered polypeptide has an amino acid sequence that includes one or more sequence elements that do(es) not occur in nature. In some embodiments, an engineered polypeptide has an amino acid sequence that includes one or more sequence elements that does occur in nature, but that is present in the engineered polypeptide in a different sequence context (e.g., separated from at least one sequence to which it is linked in nature and/or linked with at least one sequence element to which it is not linked in nature) from that in which it occurs in nature. In some embodiments, an engineered polypeptide is one in which naturally-occurring sequence element(s) is/are separated from at least one sequence with which they/it is associated (e.g., linked) in nature and/or is otherwise manipulated to comprise a polypeptide that does not exist in nature. In various embodiments, an engineered polypeptide comprises two or more covalently linked polypeptide domains. Typically such domains will be linked via peptide bonds, although the present invention is not limited to engineered polypeptides comprising polypeptide domains linked via peptide bonds, and encompasses other covalent linkages known to those skilled in the art. One or more covalently linked polypeptide domains of engineered polypeptides may be naturally occurring. Thus, in certain embodiments, engineered

polypeptides of the present invention comprise two or more covalently linked domains, at least one of which is naturally occurring. In certain embodiments, two or more naturally occurring polypeptide domains are covalently linked to generate an engineered polypeptide. For example, naturally occurring polypeptide domains from two or more different polypeptides may be covalently linked to generate an engineered polypeptide. In certain embodiments, naturally occurring polypeptide domains of an engineered polypeptide are covalently linked in nature, but are covalently linked in the engineered polypeptide in a way that is different from the way the domains are linked nature. For example, two polypeptide domains that naturally occur in the same polypeptide but which are separated by one or more intervening amino acid residues may be directly covalently linked (e.g., by removing the intervening amino acid residues) to generate an engineered polypeptide of the present invention. Additionally or alternatively, two polypeptide domains that naturally occur in the same polypeptide which are directly covalently linked together (e.g., not separated by one or more intervening amino acid residues) may be indirectly covalently linked (e.g., by inserting one or more intervening amino acid residues) to generate an engineered polypeptide of the present invention. In certain embodiments, one or more covalently linked polypeptide domains of an engineered polypeptide may not exist naturally. For example, such polypeptide domains may be engineered themselves.

[0025] *Fatty acid linkage domain*: The term "fatty acid linkage domain" as used herein refers to a polypeptide domain that covalently links a fatty acid to an amino acid to form an acyl amino acid. In some embodiments, a fatty acid linkage domain is a condensation domain; in some embodiments such a fatty acid linkage domain is part of a single polypeptide or a polypeptide complex with at least or only an adenylkation domain, a thiolation domain, or both. A variety of fatty acid linkage domains are known in the art, such as for example, fatty acid linkage domains present in various peptide synthetase complexes that produce lipopeptides. In certain embodiments, a fatty acid linkage domain links a beta-hydroxy fatty acid to an amino acid; in some embodiments, a fatty acid linkage domain links a beta-amino fatty acid to an amino acid; in some embodiments, a fatty acid linkage domain links a fatty acid that is unmodified at the beta position to an amino acid. In some embodiments, a fatty acid linkage domain catalyzes condensation of a fatty acid and an amino acid so that an amide both is formed, for example between a carboxylic acid moiety on a fatty acid and an amino moiety on an amino acid.

[0026] *Homology:* As used herein, the term "homology" refers to the overall relatedness between polymeric molecules, e.g., between nucleic acid molecules (*e.g.*, DNA molecules and/or RNA molecules) and/or between polypeptide molecules. In some embodiments, polymeric molecules are considered to be "homologous" to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identical. In some embodiments, polymeric molecules are considered to be "homologous" to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% similar (e.g., containing residues with related chemical properties at corresponding positions). For example, as is well known by those of ordinary skill in the art, certain amino acids are typically classified as similar to one another as "hydrophobic" or "hydrophilic"amino acids, and/or as having "polar" or "non-polar" side chains. Substitution of one amino acid for another of the same type may often be considered a "homologous" substitution. Typical amino acid categorizations are summarized below:

Alanine	Ala	A	nonpolar	neutral	1.8
Arginine	Arg	R	polar	positive	-4.5
Asparagine	Asn	N	polar	neutral	-3.5
Aspartic acid	Asp	D	polar	negative	-3.5
Cysteine	Cys	C	nonpolar	neutral	2.5
Glutamic acid	Glu	E	polar	negative	-3.5
Glutamine	Gln	Q	polar	neutral	-3.5
Glycine	Gly	G	nonpolar	neutral	-0.4
Histidine	His	H	polar	positive	-3.2
Isoleucine	Ile	I	nonpolar	neutral	4.5
Leucine	Leu	L	nonpolar	neutral	3.8
Lysine	Lys	K	polar	positive	-3.9
Methionine	Met	M	nonpolar	neutral	1.9
Phenylalanine	Phe	F	nonpolar	neutral	2.8
Proline	Pro	P	nonpolar	neutral	-1.6
Serine	Ser	S	polar	neutral	-0.8
Threonine	Thr	T	polar	neutral	-0.7
Tryptophan	Trp	W	nonpolar	neutral	-0.9
Tyrosine	Tyr	Y	polar	neutral	-1.3
Valine	Val	V	nonpolar	neutral	4.2

Ambiguous Amino Acids	3-Letter	1-Letter
Asparagine or aspartic acid	Asx	B
Glutamine or glutamic acid	Glx	Z
Leucine or Isoleucine	Xle	J
Unspecified or unknown amino acid	Xaa	X

As will be understood by those skilled in the art, a variety of algorithms are available that permit

comparison of sequences in order to determine their degree of homology, including by permitting gaps of designated length in one sequence relative to another when considering which residues "correspond" to one another in different sequences. Calculation of the percent homology between two nucleic acid sequences, for example, can be performed by aligning the two sequences for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second nucleic acid sequences for optimal alignment and non-corresponding sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or substantially 100% of the length of the reference sequence. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position; when a position in the first sequence is occupied by a similar nucleotide as the corresponding position in the second sequence, then the molecules are similar at that position. The percent homology between the two sequences is a function of the number of identical and similar positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which needs to be introduced for optimal alignment of the two sequences. Representative algorithms and computer programs useful in determining the percent homology between two nucleotide sequences include, for example, the algorithm of Meyers and Miller (CABIOS, 1989, 4: 11-17), which has been incorporated into the ALIGN program (version 2.0) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. The percent homology between two nucleotide sequences can, alternatively, be determined for example using the GAP program in the GCG software package using an NWSgapdna.CMP matrix.

[0027] *Identity:* As used herein, the term "identity" refers to the overall relatedness between polymeric molecules, *e.g.*, between nucleic acid molecules (e.g., DNA molecules and/or RNA molecules) and/or between polypeptide molecules. In some embodiments, polymeric molecules are considered to be "substantially identical" to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identical. As will be understood by those skilled in the art, a variety of algorithms are available

that permit comparison of sequences in order to determine their degree of homology, including by permitting gaps of designated length in one sequence relative to another when considering which residues "correspond" to one another in different sequences. Calculation of the percent identity between two nucleic acid sequences, for example, can be performed by aligning the two sequences for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second nucleic acid sequences for optimal alignment and non-corresponding sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or substantially 100% of the length of the reference sequence. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which needs to be introduced for optimal alignment of the two sequences. Representative algorithms and computer programs useful in determinng the percent identity between two nucleotide sequences include, for example, the algorithm of Meyers and Miller (CABIOS, 1989, 4: 11-17), which has been incorporated into the ALIGN program (version 2.0) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. The percent identity between two nucleotide sequences can, alternatively, be determined for example using the GAP program in the GCG software package using an NWSgapdna.CMP matrix.

[0028] *Isolated:* As used herein, the term "isolated" refers to a substance and/or entity that has been (1) separated from at least some of the components with which it was associated when initially produced (whether in nature and/or in an experimental setting), and/or (2) designed, produced, prepared, and/or manufactured by the hand of man. Isolated substances and/or entities may be separated from about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% of the other components with which they were initially associated. In some embodiments, isolated agents are about 80%,

about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% pure. As used herein, a substance is "pure" if it is substantially free of other components. In some embodiments, as will be understood by those skilled in the art, a substance may still be considered "isolated" or even "pure", after having been combined with certain other components such as, for example, one or more carriers or excipients (*e.g.*, buffer, solvent, water, etc.); in such embodiments, percent isolation or purity of the substance is calculated without including such carriers or excipients. In some embodiments, isolation involves or requires disruption of covalent bonds (e.g., to isolate a polypeptide domain from a longer polypeptide and/or to isolate a nucleotide sequence element from a longer oligonucleotide or nucleic acid).

[0029] *Naturally occurring*: The term "naturally occurring", as used herein, refers to an agent or entity that is known to exist in nature.

[0030] *Nucleic acid:* As used herein, the term "nucleic acid," in its broadest sense, refers to any compound and/or substance that is or can be incorporated into an oligonucleotide chain. In some embodiments, a nucleic acid is a compound and/or substance that is or can be incorporated into an oligonucleotide chain via a phosphodiester linkage. As will be clear from context, in some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.*, nucleotides and/or nucleosides); in some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising individual nucleic acid residues. In some embodiments, a "nucleic acid" is or comprises RNA; in some embodiments, a "nucleic acid" is or comprises DNA. In some embodiments, a nucleic acid is, comprises, or consists of one or more natural nucleic acid residues. In some embodiments, a nucleic acid is, comprises, or consists of one or more nucleic acid analogs. In some embodiments, a nuclic acid analog differs from a nucleic acid in that it does not utilize a phosphodiester backbone. For example, in some embodiments, a nucleic acid is, comprises, or consists of one or more "peptide nucleic acids", which are known in the art and have peptide bonds instead of phosphodiester bonds in the backbone, are considered within the scope of the present invention. Alternatively or additionally, in some embodiments, a nucleic acid has one or more phosphorothioate and/or 5'-N-phosphoramidite linkages rather than phosphodiester bonds. In some embodiments, a nucleic acid is, comprises, or consists of one or

more natural nucleosides (e.g., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine). In some embodiments, a nucleic acid is, comprises, or consists of one or more nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, C-5 propynyl-cytidine, C-5 propynyl-uridine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, 2-thiocytidine, methylated bases, intercalated bases, and combinations thereof). In some embodiments, a nucleic acid comprises one or more modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose) as compared with those in natural nucleic acids. In some embodiments, a nucleic acid has a nucleotide sequence that encodes a functional gene product such as an RNA or protein. In some embodiments, a nucleic acid includes one or more introns. In some embodiments, nucleic acids are prepared by one or more of isolation from a natural source, enzymatic synthesis by polymerization based on a complementary template (*in vivo* or *in vitro*), reproduction in a recombinant cell or system, and chemical synthesis. In some embodiments, a nucleic acid is at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 20, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000 or more residues long.

[0031] *Peptide synthetase complex*: The term "peptide synthetase complex" as used herein refers to an enzyme that catalyzes the non-ribosomal production of peptides. As will be appreciated by those of ordinary skill in the art, peptide synthetase complexes are modular, and are comprised of individual peptide synthetase modules that perform different steps in the synthesis of the ultimate peptide; typically, each module performs one step (e.g., adds a single amino acid). A peptide synthetase complex may comprise a single enzymatic subunit (e.g., a single polypeptide), or may comprise two or more enzymatic subunits (e.g., two or more polypeptides). A peptide synthetase complex typically comprises at least one peptide synthetase domain, and may further comprise one or more additional domains such as for example, a fatty acid linkage domain, a thioesterase domain, a reductase domain, etc. Peptide synthetase domains of a peptide synthetase complex may comprise two or more enzymatic subunits, with two or

more peptide synthetase domains present in a given enzymatic subunit. For example the surfactin peptide synthetase complex (also referred to herein simply as "surfactin synthetase complex") comprises three distinct polypeptide enzymatic subunits: the first two subunits comprise three peptide synthetase domains, while the third subunit comprises a single peptide synthetase domain.

[0032] *Peptide synthetase domain*: The term "peptide synthetase domain" as used herein refers to a polypeptide domain that minimally comprises three domains: an adenylation (A) domain, responsible for selectively recognizing and activating a specific amino acid, a thiolation (T) domain, which tethers the activated amino acid to a cofactor via thioester linkage, and a condensation (C) domain, which links amino acids joined to successive units of the peptide synthetase by the formation of amide bonds. A peptide synthetase domain typically recognizes and activates a single, specific amino acid, and in the situation where the peptide synthetase domain is not the first domain in the pathway, links the specific amino acid to the growing peptide chain.

[0033] *Polypeptide*: The term "polypeptide" as used herein refers to a series of amino acids joined together in peptide linkages. In some embodiments, a "polypeptide" has a structure as achieve through synthesis by ribosomal machinery in naturally occurring organisms. In some embodiments a "polpeptide" has a structure as achieved through chemical synthesis (e.g., *in vitro*). In some embodiments, a "polypeptide" has a structure as achieved through joining of a series of amino acids joined together by non-ribosomal machinery, such as by way of non-limiting example, polypeptides synthesized by peptide synthetases. Such non-ribosomally produced polypeptides exhibit a greater diversity in covalent linkages than polypeptides synthesized by ribosomes (although those skilled in the art will understand that the amino acids of ribosomally-produced polypeptides may also be linked by covalent bonds that are not peptide bonds, such as the linkage of cystines via di-sulfide bonds). In some embodiments, the term is used to refer to specific functional classes of polypeptides, such as, for example, autoantigen polypeptides, nicotinic acetylcholine receptor polypeptides, alloantigen polypeptides, etc. For each such class, the present specification provides several examples of amino acid sequences of known exemplary polypeptides within the class; in some embodiments, such known polypeptides are reference polypeptides for the class. In such embodiments, the term "polypeptide" refers to

any member of the class that shows significant sequence homology or identity with a relevant reference polypeptide. In many embodiments, such member also shares significant activity with the reference polypeptide. For example, in some embodiments, a member polypeptide shows an overall degree of sequence homology or identity with a reference polypeptide that is at least about 30-40%, and is often greater than about 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more and/or includes at least one region (i.e., a conserved region, often including a characteristic sequence element) that shows very high sequence identity, often greater than 90% or even 95%, 96%, 97%, 98%, or 99%. Such a conserved region usually encompasses at least 3-4 and often up to 20 or more amino acids; in some embodiments, a conserved region encompasses at least one stretch of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids. Polypeptides can be two or more amino acids in length, although most polypeptides produced by ribosomes and peptide synthetases are longer than two amino acids. For example, in some embodiments, polypeptides may be 2, 3, 4, 5, 6, 7, 8,9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000 or more amino acids in length.

[0034] *Reductase Domain*: The term "reductase domain" as used herein refers to a polypeptide domain that catalyzes release of an acyl amino acid produced by a peptide synthetase complex from the peptide synthetase complex. In certain embodiments, a reductase domain is covalently linked to a peptide synthetase domain and a fatty acid linkage domain such as a beta-hydroxy fatty acid linkage domain to generate an engineered polypeptide useful in the synthesis of an acyl amino acid. A variety of reductase domains are found in nature in nonribosomal peptide synthetase complexes from a variety of species. A non-limiting example of a reductase domain that may be used in accordance with the present invention includes the reductase domain from linear gramicidin (ATCC8185). However, any reductase domain that releases an acyl amino acid produced by a peptide synthetase complex from the peptide synthetase complex may be used in accordance with the present invention. In some embodiments, reductase domains are characterized by the presence of the consensus sequence: [LIVSPADNK]-x(9)-{P}-x(2)-Y-[PSTAGNCV]-[STAGNQCIVM]-[STAGC]-K-{PC}-[SAGFYR]-[LIVMSTAGD]-x-{K}-[LIVMFYW]-{D}-x-{YR}-[LIVMFYWGAPTHQ]-

[GSACQRHM] (SEQ ID NO: 1), where square brackets ("[]") indicate amino acids that are typically present at that position, squiggly brackets ("{}") indicate amino acids that amino acids that are typically not present at that position, and "x" denotes any amino acid or a gap. X(9) for example denotes any amino acids or gaps for nine consecutive positions. Those skilled in the art will be aware of methods to determine whether a give polypeptide domain is a reductase domain.

[0035] *Small molecule:* As used herein, the term "small molecule" means a low molecular weight organic compound that may serve as an enzyme substrate or regulator of biological processes. In general, a "small molecule" is a molecule that is less than about 5 kilodaltons (kD) in size. In some embodiments, provided nanoparticles further include one or more small molecules. In some embodiments, the small molecule is less than about 4 kD, 3 kD, about 2 kD, or about 1 kD. In some embodiments, the small molecule is less than about 800 daltons (D), about 600 D, about 500 D, about 400 D, about 300 D, about 200 D, or about 100 D. In some embodiments, a small molecule is less than about 2000 g/mol, less than about 1500 g/mol, less than about 1000 g/mol, less than about 800 g/mol, or less than about 500 g/mol. In some embodiments, one or more small molecules are encapsulated within the nanoparticle. In some embodiments, small molecules are non-polymeric. In some embodiments, in accordance with the present invention, small molecules are not proteins, polypeptides, oligopeptides, peptides, polynucleotides, oligonucleotides, polysaccharides, glycoproteins, proteoglycans, *etc*. In some embodiments, a small molecule is a therapeutic. In some embodiments, a small molecule is an adjuvant. In some embodiments, a small molecule is a drug.

[0036] *Surfactin:* Surfactin is cyclic lipopeptide that is naturally produced by certain bacteria, including the Gram-positive endospore-forming bacteria Bacillus subtilis. Surfactin is an amphiphilic molecule (having both hydrophobic and hydrophilic properties) and is thus soluble in both organic solvents and water. Surfactin exhibits exceptional surfactant properties, making it a commercially valuable molecule. Due to its surfactant properties, surfactin also functions as an antibiotic. For example, surfactin is known to be effective as an anti-bacterial, anti-viral, anti-fungal, anti-mycoplasma and hemolytic compound. Surfactin is capable of penetrating the cell membranes of all types of bacteria, including both Gram-negative and Gram-positive bacteria, which differ in the composition of their membrane. Gram-positive bacteria have a thick peptidoglycan layer on the outside of their phospholipid bilayer. In contrast, Gram-

negative bacteria have a thinner peptidoglycan layer on the outside of their phospholipid bilayer, and further contain an additional outer lipopolysaccharide membrane. Surfactin's surfactant activity permits it to create a permeable environment for the lipid bilayer and causes disruption that solubilizes the membrane of both types of bacteria. In order for surfactin to carry out minimal antibacterial effects, the minimum inhibitory concentration (MIC) is in the range of 12–50 μg/ml. In addition to its antibacterial properties, surfactin also exhibits antiviral properties, and its known to disrupt enveloped viruses such as HIV and HSV. Surfactin not only disrupts the lipid envelope of viruses, but also their capsids through ion channel formations. Surfactin isoforms containing fatty acid chains with 14 or 15 carbon atoms exhibited improved viral inactivation, thought to be due to improved disruption of the viral envelope. Surfactin consists of a seven amino acid peptide loop, and a hydrophobic fatty acid chain (beta-hydroxy myristic acid) that is thirteen to fifteen carbons long. The fatty acid chain allows permits surfactin to penetrate cellular membranes. The peptide loop comprises the amino acids L-asparagine, L-leucine, glycine, L-leucine, L-valine and two D-leucines. Glycine and asparagine residues at positions 1 and 6 respectively, constitute a minor polar domain. On the opposite side, valine residue at position 4 extends down facing the fatty acid chain, making up a major hydrophobic domain. Surfactin is synthesized by the surfactin synthetase complex, which comprises the three surfactin synthetase polypeptide subunits SrfA-A, SrfA-B, and SrfA-C. The surfactin synthetase polypeptide subunits SrfA-A and SrfA-B each comprise three peptide synthetase domains, each of which adds a single amino acid to the growing surfactin peptide, while the monomodular surfactin synthetase polypeptide subunit SrfA-C comprises a single peptide synthetase domain and adds the last amino acid residue to the heptapeptide. Additionally the SrfA-C subunit comprises a thioesterase domain, which catalyzes the release of the product via a nucleophilic attack of the beta-hydroxy of the fatty acid on the carbonyl of the C-terminal Leu of the peptide, cyclizing the molecule via formation of an ester. The spectrum of the beta-hydroxy fatty acids was elucidated as iso, anteiso C13, iso, normal C14 and iso, anteiso C15, and a recent study has indicated that surfactin retains an R configuration at C-beta (Nagai et al., Study on surfactin, a cyclic depsipeptide. 2. Synthesis of surfactin B2 produced by Bacillus natto KMD 2311. Chem Pharm Bull (Tokyo) 44: 5-10, 1996).

[0037] Surfactin is a lipopeptide synthesized by the surfactin synthetase complex. Surfactin comprises seven amino acids, which are initially joined by peptide bonds, as well as a beta-hydroxy fatty acid covalently linked to the first amino acid, glutamate. However, upon addition the final amino acid (leucine), the polypeptide is released and the thioesterase domain of the SRFC protein catalyzes the release of the product via a nucleophilic attack of the beta-hydroxy of the fatty acid on the carbonyl of the C-terminal Leu of the peptide, cyclizing the molecule via formation of an ester, resulting in the C-terminus carboxyl group of leucine attached via a lactone bond to the b-hydroxyl group of the fatty acid.

[0038] *Thioesterase domain*: The term "thioesterase domain" as used herein refers to a polypeptide domain that catalyzes release of an acyl amino acid produced by a peptide synthetase complex from the peptide synthetase complex. A variety of thioesterase domains are found in nature in nonribosomal peptide synthetase complexes from a variety of species. A non-limiting example of a thioesterase domain that may be used in accordance with the present invention includes the thioesterase domain from the *Bacillus subtilis* surfactin synthetase complex, present in Srf-C subunit. However, any thioesterase domain that releases an acyl amino acid produced by a peptide synthetase complex from the peptide synthetase complex may be used in accordance with the present invention. In some embodiments, thioesterase domains are characterized by the presence of the consensus sequence: [LIV]-{KG}-[LIVFY]-[LIVMST]-G-[HYWV]-S-{YAG}-G-[GSTAC] (SEQ ID NO: 2), where square brackets ("[]") indicate amino acids that are typically present at that position, and squiggly brackets ("{}") indicate amino acids that amino acids that are typically not present at that position. Those skilled in the art will be aware of methods to determine whether a give polypeptide domain is a thioesterase domain.

Engineered Polypeptides Useful in the Generation of Acyl Amino Acids

[0039] The present invention provides compositions and methods for the generation of acyl amino acids. In certain embodiments, compositions of the present invention comprise engineered polypeptides that are useful in the production of acyl amino acids. In certain embodiments, engineered polypeptides of the present invention comprise a peptide synthetase domain.

[0040] In one aspect, the present invention encompasses the recognition that a single peptide synthetase domain, not associated (e.g., not associated covalently and/or not otherwise associated) with, for example, another domain typically found in a peptide synthetase complex (e.g., a fatty acid linkage domain, a thioesterase domain, a reductase domain, etc. and/or a combination thereof), can be sufficient to produce an acyl amino acid as described herein.

[0041] In accordance with many embodiments of the present invention, peptide synthetase domains useful for the production of acyl amino acids as described herein, correspond and/or show significant homology and/or identity to a first peptide synthetase domain found in a naturally-occurring peptide synthetase complex. That is, as is known in the art, some peptide synthetase domains (i.e., some polypeptides comprising adenylation (A), thiolation (T), and condensation (C) domains) catalyze condensation of a fatty acid with an amino acid, and some catalyze condensation of two amino acids with one another. In accordance with the some embodiments of the present invention, peptide synthetase domains useful for the production of acyl amino acids as described herein are those that catalyze condensation of an amino acid with a fatty acid; such peptide synthetase domains are typically utilized herein in a form (e.g., as part of a polypeptide) that is separated from and/or does not include another peptide synthetase domain. Many naturally-occurring peptide synthetase domains are found in nature within peptide synthetase complexes that synthesize lipopeptides. Such peptide synthetase complexes are multienzymatic complexes found in both prokaryotes and eukaryotes, and comprising one or more enzymatic subunits that catalyze the non-ribosomal production of a variety of peptides (see, for example, Kleinkauf et al., Annu. Rev. Microbiol. 41:259-289, 1987; see also U.S. Patent Number 5,652,116 and U.S. Patent Number 5,795,738). Non-ribosomal synthesis is also known as thiotemplate synthesis (see e.g., Kleinkauf et al.). Peptide synthetase complexes typically include one or more peptide synthetase domains that recognize specific amino acids and are responsible for catalyzing addition of the amino acid to the polypeptide chain.

[0042] The catalytic steps in the addition of amino acids typically include: recognition of an amino acid by the peptide synthetase domain, activation of the amino acid (formation of an amino-acyladenylate), binding of the activated amino acid to the enzyme via a thioester bond between the carboxylic group of the amino acid and an SH group of an enzymatic co-factor,

which cofactor is itself bound to the enzyme inside each peptide synthetase domain, and formation of the peptide bonds among the amino acids.

[0043] A peptide synthetase domain comprises subdomains that carry out specific roles in these steps to form the peptide product. One subdomain, the adenylation (A) domain, is responsible for selectively recognizing and activating the amino acid that is to be incorporated by a particular unit of the peptide synthetase. The activated amino acid is joined to the peptide synthetase through the enzymatic action of another subdomain, the thiolation (T) domain, that is generally located adjacent to the A domain. Amino acids joined to successive units of the peptide synthetase are subsequently linked together by the formation of amide bonds catalyzed by another subdomain, the condensation (C) domain.

[0044] Peptide synthetase domains that catalyze the addition of D-amino acids often also have the ability to catalyze the recemization of L-amino acids to D-amino acids. Peptide synthetase complexes also typically include a conserved thioesterase domain that terminates the growing amino acid chain and releases the product.

[0045] The genes that encode peptide synthetase complexes have a modular structure that parallels the functional domain structure of the complexes (see, for example, Cosmina et al., Mol. Microbiol. 8:821, 1993; Kratzxchmar et al., J. Bacteriol. 171:5422, 1989; Weckermann et al., Nuc. Acids res. 16:11841, 1988; Smith et al., EMBO J. 9:741, 1990; Smith et al., EMBO J. 9:2743, 1990; MacCabe et al., J. Biol. Chem. 266:12646, 1991; Coque et al., Mol. Microbiol. 5:1125, 1991; Diez et al., J. Biol. Chem. 265:16358, 1990).

[0046] Hundreds of peptides are known to be produced by peptide synthetase complexes. Such nonribosomally-produced peptides often have non-linear structures, including cyclic structures exemplified by the peptides surfactin, cyclosporin, tyrocidin, and mycobacillin, or branched cyclic structures exemplified by the peptides polymyxin and bacitracin. Moreover, such nonribosomally-produced peptides may contain amino acids not usually present in ribosomally-produced polypeptides such as for example norleucine, beta-alanine and/or ornithine, as well as D-amino acids. Additionally or alternatively, such nonribosomally-produced peptides may comprise one or more non-peptide moieties that are covalently linked to the peptide. As one non-limiting example, surfactin is a cyclic lipopeptide that comprises a beta-hydroxy fatty acid covalently linked to the first glutamate of the lipopeptide. Other non-peptide

moieties that are covalently linked to peptides produced by peptide synthetase complexes are known to those skilled in the art, including for example sugars, chlorine or other halogen groups, N-methyl and N-formyl groups, glycosyl groups, acetyl groups, etc.

[0047] Typically, each amino acid of the non ribosomally-produced peptide is specified by a distinct peptide synthetase domain. For example, the surfactin synthetase complex which catalyzes the polymerization of the lipopeptide surfactin consists of three enzymatic subunits. The first two subunits each comprise three peptide synthetase domains, whereas the third has only one. These seven peptide synthetase domains are responsible for the recognition, activation, binding and polymerization of L-Glu, L-Leu, D-Leu, L-Val, L-Asp, D-Leu and L-Leu, the amino acids present in surfactin.

[0048] A similar organization in discrete, repeated peptide synthetase domains occurs in various peptide synthetase genes in a variety of species, including bacteria and fungi, for example srfA (Cosmina et al., Mol. Microbiol. 8, 821-831, 1993), grsA and grsB (Kratzxchmar et al., J. Bacterial. 171, 5422-5429, 1989) tycA and tycB (Weckermann et al., Nucl. Acid. Res. 16, 11841-11843, 1988) and ACV from various fungal species (Smith et al., EMBO J. 9, 741-747, 1990; Smith et al., EMBO J. 9, 2743-2750, 1990; MacCabe et al., J. Biol. Chem. 266, 12646-12654, 1991; Coque et al., Mol. Microbiol. 5, 1125-1133, 1991; Diez et al., J. Biol. Chem. 265, 16358-16365, 1990). The peptide synthetase domains of even distant species contain sequence regions with high homology, some of which are conserved and specific for all the peptide synthetases. Additionally, certain sequence regions within peptide synthetase domains are even more highly conserved among peptide synthetase domains which recognize the same amino acid (Cosmina et al., Mol. Microbiol. 8, 821-831, 1992).

[0049] Exemplary lipopeptides synthesized by peptide synthetase complexes in nature are listed below in Table 1 (See also the NORINE database, which provides access to information on peptides and lipopeptides that are known to be, or in some cases believed to be, produced by peptide synthetase enzymes; still further, see Segolene et al. (Ref 4)).

Table 1

Exemplary Lipopeptides Synthesized by Peptide Synthetases

Lipopeptide Name	Fatty Acid Component	Fatty Acid Component name
[Ala4]surfactin aC15	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Ala4]surfactin iC14	iC14:0-OH(3)	3-hydroxy-12-methyl-tridecanoic acid
[Ala4]surfactin iC15	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic acid
[Ala4]surfactin nC14	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
[Ala4]surfactin nC15	C15:0-OH(3)	3-hydroxy-pentadecanoic acid
[Gln1]surfactin	C15:0-OH(3)	3-hydroxy-pentadecanoic acid
[Gln1]surfactin aC15	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Gln1]surfactin iC15	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic acid
[Ile2.4.7]surfactin	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Ile4.7]surfactin	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Ile4]surfactin	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Ile7]surfactin	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Leu4]surfactin	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
[Phe25]syringopeptin 25A	C10:0-OH(3)	3-hydroxy-decanoic acid
[Val7]surfactin	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
A21978C1	aC11:0	8-methyldecanoic acid
A21978C2	iC12:0	10-methylundecanoic acid
A21978C3	aC13:0	10-methyldodecanoic acid
A54145 A	iC10:0	decanoic acid
A54145 A1	C10:0	decanoic acid
A54145 B	C10:0	decanoic acid
A54145 B1	iC10:0	decanoic acid
A54145 C	aC11:0	8-methyldecanoic acid
A54145 D	aC11:0	8-methyldecanoic acid
A54145 E	aC11:0	8-methyldecanoic acid
A54145 F	iC10:0	decanoic acid
amphibactin B	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
amphibactin C	C16:1(9)-OH(3)	3-hydroxy-9-hexadecenoic acid
amphibactin D	C14:0	tetradecanoic acid
amphibactin E	C16:1(9)	9-hexadecenoic acid
amphibactin F	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
amphibactin G	C18:1(9)-OH(3)	3-hydroxy-9-octadecenoic acid
amphibactin H	C16:0	hexadecanoic acid
amphibactin I	C18:1(9)	9-octadecenoic acid

amphisin	C10:0-OH(3)	3-hydroxy-decanoic acid
amphomycin A1437 A	iC13:1(3)	11-methyl-3-dodecenoic acid
amphomycin A1437 B	iC14:1(3)	12-methyl-3-tridecenoic acid
amphomycin A1437 D	aC15:1(3)	12-methyl-3-tetradecenoic acid
amphomycin A1437 E	aC13:1(3)	10-methyl-3-dodecenoic acid
apramide A	C8:0:1(7)-Me(2)	2-methylact-7-ynoic acid
apramide B	C8:0:1(7)	oct-7-ynoic acid
apramide C	C9:1(8)-Me(2)	2-methyl-8-noneic acid
apramide D	C8:0:1(7)-Me(2)	2-methylact-7-ynoic acid
apramide E	C8:0:1(7)	oct-7-ynoic acid
apramide F	C9:1(8)-Me(2)	2-methyl-8-noneic acid
apramide G	C8:0:1(7)-Me(2)	2-methylact-7-ynoic acid
aquachelin A	C12:1(5)	2-methyl-5-dodecenoic acid
aquachelin B	C12:0	dodecanoic acid
aquachelin C	C14:1(7)	7-tetradecenoic acid
aquachelin D	C14:0	tetradecanoic acid
arthrofactin	C10:0-OH(3)	3-hydroxy-decanoic acid
arylomycin A1	iC11:0	9-methyldecanoic acid
arylomycin A2	iC12:0	10-methylundecanoic acid
arylomycin A3	C12:0	dodecanoic acid
arylomycin A4	aC13:0	10-methyldodecanoic acid
arylomycin A5	iC14:0	12-methyl-tridecanoic acid
arylomycin B1	iC11:0	9-methyldecanoic acid
arylomycin B2	iC12:0	10-methylundecanoic acid
arylomycin B3	C12:0	dodecanoic acid
arylomycin B4	aC13:0	10-methyldodecanoic acid
arylomycin B5	iC13:0	11-methyldodecanoic acid
arylomycin B6	iC14:0	12-methyl-tridecanoic acid
arylomycin B7	aC15:0	12-methyltetradecanoic acid
bacillomycin D-1	C14:0-NH2(3)	3-amino-tetradecanoic acid
bacillomycin D-2	iC15:0-NH2(3)	3-amino-13-methyl-tetradecanoic acid
bacillomycin D-3	aC15:0-NH2(3)	3-amino-12-methyl-tetradecanoic acid
bacillomycin D-4	C16:0-NH2(3)	3-amino-hexadecanoic acid
bacillomycin D-5	iC16:0-NH2(3)	3-amino-14-methyl-pentadecanoic acid
bacillomycin F-1	iC15:0-NH2(3)	3-amino-13-methyl-tetradecanoic acid
bacillomycin F-2	aC15:0-NH2(3)	3-amino-12-methyl-tetradecanoic acid
bacillomycin F-3	iC16:0-NH2(3)	3-amino-14-methyl-pentadecanoic acid
bacillomycin F-4	C16:0-NH2(3)	3-amino-hexadecanoic acid
bacillomycin F-5	iC17:0-NH2(3)	3-amino-15-methyl-hexadecanoic acid
bacillomycin F-6	aC17:0-NH2(3)	3-amino-14-methyl-hexadecanoic acid
bacillomycin L-1	C14:0-NH2(3)	3-amino-tetradecanoic acid

bacillomycin L-2	iC15:0-NH2(3)	3-amino-13-methyl-tetradecanoic acid
bacillomycin L-3	aC15:0-NH2(3)	3-amino-12-methyl-tetradecanoic acid
bacillomycin L-4	C16:0-NH2(3)	3-amino-hexadecanoic acid
bacillomycin L-5	iC16:0-NH2(3)	3-amino-14-methyl-pentadecanoic acid
beauverolide A	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide B	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide Ba	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide C	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide Ca	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide D	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide E	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide Ea	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide F	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide Fa	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide H	C9:0-OH(3)	3-hydroxy-nonanoic acid
beauverolide I	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide II	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide III	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide IV	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide Ja	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide Ka	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide L	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide La	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
beauverolide M	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide N	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide V	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide VI	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide VII	C8:0-Me(4)-OH(3)	4-methyl-3-hydroxy-octanoic acid
beauverolide VIII	C10:0-Me(4)-OH(3)	3-hydroxy-4-methyl-decanoic acid
callipeltin A	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin C	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin D	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin F	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin G	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin H	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin I	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin J	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
callipeltin K	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid

callipeltin L	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
carmabin A	C10:0:1(9)-Me(2.4)	2,4-dimethyl-dec-9-ynoic acid
carmabin B	C10:0-Me(2.4)-oxo(9)	9-oxo-2,4-dimethyldecanoic acid
CDA1b	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA2a	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA2b	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA2d	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA2fa	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA2fb	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA3a	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA3b	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA4a	C6:0-Ep(2)	2-epoxy-hexanoic acid
CDA4b	C6:0-Ep(2)	2-epoxy-hexanoic acid
cormycin A	C16:0-OH(3.4)	3,4-dihydroxy-hexadecanoic acid
corpeptin A	C10:0-OH(3)	3-hydroxy-decanoic acid
corpeptin B	C12:1(5)-OH(3)	3-hydroxy-5-dodecenoic acid
corrugatin	C8:0	octanoic acid
daptomycin	C10:0	decanoic acid
enduracidin A	iC12:2(2.t4)	10-methyl-2,trans4-undecanoic acid
enduracidin B	aC13:2(2.t4)	10-methyl-2,trans4-dodecenoic acid
fengycin A	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
fengycin B	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
friulimicin A	iC13:1(3)	11-methyl-3-dodecenoic acid
friulimicin B	iC14:1(3)	12-methyl-3-tridecenoic acid
friulimicin C	aC13:1(3)	10-methyl-3-dodecenoic acid
friulimicin D	aC15:1(3)	12-methyl-3-tetradecenoic acid
fuscopeptin A	C8:0-OH(3)	3-hydroxy-octanoic acid
fuscopeptin B	C10:0-OH(3)	3-hydroxy-decanoic acid
Ile-polymyxin B1	aC9:0	6-methyloctanoic acid
Ile-polymyxin E1	aC9:0	6-methyloctanoic acid
Ile-polymyxin E2	iC8:0	6-methylheptanoic acid
Ile-polymyxin E8	aC10:0	8-methyldecanoic acid
iturin A-1	C13:0-NH2(3)	3-amino-tridecanoic acid
iturin A-2	C14:0-NH2(3)	3-amino-tetradecanoic acid
iturin A-3	aC15:0-NH2(3)	3-amino-12-methyl-tetradecanoic acid
iturin A-4	iC15:0-NH2(3)	3-amino-13-methyl-tetradecanoic acid
iturin A-5	C15:0-NH2(3)	3-amino-pentadecanoic acid
iturin A-6	iC16:0-NH2(3)	3-amino-14-methyl-pentadecanoic acid
iturin A-7	C16:0-NH2(3)	3-amino-hexadecanoic acid
iturin A-8	aC17:0-NH2(3)	3-amino-14-methyl-hexadecanoic acid
iturin C-1	iC14:0-NH2(3)	3-amino-12-methyl-tridecanoic acid
iturin C-2	aC15:0-NH2(3)	3-amino-12-methyl-tetradecanoic acid
iturin C-3	iC16:0-NH2(3)	3-amino-14-methyl-pentadecanoic

		acid
iturin C-4	aC17:0-NH2(3)	3-amino-14-methyl-hexadecanoic acid
kulomo opunalide 1	C8:0:1(7)-Me(2)-OH(3)	2-methyl-3-hydroxy-7-octynoic acid
kulomo opunalide 2	C8:0:1(7)-Me(2)-OH(3)	2-methyl-3-hydroxy-7-octynoic acid
lichenysin A aC13	aC13:0-OH(3)	3-hydroxy-10-methyl-dodecanoic acid
lichenysin A aC15	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
lichenysin A aC17	aC17:0-OH(3)	3-hydroxy-14-methyl-hexadecanoic acid
lichenysin A iC12	iC12:0-OH(3)	3-hydroxy-10-methyl-undecanoic acid
lichenysin A iC13	iC13:0-OH(3)	3-hydroxy-11-methyl-dodecanoic acid
lichenysin A iC14	iC14:0-OH(3)	3-hydroxy-12-methyl-tridecanoic acid
lichenysin A iC15	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic acid
lichenysin A iC16	iC16:0-OH(3)	3-hydroxy-14-methyl-pentadecanoic acid
lichenysin A iC17	iC17:0-OH(3)	3-hydroxy-15-methyl-hexadecanoic acid
lichenysin A nC12	C12:0-OH(3)	3-hydroxy-dodecanoic acid
lichenysin A nC13	C13:0-OH(3)	3-hydroxy-tridecanoic acid
lichenysin A nC14	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
lichenysin A nC15	C15:0-OH(3)	3-hydroxy-pentadecanoic acid
lichenysin A nC16	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
lokisin	C10:0-OH(3)	3-hydroxy-decanoic acid
marinobactin A	C12:0	dodecanoic acid
marinobactin B	C14:1(7)	7-tetradecenoic acid
marinobactin C	C14:0	tetradecanoic acid
marinobactin D1	C16:1(9)	9-hexadecenoic acid
marinobactin D2	C16:1(7)	7-hexadecenoic acid
marinobactin E	C16:0	hexadecanoic acid
massetolide A	C10:0-OH(3)	3-hydroxy-decanoic acid
massetolide B	C11:0-OH(3)	3-hydroxy-undecanoic acid
massetolide C	C12:0-OH(3)	3-hydroxy-dodecanoic acid
massetolide D	C10:0-OH(3)	3-hydroxy-decanoic acid
massetolide E	C10:0-OH(3)	3-hydroxy-decanoic acid
massetolide F	C10:0-OH(3)	3-hydroxy-decanoic acid
massetolide G	C11:0-OH(3)	3-hydroxy-undecanoic acid
massetolide H	C12:0-OH(3)	3-hydroxy-dodecanoic acid
massetolide L	C10:0-OH(3)	3-hydroxy-decanoic acid
mycosubtilin 1	C16:0-NH2(3)	3-amino-hexadecanoic acid
mycosubtilin 2	iC16:0-NH2(3)	3-amino-14-methyl-pentadecanoic acid
mycosubtilin 3	iC17:0-NH2(3)	3-amino-15-methyl-hexadecanoic acid
mycosubtilin 4	aC17:0-NH2(3)	3-amino-14-methyl-hexadecanoic acid

neamphamide A	iC8:0-Me(2.4)-OH(3)	2,4,6-trimethyl-3-hydroxy-heptanoic acid
Nva-polymyxin E1	aC9:0	6-methyloctanoic acid
papuamide A	aC11:2(4.6)-Me(2.6)-OH(2.3)	2,3-dihydroxy-2,6,8-trimethyldeca-(4Z,6E)-dienoic acid
papuamide B	aC11:2(4.6)-Me(2.6)-OH(2.3)	2,3-dihydroxy-2,6,8-trimethyldeca-(4Z,6E)-dienoic acid
papuamide C	aC11:2(4.6)-Me(2.6)-OH(2.3)	2,3-dihydroxy-2,6,8-trimethyldeca-(4Z,6E)-dienoic acid
papuamide D	aC11:2(4.6)-Me(2.6)-OH(2.3)	2,3-dihydroxy-2,6,8-trimethyldeca-(4Z,6E)-dienoic acid
pholipeptin	C10:0-OH(3)	3-hydroxy-decanoic acid
plusbacin A1	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
plusbacin A2	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic acid
plusbacin A3	iC16:0-OH(3)	3-hydroxy-14-methyl-pentadecanoic acid
plusbacin A4	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
plusbacin B1	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
plusbacin B2	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic acid
plusbacin B3	iC16:0-OH(3)	3-hydroxy-14-methyl-pentadecanoic acid
plusbacin B4	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
polymyxin B1	aC9:0	6-methyloctanoic acid
polymyxin B2	iC8:0	6-methylheptanoic acid
polymyxin B3	C8:0	octanoic acid
polymyxin B4	C7:0	heptanoic acid
polymyxin B5	C9:0	nonanoic acid
polymyxin B6	aC9:0-OH(3)	3-hydroxy-6-methyloctanoic acid
polymyxin E1	aC9:0	6-methyloctanoic acid
polymyxin E2	iC8:0	6-methylheptanoic acid
polymyxin E3	C8:0	octanoic acid
polymyxin E4	C7:0	heptanoic acid
polymyxin E7	iC9:0	7-methyloctanoic acid
polymyxin M	aC9:0	6-methyloctanoic acid
pseudomycin A	C14:0-OH(3.4)	3,4-dihydroxy-tetradecanoic acid
pseudomycin B	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
pseudomycin C	C16:0-OH(3.4)	3,4-dihydroxy-hexadecanoic acid
pseudomycin C2	C16:0-OH(3)	3-hydroxy-hexadecanoic acid
pseudophomin A	C10:0-OH(3)	3-hydroxy-decanoic acid
pseudophomin B	C12:0-OH(3)	3-hydroxy-dodecanoic acid
putisolvin I	C6:0	hexanoic acid
putisolvin II	C6:0	hexanoic acid
putisolvin III	C6:0	hexanoic acid
ramoplanin A1	C8:2(2.t4)	2,trans4-octenoic acid
ramoplanin A2	iC9:2(2.t4)	2,trans4-7-methyl-octenoic acid
ramoplanin A3	iC10:2(2.t4)	2,trans4-8-methyl-noneoic acid
serrawettin W1	C10:0-OH(3)	3-hydroxy-decanoic acid

serrawettin W2	C10:0-OH(3)	3-hydroxy-decanoic acid
surfactin aC13	aC13:0-OH(3)	3-hydroxy-10-methyl-dodecanoic acid
surfactin aC15	aC15:0-OH(3)	3-hydroxy-12-methyl-tetradecanoic acid
surfactin iC12	iC12:0-OH(3)	3-hydroxy-10-methyl-undecanoic acid
surfactin iC14	iC14:0-OH(3)	3-hydroxy-12-methyl-tridecanoic acid
surfactin iC15	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic acid
surfactin iC16	iC16:0-OH(3)	3-hydroxy-14-methyl-pentadecanoic acid
surfactin nC13	C13:0-OH(3)	3-hydroxy-tridecanoic acid
surfactin nC14	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
surfactin nC15	C15:0-OH(3)	3-hydroxy-pentadecanoic acid
syringafactin A	C10:0-OH(3)	3-hydroxy-decanoic acid
syringafactin B	C10:0-OH(3)	3-hydroxy-decanoic acid
syringafactin C	C10:0-OH(3)	3-hydroxy-decanoic acid
syringafactin D	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringafactin E	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringafactin F	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringomycin A1	C10:0-OH(3)	3-hydroxy-decanoic acid
syringomycin E	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringomycin G	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
syringopeptin 22 PhvA	C10:0-OH(3)	3-hydroxy-decanoic acid
syringopeptin 22 PhvB	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringopeptin 22A	C10:0-OH(3)	3-hydroxy-decanoic acid
syringopeptin 22B	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringopeptin 25A	C10:0-OH(3)	3-hydroxy-decanoic acid
syringopeptin 25B	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringopeptin 508A	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringopeptin 508B	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
syringopeptin SC 1	C10:0-OH(3)	3-hydroxy-decanoic acid
syringopeptin SC 2	C12:0-OH(3)	3-hydroxy-dodecanoic acid
syringostatin A	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
syringostatin B	C14:0-OH(3.4)	3,4-dihydroxy-tetradecanoic acid
syringotoxin B	C14:0-OH(3)	3-hydroxy-tetradecanoic acid
tensin	C10:0-OH(3)	3-hydroxy-decanoic acid
tolaasin A	Pda	pentanedioic acid
tolaasin B	C8:0-OH(3)	3-hydroxy-octanoic acid
tolaasin C	C8:0-OH(3)	3-hydroxy-octanoic acid
tolaasin D	C8:0-OH(3)	3-hydroxy-octanoic acid
tolaasin E	C8:0-OH(3)	3-hydroxy-octanoic acid
tolaasin I	C8:0-OH(3)	3-hydroxy-octanoic acid
tolaasin II	C8:0-OH(3)	3-hydroxy-octanoic acid
tripropeptin A	iC13:0-OH(3)	3-hydroxy-11-methyl-dodecanoic acid
tripropeptin B	iC14:0-OH(3)	3-hydroxy-12-methyl-tridecanoic acid
tripropeptin C	iC15:0-OH(3)	3-hydroxy-13-methyl-tetradecanoic

		acid
tripropeptin D	iC16:0-OH(3)	3-hydroxy-14-methyl-pentadecanoic acid
tripropeptin E	iC17:0-OH(3)	3-hydroxy-15-methyl-hexadecanoic acid
tripropeptin Z	iC12:0-OH(3)	3-hydroxy-10-methyl-undecanoic acid
Val-polymyxin E1	aC9:0	6-methyloctanoic acid
Val-polymyxin E2	iC8:0	6-methylheptanoic acid
viscosin	C10:0-OH(3)	3-hydroxy-decanoic acid
viscosinamide	C10:0-OH(3)	3-hydroxy-decanoic acid
White Line Inducing Principle	C10:0-OH(3)	3-hydroxy-decanoic acid

[0050] The present invention appreciates that, typically, in peptide synthetase complexes that synthesize lipopeptides, the first active peptide synthetase domain is the one that links a fatty acid to an amino acid; subsequent peptide synthetase domains typically add additional amino acids. In accordance with certain embodiments of the present invention, an acyl amino acid is prepared through use of an engineered peptide synthetase that comprises a first peptide synthetase domain found in a peptide synthetase complex that synthesizes a lipopeptide, and is engineered in that it is separated from at least some other domains found in the peptide synthetase complex.

[0051] Fatty acids utilized by naturally-occurring peptide synthetases can be β-hydroxy fatty acids (e.g., as found in surfactin and other β-hydroxy lipo-peptides described in Table 1). In other cases, utilized fatty acids are a β-amino fatty acid (for example, Iturin; see Table 1). In certain instances, utilized fatty acids are unmodified at the β-position (e.g., as in daptomycin and certain other lipo-peptides described in Table 1).

[0052] As described herein, the present invention encompasses the appreciation that, for all three types of fatty acids utilized by peptide synthetases that synthesize lipopeptides, the the first protein domain of the first module of the relevant peptide synthetase complex typically plays a critical role in lipo-initiation. However, the precise mechanism of lipo-initiation differs for each of the three types of fatty acid. In general terms, the first modules of a peptide synthetase enzyme, which naturally creates a lipo-peptide, has a particular organization. Each module begins with a condensation domain that is required for the lipo-initiation reaction. The condensation domain is followed by an adenylation domain, which is followed by a thiolation domain (also known as a peptidyl carrier protein (PCP) domain). The adenylation domain

selects the 1st amino acid that will be incorporated into the lipo-peptide and creates an amino acid adenylate. Subsequent to adenylation, the amino acid becomes tethered to the enzyme via linkage to a phosphopantethione moiety, which is attached to the thiolation domain. The chemical reaction that results in tethering of the amino acid releases AMP as a byproduct.

[0053] For synthetases that attach a β-hydroxy fatty acid to the bound amino acid, the condensation domain of the first module utilizes β-hydroxy fatty acid CoA as a substrate, and transfers the fatty acid to the N-terminus of the amino acid substrate, which is tethered to the thiolation domain. No enzyme activity, other than the activity of the C-domain itself, is required for this particular reaction, although it has been reported that the srfD protein stimulates the lipo-initiation reaction (see Steller et al., which was cited in 7,981,685) (Ref 5).

[0054] For synthetases that attach a β-amino group to the fatty acid, the condensation domain has several sub-domains, each of which has a particular function (see Figure 6 of Duitman et al.) (Ref 6). Considering the iturin synthetase as a specific examples (also known as the mycosubtilin synthetase), the mechanism of lipo-initiation is the following (see Hansen et al., (Ref 7) and Aron et al., (Ref 8) for details): the acyl ligase domain adenylates a long-chain fatty acid (in this case myristic) and the fatty acid is then transferred to an enzyme-linked 4-phosphopantetheine and AMP is released, in a separate reaction, the fenF gene product catalyzes the transfer of malonate (from manonyl-CoA) to a second acyl carrier domain (located within module 1). The β-ketoacyl synthetase domain catalyzes the condensation of the malonyl and acyl thioesters, creating a β-keto thioester, the B-keto thioester is converted into a β-amino fatty acid by a transaminase domain homologous to amino acid transferases, the β amino fatty acid is transferred to a thiolation domain and is then joined to the substrate amino acid (in this case asparagine), which was previously linked to the enzyme via the action of the module 1 adenylation domain. This series of reactions results in the joining of a beta-amino fatty acid to an amino acid.

[0055] For synthetases that attach fatty acids that are unmodified at the β-position, the condensation domain of the 1st module catalyzes the transfer of the fatty acid to the N-terminus of the amino acid substrate, which is tethered to the thiolation domain. Considering the daptomycin synthetase as an example, two additional proteins are involved: an acyl-CoA ligase (DptE) (sequence listing GenBank: AAX31555.1) and an acyl carrier protein (DptF) (sequence

listing GenBank: AAX31556.1). DptE activates the fatty acid substrate by linking it to CoA, and the activated fatty acid is then transferred to DptF, and subsequently transferred to the enzyme-bound amino acid substrate (see Wittmann et al.) (Ref 9). Note that studies conducted *in vitro* have confirmed that DptE transfers the fatty acid to DptF, but experiments aimed at demonstrating the involvement of the condensation domain in subsequent transfer of the fatty acid from DptF to the amino acid substrate appears not to have been reported in the literature.

[0056] Phylogenetic analysis of peptide synthetase condensation domains is described in Roongsawang et al. (Ref 2), and in Rausch et al. (Ref 3). Those of ordinary skill in the art, guided by the present disclosure, and optionally in consultation with such references, can readily identify, select, and/or engineer appropriate peptide synthetase condensation domains for use in designing, constructing, producing, and/or otherwise providing engineered peptide synthetases for production of acyl amino acids in accordance with the present invention.

[0057] Non-limiting examples of peptide synthetase complexes that may contain peptide synthetase domains useful in the identification, selection, design, and/or production of engineered peptide synthetases as described herein include, for example, surfactin synthetase, fengycin synthetase, arthrofactin synthetase, lichenysin synthetase, syringomycin synthetase, syringopeptin synthetase, saframycin synthetase, gramicidin synthetase, cyclosporin synthetase, tyrocidin synthetase, mycobacillin synthetase, polymyxin synthetase, bacitracin synthetase, and combinations thereof.

[0058] Thus, the present invention provides engineered peptide synthetases, which in some embodiments comprise or consist of isolated peptide synthetase domains from reference peptide synthetase complexes that synthesize lipopeptides. In some embodiments, such reference peptide synthetase complexes are known peptide synthetase complexes. In some embodiments, such reference peptide synthetase complexes are naturally occurring peptide synthetase complexes. In some embodiments, provided engineered peptide synthetases comprise or consist of a single peptide synthetase domain. In some embodiments, provided engineered peptide synthetases comprises or consist of a first peptide synthetase domain from a peptide synthetase complex that synthesizes a lipopeptide.

[0059] In some embodiments, an engineered peptide synthetase, peptide synthetase domain, or component thereof (e.g., adenylation (A) domain, thiolation (T) domain, and/or condensation

(C) domain) may contain one or more sequence modifications as compared with a reference peptide synthetase, domain, or component. Typically, however, an engineered peptide synthetase, peptide synthetase domain, or component thereof shows a high overall degree of sequence identitiy and/or homology with its reference peptide synthetase, domain, or component.

[0060] In some embodiments, an engineered peptide synthetase, peptide synthetase domain, or component thereof contains insertions, deletions, substitutions or inversions of 1, 2, 3, 4, 5, 6,7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or more amino acids as compared to its relevant reference.

[0061] In certain embodiments, such amino acid substitutions result in an engineered polypeptide that comprises an amino acid whose side chain contains a structurally similar side chain as compared to the corresponding amino acid in the relevant reference. For example, amino acids with aliphatic side chains, including glycine, alanine, valine, leucine, and isoleucine, may be substituted for each other; amino acids having aliphatic-hydroxyl side chains, including serine and threonine, may be substituted for each other; amino acids having amide-containing side chains, including asparagine and glutamine, may be substituted for each other; amino acids having aromatic side chains, including phenylalanine, tyrosine, and tryptophan, may be substituted for each other; amino acids having basic side chains, including lysine, arginine, and histidine, may be substituted for each other; and amino acids having sulfur-containing side chains, including cysteine and methionine, may be substituted for each other.

[0062] In certain embodiments, amino acid substitutions result in an engineered polypeptide that comprises an amino acid whose side chain exhibits similar chemical properties to a corresponding amino acid present in a relevant reference. For example, in certain embodiments, amino acids that comprise hydrophobic side chains may be substituted for each other. In some embodiments, amino acids may be substituted for each other if their side chains are of similar molecular weight or bulk. For example, an amino acid in an engineered domain may be substituted for an amino acid present in the relevant reference if its side chains exhibits a minimum/maximum molecular weight or takes up a minimum/maximum amount of space.

[0063] In certain embodiments, an engineered polypeptide shows at least about 50, 55, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% homology or identity with a relevant

reference (e.g., over a portion that spans at least 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more amino acids).

[0064] In certain embodiments, engineered polypeptides of the present invention comprise two or more polypeptide domains that occur in one or more naturally occurring or other known reference polypeptides, but that are i) separated from one or more sequence elements with which they are associated in the reference polypeptide; ii) associated with one or more sequence elements with which they are not associated in the reference polypeptide(s); and/or iii) associated in a different way (e.g., in a different order or via a different linkage) with one or more sequence elements with which they are associated in the reference polypeptide. As a non-limiting example, two naturally occurring polypeptide domains that are directly covalently linked in nature may be separated in an engineered polypeptide by one or more intervening amino acid residues. Additionally or alternatively, two naturally occurring polypeptide domains that are indirectly covalently linked in nature may be directly covalently linked in an engineered polypeptide, e.g. by removing one or more intervening amino acid residues.

[0065] In certain embodiments, two naturally occurring peptide domains that are from different peptide synthetases are covalently joined to generate an engineered polypeptide of the present invention.

[0066] In some embodiments, engineered peptide synthetases provided by and/or for use in accordance with the present invention do not include thioesterase and/or reductase domains. Such domains are known to function in the release of peptides and lipopeptides from the nonribosomal peptide synthetase complexes that produce them. In one aspect, the present invention provides the surprising finding that, notwithstanding their central role in release of lipopeptides from peptide synthetase complexes, such domains are often not required for release of acyl amino acids from engineered peptide synthetases as described herein. This thioesterase and/or reductase domains may optionally be included in some embodiments of the present invention, but are specifically excluded in some embodiments.

[0067] In certain embodiments, compositions and methods of the present invention are useful in large-scale production of acyl amino acids. In certain embodiments, acyl amino acids are produced in commercially viable quantities using compositions and methods of the present invention. For example, engineered polypeptides of the present invention may be used to

produce acyl amino acids to a level of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000 mg/L or higher. As will be appreciated by those skilled in the art, biological production of acyl amino acids using engineered polypeptides of the present invention achieves certain advantages over other methods of producing acyl amino acids. For example, as compared to chemical production methods, production of acyl amino acids using compositions and methods of the present invention utilizes more readily available and starting materials that are easier to store, reduces the necessity of using harsh and sometimes dangerous chemical reagents in the manufacturing process, reduces the difficulty and efficiency of the synthesis itself by utilizing host cells as bioreactors, and reduces the fiscal and environmental cost of disposing of chemical by-products. Other advantages will be clear to practitioners who utilize compositions and methods of the present invention.

Acyl Amino Acids and Compositions

[0068] The present invention provides compositions comprising acyl amino acids produced by engineered peptide synthetases as described herein. In some embodiments, such compositions comprise a collection of individual acyl amino acid molecules, that are related to one another in that they are each synthesized by the same engineered peptide synthetase and together represent a distribution of chemical entities, varied in precise chemical structure (e.g., due to varying length and/or composition of acyl chains, linkages within such acyl chains and/or between an acyl chain and the amino acid, etc), that are synthesized by the relevant engineered peptide synthetase, under the conditions of synthesis (e.g., *in vivo* or *in vitro*). In some embodiments, a provided composition includes straight-chain acyl moieties, branched acyl moieties, and/or combinations thereof.

[0069] That is, it will be appreciated by those skilled in the art that, in some embodiments, one feature of engineered production of acyl amino acids is that engineered peptide synthetases may not generate pure populations of single chemical entities, particularly when acting *in vivo*. Thus, as noted above, the present invention provides acyl amino acid compositions comprising distributions of chemical entities. In some embodiments, the present invention provides acyl

amino acid compositions in which substantially all acyl amino acids comprise the same amino acid moiety, but the composition includes a distribution of acyl moieties.

[0070] As described herein, the present invention provides a wide variety of acyl amino acids and compositions. In some embodiments, the present invention provides acyl amino acids and compositions in which the amino acid moiety is or comprises one found in an amino acid selected from the group consisting of alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and/or valine. Alternatively or additionally, in some embodiments, the present invention provides acyl amino acids and compositions in which the amino acid moiety is or comprises one found in an amino acid selected from the group consisting of selenocysteine and/or pyrrolysine. In some embodiments, the present invention provides acyl amino acids and compositions in which the amino acid moiety is or comprises one found in an amino acid selected from the group consisting of norleucine, beta-alanine and/or ornithine. In some embodiments, the present invention provides acyl amino acids and compositions in which the amino acid moiety is or comprises one found in an amino acid selected from the group consisting of L-amino acids. In some embodiments, the present invention provides acyl amino acids and compositions in which the amino acid moiety is or comprises one found in an amino acid selected from the group consisting of D-amino acids. In some embodiments,the present invention provides acyl amino acids and compositions in which the amino acid moiety is or comprises or comprises one found in an amino acid D-glu or D-diaminopropionic acid. Those skilled in the art will be aware of appropriate amino acid substrates, usable by peptide synthetases as described herein (and particularly by engineered peptide synthetases as described herein) to generate acyl amino acids containing such amino acid moieties. In some embodiments, the amino acid substrate is or comprises the recited amino acid. In some embodiments, the present invention provides acyl amino acids and compositions in which the acyl group is found in a saturated fatty acid such as butryic acid, caproic acid, caprylic acid, capric acid, lauric acid, myristic acid, palmitic acid, stearic arachidic acid, behenic acid, and/or lignoceric acid. the present invention provides acyl amino acids and compositions in which the acyl group is found in an unsaturated fatty acids such as, without limitation, myristoleic acid, palmitoleic acid, oliec acid, linoleic acid, alpha-linolenic acid, arachidonic acid,

eicosapentaenoic acid, erucic acid, and/or docosahexaenoic acid. Other saturated and unsaturated fatty acids whose acyl moieties may be used in accordance with the present invention will be known to those of ordinary skill in the art. In certain embodiments, acyl amino acids and compositions provided by present invention comprise beta-hydroxy fatty acids as the fatty acid moiety. As is understood by those of ordinary skill in the art, beta-hydroxy fatty acids comprise a hydroxy group attached to the third carbon of the fatty acid chain, the first carbon being the carbon of the carboxylate group.

[0071] In some embodiments, the present invention provides acyl amino acids and compositions in which the acyl group comprises or consists of fatty acid chains with a length within a range bounded by a shorter length selected from the group consisting of C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C27, C28, C29, C30, and an upper length selected from the group consisting of C30, C29, C28, C27, C26, C25, C24, C23, C22, C21, C20, C19, C18, C17, C16, C15, C14, C13, C12, C11, C10, C9, C8, C7, C6, C5, C4, C3, C2, and C1, wherein the upper length is the same as or larger than the lower length. In some particular embodiments, the present invention provides acyl amino acids and compositions in which the acyl group comprises or consists of C10-C14 fatty acid chains, C13-16 fatty acid chains, C13-15 fatty acid chains, C16-24 fatty acid chains, C18-22 fatty acid chains, C18-24 fatty acid chains, C8-C16 fatty acid chains. In some embodiments, the present invention provides acyl amino acids and compositions in which the acyl group comprises, consists predominantly of, or consists of C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, and/or C20 fatty acid chains. In some embodiments, the present invention provides acyl amino acids and compositions in which the acyl group comprises, consists predominantly of, or consists of comprises, consists predominantly of, or consists of C8, C9, C10, C11, C12, C13, C14, C15, and/or C16 fatty acid chains. In some embodiments, the present invention provides acyl amino acids and compositions in which the acyl group comprises, consists predominantly of, or consists of comprises, consists predominantly of, or consists of C12, C13, C14, C15, and/or C16 fatty acid chains.

[0072] In some embodiments, the present invention provides acyl amino acid compositions in which all acyl amino acids comprise the same amino acid moiety (or comprise an amino acid moiety from the same amino acid.

[0073] In some embodiments, the present invention provides acyl amino acid compositions in which different acyl amino acids within the composition have different acyl moieties (e.g., acyl moieties that differ, in composition, structure, branching, and/or length (of one or more chains). In some embodiments, such compositions predominantly include acyl moieties of a length (or within a range of lengths) as set forth above. In some such embodiments, such predominant acyl moieties are present in the composition at a level of at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%< 98%, or 99%. The Figures and Examples herein depict and/or describe certain particular acyl amino acids and/or acyl amino acid compositions that are provided by and can be prepared in accordance with certain embodiments of the present invention. To give but a few particular examples, the present invention specifically exemplifies and/or otherwise provides certain acyl amino acids and/or acyl amino acid compositions comprising, consisting predominantly of, or consisting of 2,4 diaminobutyric acid, (2S)-2,3-diaminobutyric acid, 2, 3-diaminoproprionic acid, β-hydroxy myristoyl glutamate, β-hydroxy myristoyl diaminopropionic acid, betaines, cocyl glycinate, gycine laureate, glutamine laureate, etc. For example, in some particular embodiments, the present invention provides acyl amino acid compositions in which the amino acid moiety within acyl amino acids in the composition is from glycine or glutamate, and the fatty acid moiety is predominantly a C12 fatty acid (i.e.g, is from lauric acid).; in some such embodiments, all acyl amino acids in the composition have the same amino acid moiety.

Host Cells

[0074] Engineered polypeptides of the present invention may be introduced in any of a variety of host cells for the production of acyl amino acids. As will be understood by those skilled in the art, engineered polypeptides will typically be introduced into a host cell in an expression vector. So long as a host cell is capable of receiving and propagating such an expression vector, and is capable of expressing the engineered polypeptide, such a host cell is encompassed by the present invention. An engineered polypeptide of the present invention may be transiently or stably introduced into a host cell of interest. For example, an engineered

polypeptide of the present invention may be stably introduced by integrating the engineered polypeptide into the chromosome of the host cell. Additionally or alternatively, an engineered polypeptide of the present invention may be transiently introduced by introducing a vector comprising the engineered polypeptide into a host cell, which vector is not integrated into the genome of the host cell, but is nevertheless propagated by the host cell.

[0075] In certain embodiments, a host cell is a bacterium. Non-limiting examples of bacteria that are useful as host cells of the present invention include bacteria of the genera *Escherichia*, *Streptococcus*, *Bacillus*, and a variety of other genera known to those skilled in the art. In certain embodiments, an engineered polypeptide of the present invention is introduced into a host cell of the species *Bacillus subtilis*.

[0076] Bacterial host cells of the present invention may be wild type. Alternatively, bacterial host cells of the present invention may comprise one or more genetic changes as compared to wild type species. In certain embodiments, such genetic changes are beneficial to the production of acyl amino acids in the bacterial host. For example, such genetic changes may result in increased yield or purity of the acyl amino acid, and/or may endow the bacterial host cell with various advantages useful in the production of acyl amino acids (e.g., increased viability, ability to utilize alternative energy sources, etc.).

[0077] In certain embodiments, the host cell is a plant cell. Those skilled in the art are aware of standard techniques for introducing an engineered polypeptide of the present invention into a plant cell of interest such as, without limitation, gold bombardment and agrobacterium transformation. In certain embodiments, the present invention provides a transgenic plant that comprises an engineered polypeptide that produces an acyl amino acid of interest. Any of a variety of plants species may be made transgenic by introduction of an engineered polypeptide of the present invention, such that the engineered polypeptide is expressed in the plant and produces an acyl amino acid of interest. The engineered polypeptide of transgenic plants of the present invention may be expressed systemically (e.g. in each tissue at all times) or only in localized tissues and/or during certain periods of time. Those skilled in the art will be aware of various promoters, enhancers, etc. that may be employed to control when and where an engineered polypeptide is expressed.

[0078] Insects, including insects that are threats to agriculture crops, produce acyl amino acids that are likely to be important or essential for insect physiology. For example, an enzyme related to peptide synthetases produces the product of the Drosophila Ebony genes, which product is important for proper pigmentation of the fly, but is also important for proper function of the nervous system (see e.g., Richardt et al., Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in Drosophila, J. Biol. Chem., 278(42):41160-6, 2003). Acyl amino acids are also produced by certain Lepidoptera species that are a threat to crops. Thus, compositions and methods of the present invention may be used to produce transgenic plants that produce an acyl amino acid of interest that kills such insects or otherwise disrupts their adverse effects on crops. For example, an engineered polypeptide that produces an acyl amino acid that is toxic to a given insect species may be introduced into a plant such that insects that infest such a plant are killed. Additionally or alternatively, an engineered polypeptide that produces an acyl amino acid that disrupts an essential activity of the insect (e.g., feeding, mating, etc.) may be introduced into a plant such that the commercially adverse effects of insect infestation are minimized or eliminated. In certain embodiments, an acyl amino acid of the present invention that mitigates an insect's adverse effects on a plant is an acyl amino acid that is naturally produced by such an insect. In certain embodiments, an acyl amino acid of the present invention that mitigates an insect's adverse effects on a plant is a structural analog of an acyl amino acid that is naturally produced by such an insect. Compositions and methods of the present invention are extremely powerful in allowing the construction of engineered polypeptides that produce any of a variety of acyl amino acids, which acyl amino acids can be used in controlling or eliminating harmful insect infestation of one or more plant species.

Producing Acyl Amino Acids and Compositions

[0079] Acyl amino acids and compositions may be produced by engineered peptide synthetases as described herein. In some embodiments, acyl amino acids are produced *in vitro*. In some embodiments, acyl amino acids are produced *in vivo*, for example in host cells engineered to express an engineered peptide synthetase or component or domain thereof. In some embodiments, acyl amino acids are produced in association with one or more components of a cell and/or with an engineered peptide synthetase. In some embodiments, acyl

amino acid compositions are subjected to one or more isolation procedures, for example as is known in the art, e.g., to separate produced acyl amino acid compounds from one or more components of their production system (e.g., from an engineered peptide synthetase or component or domain thereof, and/or from one or more components of a cell such as an engineered cell.

Exemplification

Example 1: Engineering Peptide Synthetases to Produce Acyl Amino Acids with β-Hydroxy Amino Acids

[0080] In some embodiments of the present invention, an engineered peptide synthetase that produces an acyl amino acid is designed and/or produced by isolating and/or otherwise engineering a known peptide synthetase domain (e.g., by separating a first peptide synthetase domain that is found in a peptide synthetase complex that synthesizes a lipopeptide from other elements, domains, or components of the lipopeptide-synthesizing complex) to produce the acyl amino acid.

[0081] For example, an acyl amino acid with a β-hydroxy fatty acid can be created by expressing Module 1 of a synthetase, such as the srf (surfactin) synthetase in an appropriate host organism. Since Module 1 of the srfAA (sequence listing srfAA module 1) is glutamate-specific, the expression of Module 1 in an appropriate host leads to the production of β-hydroxyl myristoyl glutamate.

[0082] The same approach can be used to link fatty acids to a variety of different amino acids since there are known (sequenced) "Module 1 DNA segments", which can be cloned from various natural systems, with adenylation domains specific for four distinct amino acids (Leu, Glu, Ser or Dhb; see Table). In addition, a variety of naturally occurring β-hydroxy lipo-peptides (which are believed to be produced by peptide synthetase enzymes) have been reported, for which the gene cluster encoding the synthetase responsible for their production has not been sequenced. A new β -hydroxy acyl amino acid can be produced by using standard molecule

biology techniques to specifically identify "Module 1" of one of those synthetases (which belongs to the set "Module 1's" that have not yet been sequenced) and expressing that Module 1 in an appropriate host. This approach would lead to the generation of additional new β-hydroxy acyl amino acids, including β -hydroxy acyl : Phe, D-Ala, 2,3-dehydro-2-aminobutyric acid, NMe-Ile, Gly, Thr and D-allo-threonine. The Table below summarizes various attributes of known lipopeptides and the peptide synthetases that synthesize them in nature, including the amino acid acyl group and amino acid specificity of the relevant Module 1.

number	lipopeptide	length of fatty acid chain	Reference for fatty acid information	name of module 1	reference for gene information	amino acid specifcity of the module	gene encoding adenylate-forming enzyme	gene encoding ACP	gene encoding malonyl-CoA transacylase
The "fatty acid adding" domain of these 18 synthetases adds β-hydroxy fatty acids to the amino acid									
1	amphisin (one form is arthrofactin)	C-10	A New Lipopeptide Biosurfactant Produced by Arthrobacter sp. Strain MIS38 (database 692)	arfA module1	Cloning and Characterization of the Gene Cluster Encoding Arthrofactin Synthetase from Pseudomonas sp. MIS38 (database 691)	Leu	N/A	N/A	N/A
2	beauverolide	C8 to C-10	Extraribosomal cyclic tetradepsipeptides beauverolides: profiling and modeling the fragmentation pathways (citation from PubMed)	ND	synthetase genes have not been identified	Phe	N/A	N/A	N/A
3	callipeltin	C-8	Isolation of callipeltins A-C and of two new open-chain derivatives of callipeltin A from the	ND	synthetase genes have not been identified	D-Ala	N/A	N/A	N/A

			marine sponge Latrunculia sp. A revision of the stereostructure of callipeltins (ref from Norine database)						
4	corpeptin	C-10 to C-12	Zampella A, Randazzo A, Borbone N, Luciani S, Trevisi L, Debitus C, D Auria MV, *Tetrahedron Letters*, 2002,43 (35), pp. 6163-6166	ND	synthetase genes have not been identified	2,3-dehydro-2-aminobutyric acid	N/A	N/A	N/A
5	fengycin	C-14 to C-18	Application of electrospray ionization mass spectrometry in rapid typing of fengycin homologues produced by Bacillus subtilis	fenC1	Functional and Transcriptional Analyses of a Fengycin Synthetase Gene, fenC, from Bacillus subtilis	Glu	N/A	N/A	N/A
6	fuscopeptin	C-8 to C-10	Structure of fuscopeptins, phytotoxic metabolites of Pseudomonas fuscovaginae.	ND	synthetase genes have not been identified	2,3-dehydro-2-aminobutyric acid	N/A	N/A	N/A
7	kulomo opunalide	2-hydroxyisovaleric and C8:01 (7)-Me(2)-OH(3) and 2-hydroxyisovaleric	More Peptides and Other Diverse Constituents of the Marine Mollusk Philinopsis speciosa	ND	synthetase genes have not been identified	NMe-Ile	N/A	N/A	N/A
8	lichenysin	C-15	Structural and Immunological Characterization of a Biosurfactant Produced by Bacillus licheniformis JF-2	licA module 1	Molecular and Biochemical Characterization of the Protein Template Controlling Biosynthesis of the Lipopeptide	Glu	N/A	N/A	N/A

					Lichenysin				
9	papuamdie	C-11	Papuamides A-D, HIV-inhibitory and cytotoxic depsipeptides from the sponges Theonella mirabilis and Theonella swinhoei collected in Papua New Guinea,	ND	synthetase genes have not been identified	Gly	N/A	N/A	N/A
10	plusbacin	C-14 to C-16	Structures of new peptide antibiotics, plusbacins A1-A4 and B1-B4,	ND	synthetase genes have not been identified	Thr	N/A	N/A	N/A
11	serrawettin	C-10	A Novel Extracellular Cyclic Lipopeptide Which Promotes Flagellum-Dependent and -Independent Spreading Growth of Serratia marcescens	Serratia marcescens gene required for surfactant serrawettin W1 production encodes putative aminolipid synthetase belonging to nonribosomal peptide synthetase family	D-Leu	N/A	N/A	N/A	
12	surfactin	C13 to C15	Separation and Characterization of Surfactin Isoforms Produced by Bacillus subtilis OKB 105	srfA module 1	Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis. I do not have a copy of this paper. It is not in the database.	Glu	N/A	N/A	N/A

13	syringafactin	C-10 to C-12	Identification of a biosynthetic gene cluster and the six associated lipopeptides involved in swarming motility of Pseudomonas syringae pv. tomato DC3000.	SyfA module 1	Identification of a biosynthetic gene cluster and the six associated lipopeptides involved in swarming motility of Pseudomonas syringae pv. tomato DC3000.	Leu	N/A	N/A	N/A
14	syringomycin	C12 to C14	The structure of syringomycins A1, E and G	SyrE1	Characterization of the Syringomycin Synthetase Gene Cluster	Ser	N/A	N/A	N/A
15	syringopeptin	C10 to C14	Novel Cyclic Lipodepsipeptide from Pseudomonas syringae pv. lachrymans Strain 508 and Syringopeptin Antimicrobial Activities	SypA-M1	the sypa sypb sypc synthetase genes encod twenty-two modules invovled nonribosomal peptide synthesis syringopeptin pseudomonas syringae	Dhb	N/A	N/A	N/A
16	tolaasin	C8 and glutaric (pentadecanoic)	tolaasins A-E, five new lipodepsipeptides produced by Pseudomonas tolaasii	synthetase genes have not been identified	2,3-dehydro-2-aminobutyric acid (dhb)	N/A	N/A	N/A	
17	tripropeptin	C12 to C17	tripropeptins, novel antimicrobial agents produced by Lysobacter sp	synthetase genes have not been identified	D-allo-threonine	N/A	N/A	N/A	
18	Viscosin	C10 to C12	Massetolides A-H, antimycobacterial cyclic depsipeptides produced by two pseudomonads isolated from marine habitats	Massatolide A biosynthesis in Pseudomonas fluorescens	L-leu	N/A	N/A	N/A	
The "fatty acid adding" domain of these 14 synthetases adds fatty acids to the amino acid (no β-hydroxy)									

19	A54145	C10 to C11	A54145, a new lipopeptide antibiotic complex: isolation and characteriztion		the lipopeptide antibiotic A54145 biosynthetic gene cluster from Streptomyces fradiae	Trp	lptEF	not identified	N/A
20	apramide	C8 to C9	Apramides A-G novel lipopeptides from the marine cyanobacterium Lyngbya majuscula		synthetase genes have not been identified	Nme-Ala	not identified	not identified	N/A
21	aquachelin	C12 to C14	Structure and membrane affinity of a suite of amphiphilic siderophores produced by a marine bacterium		synthetase genes have not been identified	D-OH-Asp	not identified	not identified	N/A
22	arylomycin	C11 to C15	Arylomycins A and B, new biaryl-lipopeptide antibiotics produced by Streptomyces sp. Tu 6075.II Structure elucidation		synthetase genes have not been identified	D-Nme-Ser	not identified	not identified	N/A
23	CDA1b through CDA4B	2-epoxy-hexanoic acid	Structure biosynthetic origin and enggineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor		Structure biosynthetic origin and enggineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor	Ser	ACS (acyl-CoA synthetase)	SCO3249	N/A
24	carmabin	C10	Carmabins A and B new lipopeptides from the Caribean cyanobacterium Lyngbya majuscula		synthetase genes have not been identified	NMe-Phe	not identified	not identified	N/A
25	corrugatin	C8	Corrugatin A lipopeptide siderophore from Pseudomonas corrugata		synthetase genes have not been identified	OH-His	not identified	not identified	N/A
26	daptomycin	C10	A21978C a		Daptomycin	Trp	DptE	DptF	N/A

		to C13	complex of new acidic peptide antibiotics: isolation, chemistry, and mass spectral structure elucidation		biosynthesis in streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry				
27	enduracidin	C12 to C13			The enduracidin biosynthetic gene cluster from Streptomyces fungicidicus	Asp	Orf45	Orf35	N/A
28	friulimicin	C13 to C15	Friulimicins: novel lipopeptide antibiotics with peptidoglycan synthesis inhibiting activity from Actinoplanes friuliensis sp. nov. II. Isolation and structural characterization		Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic Friulimicin in Actinoplanes friuliensis	Asp or Asn	LipA	LipD	N/A
29	marinobactin	C12 to C16	Membrane affinity of the amphiphilic marinobactin siderophores		synthetase genes have not been identified	D-OH-Asp	not identified	not identified	N/A
30	polymyxin	C7 to C9	CONTRIBUTION TO THE ELUCIDATION OF THE STRUCTURE OF POLYMYXIN B1		Identification of a Polymyxin Synthetase Gene Cluster of Paenibacillus polymyxa and Heterologous Expression of the Gene in Bacillus subtilis	2,4 diaminobutyric acid	not identified	not identified	N/A
31	putisolvin	C6	Characterization of two Pseudomonas putida lipopeptide biosurfactants, putisolvin I and II, which inhibit biofilm formation and break down existing biofilms		Genetic and functional characterization of the gene cluster directing the biosynthesis of putisolvin I and II in Pseudomonas putida strain PCL1445	Leu	not identified	not identified	N/A
32	ramoplanin	C8 to	Studies on the		Chemistry and	Asn	Ramo 26	Ramo 11	N/A

		C10	biosynthesis of the lipodepsipeptide antibiotic Ramoplanin A2		biology of the ramoplanin family of peptide antibiotics				
The "fatty acid adding" domain of this synthetase adds both β-hydroxy and "normal" (not β-hydroxy) fatty acids to the amino acid									
33	Amphibactin	C14 to C18	Structure and membrane affinity of a suite of amphiphilic siderophores produced by a marine bacterium		synthetase genes have not been identified	N-acetyl-Hydroxy Ornithine	not identified	not identified	N/A
The "fatty acid adding" domain of this synthetase adds β-amine fatty acids to the amino acid									
34	iturin	C14 to C17	Revised structure of mycosubtilin, a peptidolipid antibiotic from Bacillus subtilis	MycA	Cloning, sequencing, and characterization of the iturin A operon	Asn	N/A	N/A	fen F

[0083] As is specifically described in Examples herein, additional new β-hydroxy acyl amino acids can be produced by operationally linking a condensation domain, which specifies the addition of a β-hydroxy fatty acid, to an adenylation domain which specifies a particular desired amino acid. In Example XXX, a condensation domain is operationally linked to an adenylation domain that is specific for glycine and, upon expression of the chimera in an appropriate host, β-hydroxy myristyl glycine is produced. One who is skilled in the art will appreciate that this approach can be used to create any desired β-hydroxy acyl amino acid, as long as an adenylation domain is available that is specific for the desired amino acid.

[0084] Naturally occurring peptide synthetase modules are available that specify the use of each of the standard 20 amino acids, and in addition adenylation domains are available that are specific for about 300 additional amino acids, or amino acid-like molecules (Kleinkauf et al) (Ref 10). This approach can be used to link a β-hydroxy fatty acid to any of these amino acids, or amino acid-like molecules.

Example 2: Engineered Peptide Synthetases Comprising or Consisting of Mycosubtilin Module 1 (MycA)

[0085] Strategies analogous to those described above can be used to link a β-amino fatty acid to any desired amino acid. One approach is to identify a naturally occurring "Module 1" (such as MycA of the mycosubtilin synthetase, see Duitman et. al.) (Ref 6) and to express the module in an appropriate host. In this specific example, the FenF gene is desirably also be expressed in the host (sequence listing AAF08794.1).

[0086] In general, a particular β-amino fatty acid can be produced in an appropriate host by expressing a module known to specify the joining of a β -amino fatty acid to a particular amino acid, along with any gene or genes that encode critical additional functions that are not naturally found in the host organism (such as for example FenF). Additional new β -amino acyl amino acids can be produced by operationally linking a condensation domain, which specifies the addition of a β -amino fatty acid, to an adenylation domain which specifies a particular desired amino acid. Again, and genes that encode additional required factors (such as homologs of FenF) can also be expressed in the host. This approach can be used to link a β-amino fatty acid to any amino acid, as long as an adenylation domain is available that is specific for the desired amino acid.

Example 3: Engineered Peptide Synthetases Comprising or Consisting of Daptomycin Synthetase Module 1

[0087] Strategies analogous to those described above can be used to link a fatty acid (which is unmodified at the β-position) to any desired amino acid. One approach is to identify a naturally occurring "Module 1" (such as the Trp1 module of the daptomycin synthetase, see Miao et. al.) (Ref 11) and to express the module in an appropriate host (Sequence listing: dptA1 module 1 of daptomycin synthetase). In addition, in this specific example, the DptE and DptF genes should also be express in the host.

[0088] In general, a particular acyl amino acid (unmodified at the β-position) can be produced in an appropriate host by expressing a module known to specify the joining of a fatty

acid to a particular amino acid, along with any gene or genes that encode critical additional functions that are not naturally found in the host organism (such as for example DptE and DptF). Additional new acyl amino acids can be produced by operationally linking a condensation domain, which specifies the addition of a fatty acid, to an adenylation domain which specifies a particular desired amino acid. For example, fatty acid that is unmodified at the beta position can be attached to glycine using a chimeric synthetase composed of the condensation domain of dptA1 module 1 linked to that adenylation and thiolation domains of dptA1 module 5 (which is specific for glycine) (sequence listing dptA1 Module 5)

Example 4: Additional Genes Useful or Necessary for Some Embodiments

[0089] For the Calcium-Dependent Antibiotic (CDA) system, it is believed that specific locus-associated fatty acid synthases produce the hexanoic acid, which is joined to the first amino acid of CDA; in particular, the ACP (SCO3249), FabH4 (SCO3246), FabF3 (SCO3248) gene products are believed to be important for production of the hexanoic acid, which is then joined to the amino acid substrate, in this case Ser (Ref 12).

Example 5: FA-Glu Compositions

[0090] In some embodiments, the distribution of fatty acids produced by a typical engineered strain that utilizes an engineered peptide synthetase to synthesisze FA-Glu is composed of fatty acids that all have a β-hydroxyl but that have varying chain lengths. In some particular embodiments, the chain lengths vary in the following manner: C12, 1.6%; C13, 16.2%; C14, 55%; C15, 25.9%; C16, 1.2% and C17, 0.01%.

[0091] In some embodiments, some of the even numbered fatty acids are branched and some are straight chain.

[0092] In some embodiments, none of the odd numbered fatty acids are straight chain (i.e, they are all branched). Odd numbered chains can be either iso or anteiso; in some embodiments, the present invention provides different compositions with different relative amounts (e.g., ratios) of these forms. Branching nomenclature is well presented in Figure 1 of Ref 16. *Fatty*

Acids of the Genus Bacillus: an example of branched-chain preference, Toshi Kaneda, Bacteriological Review, 1977,Vol 41(2), 391-418.

[0093] In some embodiments, for an engineered strain that produces FA-Glu with an engineered peptide synthetase, the fatty acid chain distribution changes when particular keto acids are fed to the strain (see Table 1 below). Dramatic changes in fatty acid chain distribution can be generated when the enzyme that synthesizes the keto acids used to initiate fatty acid synthesis in *Bacillus* is knocked out and single keto acids are fed to the strain. In some embodiments, as the concentration of the keto acid is changed, the pattern of fatty acid species is altered.

[0094] In some embodiments, compositions are provided containing FA-Glu with 95% C14 fatty acid by feeding 20 mM isobutyric to the mutant.

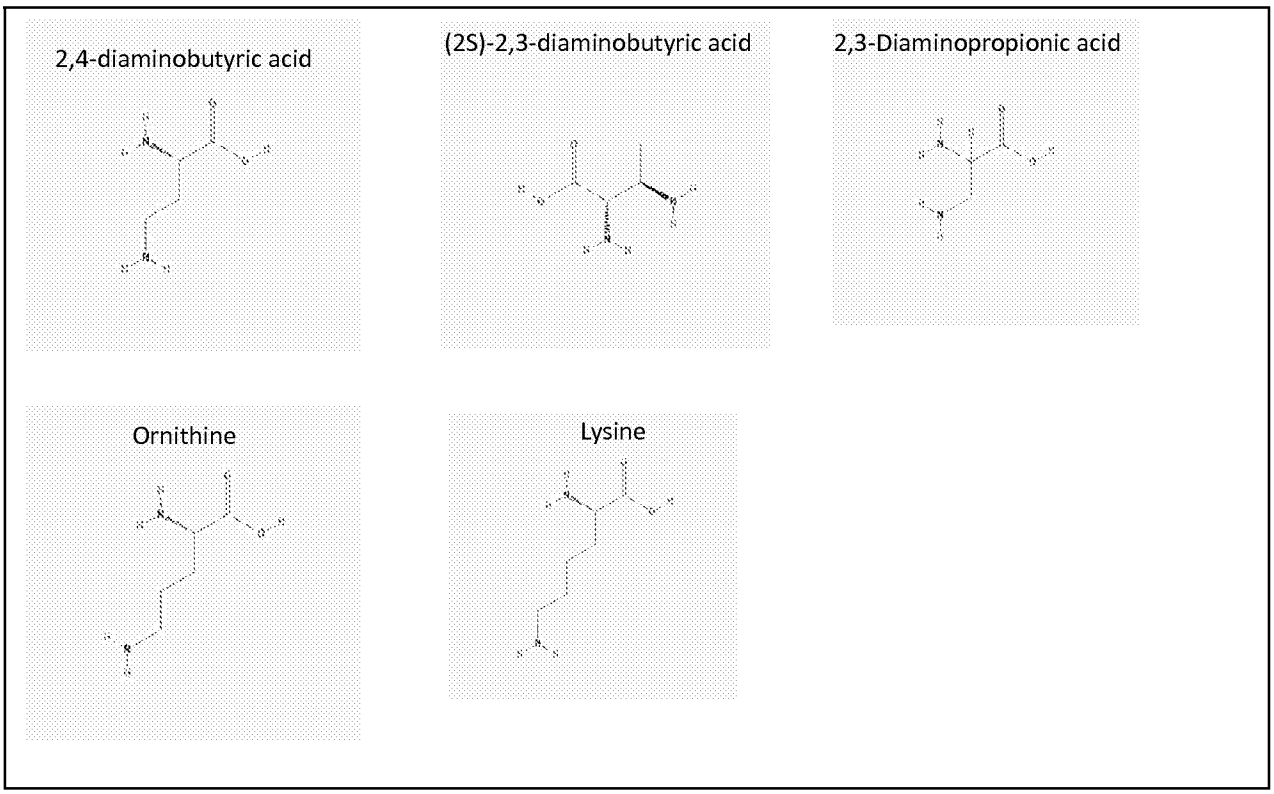
[0095] In some embodiments, feeding of low levels of keto acids that can only be used to produce branched fatty acids with odd number chains, is utilized to produce a population of fatty acids with about 80% (100 uM 2-methylbutyric or 100 uM isovaleric) surfactant with C14 length fatty acid.

[0096] Significantly, since the mutant cannot synthesize its own keto acid starters for even numbered branched chain fatty acid synthesis, feeding of low concentrations of either of these ketos acids (100 uM 2-methylbutyric or 100 uM isovaleric) allows the production of a population of surfactant that is predominantly even numbered and straight chain. Thus, the present invention surprising provides methods and compositions for generating, and compositions comprising mostly straight chain (rather than branched) fatty acid, produced by *B. subtilis*. Indeed, the present invention specifically describes strategies for generating a *Bacillus* strain (and strains so generated) that exclusively produces straight chain fatty acid.

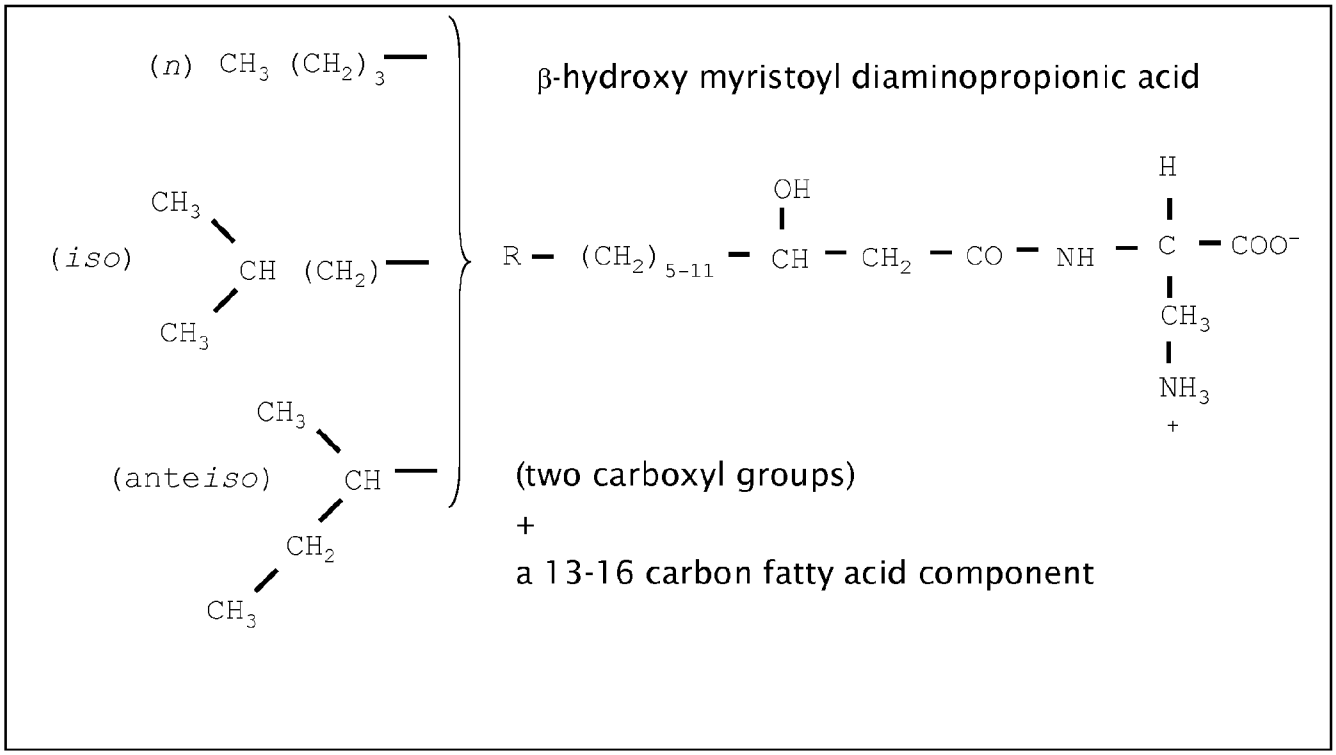
Example 6: Production of amphoteric surfactants

[0097] The present Example describes use of engineered peptide synthetases (in engineered host cells) to produce amphoteric surfactants with one region or regions that harbor a negative charge and another region or regions that harbor a positive charge. Examples of amino acids that

can be used to produce such surfactants are shown in the Figure below. The amino acids all have two amino groups and include: 2,4-diaminobutyric acid, (2S)-2,3-diaminobutyric acid, 2,3-diaminopropionic acid, ornithine and lysine.



[0098] One particular example of a surfactant of this sort is shown in the Figure below, it is β-hydroxy myristoyl diaminopropionic acid:

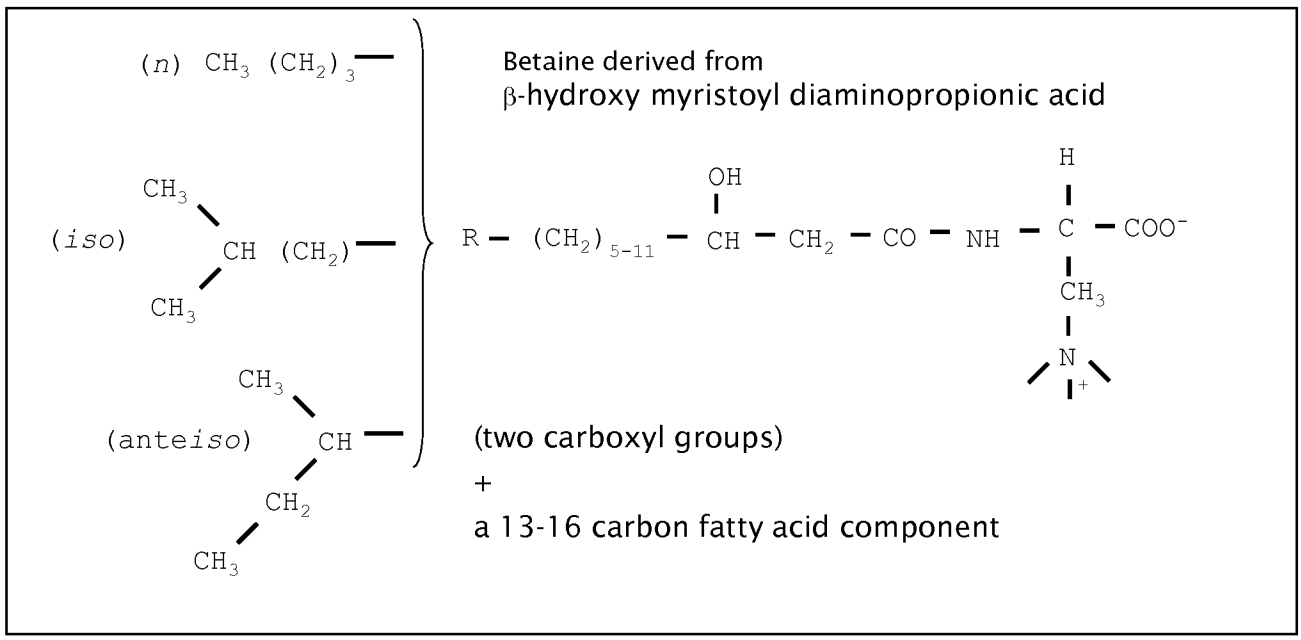


[0099] This surfactant will be zwitterionic at physiological pH given that the pKa of the beta amine of 2,3 diaminoprionic acid is 9.57 and the pKa of an alpha carboxyl is about 2.2. To generate this surfactant, a condensation domain capable of directing the linkage of β-hydroxyl fatty acid to an amino acid (such as the condensation domain of SRFAA module 1) (sequence listing srfAA Module 1) is linked to the adenylation and thiolation domain of a module that is specific for 2,3-diaminopropionic acid (DAP). Felnagle et al., described a peptide synthetase that incorporates DAP. The synthetase is found in *Saccharothrix mutabilis* subsp. *capreolus* ATCC 23892. The DAP-specific module is the second module of CmnA (Sequence listing CmaA, A2).

[00100] *Bacillus subtilis* 168 does not synthesize DAP. Two genes need to be added to *Bacillus* in order to enable conversion of serine to DAP. The genes are described in references cited below. The genes are found in *Staphylococcus aureus* and other bacteria. The genes are called sbnA and sbnB. For example, the genes are present in *Staphylococcus aureus* strain JH9, and also in *Staphylococcus aureus* strain Mu50 / ATCC 700699. The sbnA gene (sequence listing sbnA) is also known as SaurJH9 0103. The sbnB gene (sequence listing sbnB) is also known as SaurJH9 0104.

[00101] Homologues of the sbnA and sbnB genes can be used instead of, or in addition to, sbnA and sbnB. For example, *Bacillus cereus* strains that synthesize zwittermicin encode homologues of sbnA and sbnB, called ZmaU (sequence listing ZmaU) and ZmaV (sequence listing ZmaV), respectively.

[00102] The charge of the primary amine of the surfactant shown in the Figure above will depend on pH, and will be positive in the vicinity of pH 7.0. As the pH is elevated, the amine will lose a hydrogen and become neutral in charge. A surfactant with a positive charge that is independent of pH can be produced by converting the surfactant shown above into a betaine (which harbors a quaternary ammonium group) see Figure below.



[00103] This can be done *in vitro* using a method described by Simon and Shokat. (see reference in reference list). 100 mg of (2-bromoethyl) trimethylammonium bromide are added to a microfuge tube. 1 mL of a solution of the fatty Acid-DAP (FA-DAP) surfactant is added to the tube. The mixture is shaken at 50°C until the solid dissolves. Reaction proceeds for about 5 hours. To consume the remaining alkylating agent, the reaction is quenched with 50 µl 20mercaptoethanol and incubated at room temperature for 30 minutes.

[00104] Alternatively or additionally, methylation can be accomplished *in vivo* using a methyltransferase. One of the symbols did not translate it's shown as a boxBacterial -N-methyltrasferases have been described by Zhang, et al. As example, genes encoding methyltrasnferases can be obtained from *Bacillus subtilis* (sequence listing Bacillus prmA) or *E. coli* (sequence listing *E. coli* prmA). A methyltransferase that modifies cypemycin can be used (sequencing listing cypemycin methyltrasferase) ; the gene is found in Streptomyces sp. OH-4156. A gene encoding a similar protein (76% identical) can be obtained from *Streptomyces griseus* subsp. *griseus* NBRC 13350 (sequencing listing Streptomyces griseus methyltransferase).

Example 7: Production of fatty acids and fatty acid derivatives with particular fatty acid branching patterns

[00105] Naturally occurring fatty acids produced by living organisms typically have two sorts of modifications that affect the melting temperature of the fatty acids and their derivatives. These modifications are branching and desaturation (i.e., the presence of particular double bonds), and both modifications lower the melting point of the fatty acid.

[00106] Certain organisms, including particular gram positive and gram negative bacteria, as well as typical eukaryotes such as yeast, control the fluidity of membranes by desaturation of fatty acids. The ability to introduce desaturated fatty acids into membranes is important with regard to maintenance of membrane fluidity as temperature decreases. Certain bacteria, such as *Bacillus subtilis*, do not rely on desaturation to increase membrane fluidity. Instead, these

bacteria control membrane fluidity via the synthesis of branched fatty acids (for a list of representative bacterial genera that synthesize branched fatty acids, see Table 3 of Ref 13.).

[00107] Given the general need of organisms to control membrane fluidity, biologically produced oils typically contain branches, double bonds, or both. From the perspective of commercial production of fatty acids and their derivatives, there is a need to control these branching and desaturation reactions in order to produce fatty acids with particular characteristics that provide specific benefits to customers. Methods for controlling branching and desaturation are described below.

[00108] As background information, we will consider *E. coli* as an example of an organism that synthesizes straight chain fatty acids (i.e., fatty acids that lack branching), fatty acid synthesis initiates when the enzyme fadH (β-ketoacyl-ACP synthase III) catalyzes condensation of acetyl-coenzyme A (acetyl CoA) with malonyl-acyl carrier protein (malonyl-ACP)(Ref 14). This condensation produces an acetoacetyl-ACP that is then elongated by the iterative action of the *E. coli* fatty acid synthesis machinery.

[00109] Initiation of fatty acid synthesis in *Bacillus subtilis* occurs by a different, but similar, mechanism. *Bacillus subtilis* encodes two β-ketoacyl-ACP synthase III enzymes (fadHA and fadHB). Although these enzymes will utilize acetyl-CoA as a substrate, they prefer to use branched substrates such as isobutyryl-CoA, 2-methylbutyryl-CoA and isovaleryl-CoA (REF 15). These CoA derivatives are produced from the amino acids L-valine, L-isoleucine and L-leucine, respectively (REF 16).

[00110] Initiation of fatty acid synthesis with a branched starter unit leads to the syntheses of a terminally branched fatty acid. The precise chemical composition of the branched starter impacts the length and specific branching of the synthesized fatty acid. For example,initiation with isobutyrate in *Bacillus* leads to production of "iso" fatty acids with even number lengths, such as 14 carbons (C14) and 16 carbons (C16). Initiation with 2-methyl butyrate leads to synthesis of odd numbered "anteiso" fatty acids (e.g., C15 and C17). Initiation with isovalerate

[00111] The enzymatic activity responsible for conversion of particular amino acids (L-valine, L-isoleucine and L-leucine) to their respective keto acids is α-keto acid dehydrogenase. Mutant Bacillus cells that lack α-keto acid dehydrogenase activity require the addition of at least one keto acid for growth (isobutyrate, 2-methyl butyrate or isovalerate) (Ref 17). Feeding a specific keto acid to a strain that lacks β-keto acid dehydrogenase activity not only rescues the growth deficiency of the mutant strain but also specifically affects the fatty acid composition of the cells. For example, feeding isobutyrate to the mutant leads to the exclusive synthesis of fatty acids with even numbered chain length. These fatty acid chains include fatty acids derived from the isobutyrate starter (i14:0, 33%; i16:0, 51%) and also straight chain fatty acids produced using de novo synthesized acetate as a starter (14:0, 2%; 16:0, 13%) (see Ref 17). Furthermore, note that the odd numbered fatty acids are eliminated when a strain that lacks β-keto acid dehydrogenase activity is fed isobutyrate (but not fed 2-methyl butyrate and/or isovalerate).

[00112] Feeding of 2-methyl butyrate leads to the production of a15:0, 51% and a17:0, 39%, with some straight chain even numbered fatty acid still produced via utilization of de novo produced acetate (14:0, 2%; 16:0, 8%) (Ref 17).

[00113] Feeding of isovalerate leads to the following pattern: i15:0, 56%; a15:0, 7%; i17:0, 12%; a17:0, 2%; 14:0, 3% and 16:0, 16%). The presence of anteiso fatty acids is unexpected and suggests that the isovalerate used in the study was contaminated with a keto acid such as 2-methyl butyrate. The straight chain even numbered fatty acids are produced utilizing de novo produced acetate (these data are taken from Ref 17).

[00114] There is a commercial need to produce fatty acids and fatty acid derivatives with precise lengths and branching. In Examples herein, we describe methods for producing particular populations of fatty acids and fatty acid derivatives, such as acyl amino acid surfactants.

[00115] In addition to specifically controlling the branching of fatty acids in organisms such as *Bacillus*, it is advantageous in certain cases to eliminate branching in organisms such as *Bacillus*, for example in order to produce surfactants with straight chain rather than branched fatty acid tails. This can be accomplished by expressing a β-ketoacyl-ACP synthase III enzyme

in *Bacillus* that prefers to use straight chain starts, such as acetyl CoA. As an example of this, Li and coworkers converted a strain of *Streptomyces coelicolor* (which typically predominantly synthesizes branched fatty acids) into a strain that synthesizes 86% straight chain fatty acids by replacing the endogenous β-ketoacyl-ACP synthase III enzyme with E. coli fabH (Ref 18). A general method can be followed to identify enzymes that function in a manner analogous to *E. coli* fadH, that is they initiate fatty acid synthesis using predominantly straight chain starter units, such as acetyl CoA, which will result in the synthesis of straight chain fatty acids.

[00116] Methods such as gas-liquid-chromatography can be used to determine whether an organism synthesizes straight chain fatty acids, or instead synthesizes a mixture of straight chain and branched fatty acids. For example, Kaneda (Ref 16) used gas-liquid-chromatography to characterize the fatty acids of sixteen species of *Bacillus*, and found that all sixteen species synthesized a mixture of straight chain and branched fatty acids. In contrast, a similar study reported by Kaneda and Smith (Ref 19) showed that certain bacteria and yeasts exclusively synthesize straight chain fatty acids, and indeed it is true that most organisms synthesize exclusively straight chain fatty acids. Kaneda and Smith reported that the bacteria *E. coli* and *Pseudomonas fluorescens* exclusively synthesize straight chain fatty acids. Other examples of organisms that exclusively synthesize straight chain fatty acids are reported in Ref 20 and include various *Streptococcus* and *Enterococcus* species, and other species.

[00117] Once an organism has been identified that exclusively synthesizes straight chain fatty acids, assuming the genome of the organism has been sequenced, comparative sequence analysis can be used to determine whether the organism encodes a protein similar to *E. coli* fabH. For example, the gene encoding the Streptococcus pneumonia fabH homologue is 39% identical to E. coli *fabH*. The *Streptococcus* fabH has been cloned and, when the enzyme was produced and studied in vitro, it was found to prefer to utilize short straight CoA primers and to synthesize straight chain fatty acids (Ref 21)(SEE SEQUENCE LISTING AF384041).

[00118] In certain instances, an organism that exclusively or predominantly synthesizes straight chain fatty acids will encode an enzyme that is functionally equivalent to E. coli fabH, but that does not have homology to fabH. As an example, the *Pseudomonas aeruginosa* PA5174

gene encodes a fabY enzyme that is not homologous to fadH, but serves the same function and prefers to use acetyl CoA as the starter for fatty acid synthesis(see this Ref 22 Fatty Acid Biosynthesis in *Pseudomonas aeruginosa* is initiated by the FabY Class of -Ketoacyl Acyl Carrier Protein Synthases). Genes homologus to PA5174, that can be used for this purpose, include the following genes and their homologues--see Sequence listing: Pmen_0396, MDS_0454, Psefu_4068, Avin_05510, PSPA7_5914, PLES_55661 and PA14_68360.

[00119] In order to convert a strain that produces branched fatty acids (such as *Bacillus subtilis*) into a strain the produces predominantly or exclusively straight chain fatty acids a gene such as *E. coli* fabH or *Pseudomonas aeruginosa* PA5174 is introduced into the strain such that it is expressed at the correct time and level. In the specific case of *Bacillus subtilis*, to ensure that the heterologous enzyme, which prefers straight chain starters, is expressed at the correct time and at the correct level, it is advantageous to place the heterologous gene that encodes the β-ketoacyl-ACP synthase III enzyme under the control of the promoter that that usually controls the expression of *Bacillus* fadHA (the fadhA promoter, see sequence listing "fabhA promoter

[00120] Once the heterologous β-ketoacyl-ACP synthase III enzyme is being expressed in Bacillus, branched fatty acid synthesis can be further reduced by reducing, altering or eliminating β-keto acid dehydrogenase activity. In addition, the level of branched fatty acid can be reduced by reducing, altering or eliminating the activity of the endogenous Bacillus fadHA and/or fadHB genes (also known as fadH1 or fadH2).

[00121] When engineered strains are developed with lower levels of branched fatty acids, it is advantageous to express a desaturase enzyme in *Bacillus* in order to introduce sufficient double bonds into a subset of the Bacillus fatty acids to enable the Bacillus to maintain membrane fluidity. Examples of deasturases that can be used include 9-fatty acid desaturase from Psychrobacter urativorans (Ref 23)(sequence listing EF617339) and the 9-fatty acid desaturase from Mortierella alpine (Ref 24) (sequence listing AB015611).

[00122] Alternatively or additionally, genetic changes can be made that result in the constitutive expression of the endogenous *Bacillus* desaturase, des (Ref 25) (sequence listing AF037430). For example, constitutive des expression can be enabled via deletion of desk (Seq

listing DesK gen)(Ref 26). It has been demonstrated that strains with a lipA (yutB) knockout are not able to synthesize fatty acids and require both keto acids and acetate for growth Ref 26. Constitutive expression of des was achieved by knocking out desK, which leads to overexpression of the transcriptional activator DesR, resulting in constitutive expression of des. Overexpression of des led to desaturation of about 13% of the Bacillus fatty acids and eliminated the keto acid requirement, indicating that the growth defect caused by an inability to produce branched fatty acids can be overcome by desaturation of a certain population of *Bacillus* fatty acids.

[00123] An alternative strategy to produce acyl amino acid surfactants with straight chain fatty acids is to express the peptide synthetase enzyme that produces the acyl amino acid in a strain that does not produce branched fatty acids, such as *E. coli*. It has been reported that the srfA operon required for production of surfactin has been cloned and expressed in E. coli (Ref 27). However, the lipopoetide was not characterized directly, rather the authors report that the engineered strain produces a new hydrophobic compound, which was analyzed by TCL using surfactin as a control. Surfactin's Rf value was 0.63 and the new hydrophobic compound showed an Rf value of 0.52. The authors did not speculate on why the Rf values differed.

[00124] An acyl amino acid with a straight chain fatty acid can be produced by cloning a gene that encodes a peptide synthetase enzyme capable of directing the synthesis of an acyl amino acid (such as Module 1 of srfAA) into an *E. coli* plasmid under the control of a promoter such as the T7 promoter and introducing the cloned gene into *E. coli*. It is also necessary to clone and express a gene such as Bacillus sfp, which is a phosphpantetheinyl transferase needed to modify peptide synthetase, enzymes in order to functionally activate those enzymes (see Ref 28). The amount of surfactant produced, and the length of the fatty acid tails present on the population of surfactant molecules, can be determine using LCMS as described in Ref 29.

[00125] Once a strain is generated that produces a desired acyl amino acid, the strain can be further modified to increase the yield of the acyl amino acid. One strategy for increasing yield is to inactivate (e.g., delete) genes that limit production of the acyl amino acid. Once genes are identified that, when deleted, increase yield of an acyl amino acid, a strain harboring multiple such deletions can be generated. In addition, genes that either do not affect surfactant yield, or

that negatively affect surfactant yield, can be replaced with genes that stimulate acyl amino acid production. Examples herein describe genes that, when deleted, increase yield of an acyl glutamate surfactant referred to as FA-Glu

Example 8: Production of β-hydroxy myristoyl glycinate by fermentation

[00126] As described in US 7,981,685, Modular Genetics, Inc. (Modular) has shown that an engineered peptide synthetase enzyme can be used to produce an acyl amino acid (β-hydroxy myristoyl glutamate). This approach has been expanded to produce β-hydroxy myristoyl glycinate. Here is the detailed information on production of β-hydroxy myristoyl glycinate.

[00127] Engineering of a FA-GLY-TE Construct Using a Fusion Between DNA encoding the condensation domain of srfAA module 1 and DNA encoding the adenylation domain of Module 2 of Linear Gramicidin

[00128] In this Example, we amplified the genomic DNA from OKB105Δ(upp)SpectRFA-GLU-TE-MG that encodes for the genes responsible for FA-GLU production, and this region was amplified using primers 35664-C4:5'-TTGTACTGAGAGTGCACCATAtATCGACAAAAATGTCATGAAAGAATCG-3'and 35664-D4:5'-ACGCCAAGCTTGCATGCCtTTATGAAACCGTTACGGTTTGTGTATT-3'. This fragment was annealed to the PCR product obtained from the template pUC19 and primers 35664-B4:5'-AGGCATGCAAGCTTGGCGtAATCATGGTCATAGCTGTTTCCTGTG-3' and 35664-A4:5'-ATATGGTGCACTCTCAGTACAaTCTGCTCTGATGCCGCATAGTT-3'. The annealed mixture was transformed into SURE cells to produce the plasmid Psrf-Glu-TE-pUC19.

[00129] Psrf-Glu-TE-pUC19 was used as a template to engineer a variant of this plasmid that contained a fusion of the condensation domain of srfAA module 1 to the adenylation domain of Module 2 of Linear Gramicidin (which adenylation domain is specific for the amino acid glycine), followed by the TE.

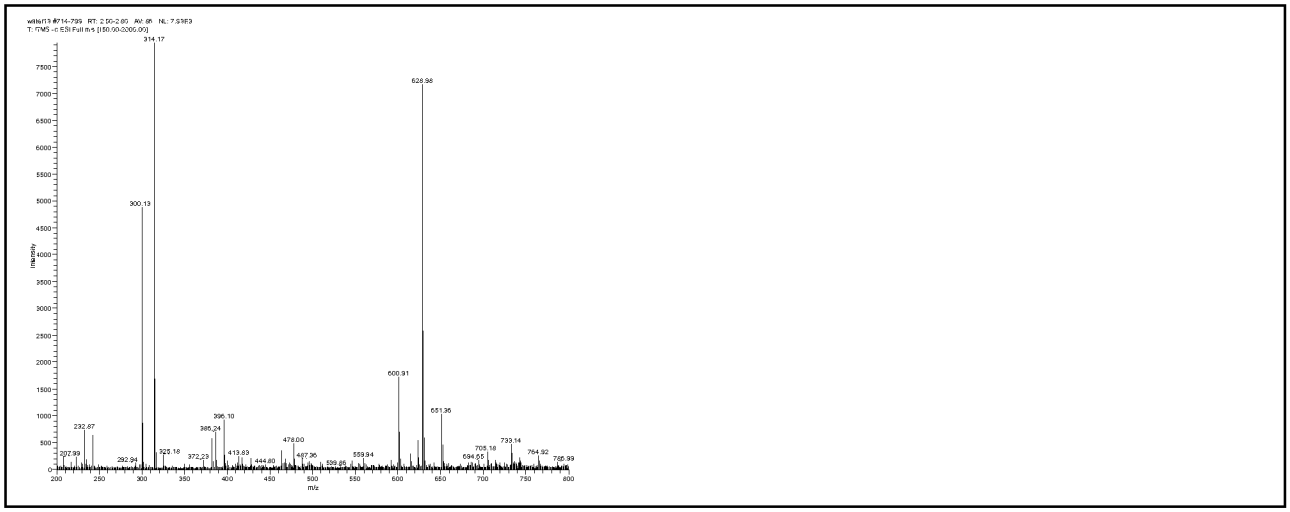
[00130] The DNA sequence corresponding to Module 2 of Linear Gramicidin was amplified from genomic DNA of strain Bacillus brevis (ATCC 8185) using primers 35664-G4:5'-GCTTGCTTGCGGAGCAGATCA-3' and 35664-H4:5'-TCGAATCTCCGCCCAGTTCGA-3'.

The resulting PCR was used as a template for primers 35664-H2:5'-CACTGATTTCTGATGCGGAgAAACGCGATTTGTTTTTGCGG-3' and 35664-F2:5'-CTCCGAGCGCAAAGAAATcGTCGCGAATCCCGATCCG-3'.

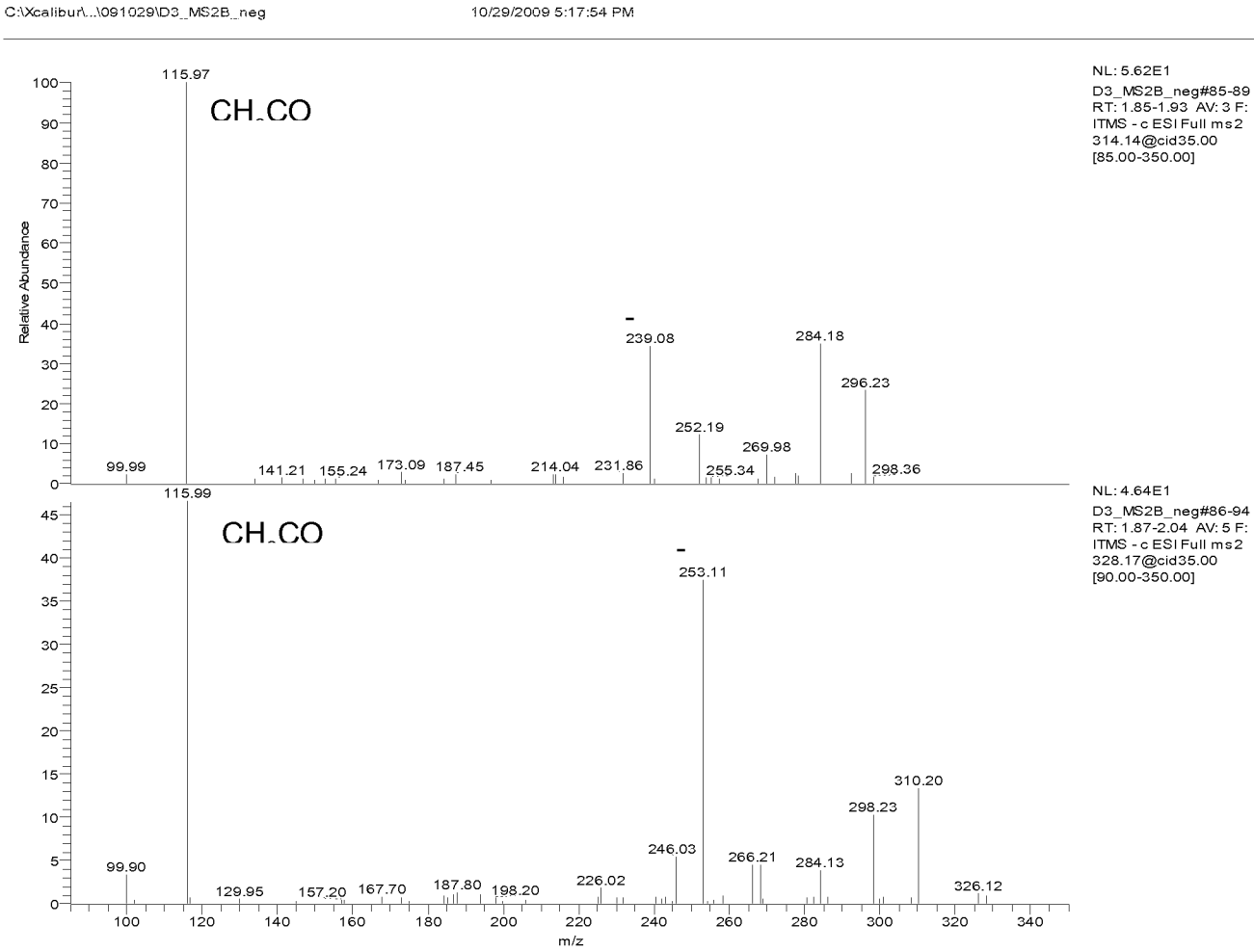
[00131] This fragment was annealed to the PCR product obtained from the template Psrf-Glu-TE-pUC19 using primers 35664-C7:5'-GATTTCTTTGCGCTCGGAgGGCATTCCTTGAAGGCCATGA-3' and 35664-E7:5'-CTCCGCATCAGAAATCAGTgTTAATTCATCAATTGTATGTTCTGGATGC-3'. The annealed mixture was transformed into SURE cells to produce the plasmid Psrf-Gly-lgr_m2-F3-TE-pUC19. This plasmid was used to transform 23844-d1 OKB105Δ(upp)SpectR(Δ mod(2-7))upp+KanR. The resulting strain was named OKB105Δ(upp)SpectRFA-GLY-TE.

[00132] One strain derived from this strategy, which had the correct sequence to produce FA-GLY, was named 37237-d3. Analysis of the production of FA-GLY by strain OKB105Δ(upp)SpectRFA-GLY-TE shows that the strain was able to produce detectable amounts of FA-GLY. Data was obtained using LC-MS analysis. MS-MS analysis of the material derived from OKB105Δ(upp)SpectRFA-GLY-TE revealed that the product was indeed FA-GLY. (sequence listing Psrf-Gly-lgr_m2-F3-TE-pUC19).

[00133] LCMS analysis of FA-Gly. The 300 Dalton species is FA-Glu with a 14 carbon fatty acid tail. The 600 Dalton species is a dimer of the 300 Dalton species. The 314 Dalton species is FA-Glu with a 15 carbon fatty acid tail. The 628 Dalton species is a dimer of the 314 Dalton species.



MS/MS analysis of the 314 Dalton and 328 Dalton species: The 314 species fragments into one species that has Gly+CH_3CO and a second species that is the expected size of the remainder of the fatty acid (labeled "-Gly"). The 328 species fragments into one species that has Gly+CH_3CO and a second species that is the expected size of the remainder of the fatty acid (labeled "-Gly").


C:\Xcalibur\...\091029\D3_MS2B_neg 10/29/2009 5:17:54 PM
115.97
CH3CO
239.08
284.18
296.23
252.19
269.98
99.99
141.21
155.24
173.09
187.45
214.04
231.86
255.34
298.36
NL: 5.62E1
D3_MS2B_neg#85-89
RT: 1.85-1.93 AV: 3 F:
ITMS - c ESI Full ms2
314.14@cid35.00
[85.00-350.00]
115.99
CH3CO
253.11
310.20
298.23
246.03
266.21
284.13
99.90
129.95
157.20
167.70
187.80
198.20
226.02
326.12
NL: 4.64E1
D3_MS2B_neg#86-94
RT: 1.87-2.04 AV: 5 F:
ITMS - c ESI Full ms2
328.17@cid35.00
[90.00-350.00]
Relative Abundance
m/z

Example 9

[00134] The Bacillus α-keto acid dehydrogenase activity was knocked out by deleting the genes that encode two enzymes bkdAA and bkdAB. These genes encode the *Bacillus* E1 and E1β components of α-keto acid dehydrogenase (also known as branched chain -oxo acid dehydrogenase) see **Ref 30**. These genes were knocked out in a strain that produces an acyl amino acid surfactant called FA-Glu, which is composed of fatty acid (FA) linked to the amino acid glutamic acid (Glu).

[00135] As is shown in Table A for the control strain (which retains α-keto acid dehydrogenase activity), the surfactant is composed of a population of molecules with fatty acid tails that vary in length from C12 to C17, with C14 predominant (55%). When the mutant strain (which lacks α-keto acid dehydrogenase activity) is fed 20mM isobutyrate the fatty acid composition of the surfactant population narrows to about 95% C14. Surfactants with a fatty acid tail length of C14 are particularly useful for certain applications, such as use in personal care products such as shampoos, body washes and other products. The population of surfactant fatty acid tail lengths can be specifically modified by feeding the mutant strain a starter keto acid that results in production of odd numbered branched fatty acids. Specifically, a population of surfactant molecules with a fatty acid tail composition of C13:0,27%;C15:0,65% was produced upon feeding the mutant 20mM 2-methylbutyric acid. Thus, the strain produced surfactant with over 90% odd numbered branched fatty acid tails (presumably anteiso). A population of surfactant molecules with a fatty acid tail composition of C12:0,3.71%;C14:0,76.04%;C16:0,2.20% was produced upon feeding the mutant 100μM 2-methylbutyric acid. Thus, the strain produced surfactant with over 80% even numbered fatty acid tails. Given that the mutant strain is incapable of producing branched fatty acids with even numbered chain lengths, and was fed a keto acid that can only be used to produce odd numbered branched fatty acids this population of even numbered fatty acid molecules is comprised of straight chain (unbranched) fatty acids. Feeding of 20mM isovaleric produced surfactant with over 90% odd numbered branched fatty acid tails (presumably iso). Feeding of 100μM isovaleric produced surfactant with over 80% even numbered (straight chain) fatty acid tails.

[00136] We have demonstrated previously that acylases can be used to specifically cleave an acyl amino acid surfactant to generate a free fatty acid and an amino acid. This approach can be used with the surfactant populations described above to produce particular purified populations of fatty acids, for example a population composed of over 90% C14 fatty acid or a population composed of over 90% anteiso C13 and C15 or over 90% iso C13 and C15, or over 80% straight chain (even numbered fatty acids).

Experimental Details:

[00137] In this example, we amplified the genomic region of *B. subtilis* strain OKB105 encoding for the bkdAA and bkdAB genes and upstream and downstream flanking genes (buk, lpdV, bkdB, bmrR, and bmr) using primers 47014:5'-AATATCGTATTGAATAGACAGACAGG-3' and 47015:5'-ATCTTTATTTGCATTATTCGTGGAT-3'. The resulting PCR was used as a template to amplify both upstream and downstream fragments.

[00138] The upstream fragment was amplified using primers 47020:5'-GTGTAAATCATTTAATGAAAAAAGGAAAAATTGACGTG-3' and 47023:5'-ATCATTAAGCCTTCCTGGCAGTCAGCCCTAGTGCTTGATGTCGGTTTG-3'. The downstream fragment was amplified using primers 47026:5'-AATTAAAAGCCATTGAGGCAGACGTAAGGGAGGATACAATCATGGCAATT-3' and 47021:5'-GGTATTCTTGCTGACAACGGTACATTCATATG-3'. The genes encoding for UPP/Kan were amplified from the template pUC19-UPP-KAN using primers 47024:5'-ACACGATATAGCCAGGAAGGCGGGTTTTTTGACGATGTTCTTGAAACTC-3' and 47025:5'-AATTAAAAGCCACAAAGGCCTAGGTACTAAAACAATTCATCCAGTAA-3'.

[00139] The upstream, downstream and UPP/Kan fragments were all digested to completion with restriction endonuclease BglI. All 3 fragments were subsequently ligated together with T4 DNA ligase. The ligated DNA mix was transformed into FA-Glu producing strain 43074-B2 and transformants were selected for ability to grow on LB agar supplemented with Kanamycin (30ug/mL) and Isobutyric, Isovaleric and 2-methylbutyric acids (100uM). One strain derived

from this strategy, which had the correct sequence to replace bkdAA and bkdAB with UPP/Kan, was named 47392-A6 and was used in subsequent experiments.

[00140] 47392-A6 was grown alongside 43074-B2 in S7(Phos7.5) (minimal media containing 100mM Potassium Phosphate Buffer pH7.5, 10mM Ammonium Sulfate, 20mM Monosodium Glutamate, 2% Glucose and trace metals) supplemented with 0, 100uM, 1mM, 5mM or 20mM 2-methylbutyric, Isovaleric, Isobutyric acids (all neutralized to pH7.5) in 10mM cultures for 4 days at 37C.

Table A

		344=C12	358=C13	372=C14	386=C15	400=C16	414=C17	FA-Glu (mg/L)
Control	No Acid	1.60%	16.29%	54.78%	26.02%	1.19%	0.12%	439.2
	100uM 2-methylbutyric	1.76%	18.27%	52.08%	26.68%	1.09%	0.12%	397.1
	1mM 2-methylbutyric	1.25%	23.84%	34.54%	39.28%	0.74%	0.35%	443.8
	5mM 2-methylbutyric	0.99%	26.91%	22.05%	49.22%	0.38%	0.46%	409.6
	20mM 2-methylbutyric	0.57%	26.79%	16.49%	55.19%	0.30%	0.65%	333.6
	100uM Isovaleric	1.66%	17.42%	53.04%	26.70%	1.05%	0.12%	451.4
	1mM Isovaleric	1.15%	24.84%	39.84%	33.28%	0.75%	0.15%	437.6
	5mM Isovaleric	0.64%	34.26%	19.87%	44.67%	0.33%	0.22%	434.4
	20mM Isovaleric	0.53%	34.06%	8.55%	56.54%	0.14%	0.19%	338.5
	100uM Isobutyric	1.72%	15.64%	58.19%	23.08%	1.23%	0.13%	457.1
	1mM Isobutyric	1.53%	11.44%	63.98%	21.51%	1.45%	0.10%	470.1
	5mM Isobutyric	1.55%	9.43%	69.63%	17.76%	1.53%	0.09%	433.2
	20mM Isobutyric	1.33%	9.09%	69.83%	17.86%	1.82%	0.07%	434.5
Mutant	No Acid	no growth observed						
	100uM 2-methylbutyric	3.71%	10.41%	76.04%	7.56%	2.20%	0.07%	401.4
	1mM 2-methylbutyric	2.38%	25.73%	32.49%	38.46%	0.57%	0.36%	441.4
	5mM 2-methylbutyric	1.00%	31.76%	10.00%	56.32%	0.21%	0.71%	415.2
	20mM 2-methylbutyric	0.68%	27.28%	6.37%	64.77%	0.17%	0.73%	307.2
	100uM Isovaleric	3.53%	8.30%	78.33%	7.89%	1.93%	0.02%	417.9
	1mM Isovaleric	1.28%	22.86%	36.65%	38.72%	0.43%	0.06%	370.8
	5mM Isovaleric	0.48%	38.41%	11.76%	49.02%	0.20%	0.13%	425.8
	20mM Isovaleric	0.31%	36.41%	4.14%	58.89%	0.09%	0.16%	334.9

	100uM Isobutyric	2.88%	5.96%	84.74%	4.67%	1.72%	0.03%	250.1
	1mM Isobutyric	2.34%	3.37%	90.10%	2.08%	2.08%	0.02%	420.3
	5mM Isobutyric	1.82%	0.66%	94.03%	1.01%	2.48%	0.01%	433.0
	20mM Isobutyric	1.68%	0.30%	94.50%	0.81%	2.69%	0.02%	390.7

BKD up-U/K-down sequence using restriction sites:

AATATCGTATTGAATAGACAGACAGGAGTGAGTCACCAT
GGCAACTGAGTATGACGTAGTCATTCTGGGCGGCGGTACCGGCGGTTATG
TTGCGGCCATCAGAGCCGCTCAGCTCGGCTTAAAAACAGCCGTTGTGGAA
AAGGAAAAACTCGGGGGAACATGTCTGCATAAAGGCTGTATCCCGAGTAA
AGCGCTGCTTAGAAGCGCAGAGGTATACCGGACAGCTCGTGAAGCCGATC
AATTCGGAGTGGAAACGGCTGGCGTGTCCCTCAACTTTGAAAAAGTGCAG
CAGCGTAAGCAAGCCGTTGTTGATAAGCTTGCAGCGGGTGTAAATCATTT
AATGAAAAAAGGAAAAATTGACGTGTACACCGGATATGGACGTATCCTTG
GACCGTCAATCTTCTCTCCGCTGCCGGGAACAATTTCTGTTGAGCGGGGA
AATGGCGAAGAAAATGACATGCTGATCCCGAAACAAGTGATCATTGCAAC
AGGATCAAGACCGAGAATGCTTCCGGGTCTTGAAGTGGACGGTAAGTCTG
TACTGACTTCAGATGAGGCGCTCCAAATGGAGGAGCTGCCACAGTCAATC
ATCATTGTCGGCGGAGGGGTTATCGGTATCGAATGGGCGTCTATGCTTCA
TGATTTTGGCGTTAAGGTAACGGTTATTGAATACGCGGATCGCATATTGC
CGACTGAAGATCTAGAGATTTCAAAAGAAATGGAAAGTCTTCTTAAGAAA
AAAGGCATCCAGTTCATAACAGGGGCAAAAGTGCTGCCTGACACAATGAC
AAAAACATCAGACGATATCAGCATACAAGCGGAAAAAGACGGAGAAACCG
TTACCTATTCTGCTGAGAAAATGCTTGTTTCCATCGGCAGACAGGCAAAT
ATCGAAGGCATCGGCCTAGAGAACACCGATATTGTTACTGAAAATGGCAT
GATTTCAGTCAATGAAAGCTGCCAAACGAAGGAATCTCATATTTATGCAA
TCGGAGACGTAATCGGTGGCCTGCAGTTAGCTCACGTTGCTTCACATGAG
GGAATTATTGCTGTTGAGCATTTTGCAGGTCTCAATCCGCATCCGCTTGA
TCCGACGCTTGTGCCGAAGTGCATTTACTCAAGCCCTGAAGCTGCCAGTG
TCGGCTTAACCGAAGACGAAGCAAAGGCGAACGGGCATAATGTCAAAATC
GGCAAGTTCCCATTTATGGCGATTGGAAAAGCGCTTGTATACGGTGAAAG
CGACGGTTTTGTCAAAATCGTGGCTGACCGAGATACAGATGATATTCTCG
GCGTTCATATGATTGGCCCGCATGTCACCGACATGATTTCTGAAGCGGGT
CTTGCCAAAGTGCTGGACGCAACACCGTGGGAGGTCGGGCAAACGATTCA
CCCGCATCCAACGCTTTCTGAAGCAATTGGAGAAGCTGCGCTTGCCGCAG
ATGGCAAAGCCATTCATTTTTAAAAGCATAAAGGAGGGGCTTGAATGAGT
ACAAACCGACATCAAGCACTAGGGCTGACTGCCAGGAAGGC
GGGTTTTTTGACG 1200
1201 ATGTTCTTGAAACTCAATGTCTTTTTTTGTAGAATCAATAGAAGTGTGTA 1250
1251 ATTGTTGATGGGACAATAAAAAAGGAGCTGAAACACAGTATGGGAAAGGT 1300
1301 TTATGTATTTGATCATCCTTTAATTCAGCACAAGCTGACATATATACGGA 1350
1351 ATGAAAATACAGGTACGAAGGATTTTAGAGAGTTAGTAGATGAAGTGGCT 1400

```
1401  ACACTCATGGCATTTGAAATTACCCGCGATCTTCCTCTGGAAGAAGTGGA  1450
1451  TATCAATACACCGGTTCAGGCTGCGAAATCGAAAGTCATCTCAGGGAAAA  1500
1501  AACTCGGAGTGGTTCCTATCCTCAGAGCAGGATTGGGAATGGTTGACGGC  1550
1551  ATTTTAAAGCTGATTCCTGCGGCAAAAGTGGGACATGTCGGCCTTTACCG  1600
1601  TGATCCAGAAACCTTAAAACCCGTGGAATACTATGTCAAGCTTCCTTCTG  1650
1651  ATGTGGAAGAGCGTGAATTCATCGTGGTTGACCCGATGCTCGCTACAGGC  1700
1701  GGTTCCGCAGTTGAAGCCATTCACAGCCTTAAAAAACGCGGTGCGAAAAA  1750
1751  TATCCGTTTCATGTGTCTTGTAGCAGCGCCGGAGGGTGTGGAAGAATTGC  1800
1801  AGAAGCATCATTCGGACGTTGATATTTACATTGCGGCGCTAGATGAAAAA  1850
1851  TTAAATGAAAAAGGATATATTGTTCCAGGTCTCGGAGATGCGGGTGACCG  1900
1901  CATGTTTGGAACAAAATAAAAAATGAAATCCCCAAAAGGGGGTTTCATTT  1950
1951  TTTTATCCAGTTTTTTGCTATTCGGTGAATCTGTATACAATTATAGGTGA  2000
2001  AAATGTGAACATTCTGGGATCCGATAAACCCAGCGAACCATTTGAGGTGA  2050
2051  TAGGTAAGATTATACCGAGGTATGAAAACGAGAATTGGACCTTTACAGAA  2100
2101  TTACTCTATGAAGCGCCATATTTAAAAAGCTACCAAGACGAAGAGGATGA  2150
2151  AGAGGATGAGGAGGCAGATTGCCTTGAATATATTGACAATACTGATAAGA  2200
2201  TAATATATCTTTTATATAGAAGATATCGCCGTATGTAAGGATTTCAGGGG  2250
2251  GCAAGGCATAGGCAGCGCGCTTATCAATATATCTATAGAATGGGCAAAGC  2300
2301  ATAAAAACTTGCATGGACTAATGCTTGAAACCCAGGACAATAACCTTATA  2350
2351  GCTTGTAAATTCTATCATAATTGTGGTTTCAAAATCGGCTCCGTCGATAC  2400
2401  TATGTTATACGCCAACTTTCAAAACAACTTTGAAAAAGCTGTTTTCTGGT  2450
2451  ATTTAAGGTTTTAGAATGCAAGGAACAGTGAATTGGAGTTCGTCTTGTTA  2500
2501  TAATTAGCTTCTTGGGGTATCTTTAAATACTGTAGAAAAGAGGAAGGAAA  2550
2551  TAATAAATGGCTAAAATGAGAATATCACCGGAATTGAAAAAACTGATCGA  2600
2601  AAAATACCGCTGCGTAAAAGATACGGAAGGAATGTCTCCTGCTAAGGTAT  2650
2651  ATAAGCTGGTGGGAGAAAATGAAAACCTATATTTAAAAATGACGGACAGC  2700
2701  CGGTATAAAGGGACCACCTATGATGTGGAACGGGAAAAGGACATGATGCT  2750
2751  ATGGCTGGAAGGAAAGCTGCCTGTTCCAAAGGTCCTGCACTTTGAACGGC  2800
2801  ATGATGGCTGGAGCAATCTGCTCATGAGTGAGGCCGATGGCGTCCTTTGC  2850
2851  TCGGAAGAGTATGAAGATGAACAAAGCCCTGAAAAGATTATCGAGCTGTA  2900
2901  TGCGGAGTGCATCAGGCTCTTTCACTCCATCGACATATCGGATTGTCCCT  2950
2951  ATACGAATAGCTTAGACAGCCGCTTAGCCGAATTGGATTACTTACTGAAT  3000
3001  AACGATCTGGCCGATGTGGATTGCGAAAACTGGGAAGAAGACACTCCATT  3050
3051  TAAAGATCCGCGCGAGCTGTATGATTTTTTAAAGACGGAAAAGCCCGAAG  3100
3101  AGGAACTTGTCTTTTCCCACGGCGACCTGGGAGACAGCAACATCTTTGTG  3150
3151  AAAGATGGCAAAGTAAGTGGCTTTATTGATCTTGGGAGAAGCGGCAGGGC  3200
3201  GGACAAGTGGTATGACATTGCCTTCTGCGTCCGGTCGATCAGGGAGGATA  3250
3251  TCGGGGAAGAACAGTATGTCGAGCTATTTTTTGACTTACTGGGGATCAAG  3300
3301  CCTGATTGGGAGAAAATAAAATATTATATTTTACTGGATGAATTGTTTTA  3350
3351  GTACCTAGGCCTTTG
```

AGGCAGACGTAAGGGAGGAT

ACAATCATGGCAATTGAACAAATGACGATGCCGCAGCTTGGAGAAAGCGT

AACAGAGGGGACGATCAGCAAATGGCTTGTCGCCCCCGGTGATAAAGTGA

ACAAATACGATCCGATCGCGGAAGTCATGACAGATAAGGTAAATGCAGAG

GTTCCGTCTTCTTTTACTGGTACGATAACAGAGCTTGTGGGAGAAGAAGG

CCAAACCCTGCAAGTCGGAGAAATGATTTGCAAAATTGAAACAGAAGGCG

CGAATCCGGCTGAACAAAAACAAGAACAGCCAGCAGCATCAGAAGCCGCT

GAGAACCCTGTTGCAAAAAGTGCTGGAGCAGCCGATCAGCCCAATAAAAA

GCGCTACTCGCCAGCTGTTCTCCGTTTGGCCGGAGAGCACGGCATTGACC

TCGATCAAGTGACAGGAACTGGTGCCGGCGGGCGCATCACACGAAAAGAT

ATTCAGCGCTTAATTGAAACAGGCGGCGTGCAAGAACAGAATCCTGAGGA

GCTGAAAACAGCAGCTCCTGCACCGAAGTCTGCATCAAAACCTGAGCCAA

AAGAAGAGACGTCATATCCTGCGTCTGCAGCCGGTGATAAAGAAATCCCT
GTCACAGGTGTAAGAAAAGCAATTGCTTCCAATATGAAGCGAAGCAAAAC
AGAAATTCCGCATGCTTGGACGATGATGGAAGTCGACGTCACAAATATGG
TTGCATATCGCAACAGTATAAAAGATTCTTTTAAGAAGACAGAAGGCTTT
AATTTAACGTTCTTCGCCTTTTTTGTAAAAGCGGTCGCTCAGGCGTTAAA
AGAATTCCCGCAAATGAATAGCATGTGGGCGGGGGACAAAATTATTCAGA
AAAAGGATATCAATATTTCAATTGCAGTTGCCACAGAGGATTCTTTATTT
GTTCCGGTGATTAAAAACGCTGATGAAAAAACAATTAAAGGCATTGCGAA
AGACATTACCGGCCTAGCTAAAAAAGTAAGAGACGGAAAACTCACTGCAG
ATGACATGCAGGGAGGCACGTTTACCGTCAACAACACAGGTTCGTTCGGG
TCTGTTCAGTCGATGGGCATTATCAACTACCCTCAGGCTGCGATTCTTCA
AGTAGAATCCATCGTCAAACGCCCGGTTGTCATGGACAATGGCATGATTG
CTGTCAGAGACATGGTTAATCTGTGCCTGTCATTAGATCACAGAGTGCTT
GACGGTCTCGTGTGCGGACGATTCCTCGGACGAGTGAAACAAATTTTAGA
ATCGATTGACGAGAAGACATCTGTTTACTAAATAAGCAAAAAGAGCATTT
TTTGAAGTTTTGTTTCAAAAAATGCTCTTTTTCTATGCTTTATTATTCAG
CGATCCGTATTTTCATTTCGACTCGATATTCTTCTTGTTTTTTCGGGGAG
TAATGAATCGGTATGATTAACTCGTATACATCACTGACAACTGTTAATTG
GCGGTCCGCGATATATTTGATAAGCTTCTGTAAGTTGAGAAAATAATGTT
CAGGCGAAAAATTATACGCGATACACGCATACCTCCCTTTAGGGATCGTT
GTGATTTCCATATCCGGCGTAATTGATGAAATCTGTTTATTCGTCAATAC
AGGTGTGAAAATATGACGGTAAGTCATTTCATCAATGCTGGTGTAGGGCT
GAAAAGAGAAAGTAGCGCCGTAGCTATTGTTCGTAAATCCATCTGCTGAC
TCGATAAATTTTTTTAATTTGCTGTAGGAGGCGTTGAGCACGTTTTCAGG
CCCGATTCCTTCTGCCTCTGTCTGAATGATCCGTATTTCTTCTTCATCTA
AAACAAACACCTCACCGAGCGCGGGATATTCCATCTGCCGTTTCATCCGC
TTTTTCACCAATGAAATGGTTTGCTCCAGGGCTGATAAAAAGTCTAATTT
CTCCCTGATTTGCCTCTCCTGCTCTGTATAAAAAGCAAACAGTTCTTCCA
TCTCTAAGTCCTGTGCTTTTTTCATCTCTTCTAAAGGTGTGCCGATATAT
TTCAATGATTTGATCAAATCCAGATGAATGAGCTGAGAATCTGTATAATA
GCGGTAGCTGGTATCCGGGTCGACGTAGGCTGGTTTAAATAAATCAATTT
TATCGTAATAACGGAGCGCTTTTATCGACACGTTTGCCAGTTTTGATACT
TCCCCAATTGAGTAATACGATTCCTTCATGCCATCACTCCTTCTATCATC
AGTATAAAGAAGAAGCGCATTCTTTGCAGTACACAAAGAATGCGCTTCTT
ATCACGTGCTGGCTTTAAGATGTGCAGGCGCTTTCCAAGCAATGGTCAGT
GCAATCCCTATGGCTAAGGTGACCGTTGCAAAGTAGAAAGGATAGTTTAC
ATCTATATCGAACAGCATTCCGCCGATAATAGGCCCGAATACATTGCCGA
TACTTGTAAACATTGAATTCATACCGCCGGCAAACCCCTGTTCATTTCCC
GCAATCTTTGACAGGTAAGTCGTTACCGCAGGCCGCATGAGATCAAATCC
GACAAATACGGTGACTGTCACCAGCAGAATCGCAACATATGAATGTACCG
TTGTCAGCAAGAATACCAGACTCGTCGAGAGAATTAAGCTGTACCGAATT
AAATGAATTTCGCCAAACCATCTTGTGAAGCGGTCGAATAAGACGACTTG
CGTAATGGCGCCAACAATCGCTCCTCCTGTAATCATAATGGCAATGTCGC
TGGCCGTAAATCCGAATTTATGATCCACGAATAATGCAAATAAAGAT

Example 10

[00141] The following genes were deleted by replacing the coding sequence of each gene with a upp/kan cassette. The effect on FA-Glu yield is shown in Table #: Maf, Abh, RocG, degU, RapC, eps, yngF, yhaR, mmgB. spxA.

[00142] An additional copy of each of the following gene was introduced into Bacillus under the control of either a constitutive promoter (e.g., PgroEL or under the control of the Psrf promoter, which normally controls expression of genes in the srf operon (which genes are required for production of surfactin). The effect on FA-Glu yield is shown in Table B:

Single Knockouts	FA-Glu Increase relative to parental strain	Ave
RapC	34.1%	25.9%
	17.6%	
plip	21.2%	21.2%
yqxM	19.7%	19.7%
eps	19.1%	19.1%
degU	13.3%	18.0%
	22.8%	
yngF	14.5%	14.5%
RocG	12.0%	12.0%
yhaR	13.3%	11.5%
	9.6%	
mmgB	11.4%	11.4%
abh	6.2%	9.7%
	16.0%	
	6.9%	
maf	15.6%	8.0%
	0.5%	
spoIIAC	7.8%	7.8%
fapR	3.3%	4.8%
	6.3%	
spxA	2.7%	2.7%

Knockin	FA-Glu Increase relative to parental strain	
eps->pGroEL-lcfA	11.5%	
amyE->Pspac-srfD	12.7%	
amyE->PgroEL-sfp-srfD	44.3%	
phe+	79.2%	
NOTE: All Single knockouts are in the 43074-B2 background that contains 1) plip KO, 2) phe+, 3)amyE->PgroEL-sfp-srfD and 4) spoIIAC KO		

References

1. Krass et al., "Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation" *Chemistry & Biology*, 17:872-880, 2010.
2. Roongsawang et al., "Phylogeneticv analysis of condensation domains in nonribosomal peptide synthetases" *FEMS Microbiology Letters*, 252:143-151, 2005.
3. Rausch et al., "Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution" *BMC Evolutionary Biology* 7(78): 1-15, 2007.
4. Segolene et al., "NORINE: a database of nonribosomal peptides." *Nucleic Acids Research*, 36: D327-D331, 2008.
5. Steller, et al., "initiation of Surfactin Biosynthesis and the Role of the SrfD-Thioesterase Protein." *Biochemistry*, 43:11331-11343, 2004.
6. Duitman et al., "The Mycosubtilin synthetase of Bacillus subtilis ATCC6633: A Multifunctional Hybrid Between a Peptide Synthetase, an Amino Transferase, and a fatty Acid Synthase" *PNAS*, 96(23):13294-13299, 1999.
7. Hansen et al., "The Loading Module of Mycosubtilin: An adenylation Domain with fatty Acid Selectivity" J Am Chem Soc, 129(20): 6366-6367, 2007.
8. Aron et al., "FenF: Servicing the Mycosubtilin Synthetase Assembly Line *in trans*" *ChemBioChem*, 8: 613-616, 2007.
9. Wittmann et al., "Role of DptE and DptF in the lipidation reaction of daptomycin" *FEBS Journal*, 275:5343-5353, 2008.
10. Kleinkauf et al., "A nonribosomal system of peptide biosynthesis" *Eur J Biochem*, 236: 335-351, 1996.
11. Miao et al., "Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry" *Microbiology*, 151: 1507-1523, 2005.

12. Zohreh et al., "Structure, biosynthetic origin, and engineering Biosynthesis of calcium-Dependent Antibiotics from Streptomyces coelicolor" *Chemistry & Biology*, 9:1175-1187, 2002.
13. Kaneda, "Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance," *Microbiological Reviews,* 55(2):288-302, 1991.
14. Isolation and characterization of the -ketoacyl-acyl carrier protein synthatse III gene (fabH) from *Escherichia coli* K-12, Tsay, et al., JBC, 267(10), 6807-6814, 1992.
15. Choi et al., " -ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis," *Journal of Bacteriology*, 182(2):365-370, 2000.
16. Kaneda, "Fatty acids of the genus *Bacillus*: an example of branched-chain preference," *Bacteriological Reviews*, 41(2):391-418, 1977.
17. Willecke et al., "Fatty acid-requiring mutant of bacillus subtilis defective in branched chain α-keto acid dehydrogenase," *The Journal of Biological Chemistry*, 246(17):5264-5272.

18. Alteration of the fatty acid profile of Streptomyces coelicolor by replacement of the initiation enzyme 3-ketoacyl acyl carrier protein synthase III (FabH).

19. Relationship of primer specificity of fatty acid de novo synthetase to fatty acid composition in 10 species of bacteria and yeasts. Kaneda dn Smith. Can. J. Microbiol., Vol 26, 1980.

20. Application of cellular fatty acid analysis, david welch, Clinical Microbiology Reviews, Oct. 1991, 422-438.

21. Identification, substrate specificity, and inhibition of the Streptococcus pneumonia b-ketoacyl-acp carrier protein synthase III (FabH), Khandekar, et al., JBC, 276(32), 2001.

22. Fatty Acid Biosynthesis in Pseudomonas aeruginosa Is Initiated by the FabY Class of -Ketoacyl Acyl Carrier Protein Synthases

23. Identification and functional expression of a 9-fatty acid desaturase from Psychrobacter urativorans in Escherichia coli, Li et al., Lipids, 43, 207-213, 2008.

24. Δ9-fatty acid desaturase from arachidonic acid-producing fungus unique gene sequence and its heterologous expression in a fungus, *Aspergillus*, Sakuradani et al., Eur. J. Biochem., 260, 208-219, 1999.

25. A Bacillus subtilis gene induced by cold shock encodes a membrane phospholipid desaturase, Aguilar et al., Journal of bacteriology, 180(8):2194-2200, 1998.

26. Martin, et al., "a lipA (yutB) mutant, encoding lipoic acid synthase, provides insight into the interplay between branched-chain and unsaturated fatty acid biosynthesis in *Bacillus subtilis*," *Journal of Bacteriology*, 191(24):7447-7455, 2009.

27. Lee et al., "Cloning of srfA operon from Bacillus subtilis C9 and its expression in E. coli," *Appl Microbiol Biotechnol*, 75(3):567-572, 2007.

28. Quadri et al., "Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptididyl carrier domains in peptide synthetases," *Biochemistry*, 37(6):1585-1595, 1998.

29. Reznik et al., "Use of sustainable chemistry to produce an acyl amino acid surfactant,", *Appl Microbiol Biotechnol*, published online, 2010.

30. Wang, et al., "The primary structure of branched-chain a-oxo acid dehydrogenase from bacillus subtilis and its similarity to other a-oxo acid dehydrogenases," *Eur. J. Biochem.*, 213:1091-1099, 1993.

31. This reference is out of order relative to its position in the text. Felnagle, et al., "Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin," Applied and Environmantal Microbiology, 73(13):4162-4170, 2007.

32. This reference is out of order relative to where it is mentioned in the text. Beasley et al., "Mutation of L-2,3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in *Staphylococcus aureus*," *BMC Microbilogy*, 11:199, 2011.

33. This reference is out of order relative to its position in the text. Simon and Shokat, "A method to site-specifically incorporate methyl-lysine analogues into recombinant proteins," Methods in Enzymology, Volue 512, Nuclososomes: Histones & Chromatin, Part A, edited Carl Wu and C. David Allis, Elsevier, Inc., 2012.

34. This reference is out of order relative to its position in the text. Zhang et al., "Catalytic promiscuity of a bacterial -N-methyltransferase," *FEBS Letters*, 586:3391-3397, 2012.

35. This reference is out of order relative to its position in the text. Komiyama, et al., "A new antibiotic, cepemycin taxonomy, fermentation, isolation and biological characteristics," *The Journal of Antibiotics,* 46(11):1666-1671, 1993.

Sequence Listing

Proteins for synthesis of 2,3-diaminopropionic acid

sbnA
>sp|Q7A1Z6|SBNA_STAAW Probable siderophore biosynthesis protein SbnA
OS=Staphylococcus aureus (strain MW2) GN=sbnA PE=3 SV=1
MIEKSQACHDSLLDSVGQTPMVQLHQLFPKHEVFAKLEYMNPGGSMKDRPAKYIIEHGIK
HGLITENTHLIESTSGNLGIALAMIAKIKGLKLTCVVDPKISPTNLKIIKSYGANVEMVE
EPDAHGGYLMTRIAKVQELLATIDDAYWINQYANELNWQSHYHGAGTEIVETIKQPIDYF
VAPVSTTGSIMGMSRKIKEVHPNAQIVAVDAKGSVIFGDKPINRELPGIGASRVPEILNR
SEINQVIHVDDYQSALGCRKLIDYEGIFAGGSTGSIIAAIEQLITSIEEGATIVTILPDR
GDRYLDLVYSDTWLEKMKSRQGVKSE

sbnB
>tr|Q6X7U6|Q6X7U6_STAAU SbnB OS=Staphylococcus aureus GN=sbnB PE=4 SV=1
MNREMLYLNRSDIEQAGGNHSQVYVDALTEALTAHAHNDFVQPLKPYLRQDPENGHIADR
IIAMPSHIGGEHAISGIKWIGSKHDNPSKRNMERASGVIILNDPETNYPIAVMEASLISS
MRTAAVSVIAAKHLAKKGFKDLTIIGCGLIGDKQLQSMLEQFDHIERVFVYDQFSEACAR
FVDRWQQQRPEINFIATENAKEAVSNGEVVITCTVTDQPYIEYDWLQKGAFISNISIMDV
HKEVFIKADKVVVDDWSQCNREKKTINQLVLEGKFSKEALHAELGQLVTGDIPGREDDDE
IILLNPMGMAIEDISSAYFIYQQAQQQNIGTTLNLY

ZmaU
>gi|223047493|gb|ACM79820.1| ZmaU [Bacillus cereus]
MSFRYKFYLKYIRKNIYTYLSLIIFLDFNQERKQIMLKKLESLERVIGNTPMIKLEHEKINLYAKLEYYN
LMNSVKVRAAYHILKSAINRGEVNENSTIIESSSGNFAVALATLCRYIGLKFIPVIDPNINDSYENFLRA
TSYQVANVDERDETGGYLLTRLNKVKELLNTIPNAYWTNQYNNADNFEAHYQGIGGEISNDFKQLDYAFI
GVSTGGTIAGVSTRLKEKFPNIKIIAVDSQGSIIFGDKPRKRYIPGIGASMIPGMVKKALIDDVMIVPEV
HTVAGCYELFNKHAIFAGGSSGTSYYAIQKYFENRDVQNTPNVVFLCPDNGQAYTSTIYNVEWVEWLNTQ
KSVEDQLVSL

ZmaV
>gi|223047494|gb|ACM79821.1| ZmaV [Bacillus cereus]
MMYLNTKHFNEMGVNWEETINVISKAVKSLDSEDFSQPIKPYLRFDDPANRIIAMPAYIGGEFKVSGIKW
IASFPKNIEKGIQRAHSVTILNDAMTGKPFATLNTAMVSVIRTASVTGLMIREFAKLRDLNNVKVGIIGF
GPIGQMHLKMVTALLGDKIEGVYLYDINGIKDELIPEEIYSKTQKVNAYEEAYNDADIFITCTVSAEGYI
DKKPKDGALLLNVSLRDFKPDILEYTKSLVVDNWEEVCREKTDVERMHLERGLQKEDTVSIADVVIRGAL
QNFPYDKAILFNPMGMAIFDVAIAAYYYQRARENEMGVLLED

Methyltrasferases

Bacillus prmA
>gnl|BSUB|BSU25450-MONOMER ribosomal protein L11 methyltransferase
(complement(2624760..2623825)) Bacillus subtilis subtilis 168
MKWSELSIHT THEAVEPISN ILHEAGASGV VIEDPLDLIK ERENVYGEIY QLDPNDYPDE

GVIVKAYLPV NSFLGETVDG IKETINNLLL YNIDLGRNHI TISEVNEEEW ATAWKKYYHP
VKISEKFTIV PTWEEYTPVH TDELIIEMDP GMAFGTGTHP TTVLCIQALE RFVQKGDKVI
DVGTGSGILS IAAAMLEAES VHAYDLDPVA VESARLNLKL NKVSDIAQVK QNNLLDGIEG
EHDVIVANIL AEVILRFTSQ AYSLLKEGGH FITSGIIGHK KQEVKEALEQ AGFTIVEILS
MEDWVSIIAK K

E. coli prmA
>gnl|ECOLI|EG11497-MONOMER methyltransferase for 50S ribosomal subunit protein
L11 3407092..3407973 Escherichia coli K-12 substr. MG1655
MPWIQLKLNT TGANAEDLSD ALMEAGAVSI TFQDTHDTPV FEPLPGETRL WGDTDVIGLF
DAETDMNDVV AILENHPLLG AGFAHKIEQL EDKDWEREWM DNFHPMRFGE RLWICPSWRD
VPDENAVNVM LDPGLAFGTG THPTTSLCLQ WLDSLDLTGK TVIDFGCGSG ILAIAALKLG
AAKAIGIDID PQAIQASRDN AERNGVSDRL ELYLPKDQPE EMKADVVVAN ILAGPLRELA
PLISVLPVSG GLLGLSGILA SQAESVCEAY ADSFALDPVV EKEEWCRITG RKN

cypemycin methyltrasferase

>sp|E5KIC0|CYPM_STRSQ Cypemycin methyltransferase OS=Streptomyces sp. GN=cypM
PE=1 SV=1
MSDPSVYDETAIEAYDLVSSMLSPGAGLVAWVSSHRPLDGRTVLDLGCGTGVSSFALAEA
GARVVAVDASRPSLDMLEKKRLDRDVEAVEGDFRDLTFDSTFDVVTMSRNTFFLAQEQEE
KIALLRGIARHLKPGGAAFLDCTDPAEFQRAGGDARSVTYPLGRDRMVTVTQTADRAGQQ
ILSIFLVQGATTLTAFHEQATWATLAEIRLMARIAGLEVTGVDGSYAGEPYTARSREMLV
VLERQ

Streptomyces griseus methyltransferase
>gi|182440155|ref|YP_001827874.1| methyltransferase [Streptomyces griseus
subsp. griseus NBRC 13350]
MSEPTVYDAAAIDAYDLISSMLSPGAGLAAWVSSHRPLAGRTVLDLGAGTGVSSFALADAGAQVVAVDAS
RPSLDLLESRRGERKVDTVEADFRDLRLDSAFDVVTMSKNTFFLAQSHDEKIELLRAIGRHLKPGGAVFL
DCTDPVEYLRADGAAHTVTYPLGREQMVTITQNADRATQAIMSIFMVQSASTLTSFHEMATWASLPEIRL
LARAAGLEVTAVDGSYAGDAYTARSREMLVVLEAK

Proteins for initiation of straight chain fatty acid synthesis

fadH family members for initiation of straight chain fatty acid synthesis

M77744

>M77744_1(M77744|pid:none) Escherichia coli beta-ketoacyl-acyl carrier
protein synthase III (fabH) gene, complete cds.
MYTKIIGTGSYLPEQVRTNADLEKMVDTSDEWIVTRTGIRERHIAAPNETVSTMGFEAAT
RAIEMAGIEKDQIGLIVVATTSATHAFPSAACQIQSMLGIKGCPAFDVAAACAGFTYALS
VADQYVKSGAVKYALVVGSDVLARTCDPTDRGTIIIFGDGAGAAVLAASEEPGIISTHLH
ADGSYGELLTLPNADRVNPENSIHLTMAGNEVFKVAVTELAHIVDETLAANNLDRSQLDW
LVPHQANLRIISATAKKLGMSMDNVVVTLDRHGNTSAASVPCALDEAVRDGRIKPGQLVL
LEAFGGGFTWGSALVRF

AF384041

>sp|P0A3C5|FABH_STRPN 3-oxoacyl-[acyl-carrier-protein] synthase 3 OS=Streptococcus pneumoniae serotype 4 (strain ATCC BAA-334 / TIGR4) GN=fabH PE=3 SV=1
MAFAKISQVAHYVPEQVVTNHDLAQIMDTNDEWISSRTGIRQRHISRTESTSDLATEVAK
KLMAKAGITGEELDFIILATITPDSMMPSTAARVQANIGANKAFAFDLTAACSGFVFALS
TAEKFIASGRFQKGLVIGSETLSKAVDWSDRSTAVLFGDGAGGVLLEASEQEHFLAESLN
SDGSRSECLTYGHSGLHSPFSDQESADSFLKMDGRTVFDFAIRDVAKSIKQTIDESPIEV
TDLDYLLLHQANDRILDKMARKIGVDRAKLPANMMEYGNTSAASIPILLSECVEQGLIPL
DGSQTVLLSGFGGGLTWGTLILTI

fadY family members for initiation of straight chain fatty acid synthesis

PA5174

>tr|Q9HU15|Q9HU15_PSEAE Probable beta-ketoacyl synthase OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=PA5174 PE=3 SV=1
MSRLPVIVGFGGYNAAGRSSFHHGFRRMVIESMDPQARQETLAGLAVMMKLVKAEGGRYL
AEDGTPLSPEDIERRYAERIFASTLVRRIEPQYLDPDAVHWHKVLELSPAEGQALTFKAS
PKQLPEPLPANWSIAPAEDGEVLVSIHERCEFKVDSYRALTVKSAGQLPTGFEPGELYNS
RFHPRGLQMSVVAATDAIRSTGIDWKTIVDNVQPDEIAVFSGSIMSQLDDNGFGGLMQSR
LKGHRVSAKQLPLGFNSMPTDFINAYVLGSVGMTGSITGACATFLYNLQKGIDVITSGQA
RVVIVGNSEAPILPECIEGYSAMGALATEEGLRLIEGRDDVDFRRASRPFGENCGFTLAE
SSQYVVLMDDELALRLGADIHGAVTDVFINADGFKKSISAPGPGNYLTVAKAVASAVQIV
GLDTVRHASFVHAHGSSTPANRVTESEILDRVASAFGIDGWPVTAVKAYVGHSLATASAD
QLISALGTFKYGILPGIKTIDKVADDVHQQRLSISNRDMRQDKPLEVCFINSKGFGGNNA
SGVVLSPRIAEKMLRKRHGQAAFAAYVEKREQTRAAARAYDQRALQGDLEIIYNFGQDLI
DEHAIEVSAEQVTVPGFSQPLVYKKDARFSDMLD

Pmen_0396

>pmy:Pmen_0396 pyrC; dihydroorotase (EC:3.5.2.3); K01465 dihydroorotase [EC:3.5.2.3] (A)
MRTAILGARVIDPASGLDQVTDLYIDGTKLVAFGQAPAGFTADKTLNAQGLIAAPGLVDL
SVALREPGYSRKGSIATETLAAAAGGVTSLCCPPLTKPVLDTPAVAELILDRAREAGHTK
VFPIGALSKGLAGEQLAELVALRDAGCVAFGNGLDNFRSARTLRRALEYAATFDLQVIFH
SQDFDLAEGGLAHEGPTASFLGLAGIPETAETVALARDLLLVEQSGVRAHFSQITSARGA
ELIANAQARGLPVTADVALYQLILTDEALIDFSSLYHVQPPLRSRADRDGLREAVKAGVI
SAIASHHQPHERDAKLAPFAATEPGISSVQLQLPLAMSLVQDGLLDLPTLLARLSSGPAA
ALRLPAGTLSVGGAADIVLFDAQASTVAGEQWYSKGSNCPFIGHCLPGAVRYTLVDGHIS
YQS

MDS_0454

>pmk:MDS_0454 beta-ketoacyl synthase (A)
MSRLPVIVGFGGYNAAGRSSFHHGFRRTVQESLEPQARQETLAGLAQMMKLVRVVDGQYQ
DQDGQPLSLADIESRYAKQILAGTLVRRIEKQHLDPDAAHWQKSIGVTPADGTSLSFLTQ
RKQLPEPLPANWSIEELEGNEVRVTLHDSCEFKVDSYRPLAVKSAGQLPTGFEPSELYNA
RFHPRGLAMTVVGVTDALRSVGIDWQRIVQHVAPDEIAVFASCIMSQLDENGFGGMMQSR

LKGGRVTAKQLALGLNTMPADFINAYVLGSVGTTGSITGACATFLYNLQKGIEQIASGKA
RVVIVGSSEAPINQECIEGYGAMGALATEEGLRQIEGKSEVDFRRASRPFGDNCGFTLAE
ACQFVVLMDDELALELGADIHGAVPDVFINADGFKKSISAPGPGNYLTVAKAVASAVQLL
GLDAVRNRSFVHAHGSSTPANRVTESEILDRVAAAFGIEQWPVTAVKAFVGHSLATASGD
QVIGALGAFKYGIVPGIKTIDAVAGDVHQHHLSLSTEDRKVGDQALDVAFINSKGFGGNN
ASALVLAPHVTERMLRKRHGQAAFDAYLARREGTRAAAAAYDQQALQGKLDIIYNFGNDM
IDDQAISITTEEVKVPGFDQPLVFRKDARYSDMLD

Psefu_4068

>tr|F6AJT1|F6AJT1_PSEF1 Beta-ketoacyl synthase OS=Pseudomonas fulva (strain 12-X) GN=Psefu_4068 PE=3 SV=1
MKSRLPVIVGFGGYNAAGRSSFHHGFRRTVIESLDEQARQETLTGLAVMTKLVRVVDGRY
QSQDGEALSPADIERRYGAQILASTLVRRIEKQHLDPDAAHWHKSIAVGGEAGSLTFVSS
RKQLPEPLPANWTVEELGGNDVRVTLHDSCEFKVDSYRALPVKSAGQLPTGFEPGELYNS
RFHPRGLQMAVVGVTDALRATGVPWQTIVDHVAPDEIAVFAGSIMSQLDENGFGGLMQSR
LKGHRVSSKQLALGLNTMPADFINAYVLGSVGTTGSVTGACATFLYNLQKGIEQINAGKA
RVVIVGNSEAPINAECIEGYGAMGALATEDGLRLIEGKDDVDFRRASRPFGENCGFTLSE
ACQFVVLMDDELALQLGADIHGAATDVFINADGFKKSISAPGPGNYLTVAKAVAAATQLV
GIDAVRRRSFVHAHGSSTPANRVTESELLDRVAAAFAIDSWPVAAVKAFVGHSLATASGD
QVISALGTFKYGIIPGIKTIDEVAADVHQQHLSISNVDRHDQRMDVCFINSKGFGGNNAS
AVVLAPHVVERMLRKRHGEAAFSAYQQRREQTRANAQAYDEQATKGQLEIIYNFGNDLID
DTEIAIDDAQIKVPGFAQPLLYKQDDRYSDMLD

Avin_05510

>avn:Avin_05510 beta-ketoacyl synthase (A)
MSRLPVIVGFGGYNSAGRSSFHHGFRRTVIESLTPQARQETLAGLAVMMKLVSVVDGQYR
DSDGSTLTPAEIERRHGERILAATLIRRIERQYFDVDATHWHKSLTLSGEDQPLHFTTSA
KQLPEPLPANWSVEPLEEHQVRVTIHGSCEFKVDSYREMPVKSAGQLPTGFEPGELYNSR
FHPRGLQLSVVAATDALRSTGIDWQTILDHVQPDEVAVFSGSIMSQLDENGYGGLLQSRL
KGHRVSSKQLPLGFNSMPTDFINAYVLGSVGSTGSITGACATFLYNLQKGIDVITSGQAR
VVVAGNAEAPITPEIVEGYAAMGALATEEGLRHIEGRDQVDFRRASRPFGANCGFTLAEA
AQYVVLMDDSLALELGADIHGAVPDVFVNADGFKKSISAPGPGNYLTVAKAVASAMQLVG
EDGVRQRSFIHAHGSSTPANRVTESELLDRVAGAFGIADWPVAAVKAYVGHSLATASGDQ
LISALGTFKYGLLPGIKTVDRFADDVHDQHLRLSMRDVRRDDLDVCFINSKGFGGNNATG
VLLSPRVTEKMLRKRHGEAAFADYRSRREATREAARRYDEQVLQGRFDILYNFGQDMIDE
HAIEVNEEGVKVPGFKQAIRFRKDERFGDMLD

PSPA7_5914

>pap:PSPA7_5914 putative beta-ketoacyl synthase (A)
MSRLPVIVGFGGYNAAGRSSFHHGFRRMVIESMDPQARQETLAGLAVMMKLVKAEGGRYL
AEDGTPLSPEDIERRYAERIFASTLVRRIEPQYLDPDAVHWHKVLEATPAEGQALTFKAS
PKQLPEPLPGNWSVTPAADGEVLVSIHERCEFKVDSYRPLTVKSAGQLPTGFEPGELYNS
RFHPRGLQMSVVAATDAIRSTGIDWQTIVDNVQPDEIAVFSGSIMSQLDDNGFGGLMQSR
LKGHRVSAKQLPLGFNSMPTDFINAYVLGSVGMTGSITGACATFLYNLQKGIDVITSGQA
RVVIVGNSEAPILPECIEGYSAMGALATEEGLRLIEGRDEVDFRRASRPFGENCGFTLAE
SSQYVVLMDDELALRLGADIHGAVTDVFINADGFKKSISAPGPGNYLTVAKAVASAVQIV
GLDTVRHASFVHAHGSSTPANRVTESEILDRVASAFGIDGWPVTAVKAYVGHSLATASAD
QLISALGTFKYGILPGIKTIDKVADDVHQQRLSISNRDVRQDKPLEVCFINSKGFGGNNA
SGVVLSPRIAEKMLRRRHGEAAFAAYVEKREQTRGAARAYDQRALQGDLEIIYNFGQDLI
DEQAIEVSAEQVTVPGFSQPLVYKKDARFSDMLD

PLES_55661

>pag:PLES_55661 putative beta-ketoacyl synthase (A)
MYRLPVIVGFGGYNAAGRSSFHHGFRRMVIESMDPQARQETLAGLAVMMKLVKAEGGRYL
AEDGTPLSPEDIERRYAERIFASTLVRRIEPQYLDPDAVHWHKVLELSPAEGQALTFKAS
PKQLPEPLPANWTIAPAEDGEVLVSIHERCEFKVDSYRALTVKSAGQLPTGFEPGELYNS
RFHPRGLQMSVVAATDAIRSTGIDWKTIVDNVQPDEIAVFSGSIMSQLDDNGFGGLMQSR
LKGHRVSAKQLPLGFNSMPTDFINAYVLGSVGMTGSITGACATFLYNLQKGIDVITSGQA
RVVIVGNSEAPILPECIEGYSAMGALATEEGLRLIEGRDDVDFRRASRPFGENCGFTLAE
SSQYVVLMDDELALRLGADIHGAVTDVFINADGFKKSISAPGPGNYLTVAKAVASAVQIV
GLDTVRHASFVHAHGSSTPANRVTESEILDRVASAFGIDGWPVTAVKAYVGHSLATASAD
QLISALGTFKYGILPGIKTIDKVADDVHQQRLSISNRDMRQDKPLEVCFINSKGFGGNNA
SGVVLSPRIAEKMLRKRHGQAAFAAYVEKREQTRAAARAYDQRALQGDLEIIYNFGQDLI
DEHAIEVSAEQVTVPGFSQPLVYKKDARFSDMLD

PA14_68360

>tr|Q02EJ1|Q02EJ1_PSEAB Putative beta-ketoacyl synthase OS=Pseudomonas aeruginosa (strain UCBPP-PA14) GN=PA14_68360 PE=3 SV=1
MSRLPVIVGFGGYNAAGRSSFHHGFRRMVIESMDPQARQETLAGLAVMMKLVKAEGGRYL
AEDGTPLSPEDIERRYAERIFASTLVRRIEPRYLDPDAVHWHKVLELSPAEGQALTFKAS
PKQLPEPLPANWSIAPAEDGEVLVSIHERCEFKVDSYRALTVKSAGQLPTGFEPGELYNS
RFHPRGLQMSVVAATDAIRSTGIDWKTIVDNVQPDEIAVFSGSIMSQLDDNGFGGLMQSR
LKGHRVSAKQLPLGFNSMPTDFINAYVLGSVGMTGSITGACATFLYNLQKGIDVITSGQA
RVVIVGNSEAPILPECIEGYSAMGALATEEGLRLIEGRDDVDFRRASRPFGENCGFTLAE
SSQYVVLMDDELALRLGADIHGAVTDVFINADGFKKSISAPGPGNYLTVAKAVASAVQIV
GLDTVRHASFVHAHGSSTPANRVTESEILDRVASAFGIDGWPVTAVKAYVGHSLATASAD
QLISALGTFKYGILPGIKTIDKVADDVHQQRLSISNRDMRQDKPLEVCFINSKGFGGNNA
SGVVLSPRIAEKMLRKRHGQAAFAAYVEKREQTRAAARAYDQRALRGDLEIIYNFGQDLI
DEHAIEVSAEQVTVPGFSQPLVYKKDARFSDMLD

fabHA promoter

ACGCCTCCTTTCCATATACCATACTCTATGAGTAAGATGAACTGATAGTTTAGACGAATATATTGCCATGTGAAAAA
AAATAGGATAGAATTAGTACCTGATACTAATAATTGATCACAACCTGATTGATCTTCTAAATTTAAGATATAAAGGA
GTCTTCCCTA

Proteins that prefer to initiation fatty acid synthesis using short straight chain starters

fabHA

>gnl|BSUB|BSU11330-MONOMER beta-ketoacyl-acyl carrier protein synthase III 1208222..1209160 Bacillus subtilis subtilis 168
MKAGILGVGR YIPEKVLTNH DLEKMVETSD EWIRTRTGIE ERRIAADDVF SSHMAVAAAK
NALEQAEVAA EDLDMILVAT VTPDQSFPTV SCMIQEQLGA KKACAMDISA ACAGFMYGVV
TGKQFIESGT YKHVLVVGVE KLSSITDWED RNTAVLFGDG AGAAVVGPVS DDRGILSFEL
GADGTGGQHL YLNEKRHTIM NGREVFKFAV RQMGESCVNV IEKAGLSKED VDFLIPHQAN
IRIMEAARER LELPVEKMSK TVHKYGNTSA ASIPISLVEE LEAGKIKDGD VVVMVGFGGG
LTWGAIAIRW GR

fabHB

>gnl|BSUB|BSU10170-MONOMER beta-ketoacyl-acyl carrier protein synthase III (complement(1093747..1092770)) Bacillus subtilis subtilis 168
MSKAKITAIG TYAPSRRLTN ADLEKIVDTS DEWIVQRTGM RERRIADEHQ FTSDLCIEAV
KNLKSRYKGT LDDVDMILVA TTTSDYAFPS TACRVQEYFG WESTGALDIN ATCAGLTYGL
HLANGLITSG LHQKILVIAG ETLSKVTDYT DRTTCVLFGD AAGALLVERD EETPGFLASV
QGTSGNGGDI LYRAGLRNEI NGVQLVGSGK MVQNGREVYK WAARTVPGEF ERLLHKAGLS
SDDLDWFVPH SANLRMIESI CEKTPFPIEK TLTSVEHYGN TSSVSIVLAL DLAVKAGKLK
KDQIVLLFGF GGGLTYTGLL IKWGM

Desaturase enzymes

EF617339

>gi|148791377|gb|ABR12480.1| D9-fatty acid desaturase [Psychrobacter urativorans]
MIAKTAMGLPLKGLRLAIKSSDILIQTAGTQALRLKTWYEEGKANEAASEQPTATSNVNELSPANDDTSI
NTKTSASTSDNNKTLSTEKPIDIRELEFKKAPINWIPATILITTPIAAAVITPWYLFTHQVSAPVWGVFG
AFMVWTGISITAGYHRLLAHRAYKAHPIVKNFLLLGSTLAVQGSAFDWVSGHRSHHRHVDDRMDDPYSAK
RGFFFSHIGWMLKNYPSGKFDYKNIPDLTKDRTLQIQHKYYGLWVLAANVGLVAAIGWLIGDVWGTLVLA
GLLRLVLTHHFTFFINSLCHMFGSRPYTDTNTARDNFFLALFTWGEGYHNYHHFFQYDYRNGVKWWQYDP
TKWLIAGLSKVGLTTELRTIDDTTIKHAEVQMQFKKAQQQIDTVNAGGLDIPHAMKTFQDRIKFEFEAFT
QTVEEWQALKAKAIEMKKTEFADRLHEVDDKLKHEYANIEQKIHEHNDNLKVAFRSIGHNSKAA
AB015611

>tr|O94747|O94747_MORAP Delta-9 fatty acid desaturase OS=Mortierella alpina PE=2 SV=1
MATPLPPSFVVPATQTETRRDPLQHEELPPLFPEKITIYNIWRYLDYKHVVGLGLTPLIA
LYGLLTTEIQTKTLIWSIIYYYATGLGITAGYHRLWAHRAYNAGPAMSFVLALLGAGAVE
GSIKWWSRGHRAHHRWTDTEKDPYSAHRGLFFSHIGWMLIKRPGWKIGHADVDDLNKSKL
VQWQHKNYLPLVLIMGVVFPTLVAGLGWGDWRGGYFYAAILRLVFVHHATFCVNSLAHWL
GDGPFDDRHSPRDHFITAFVTLGEGYHNFHHQFPQDYRNAIRFYQYDPTKWVIALCAFFG
LASHLKTFPENEVRKGQLQMIEKRVLEKKTKLQWGTPIADLPILSFEDYQHACKNDNKKW
ILLEGVVYDVADFMSEHPGGEKYIKMGVGKDMTAAFNGGMYDHSNAARNLLSLMRVAVVE
YGGEVEAQKKNPSMPIYGTDHAKAE

AF037430

>sp|O34653|DES_BACSU Fatty acid desaturase OS=Bacillus subtilis (strain 168) GN=des PE=2 SV=1
MTEQTIAHKQKQLTKQVAAFAQPETKNSLIQLLNTFIPFFGLWFLAYLSLDVSYLLTLAL
TVIAAGFLTRIFIIFHDCCHQSFFKQKRYNHILGFLTGVLTLFPYLQWQHSHSIHHATSS
NLDKRGTGDIWMLTVNEYKAASRRTKLAYRLYRNPFIMFILGPIYVFLITNRFNKKGARR
KERVNTYLTNLAIVALAAACCLIFGWQSFLLVQGPIFLISGSIGVWLFYVQHTFEDSYFE
ADENWSYVQAAVEGSSFYKLPKLLQWLTGNIGYHHVHHLSPKVPNYKLEVAHEHHEPLKN
VPTITLKTSLQSLAFRLWDEDNKQFVSFRAIKHIPVSLPPDSPEKQKLRKNA

Regulatory factors

DesK

>gnl|BSUB|BSU19190-MONOMER DesK two-component sensory histidine kinase
2090574..2091686 Bacillus subtilis subtilis 168
MIKNHFTFQK LNGITPYIWT IFFILPFYFI WKSSSTFVII VGIILTLLFF SVYRFAFVSK
GWTIYLWGFL LIGISTASIT LFSYIYFAFF IAYFIGNIKE RVPFHILYYV HLISAAVAAN
FSLVLKKEFF LTQIPFVVIT LISAILLPFS IKSRKERERL EEKLEDANER IAELVKLEER
QRIARDLHDT LGQKLSLIGL KSDLARKLIY KDPEQAAREL KSVQQTARTS LNEVRKIVSS
MKGIRLKDEL INIKQILEAA DIMFIYEEEK WPENISLLNE NILSMCLKEA VTNVVKHSQA
KTCRVDIQQL WKEVVITVSD DGTFKGEENS FSKGHGLLGM RERLEFANGS LHIDTENGTK
LTMAIPNNSK

Peptide synthetase modules

srfAA module 1

(condensation domain, adenylation domain, thiolation domain, it is glutamate specific)

MEITFYPLTDAQKRIWYTEKFYPHTSISNLAGIGKLVSADAIDYVLVEQAIQEFIRRNDA
MRLRLRLDENGEPVQYISEYRPVDIKHTDTTEDPNAIEFISQWSREETKKPLPLYDCDLF
RFSLFTIKENEVWFYANVHHVISDGISMNILGNAIMHIYLELASGSETKEGISHSFIDHV
LSEQEYAQSKRFEKDKAFWNKQFESVPELVSLKRNASAGGSLDAERFSKDVPEALHQQIL
SFCEANKVSVLSVFQSLLAAYLYRVSGQNDVVTGTFMGNRTNAKEKQMLGMFVSTVPLRT
NIDGGQAFSEFVKDRMKDLMKTLRHQKYPYNLLINDLRETKSSLTKLFTVSLEYQVMQWQ
KEEDLAFLTEPIFSGSGLNDVSIHVKDRWDTGKLTIDFDYRTDLFSREEINMICERMITM
LENALTHPEHTIDELTLISDAEKEKLLARAGGKSVSYRKDMTIPELFQEKAELLSDHPAV
VFEDRTLSYRTLHEQSARIANVLKQKGVGPDSPVAVLIERSERMITAIMGILKAGGAYVP
IDPGFPAERIQYILEDCGADFILTESKVAAPEADAELIDLDQAIEEGAEESLNADVNARN
LAYIIYTSGTTGRPKGVMIEHRQVHHLVESLQQTIYQSGSQTLRMALLAPFHFDASVKQI
FASLLLGQTLYIVPKKTVTNGAALTAYYRKNSIEATDGTPAHLQMLAAAGDFEGLKLKHM
LIGGEGLSSVVADKLLKLFKEAGTAPRLTNVYGPTETCVDASVHPVIPENAVQSAYVPIG
KALGNNRLYILDQKGRLQPEGVAGELYIAGDGVGRGYLHLPELTEEKFLQDPFVPGDRMY
RTGDVVRWLPDGTIEYLGREDDQVKVRGYRIELGEIEAVIQQAPDVAKAVVLARPDEQGN
LEVCAYVVQKPGSEFAPAGLREHAARQLPDYMVPAYFTEVTEIPLTPSGKVDRRKLFALE
VKAVSGTAYTAPRNETEKAIAAIWQDVLNVEKAGIFDNFFETGGHSLKAMTLLTKIHKET
GIEIPLQFLFEHPTITALAEEADHRESKAFAVIEPAEKQEHYPL

dptA1 module 1 of daptomycin synthetase

MDMQSQRLGVTAAQQSVWLAGQLADDHRLYHCAAYLSLTGSIDPRTLGTAVRRTLDETEALRTRFVPQDG
ELLQILEPGAGQLLLEADFSGDPDPERAAHDWMHAALAAPVRLDRAGTATHALLTLGPSRHLLYFGYHHI
ALDGYGALLHLRRLAHVYTALSNGDDPGPCPFGPLAGVLTEEAAYRDSDNHRRDGEFWTRSLAGADEAPG
LSEREAGALAVPLRRTVELSGERTEKLAASAAATGARWSSLLVAATAAFVRRHAAADDTVIGLPVTARLT
GPALRTPCMLANDVPLRLDARLDAPFAALLADTTRAVGTLARHQRFRGEELHRNLGGVGRTAGLARVTVN
VLAYVDNIRFGDCRAVVHELSSGPVRDFHINSYGTPGTPDGVQLVFSGNPALYTATDLADHQERFLRFLD
AVTADPDLPTGRHRLLSPGTRARLLDDSRGTERPVPRATLPELFAEQARRTPDAPAVQHDGTVLTYRDLH
RSVERAAGRLAGLGLRTEDVVALALPKSAESVAILLGIQRAGAAYVPLDPTHPAERLARVLDDTRPRYLV
TTGHIDGLSHPTPQLAAADLLREGGPEPAPGRPAPGNAAYIIQTSGSTGRPKGVVVTHEGLATLAADQIR

RYRTGPDARVLQFISPGFDVFVSELSMTLLSGGCLVIPPDGLTGRHLADFLAAEAVTTTSLTPGALATMP
ATDLPHLRTLIVGGEVCPPEIFDQWGRGRDIVNAYGPTETTVEATAWHRDGATHGPVPLGRPTLNRRGYV
LDPALEPVPDGTTGELYLAGEGLARGYVAAPGPTAERFVADPFGPPGSRMYRTGDLVRRRSGGMLEFVGR
ADGQVKLRGFRIELGEVQAALTALPGVRQAGVLIREDRPGDPRLVGYIVPAPGAEPDAGELRAALARTLP
PHMVPWALVPLPALPLTSNGKLDRAALPVPAARAGGSGQRPVTPQEKTLCALFADVLGVTEVATDDVFFE
LGGHSLNGTRLLARIRTEFGTDLTLRDLFAFPTVAGLLPLLDDNGRQHTTPPLPPRPERLPLS

dptA1 module 5

IDRRPERLPLSFAQRRLWFL
SKLEGPSATYNIPVAVRLTGALDVPALRAALGDVTARHESLRTVFPDDGGEPRQLVLPHAEPPFLTHEVT
VGEVAEQAASATGYAFDITSDTPLRATLLRVSPEEHVLVVVIHHIAGDGWSMGPLVRDLVTAYRARTRGD
APEYTPLPVQYADYALWQHAVAGDEDAPDGRTARRLGYWREMLAGLPEEHTLPADRPRPVRSSHRGGRVR
FELPAGVHRSLLAVARDRRATLFMVVQAALAGLLSRLGAGDDIPIGTPVAGRGDEALDDVVGFFVNTLVL
RTNLAGDPSFADLVDRVRTADLDAFAHQDVPFERLVEALAPRRSLARHPLFQIWYTLTNADQDITGQALN
ALPGLTGDEYPLGASAAKFDLSFTFTEHRTPDGDAAGLSVLLDYSSDLYDHGTAAALGHRLTGFFAALAA
DPTAPLGTVPLLTDDERDRILGDWGSGTHTPLPPRSVAEQIVRRAALDPDAVAVITAEEELSYRELERLS
GETARLLADRGIGRESLVAVALPRTAGLVTTLLGVLRTGAAYLPLDTGYPAERLAHVLSDARPDLVLTHA
GLAGRLPAGLAPTVLVDEPQPPAAAAPAVPTSPSGDHLAYVIHTSGSTGRPKGVAIAESSLRAFLADAVR
RHDLTPHDRLLAVTTVGFDIAGLELFAPLLAGAAIVLADEDAVRDPASITSLCARHHVTVVQATPSWWRA
MLDGAPADAAARLEHVRILVGGEPLPADLARVLTATGAAVTNVYGPTEATIWATAAPLTAGDDRTPGIGT
PLDNWRVHILDAALGPVPPGVPGEIHIAGSGLARGYLRRPDLTAERFVANPFAPGERMYRTGDLGRFRPD
GTLEHLGRVDDQVKVRGFRIELGDVEAALARHPDVGRAAAAVRPDHRGQGRLVAYVVPRPGTRGPDAGEL
RETVRELLPDYMVPSAQVTLTTLPHTPNGKLDRAALPAPVFGTPAGRAPATREEKILAGLFADILGLPDV
GADSGFFDLGGDSVLSIQLVSRARREGLHITVRDVFEHGTVGALAAAALPAPADDADDTVPGTDVLPSIS
DDEFEEFELELGLEGEEEQW

Module 2 of CmnA (Sequence listing CmaA, A2)

PSPEPVAEVSRAEQR
IWLLSRLGGHPAEYAIPVALRLAGPLDVAKLKNAVDAVVRRHEGLRHVFPEVDGSPTRAVLDPGSITVAE
EANRSVREVLAEGVAALDPATGPLARFTLVNQGPQDHVLAIVLHHLIADGWSVDVLLRDIAAHYTGAPTA
TPGRYADYLALERAEEQDGALGRALEHFVTALDGVPDEVSFPPDHPRPAQRTGRGDVVRHRIDAAPVTAL
AERLRTTPFAVLLAAVGVLLHRVGGHRDVVVGTAVARRPDAGLDHLVGLCLNTLALRWPVQPHDTLGEVV
RAVTDRLADGLQHDAASFDRVVDKLAPARDSGRTPVFQVMALYEEPYETALALPDVTTTDVTVHCGSAQA
DAAFGFVPREGGIDLTLQFSTDVFTRATASRWARRLATLLAGARADTRVADLPLLPEDESQDLERWSGTT
GEAPTTTLHALAHEIAQRHPDRPAIHFGQNSLTYGEFDARSAQLAHELRARGVRAETPVVVCLERSPEAL
IAVYGVLKAGGAYVPVETSNPDLRIAELIADSGAALVLTQRRLADRLAALGAEVVVVDEPLPRHPTTDPE

PLTGPDHLAYVIYTSGSTGRPKGVMVQHGSVLNFLDALDRRFDLTPDDRLLHKSPLAFDVSVREVFWALT
RGASVVVAEPGRHADPGHLVDLVERERVTVAHFVPSSLAVFLEGLPGPGRCPTLRHVLTSGETLPVTTAR
AARDLLGARLRNMYGPTETTVEMTDHDVVDDTVDRLPIGHPFEGAVVRVLDADLRPVPPGSTGELCVGGL
PVARGYLGRPALTAERFVPDPLGPAGARLYRTGDLARLLPDGQLDFLGRNDFQVKVRGHRIEPGEVEAVL
GALPGVHGALVTAHDDRLIGYAVTDRDGEELRTALAERLPEHLVPSVVLTLDRFPLTGNGKLDRAALPTP
TGRHTGDSRPLTATEAALAAIWRDLLDVPEVRADDHFFALGGHSLLAARVAARAGAALGVALPLPTVLRF
PRLADLATAVDGTRADREPVRPRPDRRRRAPLSSAQRRLWIEENLRPGTATYTVAEAFRLRGELDEEAFA
AAVDDVLRRHDALRAHVESVEDGEPELVVAPEPRTALRVGDLPADRVRDALAAESARVFDPAGPLVATSL
HRLAPDEWLFQFTAHHLVVDGWSLDVLWRDLAACYHDRRAGRAPRPRDGLTFTDYTWWERDVRSRDLEPH
LAFWRGELAGLRPQPPADAHGPGAVLDFALGAALSDELRATAAGLGVSPFVLGLTAFALALGEDSPGAIG
VEVANRASAETADLVGLFVNHVPVRVAPRGTGRAAVAAVDEARRRVLPHEHVPFDLVVDLLGPGRAPTSV
AFSHLDVRGHSPRLDGVTATRLTPPHNGTAKFDLLLEVLDTEHGLTGAFEYRPERFTAARVAQVRNHWEA
ALLTLLADPDLPVDARRPDFA

/gene="mycA"
/coded_by="AF184956.1:3161..15076"
/transl_table=11

ORIGIN

```
        1 mytsqfqtlv dvirnrsnis drgirfiesd kietfvsyrq lfdeaqgflg ylqhigiqpk
       61 qeivfqiqen ksfvvafwac llggmipvpv sigedndhkl kvwriwniln npfllasetv
      121 ldkmkkfaad hdlqdfhhql ieksdiiqdr iydhpasqye peadelafiq fssgstgdpk
      181 gvmlthhnli hntcairnal aidlkdtlls wmplthdmgl iachlvpala ginqnlmpte
      241 lfirrpilwm kkahehkasi lsspnfgyny flkflkdnks ydwdlshirv iangaepilp
      301 elcdefltrc aafnmkrsai lnvyglaeas vgatfsnige rfvpvylhrd hlnlgerave
      361 vskedqncas fvevgkpidy cqiricnean egledgfigh iqikgenvtq gyynnpestn
      421 raltpdgwvk tgdlgfirkg nlvvtgrekd iifvngknvy phdiervaie ledidlgrva
      481 acgvydqetr sreivlfavy kksadrfapl vkdikkhlyq rggwsikeil pirklpktts
      541 gkvkryelae qyesgkfale stkikefleg hstepvqtpi heietallsi fsevmdgkki
      601 hlndhyfdmg atslqlsqia erieqkfgce ltvadlftyp siadlaaflv enhseikqtd
      661 takpsrsssk diaiigmsln vpgasnksdf whllengehg ireypaprvk daidylrsik
      721 sernekqfvr ggyldeidrf dysffglapk takfmdpnqr lflqsawhai edagyagdti
      781 sgsqlgvyvg yskvgydyer llsanypeel hhyivgnlps vlasriayfl nlkgpavtvd
      841 tacssslvav hmackalltg dcemalaggi rtsllpmrig ldmessdglt ktfskdsdgt
      901 gsgegvaavl lkplqaaird gdhiygvikg sainqdgttv gitapspaaq teviemawkd
      961 agiapetlsf ieahgtgtkl gdpvefnglc kafekvtekk qfcaigsvka nighlfeaag
     1021 ivgliksalm lnhkkippla hfnkpnplip fhsspfyvnq evmdftpedr plrggissfg
     1081 fsgtnahvvl eeytpeseya pedgndphlf vlsahteasl yelthqyrqy isddsqsslr
     1141 sicytastgr ahldyclami vssnqelidk ltsliqgern lpqvhfgykn ikemqpaekd
     1201 nlskqisdlm qhrpctkder itwlnriael yvqravidwr avysnevvqk tplplypfer
     1261 nrcwveavye sakerkekge valdinhtkt hiesflktvi snasgirade idsnahfigf
     1321 gldsimltqv kkaiadefnv dipmerffdt mnniesvvdy laenvpsaas tppqesvtaq
     1381 eelvisgaqp elehqehmld kiiasqnqli qqtlqaqlds fnllrnnshf vskeseisqd
     1441 ktslspksvt akknsaqeak pyipfqrqtl neqvnytpqq rqylesfiek yvdktkgskq
     1501 ytdetrfaha nnrnlssfrs ywkemvypii aersdgsrmw didgneyidi tmgfgvnlfg
     1561 hhpsfitqtv vdsthsalpp lgpmsnvage vadriractg vervafynsg teavmvalrl
     1621 araatgrtkv vvfagsyhgt fdgvlgvant kggaepanpl apgipqsfmn dliilhynhp
```

```
1681 dsldvirnlg nelaavlvep vqsrrpdlqp esflkelrai tqqsgtalim deiitgfrig
1741 lggaqewfdi qadlvtygki igggqplgiv agkaefmnti dggtwqygdd syptdeakrt
1801 fvagtfnthp ltmrmslavl rylqaegetl yerlnqktty lvdqlnsyfe qsqvpirmvq
1861 fgslfrfvss vdndlffyhl nykgvyvweg rncflstaht sddiayiiqa vqetvkdlrr
1921 ggfipegpds pndgghkepe tyelspeqkq lavvsqygnd asaalnqsim lkvkgavqht
1981 llkqavrniv krhdalrtvi hvddevqqvq arinveipii dftgypneqr esevqkwlte
2041 dakrpfhfhe qkplfrvhvl tskqdehliv ltfhhiiadg wsiavfvqel estyaaivqg
2101 splpshevvs frqyldwqqa qienghyeeg irywrqylse pipqailtsm sssryphgye
2161 gdrytvtldr plskaiksls irmknsvfat ilgafhlflq qltkqaglvi giptagqlhm
2221 kqpmlvgncv nmvpvkntas sestladylg hmkenmdqvm rhqdvpmtlv asqlphdqmp
2281 dmriifnldr pfrklhfgqm eaeliaypik cisydlflnv tefdqeyvld fdfntsviss
2341 eimnkwgtgf vnllkkmveg dsasldslkm fskedqhdll elyadhqlri sstldhkgvr
2401 avyeepenet elqiaqiwae llglekvgrs dhflslggns lkatlmlski qqtfnqkvsi
2461 gqffshqtvk elanfirgek nvkyppmkpv eqkafyrtsp aqqrvyflhq mepnqvsqnm
2521 fgqisiigky dekaliaslq qvmqrheafr tsfhiidgei vqqiageldf nvrvhsmdre
2581 efeayadgyv kpfrleqapl vraelikvdn eqaellidmh hiisdgysms iltnelfaly
2641 hgnplpeipf eykdfaewqn qlligevmeq qeeywleqfk qevpilqlpa dgsramewss
2701 egqrvtcslq sslirslqem aqqkgttlym vllaaynvll hkytgqediv vgtpvsgrnq
2761 pniesmigif iqtmgirtkp qankrftdyl devkrqtlda fenqdypfdw lvekvnvqre
2821 ttgkslfntm fvyqniefqe ihqdgctfrv kernpgvsly dlmltiedae kqldihfdfn
2881 pnqfeqetie qiirhytsll dslvkepeks lssvpmlsdi erhqllmgcn dtetpfphnd
2941 tvcqwfetqa eqrpddeavi fgnerctygq lnervnqlar tlrtkgvqad qfvaiicphr
3001 ielivgilav lkaggayvpi dpeypedriq ymlkdseaki vlaqldlhkh ltfdadvvll
3061 deessyhedr snleptcgan dlaymiytsg stgnpkgvli ehrglanyie wakevyvnde
3121 ktnfplyssi sfdltvtsif tplvtgntii vfdgedksav lstimqdpri diikltpahl
3181 hvlkemkiad gttirkmivg genlstrlaq svseqfkgql difneygpte avvgcmiyry
3241 dtkrdrrefv pigspaants iyvldasmnl vpvgvpgemy iggagvargy wnrpdltaek
3301 fvhnpfapgt imyktgdlak rlrdgnliyl grideqvkir ghrielgeve aamhkveavq
3361 kavvlareee dglqqlcayy vsnkpitiae ireqlslelp dymvpshyiq leqlpltsng
3421 kinrkalpap evsleqiaey vppgnevesk lavlwqemlg ihrvgikhnf fdlggnsira
3481 talaarihke ldvnlsvkdi fkfptieqla nmalrmekir yvsipsaqki syypvssaqk
3541 rmyllshteg geltynmtga msvegaidle rltaafqkli erhevlrtsf elyegepaqr
3601 ihpsieftie qiqareeeve dhvldfiksf dlakpplmrv glieltpekh vllvdmhhii
3661 sdgvsmnilm kdlnqfykgi epdplpiqyk dyavwqqtea qrqnikkqea ywlnrfhdei
3721 pvldmptdye rpairdyege sfeflipiel kqrlsqmeea tgttlymilm aaytillsky
3781 sgqedivvgt pvsgrshmdv esvvgmfvnt lvirnhpagr kifedylnev kenmlnayqn
3841 qdypleeliq hvhllkdssr nplfdtmfvl qnldqvelnl dslrftpykl hhtvakfdlt
3901 lsiqtdqdkh hglfeyskkl fkksrieals kdylhilsvi sqqpsiqieh ielsgstaed
3961 dnlihsieln f
```

Psrf-Gly-lgr_m2-F3-TE-pUC19

```
   1  TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCG    50
  51  GAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCG   100
 101  TCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG   150
 151  CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATATCGACAAAAATG   200
 201  TCATGAAAGAATCGTTGTAAGACGCTCTTCGCAAGGGTGTCTTTTTTTGC   250
 251  CTTTTTTTCGGTTTTTGCGCGGTACACATAGTCATGTAAAGATTGTAAAT   300
```

```
 301 TGCATTCAGCAATAAAAAAAGATTGAACGCAGCAGTTTGGTTTAAAAATT  350
 351 TTTATTTTTCTGTAAATAATGTTTAGTGGAAATGATTGCGGCATCCCGCA  400
 401 AAAAATATTGCTGTAAATAAACTGGAATCTTTCGGCATCCCGCATGAAAC  450
 451 TTTTCACCCATTTTTCGGTGATAAAAACATTTTTTTCATTTAAACTGAAC  500
 501 GGTAGAAAGATAAAAAATATTGAAAACAATGAATAAATAGCCAAAATTGG  550
 551 TTTCTTATTAGGGTGGGGTCTTGCGGTCTTTATCCGCTTATGTTAAACGC  600
 601 CGCAATGCTGACTGACGGCAGCCTGCTTTAATAGCGGCCATCTGTTTTTT  650
 651 GATTGGAAGCACTGCTTTTTAAGTGTAGTACTTTGGGCTATTTCGGCTGT  700
 701 TAGTTCATAAGAATTAAAAGCTGATATGGATAAGAAAGACGAAAATGCGTT  750
 751 GCACATGTTCACTGCTTATAAAGATTAGGGGAGGTATGACAATATGGAAA  800
 801 TAACTTTTTACCCTTTAACGGATGCACAAAAACGAATTTGGTACACAGAA  850
 851 AAATTTTATCCTCACACGAGCATTTCAAATCTTGCGGGGATTGGTAAGCT  900
 901 GGTTTCAGCTGATGCGATTGATTATGTGCTTGTTGAGCAGGCGATTCAAG  950
 951 AGTTTATTCGCAGAAATGACGCCATGCGCCTTCGGTTGCGGCTAGATGAA 1000
1001 AACGGGGAGCCTGTTCAATATATTAGCGAGTATCGGCCTGTTGATATAAA 1050
1051 ACATACTGACACTACTGAAGATCCGAATGCGATAGAGTTTATTTCACAAT 1100
1101 GGAGCCGGGAGGAAACGAAGAAACCTTTGCCGCTATACGATTGTGATTTG 1150
1151 TTCCGTTTTTCCTTGTTCACCATAAAGGAAAATGAAGTGTGGTTTTACGC 1200
1201 AAATGTTCATCACGTGATTTCTGATGGTATGTCCATGAATATTGTCGGGA 1250
1251 ATGCGATCATGCACATTTATTTAGAATTAGCCAGCGGCTCAGAGACAAAA 1300
1301 GAAGGAATCTCGCATTCATTTATCGATCATGTTTTATCTGAACAGGAATA 1350
1351 TGCTCAATCGAAGCGGTTTGAAAAGGACAAGGCGTTTTGGAACAAACAAT 1400
1401 TTGAATCGGTGCCTGAACTTGTTTCCTTGAAACGGAATGCATCCGCAGGG 1450
1451 GGAAGTTTAGATGCTGAGAGGTTCTCTAAAGATGTGCCTGAAGCGCTTCA 1500
1501 TCAGCAGATTCTGTCGTTTTGTGAGGCGAATAAAGTCAGTGTTCTTTCGG 1550
1551 TATTTCAATCGCTGCTCGCCGCCTATTTGTACAGGGTCAGCGGCCAGAAT 1600
1601 GATGTTGTGACGGGAACATTTATGGGCAACCGGCAAAATGCGAAAGAGAA 1650
1651 GCAGATGCTTGGCATGTTTGTTTCTACGGTTCCGCTTCGGACAAACATTG 1700
1701 ACGGCGGGCAGGCGTTTTCAGAATTTGTCAAAGACCGGATGAAGGATCTG 1750
1751 ATGAAGACACTTCGCCACCAAAAGTATCCGTATAATCTCCTAATCAACGA 1800
1801 TTTGCGTGAAACAAAGAGCTCTCTGACCAAGCTGTTCACGGTTTCTCTTG 1850
1851 AATATCAAGTGATGCAGTGGCAGAAAGAAGAGGATCTTGCCTTTTTGACT 1900
1901 GAGCCGATTTTCAGCGGCAGCGGATTAAATGATGTCTCAATTCATGTAAA 1950
1951 GGATCGATGGGATACTGGGAAACTCACCATAGATTTTGATTACCGCACTG 2000
2001 ATTTATTTTCACGTGAAGAAATCAACATGATTTGTGAGCGCATGATTACC 2050
2051 ATGCTGGAGAACGCGTTAACGCATCCAGAACATACAATTGATGAATTAAC 2100
2101 ACTGATTTCTGATGCGGAGAAACGCGATTTGTTTTTGCGGGTGAACGATA 2150
2151 CAGCCAAGGCGTATCCGAACAAGCTGATCATGTCGATGCTGGAGGATTGG 2200
2201 GCGGCGGCTACCCCTGACAAAACAGCGCTAGTCTTCCGCGAACAACGCGT 2250
2251 GACGTATCGCGAGCTGAACGAGCGGGTCAACCAGTTGGCACACACTTTGC 2300
2301 GCGAAAAAGGGGTGCAACCTGACGATCTCGTGATGCTGATGGCAGAGCGG 2350
2351 TCGGTCGAGATGATGGTGGCGATTTTCGCTGTGTTGAAAGCGGGCGGAGC 2400
2401 GTACTTGCCCATCGACCCGCACAGTCCGGCGGAGCGAATCGCCTACATTT 2450
2451 TCGCAGACAGCGGAGCCAAGCTGGTGCTGGCACAGTCGCCGTTTGTGGAA 2500
2501 AAGGCAAGCATGGCGGAAGTGGTCCTTGATCTGAACAGTGCGAGCAGCTA 2550
2551 TGCGGCGGATACGAGCAACCCGCCACTGGTCAACCAGCCAGGCGATCTGG 2600
2601 TGTATGTCATGTACACTTCCGGCTCAACGGGAAAACCAAAAGGCGTGATG 2650
2651 ATCGAGCACGGAGCGCTGCTCAATGTGCTTCACGGAATGCAGGACGAGTA 2700
2701 CCCGCTTTTGCAGGACGATGCCTTCTTGCTCAAGACAACCTACATATTCG 2750
2751 ATATTTCAGTCGCGGAAATTTTCGGGTGGGTTCCGGGTCGTGGCAAACTG 2800
2801 GTGATTTTGGAACCGGAGGCGGAAAAGAACCCGAAGGCTATTTGGCAGGC 2850
2851 GGTAGTCGGAGCGGGAATTACCCACATCAACTTCGTGCCCTCCATGCTGA 2900
2901 TCCCGTTTGTCGAGTATTTGGAAGGGCGAACAGAAGCAAATCGCTTGCGG 2950
2951 TACATCTTGGCTTGCGGCGAAGCGATGCCGGATGAACTCGTGCCAAAAGT 3000
3001 GTACGAAGTATTGCCAGAGGTGAAGCTGGAAAACATCTACGGCCCGACAG 3050
```

```
3051  AAGCGACGATTTACGCTTCCCGTTACTCGCTCGCGAAAGGCTCGCAGGAA  3100
3101  AGTCCTGTTCCAATCGGAAAGCCGCTGCCCAACTATCGCATGTATATCAT  3150
3151  CAATCGGCATGGACAACTGCAACCAATCGGCGTACCAGGAGAGCTATGCA  3200
3201  TCGCAGGAGCAAGTCTGGCGAGAGGGTATTTGAACAATCCAGCGCTGACA  3250
3251  GAAGAAAAATTCACTCCTCATCCGCTGGAGAAAGGCGAGCGGATTTATCG  3300
3301  CACGGGTGATCTCGCCCGTTATCGCGAGGATGGCAACATCGAATACCTCG  3350
3351  GACGGATGGACCATCAGGTGAAAATTCGCGGATACCGGATCGAACTGGAC  3400
3401  GAAATCCGCAGCAAGCTGATTCAGGAGGAAACGATTCAGGACGCGGTGGT  3450
3451  CGTAGCCCGAAACGATCAAAACGGCCAAGCGTACTTGTGCGCCTACCTGC  3500
3501  TGTCCGAACAGGAGTGGACAGTCGGTCAACTGCGCGAGTTGCTTCGCCGT  3550
3551  GAACTGCCTGAATACATGATTCCGGCCCATTTCGTTTTGCTGAAACAGTT  3600
3601  CCCGCTCACAGCCAATGGCAAGCTCGATCGCAAGGCTTTGCCAGAACCGG  3650
3651  ACGGCAGTGTGAAAGCGGAAGCGGAATATGCAGCGCCGCGCACGGAACTG  3700
3701  GAAGCGACTTTGGCGCACATTTGGGGCGAAGTGCTCGGAATCGAACGGAT  3750
3751  CGGGATTCGCGACGATTTCTTTGCGCTCGGAGGGCATTCCTTGAAGGCCA  3800
3801  TGACCGCCGTCCCGCATCAACAAGAGCTCGGGATTGATCTTCCAGTGAAG  3850
3851  CTTTTGTTTGAAGCGCCGACGATCGCCGGCATTTCAGCGTATTTGAAAAA  3900
3901  CGGGGGCTCTGATGGCTTGCAGGATGTAACGATAATGAATCAGGATCAGG  3950
3951  AGCAGATCATTTTCGCATTTCCGCCGGTTCTGGGCTATGGCCTTATGTAC  4000
4001  CAAAATCTGTCCAGCCGCTTGCCGTCATACAAGCTATGCGCCTTTGATTT  4050
4051  TATTGAGGAGGAAGACCGGCTTGACCGCTATGCGGATTTGATCCAGAAGC  4100
4101  TGCAGCCGGAAGGGCCTTTAACATTGTTTGGATATTCAGCGGGATGCAGC  4150
4151  CTGGCGTTTGAAGCTGCGAAAAAGCTTGAGGAACAAGGCCGTATTGTTCA  4200
4201  GCGGATCATCATGGTGGATTCCTATAAAAAACAAGGTGTCAGTGATCTGG  4250
4251  ACGGACGCACGGTTGAAAGTGATGTCGAAGCGTTGATGAATGTCAATCGG  4300
4301  GACAATGAAGCGCTCAACAGCGAAGCCGTCAAACACGGCCTCAAGCAAAA  4350
4351  AACACATGCCTTTTACTCATACTACGTCAACCTGATCAGCACAGGCCAGG  4400
4401  TGAAAGCAGATATTGATCTGTTGACTTCCGGCGCTGATTTTGACATGCCG  4450
4451  GAATGGCTTGCATCATGGGAAGAAGCTACAACAGGTGTTTACCGTGTGAA  4500
4501  AAGAGGCTTCGGAACACACGCAGAAATGCTGCAGGGCGAAACGCTAGATA  4550
4551  GGAATGCGGAGATTTTGCTCGAATTTCTTAATACACAAACCGTAACGGTT  4600
4601  TCATAAAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT  4650
4651  GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCA  4700
4701  TAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATT  4750
4751  GCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCT  4800
4801  GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC  4850
4851  GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTG  4900
4901  CGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA  4950
4951  ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG  5000
5001  CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC  5050
5051  CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA  5100
5101  CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG  5150
5151  TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT  5200
5201  CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT  5250
5251  CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC  5300
5301  CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCC  5350
5351  AACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG  5400
5401  GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT  5450
5451  GGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTG  5500
5501  CTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAA  5550
5551  ACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA  5600
5601  CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGG  5650
5651  TCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG  5700
5701  ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT  5750
5751  TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAA  5800
```

```
5801  TGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC  5850
5851  CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCT  5900
5901  TACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCG  5950
5951  GCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG  6000
6001  AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCC  6050
6051  GGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTT  6100
6101  GCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC  6150
6151  ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT  6200
6201  TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGT  6250
6251  AAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTC  6300
6301  TCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT  6350
6351  CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC  6400
6401  CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT  6450
6451  GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC  6500
6501  CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCT  6550
6551  TCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG  6600
6601  GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC  6650
6651  TCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT  6700
6701  CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGG  6750
6751  GGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCA  6800
6801  TTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTT  6850
6851  CGTC  6854
```

CoA Ligases

GenBank: AAX31555.1

acyl-CoA ligase [Streptomyces roseosporus NRRL 11379]

GenPept Graphics

```
>gi|60650930|gb|AAX31555.1| acyl-CoA ligase [Streptomyces
roseosporus NRRL 11379]
MSESRCAGQGLVGALRTWARTRARETAVVLVRDTGTTDDTASVDYGQLDEWARSIAVTLRQQ
LAPGGRAL
LLLPSGPEFTAAYLGCLYAGLAAVPAPLPGGRHFERRRVAAIAADSGAGVVLTVAGETASVH
DWLTETTA
PATRVVAVDDRAALGDPAQWDDPGVAPDDVALIQYTSGSTGNPKGVVVTHANLLANARNLAE
ACELTAAT
PMGGWLPMYHDMGLLGTLTPALYLGTTCVLMSSTAFIKRPHLWLRTIDRFGLVWSSAPDFAY
DMCLKRVT
DEQIAGLDLSRWRWAGNGAEPIRAATVRAFGERFARYGLRPEALTAGYGLAEATLFVSRSQG
LHTARVAT
AALERHEFRLAVPGEAAREIVSCGPVGHFRARIVEPGGHRVLPPGQVGELVLQGAAVCAGYW
QAKEETEQ
```

TFGLTLDGEDGHWLRTGDLAALHEGNLHITGRCKEALVIRGRNLYPQDIEHELRLQHPELES
VGAAFTVP
AAPGTPGLMVVHEVRTPVPADDHPALVSALRGTINREFGLDAQGIALVSRGTVLRTTSGKVR
RGAMRDLC
LRGELNIVHADKGWHAIAGTAGEDIAPTDHAPHPHPA

Acyl Carrier proteins

GenBank: AAX31556.1

probable acyl carrier protein [Streptomyces roseosporus NRRL 11379]

GenPept Graphics

>gi|60650931|gb|AAX31556.1| probable acyl carrier protein
[Streptomyces roseosporus NRRL 11379]
MNPPEAVSTPSEVTAWITGQIAEFVNETPDRIAGDAPLTDHGLDSVSGVALCAQVEDRYGIE
VDPELLWS
VPTLNEFVQALMPQLADRT

malonyl-CoA transacylase

/protein_id="AAF08794.1"

/gene="fenF"

/note="malonyl-CoA transacylase"

/codon_start=1

/transl_table=11

/product="FenF"

/db_xref="GI:6449054"

/translation="MNNLAFLFPGQGSQFVGMGKSFWNDFVLAKRLFEEASDAISMDV

KKLCFDGDMTELTRTMNAQPAILTVSVIAYQVYMQEIGIKPHFLAGHSLGEYSALVCA

GVLSFQEAVKLIRQRGILMQNADPEQLGTMAAITQVYIQPLQDLCTEISTEDFPVGVA

CMNSDQQHVISGHRQAVEFVIKKAERMGANHTYLNVSAPFHSSMMRSASEQFQTALNQ

YSFRDAEWPIISNVTAIPYNNGHSVREHLQTHMTMPVRWAESMHYLLLHGVTEVIEMG

PKNVLVGLLKKITNHIAAYPLGQTSDLHLLSDSAERNENIVNLRKKQLNKMMIQSIIA

RNYNKDAKTYSNLTTPLFPQIQLLKERVERKEVELSAEELEHSIHLCQLICEAKQLPT

WEQLRILK"

Claims

We claim:

1. A method of making an acyl amino acid composition by contacting an engineered peptide synthetase with an amino acid substrate and an acyl entity substrate for the engineered peptide synthetase, under conditions and for a time sufficient for an acyl amino acid composition to be made.

1A. The method of claim 1, wherein the engineered peptide sythetase includes an adenylation (A) domain, a thiolation (T) domain, and a condensation (C) domain.

1A1, The method of claim 1 or claim 1A, wherein the engineered peptide synthetase lacks thioesterase domain, and/or a reductase domain.

1A1a. The method of claim 1 or 1A1, wherein the engineered peptide synthetase contains only a single peptide synthetase domain.

1A2. The method of claim 1 or 1A1a, wherein the engineered peptide synthetase is or comprises a peptide synthetase domain found in as a first domain in a peptide synthetase that synthesizes a lipopeptide.

1B. The method of claim 1 or claim 1A, wherein acyl amino acid composition includes, as a prominent component, an acyl amino acid whose amino acid moiety is from an amino acid selected from the group consisting of glycine or glutamate and whose acyl moiety is from a fatty acid selected from the group consisting of myristic acid and or lauric acid.

1C. The method of any one of the preceding claims, wherein the step of contacting comprises providing a cell engineered to express at least one engineered peptide synthetase.

2. A cell engineered to express at least one engineered peptide synthetase that synthesizes an acyl amino acid.

3. An acyl amino acid composition produced by an engineered peptide synthetase.

3B. The composition of claim 3, wherein substantially all of the acyl amino acids in the composition contain the same amino acid component.

3C. The composition of any one of claims 3-3B, wherein acyl amino acids in the compositon comprise different acyl moieties.

4. A method of preparing a product comprising:

providing or obtaining an acyl amino acid composition prepared in an engineered microbial cell;

enriching the acyl amino acid composition for a particular acyl amino acid;

combining the enriched acyl amino acid composition with at least one other component to produce a product.

5. A method comprising steps of:

contacting an engineered peptide synthetase polypeptide that comprises a single peptide synthetase domain and lacks a thioesterase domain, and/or a reductase domain with:

an amino acid substate of the peptide synthetase polypeptide; and

an acyl moiety substrate of the peptide synthetase polypeptide,

the contacting being performed under conditions and for a time sufficient that the engineered peptide synthetase polypeptide covalently links the acyl moiety from the acyl moiety substrate to the amino acid so that an acyl amino acid is generated.

5A. The method of claim 5, wherein the engineered peptide synthetase polypeptide is produced by a cell.

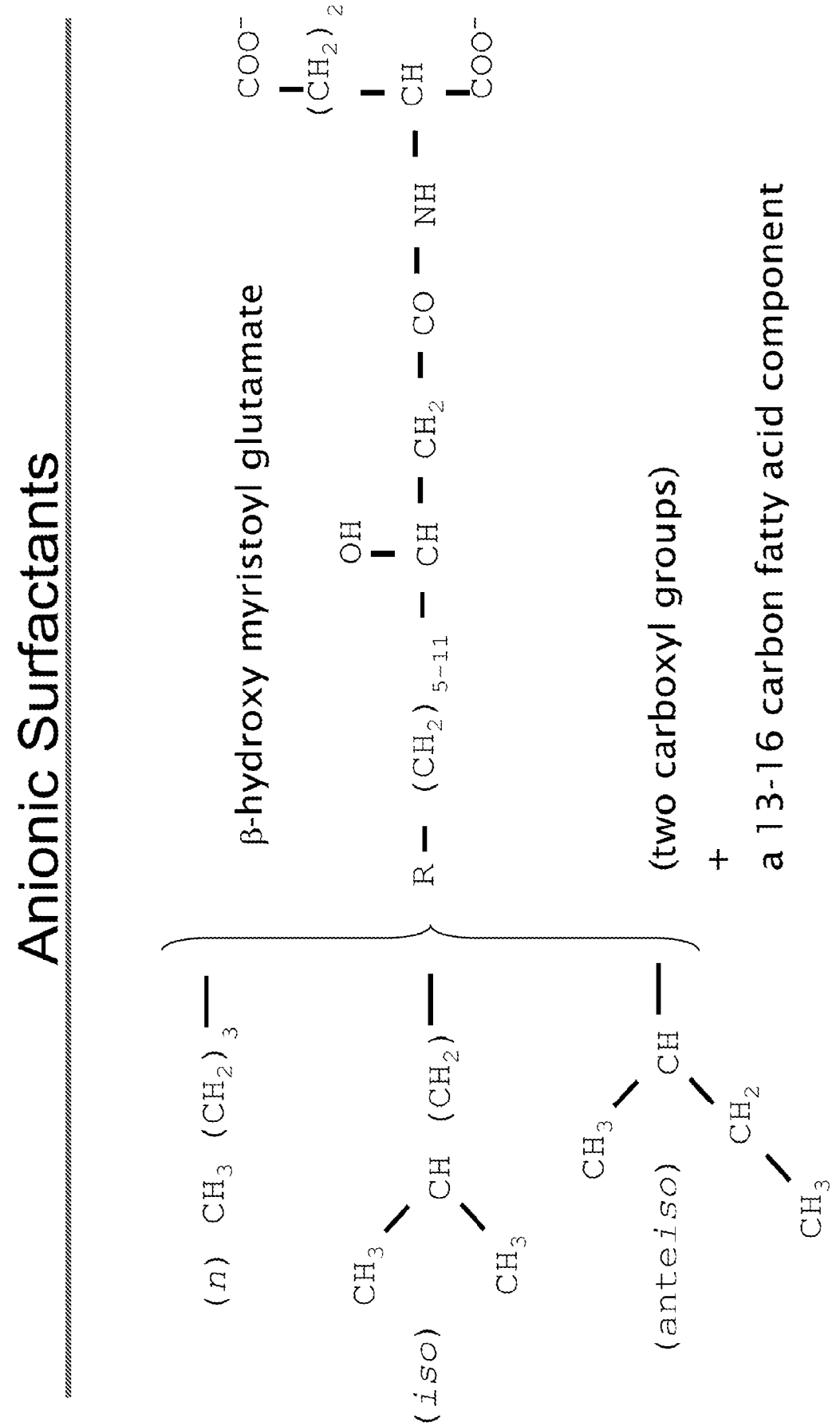
5A1. The method of claim 5A, wherein the cell is a microbial cell.

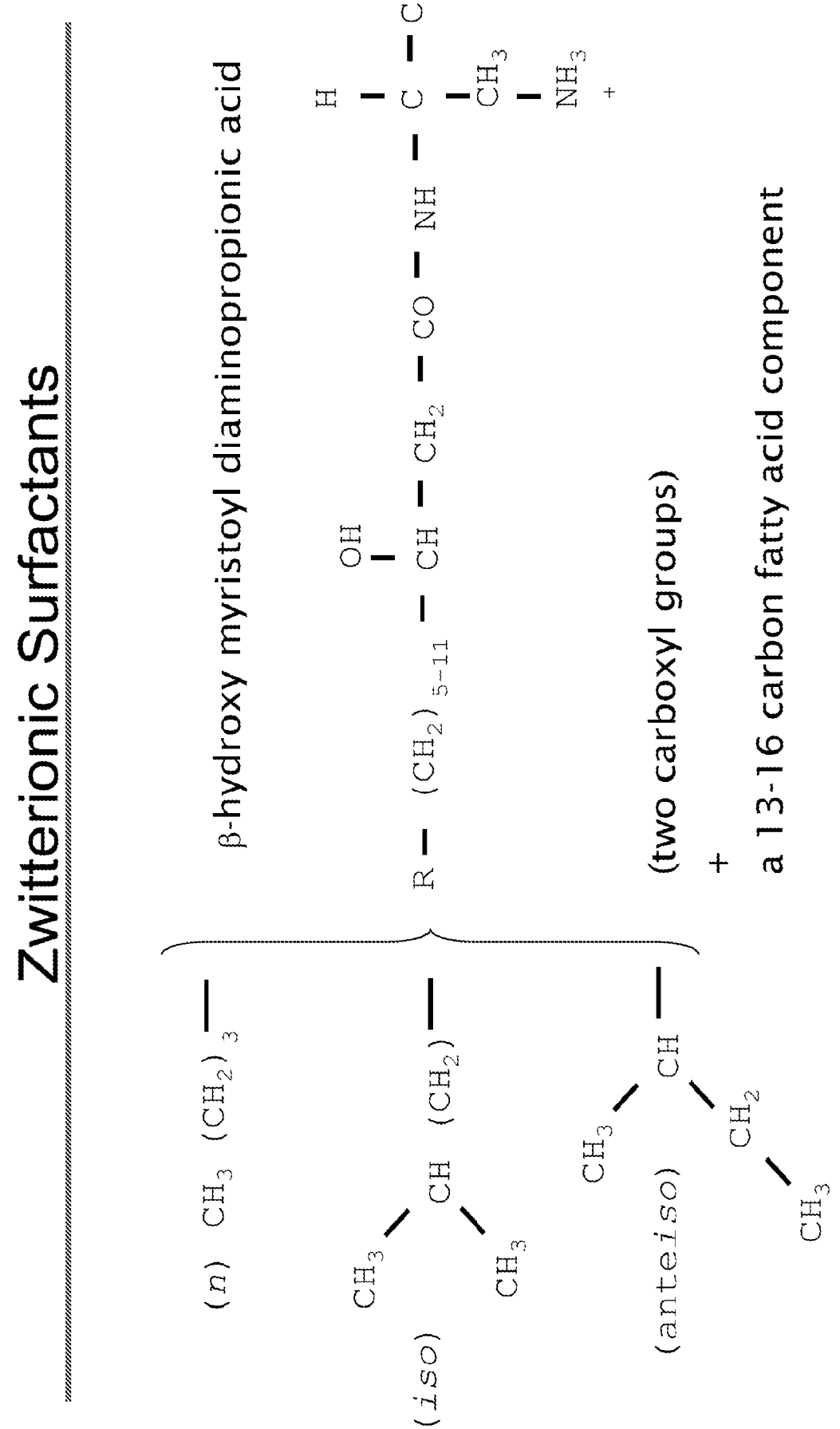
5A1a. The method of claim 5A1, wherein the cell is a bacterial cell.

5A2. The method of any one of claims 1A, 1A1, or 1A1a, wherein the step of contacting comprises contacting the cell with the substrates.

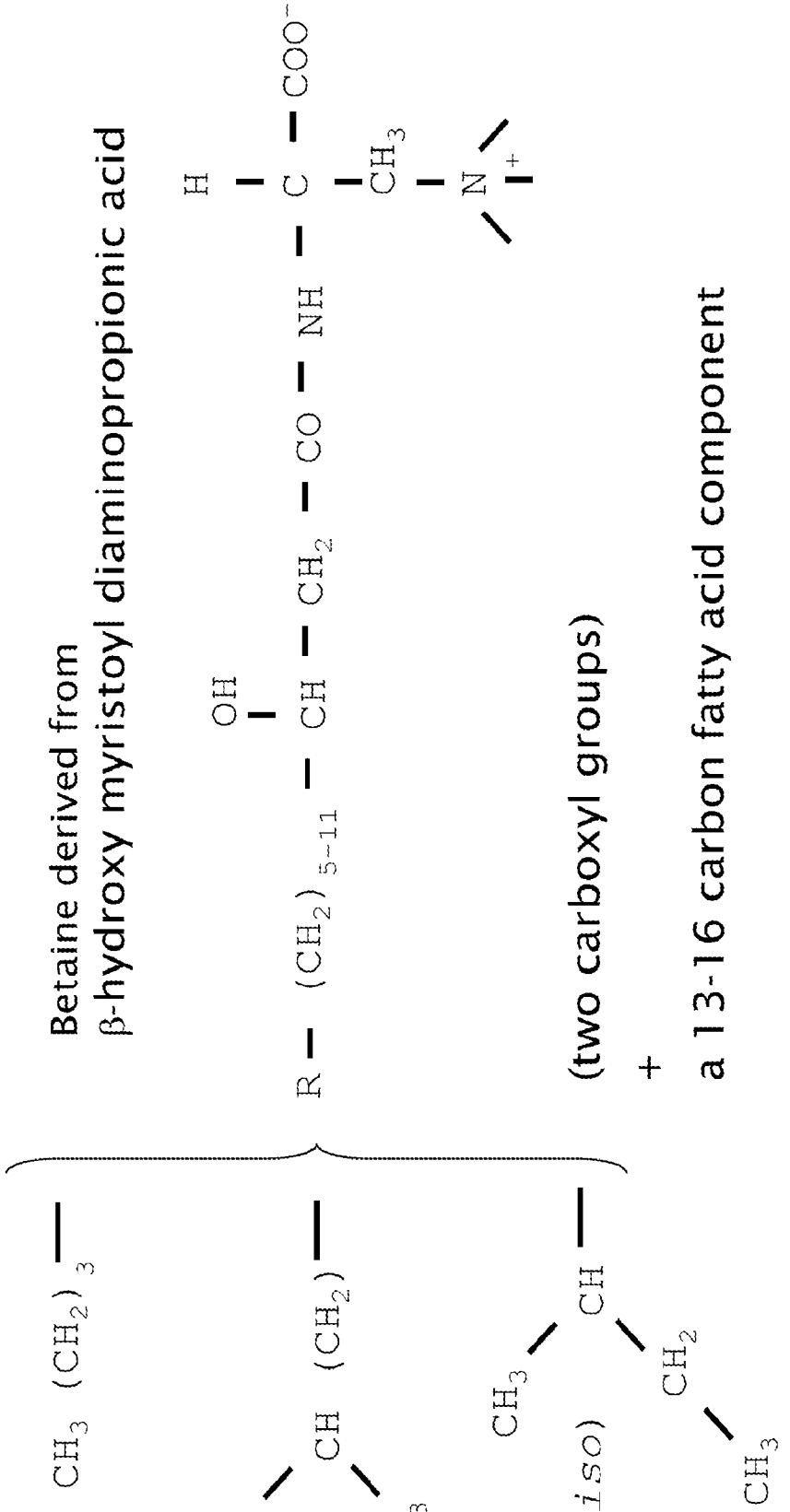
5B. The method of claim 5, wherein the cell is an engineered cell.

1C. The method of claim 5B, wherein the cell is engineered in that the peptide synthetase polypeptide is an engineered peptide synthetase polypeptide.


Anionic Surfactants
β-hydroxy myristoyl glutamate
(two carboxyl groups)
+
a 13-16 carbon fatty acid component
(n)
(iso)
(anteiso)


Zwitterionic Surfactants
β-hydroxy myristoyl diaminopropionic acid
(two carboxyl groups)
+
a 13-16 carbon fatty acid component
(n) CH3 (CH2)3
(iso) CH (CH2)
CH3
CH3
(anteiso) CH
CH3
CH2
CH3
R (CH2)5-11 CH CH2 CO NH
OH
H C COO-
CH3
NH3+

Betaines


Betaine derived from
β-hydroxy myristoyl diaminopropionic acid
(two carboxyl groups)
+
a 13-16 carbon fatty acid component
(n) CH3 (CH2)3
(iso) CH3 CH (CH2) CH3
(anteiso) CH3 CH CH2 CH3
R (CH2)5-11 CH OH CH2 CO NH C H COO- CH3 N+

Anionic Surfactants

cocoyl glycinate

(one carboxyl group)

+

a 10-14 carbon fatty acid component

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/029150

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 9/10 (2014.01)
USPC - 435/193
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12N 1/20, 1/21, 9/00, 9/10 (2014.01)
USPC - 435/183, 193, 194, 252.3, 252.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - C12N 1/20, 1/21, 9/00, 9/10 (2014.06)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatBase, Google Patents, Google, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/0030102 A1 (JARRELL et al) 03 February 2011 (03.02.2011) entire document	1-3, 6, 8-16, 18, 19
A	VAN WAGONER et al. "FeeM, an N-Acyl Amino Acid Synthase from an Uncultured Soil Microbe: Structure, Mechanism, and Acyl Carrier Protein Binding," Structure, Vol. 14, Pg. 1425-1435. 01 September 2006 (01.09.2006). entire document	1-3, 6, 8-16, 18, 19
A	US 2011/0030103 A1 (REZNIK et al) 03 February 2011 (03.02.2011) entire document	1-3, 6, 8-16, 18, 19
A	US 2012/0128603 A1 (TANAKA) 24 May 2012 (24.05.2012) entire document	1-3, 6, 8-16, 18, 19
A	US 8,318,950 B2 (NEBOLSIN et al) 27 November 2012 (27.11.2012) entire document	1-3, 6, 8-16, 18, 19

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 July 2014	07 AUG 2014
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/029150

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4, 5, 7, 17
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.