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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 January 2017 (19.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/011407 A2

(51) International Patent Classification:
C12P 7/40 (2006.01) *C12N 9/10* (2006.01)

(21) International Application Number:
PCT/US2016/041795

(22) International Filing Date:
11 July 2016 (11.07.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/191,283 10 July 2015 (10.07.2015) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: **HAGEN, Andrew, R.**; 922A 40th Street, Oakland, CA 94608 (US). **POUST, Sean**; 6714 Moeser Lane, El Cerrito, CA 94530 (US). **KATZ, Leonard**; 6389 Longcroft Drive, Oakland, CA 94611 (US). **KEASLING, Jay, D.**; 828 Contra Costa Avenue, Berkeley, CA 94707 (US).

(74) Agent: **CHIANG, Robin, C.**; Lawrence Berkeley National Laboratory, Innovation & Partnership Office, One Cyclotron Road, MS 56A-0120, Berkeley, CA 94720-8127 (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, 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, ST, 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).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: PRODUCING ADIPIC ACID AND RELATED COMPOUNDS USING HYBRID POLYKETIDE SYNTHASES

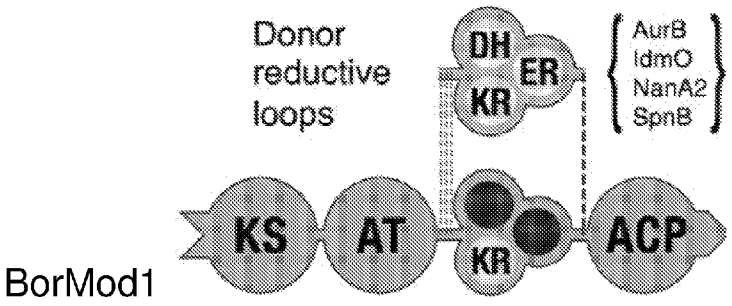

Figure 1A

(57) Abstract: The present invention provides for a polyketide synthase (PKS) capable of synthesizing a carboxylic acid, said PKS comprising a synthetic or hybrid module. The present invention also provides for a host cell comprising the PKS and when cultured produces the carboxylic acid. In some embodiments, the carboxylic acid is adipic acid.

WO 2017/011407 A2

Producing Adipic Acid and Related Compounds Using Hybrid Polyketide Synthases

Inventors: Andrew Hagen, Sean Poust, Leonard Katz, Jay D. Keasling

Related Patent Applications

[0001] The application claims priority to U.S. Provisional Patent Application Ser. No. 62/191,283, filed July 10, 2015; which is incorporated herein by reference.

Statement of Governmental Support

[0002] This invention was made with government support under Contract No. DE-AC02-05CH11231 awarded by the U.S. Department of Energy, and Grant Nos. EEC 0540879, DGE 1106400, and MCB 1341894 awarded by the National Science Foundation. The government has certain rights in the invention.

Field of the Invention

[0003] This invention relates generally to production of adipic acid and related compounds using polyketide synthases.

Background of the Invention

[0004] Dicarboxylic acids (diacids) are important compounds that are used in the manufacture of commercial polymers (e.g. polyesters, polyurethanes). The diacid adipic acid [1] is used mainly as a monomer in the production of 6,6-nylon, a polyamide generated through the reaction of [1] with hexane-1,6-diamine. Polyesters (for use in fabrics and plastics of many compositions) are formed through the polymerization of terephthalic acid [3] and a dialcohol (diol) such as ethylene glycol (to make polyethylene terephthalate), propane diol (poly(1,3-propanediol terephthalate)) or butanediol (poly(1,4-butanediolphthalate). Adipic acid is also used in the synthesis of various polyesters. Currently adipic acid is synthesized via oxidation of cyclohexane and similar petrochemicals using traditional chemical synthesis.

[0005] The large scale worldwide use of nylons and polyesters requires the production of millions of metric tons of [1] and [3] annually. These diacids are themselves synthesized from starting materials extracted from petroleum. One means of reducing the large

dependence on oil for the commercial production of polymers is to generate the diacids by a fermentation process involving the use of polyketide synthases.

[0006] The use of hybrid polyketide synthases to produce diacids with a carbon backbone with an odd number of carbon atoms is disclosed in International Patent Application No. PCT/US2009/038831, filed March 30, 2009, which published as PCT publication no. WO 2009/121066 on October 1, 2009. The use of hybrid polyketide synthases to produce diacids is disclosed in U.S. Patent Application Pub. No. 2013/0280766, now issued as U.S. Pat. No. 9,334,514.

[0007] The polyketides are one of the most diverse and chemically complicated classes of molecules known, its members frequently weighing in excess of 500 daltons and harboring numerous stereocenters. Partly owing to their antibacterial,, immunosuppressive, and anti-cancer activities, much effort has been devoted to deciphering the mechanism by which polyketide synthases (PKSs) synthesize their products. PKSs perform Claisen condensation reactions between a loaded acyl-ACP intermediate and an α-substituted (H, CH_3, C_2H_5, etc.) malonyl-CoA extender unit analogous to fatty acid biosynthesis. This is then followed by varying degrees of β-reduction by accessory domains. This condensation-reduction cycle is repeated by subsequent downstream modules until the intermediate is liberated from the enzyme, most commonly by the activity of a thioesterase domain (reviewed in (Khosla, 2009)).

[0008] Engineering of type I modular PKSs has the potential to produce an enormous variety of novel, rationally-designed compounds. Yet, more than two decades after their modular nature was discovered (Donadio et al., 1991), there are currently no commercial applications of engineered PKSs.

SUMMARY OF THE INVENTION

[0009] The present invention provides for a polyketide synthase (PKS) capable of synthesizing a carboxylic acid, said PKS comprising a synthetic module comprising the S3c variant module, or a functional variant thereof, wherein the PKS is capable of synthesizing a carboxylic acid.

[0010] The present invention also provides for a polyketide synthase (PKS) capable of synthesizing a carboxylic acid, said PKS comprising a hybrid module comprising a BorA2

KS domain, or functional variant thereof, a BorA2 AT domain, or functional variant thereof, a DH described in Example 1, or functional variant thereof, a heterologous KR domain, a heterologous ER domain, and a BorA2 ACP domain, or functional variant thereof, wherein the PKS is capable of synthesizing a carboxylic acid.

[0011] The present invention provides for a recombinant nucleic acid that encodes a polyketide synthase (PKS) of the present invention. The recombinant nucleic acid can be replicon capable of stable maintenance in a host cell. In some embodiments, the replicon is stably integrated into a chromosome of the host cell. In some embodiments, the replicon is a plasmid. The present invention also provides for a vector or expression vector comprising a recombinant nucleic acid of the present invention. The present invention provides for a host cell comprising any of the recombinant nucleic acid and/or PKS of the present invention. In some embodiments, the host cell, when cultured under a suitable condition, is capable of producing the carboxylic acid or diacid.

[0012] The present invention provides for a host cell comprising any of the recombinant nucleic acid and/or PKS of the present invention. In some embodiments, the host cell, when cultured, is capable of producing a carboxylic acid or diacid.

[0013] The present invention provides a method of producing a carboxylic acid or diacid, comprising: providing a host cell of the present invention, and culturing said host cell in a suitable culture medium such that the carboxylic acid or diacid is produced.

[0014] The present invention provides for a composition comprising a carboxylic acid or diacid isolated from a host cell from which the carboxylic acid or diacid was produced, and trace residues and/or contaminants of the host cell. Such trace residues and/or contaminants include cellular material produced by the lysis of the host cell. In some embodiments, the trace residues and/or contaminants do not or essentially do not interfere or retard a polymerization reaction involving the carboxylic acid or diacid.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The foregoing aspects and others will be readily appreciated by the skilled artisan from the following description of illustrative embodiments when read in conjunction with the accompanying drawings.

[0016] Figure 1A shows loops were introduced combinatorially into BorMod1 using two

alternative N-terminal and a single C-terminal splice sites to generate eight chimeras to be tested for adipoyl-ACP production.

[0017] Figure 1B shows the loading and first extension modules of the borrelidin PKS (hereafter referred to as "BorLM" and "BorMod1", respectively) are capable of producing a 3-hydroxy-adipoyl-ACP intermediate in vitro using succinyl-CoA as a starter substrate and the natural extender substrate, malonyl-CoA.

[0018] Figure 2A shows extension intermediate analysis of BorMod1 variants in an initial library. Variants designated by reductive loop source (A = AurB, I = IdmO, N = NanA2, S = SpnB); N-terminal junction (1, 2, 3) and BorDH2 presence (null = wildtype DH domain, t = in trans, c = in cis).

[0019] Figure 2B shows the effect of BorDH2 in trans (indicated using t). Variants designated by reductive loop source (A = AurB, I = IdmO, N = NanA2, S = SpnB); N-terminal junction (1, 2, 3) and BorDH2 presence (null = wildtype DH domain, t = in trans, c = in cis).

[0020] Figure 2C shows the effect of BorDH2 in cis (indicated using c). Variants designated by reductive loop source (A = AurB, I = IdmO, N = NanA2, S = SpnB); N-terminal junction (1, 2, 3) and BorDH2 presence (null = wildtype DH domain, t = in trans, c = in cis).

[0021] Figure 2D shows the effect of junction 3 without and with BorDH2 in cis. Variants designated by reductive loop source (A = AurB, I = IdmO, N = NanA2, S = SpnB); N-terminal junction (1, 2, 3) and BorDH2 presence (null = wildtype DH domain, t = in trans, c = in cis).

[0022] Figure 3 shows the descarboxy substrates propionyl- and CPMA-ACP were extended and fully reduced to their respective products by both S3 and S3c protein variants.

[0023] Figure 4 shows the BorMod1-TE construct produced exclusively 3-hydroxy-adipic acid whereas S3c-TE produced a mixture of the partially and fully reduced adipic acid products.

[0024] Figure 5. LC-MS/MS chromatograms of extension reactions using different starter substrates (CPMA-, CPDA-ACP) and BorMod1 variants (S3, S3c). Identity of each peak indicated by molecule appearing above it (3-cyclopentyl-3-hydroxypropanoyl intermediate

was not detected; small peak at RT approximately 9.5 mins is a contaminant found in all samples)

[0025] Figure 6. N- and C- terminal junctions for initial constructs. Arrows indicates crossover point.

[0026] Figure 7. Junctions for DH swap constructs. Arrows indicates crossover point.

[0027] Figure 8. N-terminal junctions including junction 3. Arrows and lines indicate crossover points.

[0028] Figure 9 shows a scheme for making novel polyamides or novel polyesters using diacids (using adipic acid as an example).

[0029] Figure 10 shows different domains for terminating the synthesis of the compound.

[0030] Figure 11 shows different sides chains that can be added using different extenders.

DETAILED DESCRIPTION

[0031] Before the present invention is described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0032] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0033] Unless defined otherwise, all technical and scientific terms used herein have the same

meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0034] It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a diacid" includes a plurality of such diacids, and so forth.

[0035] The term “functional variant” describes an enzyme that has a polypeptide sequence that is at least 70%, 75%, 80%, 85%, 90%, 95% or 99% identical to any one of the enzymes described herein. The “functional variant” enzyme may retain amino acids residues that are recognized as conserved for the enzyme, and may have non-conserved amino acid residues substituted or found to be of a different amino acid, or amino acid(s) inserted or deleted, but which does not affect or has insignificant effect its enzymatic activity as compared to the enzyme described herein. The “functional variant” enzyme has an enzymatic activity that is identical or essentially identical to the enzymatic activity of the enzyme described herein. The “functional variant” enzyme may be found in nature or be an engineered mutant thereof.

[0036] These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

[0037] This invention provides for an engineered enzyme capable of production of hexane 1,6-dicarboxylic acid ("adipic acid") when used in conjunction with its native loading module or provided a synthetic substrate. Successful production of adipic acid comprises one or more of the following: (1) selection of suitable reductive loop donors, (2) elucidating chimeric junction boundaries that result in high enzyme activity, (3) replacement of the dehydratase domain in the reductive loop, and (4) concatenation of a thioesterase domain.

[0038] In some embodiments, the enzyme is loaded with a succinic acid analog (presented as

succinyl-acyl carrier protein ("succinyl-ACP") from the upstream loading module or as a synthetic succinyl-n-acetyl-cysteamine ("succinyl-SNAC"), which is then condensed with malonyl-coenzyme A ("malonyl-CoA") to produce 3-keto-adipic-acyl carrier protein ("3-keto-adipic-ACP"). The engineered "reductive loop" processively reduces this intermediate with NADPH to adipic-acyl-carrier protein ("adipic-ACP"), which is hydrolytically released from the enzyme by the action of the thioesterase ("TE") domain. An aspect of the invention is the replacement of the dehydratase domain in the reductive loop.

Polyketide Synthases (PKS)

[0039] In some embodiments, the synthetic module comprises one or more of the following domains: a BorA2 KS domain, or functional variant thereof, a BorA2 AT domain, or functional variant thereof, a DH described in Example 1, or functional variant thereof, a heterologous KR domain, a heterologous ER domain, and a BorA2 ACP domain, or functional variant thereof.

[0040] In some embodiments, the heterologous KR domain is a KR domain of AurB, IdmO, NanA2, or SpnB, or a functional variant thereof. In some embodiments, the heterologous ER domain is an ER domain of AurB, IdmO, NanA2, or SpnB, or a functional variant thereof.

[0041] In some embodiments, the PKS further comprises a second module comprising a BorA1 AT domain, or a functional variant thereof, and a BorA1 ACP domain, or a functional variant thereof.

[0042] In some embodiments, the PKS further comprises one or more extender modules or domains, and a thiosterase (TE) domain, such as ery TE, or an R domain. In some embodiments, the PKS is modified as shown in Figure 10 wherein the PKS has the domains KR-ACP-ST-TE.

[0043] In some embodiments, the PKS further comprises one or more extender modules or domains between the synthetic module or hybrid module, and TE domain or R domain.

[0044] The amino acid sequence of the S3c variant is:

MAHEDKLRHLLKRVSAELDDTQRRVREMEESEREPIAIVGMSCRLPGGVNSPGEFWSLLE

AGTDAVSEFPRDRGWDVENLYDPDPDAPGRSYVREGGFLDGAGQFDAAFFGISPREALAM

DPQQRLLLECSWEAIERSRIDPKTLHGSRTGVFAGSNWQDYNTLLLNAEERSQSYLATGA

SGSVLSGRVSYTLGMEGPAITVNTACSSSLVAVHLAARSLRAGECDLALAGAVTVMSTPQ

LPVAFSRQRGLAPDGRSKAFAVSADGMGFGEGVGVLVLERLSVARRNGHRVLAVVRGSAV

NQDGASNGLTAPNGPSQQRVIRAALASAGLGPADVDVVEAHGTGTRLGDPIEAQALLATY

GRGRDAERPLWLGSVKSNIGHAQAAAGVAGVIKMVLAMEKGRLPRTLHVDEPSGEVDWDS

GAVRLLTEARDWPSEEGRLRRAGVSSFGISGTNAHVIIEEAPEEGEEPESDAGGVVPWVL

SARTEGALQAQAVQLSEFVGESSPVDVGWSLVSTRAAFEHRAVVVGRGRDELVRGLSEVA

QGRGVRGVASSASGGLAFVFAGQGSQRLGMGRGLYERFPVFAEAFDEVCGRVGPGVREVV

FGSDAGELDRTVWAQAGLFALEVALFRLLESWGVRPGCLIGHSVGELSAACVAGLWSLED

ACRVVAARARLMQALPAGGVMVAVRAEAGELAGFLGEDVVIASVNAPGQVVIAGPEGGVE

RVVAACGARSRRLAVSHAFHSPLVEPMLGEFRRVVESVAFGVPSLRVVSNVTGAWVDPEE

WGTPEYWVRQVREPVRFADGVATLLDAGVRTFVELGPAGALTSMVSHCADATATSVTAVP

TLRPDHDESRTVLSAAASLYVQGHPVDWAPLFPRARTVDLPTYPFQHQHYWMMNTGSAAE

PAELGLGDARHPLLGSVVTVAGDDKVVFAGRLALRTHPWLADHTVLDAVLLPATAFLELA

VRAGEEVSCPVVHDLTLHRPLVVPERGAVQVQMAVGAPEADGRREVRVYSRPDDDAEHEW

TLHAAGLLASAATAEPAVAAGAWPPPEAQAVDLDGFYAGLAEHGYHYGPLFQGVRAAWRL

GDDVLAEIVLPEAAGADAARYGMHPALLDAVLHAARLGAFRERSEEKYLPFAWEGVTLRT

RGATAVRARISRAGTDAIRLDVTDTADRPVLTAESLVLRSAAARRTGARRQAHQARLYRL

SWPTVQLPTSAQPPSCVLLGTSEVSADIQVYPDLRSLTAALDAGAEPPGVVIAPTPPGGG

RTADVRETTRHALDLVQGWLSDQRLNESRLLLVTQGAVAVEPGEPVTDLAQAALWGLLRS

TQTEHPDRFVLVDVPEPAQLLPALPGVLACGEPQLALRRGGAHAPRLAGLGSDDVLPVPD

GTGWRLEATRPGSLDGLALVDEPTATAPLGDGEVRIAMRAAGVNFRDALIALGMYPGVAS

LGSEGAGVVVETGPGVTGLAPGDRVMGMIPKAFGPLAVADHRMVTRIPAGWSFARAASVP

IVFLTAYYALVDLAGLRPGESLLVHSAAGGVGMAAIQLARHLGAEVYATASEDKWQAVEL

SREHLASSRTCDFEQQFLGATGGRGVDVVLNSLAGEFADASLRMLPRGGRFLELGKTDVR

DPVEVADAHPGVSYQAFDTVEAGPQRIGEMLHELVELFEGRVLEPLPVTAWDVRQAPEAL

RHLSQARHVGKLVLTMPPVWDAAGTVLVTGGTGALGAEVARHLVIERGVRNLVLVSRRGP

AASGAAELVAQLTAYGAEVSLQACDVADRETLAKVLASIPDEHPLTAVVHAAGVLDDGVS

ESLTVERLDQVLRPKVDGARNLLELIDPDVALVLFSSVSGVLGSGGQGNYAAANSFLDAL

AQQRQSRGLPTRSLAWGPWAEHGMASTLREAEQDRLARSGLLPISTEEGLSQFDAACGGA

HTVVAPVRFSRLSDGNAIKFSVLQGLVGPHRVNKAATADDAESLRKRLAALPEADRRRAV

LDLVEELVLGVLGHETRAAIGPDSSFHAIGFDSLTAVELRNLLTVRLGMKLPATLVYDHP

TLSSLADHLHEQLVIDGTPMTDTAADLLAELDALAARLAAVGLEPEARARIGRRLKDMQT

ACEPRSESSRDLKSASRTEVLDFLTNELGISR (SEQ ID NO:1)

[0045] The amino acid sequence of the S3c variant with ery TE is:

MAHEDKLRHLLKRVSAELDDTQRRVREMEESEREPIAIVGMSCRLPGGVNSPGEFWSLLE

AGTDAVSEFPRDRGWDVENLYDPDPDAPGRSYVREGGFLDGAGQFDAAFFGISPREALAM

DPQQRLLLECSWEAIERSRIDPKTLHGSRTGVFAGSNWQDYNTLLLNAEERSQSYLATGA

SGSVLSGRVSYTLGMEGPAITVNTACSSSLVAVHLAARSLRAGECDLALAGAVTVMSTPQ

LPVAFSRQRGLAPDGRSKAFAVSADGMGFGEGVGVLVLERLSVARRNGHRVLAVVRGSAV

NQDGASNGLTAPNGPSQQRVIRAALASAGLGPADVDVVEAHGTGTRLGDPIEAQALLATY

GRGRDAERPLWLGSVKSNIGHAQAAAGVAGVIKMVLAMEKGRLPRTLHVDEPSGEVDWDS

GAVRLLTEARDWPSEEGRLRRAGVSSFGISGTNAHVIIEEAPEEGEEPESDAGGVVPWVL

SARTEGALQAQAVQLSEFVGESSPVDVGWSLVSTRAAFEHRAVVVGRGRDELVRGLSEVA

QGRGVRGVASSASGGLAFVFAGQGSQRLGMGRGLYERFPVFAEAFDEVCGRVGPGVREVV

FGSDAGELDRTVWAQAGLFALEVALFRLLESWGVRPGCLIGHSVGELSAACVAGLWSLED

ACRVVAARARLMQALPAGGVMVAVRAEAGELAGFLGEDVVIASVNAPGQVVIAGPEGGVE

RVVAACGARSRRLAVSHAFHSPLVEPMLGEFRRVVESVAFGVPSLRVVSNVTGAWVDPEE

WGTPEYWVRQVREPVRFADGVATLLDAGVRTFVELGPAGALTSMVSHCADATATSVTAVP

TLRPDHDESRTVLSAAASLYVQGHPVDWAPLFPRARTVDLPTYPFQHQHYWMMNTGSAAE

PAELGLGDARHPLLGSVVTVAGDDKVVFAGRLALRTHPWLADHTVLDAVLLPATAFLELA

VRAGEEVSCPVVHDLTLHRPLVVPERGAVQVQMAVGAPEADGRREVRVYSRPDDDAEHEW

TLHAAGLLASAATAEPAVAAGAWPPPEAQAVDLDGFYAGLAEHGYHYGPLFQGVRAAWRL

GDDVLAEIVLPEAAGADAARYGMHPALLDAVLHAARLGAFRERSEEKYLPFAWEGVTLRT

RGATAVRARISRAGTDAIRLDVTDTADRPVLTAESLVLRSAAARRTGARRQAHQARLYRL

SWPTVQLPTSAQPPSCVLLGTSEVSADIQVYPDLRSLTAALDAGAEPPGVVIAPTPPGGG

RTADVRETTRHALDLVQGWLSDQRLNESRLLLVTQGAVAVEPGEPVTDLAQAALWGLLRS

TQTEHPDRFVLVDVPEPAQLLPALPGVLACGEPQLALRRGGAHAPRLAGLGSDDVLPVPD

GTGWRLEATRPGSLDGLALVDEPTATAPLGDGEVRIAMRAAGVNFRDALIALGMYPGVAS

LGSEGAGVVVETGPGVTGLAPGDRVMGMIPKAFGPLAVADHRMVTRIPAGWSFARAASVP

IVFLTAYYALVDLAGLRPGESLLVHSAAGGVGMAAIQLARHLGAEVYATASEDKWQAVEL

SREHLASSRTCDFEQQFLGATGGRGVDVVLNSLAGEFADASLRMLPRGGRFLELGKTDVR

DPVEVADAHPGVSYQAFDTVEAGPQRIGEMLHELVELFEGRVLEPLPVTAWDVRQAPEAL

RHLSQARHVGKLVLTMPPVWDAAGTVLVTGGTGALGAEVARHLVIERGVRNLVLVSRRGP

AASGAAELVAQLTAYGAEVSLQACDVADRETLAKVLASIPDEHPLTAVVHAAGVLDDGVS

ESLTVERLDQVLRPKVDGARNLLELIDPDVALVLFSSVSGVLGSGGQGNYAAANSFLDAL

AQQRQSRGLPTRSLAWGPWAEHGMASTLREAEQDRLARSGLLPISTEEGLSQFDAACGGA

HTVVAPVRFSRLSDGNAIKFSVLQGLVGPHRVNKAATADDAESLRKRLAALPEADRRRAV

LDLVEELVLGVLGHETRAAIGPDSSFHAIGFDSLTAVELRNLLTVRLGMKLPATLVYDHP

TLSSLADHLHEQLESGTPAREASSALRDGYRQAGVSGRVRSYLDLLAGLSDFREHFDGSD

GFSLDLVDMADGPGEVTVICCAGTAAISGPHEFTRLAGALRGIAPVRAVPQPGYEEGEPL

PSSMAAVAAVQADAVIRTQGDKPFVVAGHSAGALMAYALATELLDRGHPPRGVVLIDVYP

PGHQDAMNAWLEELTATLFDRETVRMDDTRLTALGAYDRLTGQWRPRETGLPTLLVSAGE

PMGPWPDDSWKPTWPFEHDTVAVPGDHFTMVQEHADAIARHIDAWLGGGNS* (SEQ ID NO:2)

[0046] The present invention provides for a polyketide synthase (PKS) capable of synthesizing a carboxylic acid or diacid. The PKS is not a naturally occurring PKS. In some embodiments, the carboxylic acid or diacid is not a compound synthesized by a naturally occurring PKS. In some embodiments, the PKS is a hybrid PKS comprising modules, domains, and/or portions thereof from two or more PKSs. Such carboxylic acids or diacids include the diketides and triketides, and polyketides of more than three ketide units, such as 4, 5, or 6 or more ketide units. The carboxylic acid or diacid can further include one or more functional groups. Such functional groups include, but are not limited to, ethyl, methyl and hydroxyl side chains, and internal olefins and ketones.

[0047] In some embodiments, the diacid is adipic acid (or hexanedioc acid), suberic acid (or octanedioc acid), or sebacic acid (or decanedioc acid). In some embodiments, the diacid is a symmetrical compound, such as a fully reduced symmetrical aliphatic compound.

[0048] Adipic acid is a six carbon chain fully reduced symmetrical aliphatic compound with no side chains, hence no chiral centers. Side chains (methyl, allyl, hydroxyl) of the carboxylic acid or diacid may be incorporated or formed, depending on the modules employed.

[0049] Complex polyketides comprise a large class of natural products that are synthesized in bacteria (mainly members actinomycete family; e.g. *Streptomyces*), fungi and plants.

Polyketides form the aglycone component of a large number of clinically important drugs, such as antibiotics (e.g. erythromycin, tylosin), antifungal agents (e.g. nystatin), anticancer agents (e.g. epothilone), immunosuppressives (e.g. rapamycin), etc. Though these compounds do not resemble each other either in their structure or their mode of action, they share a common basis for their biosynthesis, which is carried out by a group of enzymes designated polyketide synthases.

[0050] Polyketide synthases (PKS) employ short chain fatty acyl CoAs in Claisen condensation reactions to produce polyketides. Unlike fatty acid synthases which utilize acetyl CoA as the starter and malonyl CoA as the extender units, and use a single module iteratively to produce the nascent acyl chains, PKSs are composed of discrete modules, each catalyzing the chain growth of a single step. Modules can differ from each other in composition so that overall, a number of different starters (e.g. acetyl CoA, propionyl CoA) and extenders, some of which contain stereospecific methyl (or ethyl) side chains can be incorporated. In addition, PKS modules do not always reduce the 3-carbonyl formed from condensation but may leave it either unreduced (ketone), partially reduced (hydroxyl, 2,3-ene) or fully reduced (3-methylene). In some cases the terminal carboxyl group is usually removed by a decarboxylase domain present at the N-terminus of the corresponding loading domain of the PKS. Because of the correspondence between use of modules in the synthesis and the structure of the polyketide produced, it is possible to program the synthesis to produce a compound of desired structure by selection and genetic manipulation of polyketide synthases. Figure 9 shows a scheme for making novel polyamides or novel polyesters using diacids (using adipic acid as an example). Figure 11 shows the various modules and the precursor utilized by each module for incorporation into the corresponding nascent acyl (polyketide) chain to give rise to the range of compounds of interest. Table 4 provides a PKS source for each module. Each PKS source is well-known to one skilled in the art is readily available. In addition, for each module taught in Table 4, there may be other modules from other PKS that can be used.

[0051] Table 4. PKS sources of the various modules.

Module	PKS Source
A	Rifamycin PKS Module 2
B	Oligomycin PKS Module 1
C	Spiramycin PKS Module 1
D	Pikromycin PKS Module 2

E	Oligomycin PKS Module 3
F	Erythromycin PKS Module 3
G	Oligomycin PKS Module 5
H	Primaricin PKS Module 7
I	Tylosin PKS Module 1
J	Erythromycin PKS Module 1
K	Avermectin PKS Module 7
L	Rapamycin PKS Module 1
M	Erythromycin PKS Module 4
N	Pederin Module 2
O	Ascomycin Module 4
P	FK506 Module 4

[0052] All extender modules carry the β-acyl ACP synthase (commonly called the ketosynthase or KS) domain, which conducts the decarboxylative condensation step between the extender and the growing polyketide chain, and the acyl carrier protein (ACP) domain that carries the growing acyl chain and presents it to the cognate reductive domains for reduction of the β-carbonyl. Modules can differ from each other in composition so that a number of different starter and extender units, some of which contain stereospecific side chains (e.g. methyl, ethyl, propylene) can be incorporated. The acyltransferase (AT) domain of each module determines the extender unit (e.g. malonyl CoA, methylmalonyl CoA, etc.) incorporated. In addition, PKS modules do not always reduce the β-carbonyl formed from condensation but may leave it either unreduced (ketone), partially reduced (hydroxyl, 2,3-ene) or fully reduced (3-methylene). The ketoreductase (KR) domain reduces the ketone to the OH function (stereospecifically); the dehydratase (DH) domain removes water from the α and β carbons leaving an α,β *trans*-double bond; the enoylreductase (ER) domain reduces the double bond to a β-methylene center; the reductive state of the β-carbonyl, therefore, is determined by the presence of functional reductive domains in the corresponding module. Less commonly, modules are found to contain an additional C-methylation domain (yielding an additional α-methyl side chain, as in epothilone). The makeup of the PKS, therefore, determines the choice of starter and extender acyl units incorporated, the extent of reduction at each condensation step, and the total number of units added to the chain. The wide diversity of structures of polyketides seen in nature is attributed to the diversity in PKS compositions.

[0053] A partial list of sources of PKS sequences that can be used in making the PKSs of the present invention, for illustration and not limitation, includes Ambruticin (U.S. Pat. No.

7,332,576); Avermectin (U.S. Pat. No. 5,252,474; MacNeil et al., 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256; MacNeil et al., 1992, Gene 115: 119-25); Candicidin (FRO008) (Hu et al., 1994, Mol. Microbiol. 14: 163-72); Epothilone (U.S. Pat. No. 6,303,342); Erythromycin (WO 93/13663; U.S. Pat. No. 5,824,513; Donadio et al., 1991, Science 252:675-79; Cortes et al., 1990, Nature 348:176-8); FK506 (Motamedi et al., 1998, Eur. J. Biochem. 256:528-34; Motamedi et al., 1997, Eur. J. Biochem. 244:74-80); FK520 or ascomycin (U.S. Pat. No. 6,503,737; see also Nielsen et al., 1991, Biochem. 30:5789-96); Jerangolid (U.S. Pat. No. 7,285,405); Leptomycin (U.S. Pat. No. 7,288,396); Lovastatin (U.S. Pat. No. 5,744,350); Nemadectin (MacNeil et al., 1993, supra); Niddamycin (Kakavas et al., 1997, J. Bacteriol. 179:7515-22); Oleandomycin (Swan et al., 1994, Mol. Gen. Genet. 242:358-62; U.S. Pat. No. 6,388,099; Olano et al., 1998, Mol. Gen. Genet. 259:299-308); Pederin (PCT publication no. WO 2003/044186); Pikromycin (Xue et al., 2000, Gene 245:203-211); Pimaricin (PCT publication no. WO 2000/077222); Platenolide (EP Pat. App. 791,656); Rapamycin (Schwecke et al., 1995, Proc. Natl. Acad. Sci. USA 92:7839-43); Aparicio et al., 1996, Gene 169:9-16); Rifamycin (August et al., 1998, Chemistry & Biology, 5: 69-79); Soraphen (U.S. Pat. No. 5,716,849; Schupp et al., 1995, J. Bacteriology 177: 3673-79); Spiramycin (U.S. Pat. No. 5,098,837); Tylosin (EP 0 791,655; Kuhstoss et al., 1996, Gene 183:231-36; U.S. Pat. No. 5,876,991). Additional suitable PKS coding sequences are readily available to one skilled in the art, or remain to be discovered and characterized, but will be available to those of skill (e.g., by reference to GenBank). Each of the references cited is hereby specifically and individually incorporated by reference.

[0054] Of the more than thirty PKSs examined, the correspondence between use of modules in the biosynthesis and the structure of the polyketide produced is fully understood both at the level of the protein sequence of the PKS and the DNA sequence of the corresponding genes. The programming of modules into polyketide structure can be identified by sequence determination. It is possible to clone (or synthesize) DNA sequences corresponding to desired modules and transfer them as fully functioning units to heterologous, otherwise non-polyketide producing hosts such as *E. coli* (B. A. Pfeifer, et al., *Science* 291, 1790 (2001)) and *Streptomyces* (C. M. Kao, et al., *Science* 265, 509 (1994)). Additional genes employed for polyketide biosynthesis have also been identified. Genes that determine phosphopantetheine:protein transferase (PPTase) that transfer the 4-phosphopantetheine co-factor of the ACP domains, commonly present in polyketide producing hosts, have been

cloned in *E. coli* and other hosts (K. J. Weissman, et al., *Chembiochem* **5**, 116 (2004)). It is also possible to re-program polyketide biosynthesis to produce a compound of desired structure by either genetic manipulation of a single PKS or by construction of a hybrid PKS composed of modules from two or more sources (K. J. Weissman, et al., *Chembiochem* 5, 116 (2004)).

[0055] Recombinant methods for manipulating modular PKS genes to make the PKSs of the present invention are described in U.S. Pat. Nos. 5,672,491; 5,843,718; 5,830,750; 5,712,146; and 6,303,342; and in PCT publication nos. WO 98/49315 and WO 97/02358; hereby incorporated by reference. A number of genetic engineering strategies have been used with various PKSs to demonstrate that the structures of polyketides can be manipulated to produce novel polyketides (see the patent publications referenced supra and Hutchinson, 1998, Curr. Opin. Microbiol. 1:319-329, and Baltz, 1998, Trends Microbiol. 6:76-83; hereby incorporated by reference). In some embodiments, the components of the hybrid PKS are arranged onto polypeptides having interpolypeptide linkers that direct the assembly of the polypeptides into the functional PKS protein, such that it is not required that the PKS have the same arrangement of modules in the polypeptides as observed in natural PKSs. Suitable interpolypeptide linkers to join polypeptides and intrapolypeptide linkers to join modules within a polypeptide are described in PCT publication no. WO 00/47724, hereby incorporated by reference.

[0056] The vast number of polyketide pathways that have been elucidated provide a host of different options to produce these diacids as well as the large number of derivatives. While the products can be vastly different in size and functionality, all employ virtually the same strategy for biosynthesis. The exact interfaces between non-cognate enzyme partners will be determined on a case-by-case basis. ACP-linker-KS and ACP-linker-TE regions from the proteins of interest will be aligned to examine the least disruptive fusion point for the hybrid synthase. Genetic constructions will employ sequence and ligation independent cloning (SLIC) so as to eliminate the incorporation of genetic "scarring".

Nucleic Acids encoding the PKS

[0057] The present invention provides for a recombinant nucleic acid that encodes a polyketide synthase (PKS) of the present invention. The recombinant nucleic acid can be a double-stranded or single-stranded DNA, or RNA. The recombinant nucleic acid can encode

an open reading frame (ORF) of the PKS of the present invention. The recombinant nucleic acid can also comprise promoter sequences for transcribing the ORF in a suitable host cell. The recombinant nucleic acid can also comprise sequences sufficient for having the recombinant nucleic acid stably replicate in a host cell. The recombinant nucleic acid can be replicon capable of stable maintenance in a host cell. In some embodiments, the replicon is stably integrated into a chromosome of the host cell. In some embodiments, the replicon is a plasmid. The present invention also provides for a vector or expression vector comprising a recombinant nucleic acid of the present invention. The present invention provides for a host cell comprising any of the recombinant nucleic acid and/or PKS of the present invention. In some embodiments, the host cell, when cultured under a suitable condition, is capable of producing the carboxylic acid or diacid.

[0058] It will be apparent to one of skill in the art that a variety of recombinant vectors can be utilized in the practice of aspects of the invention. As used herein, "vector" refers to polynucleotide elements that are used to introduce recombinant nucleic acid into cells for either expression or replication. Selection and use of such vehicles is routine in the art. An "expression vector" includes vectors capable of expressing DNAs that are operatively linked with regulatory sequences, such as promoter regions. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those that integrate into the host cell genome.

[0059] The vectors may be chosen to contain control sequences operably linked to the resulting coding sequences in a manner that expression of the coding sequences may be effected in an appropriate host. Suitable control sequences include those that function in eukaryotic and prokaryotic host cells. If the cloning vectors employed to obtain PKS genes encoding derived PKS lack control sequences for expression operably linked to the encoding nucleotide sequences, the nucleotide sequences are inserted into appropriate expression vectors. This can be done individually, or using a pool of isolated encoding nucleotide sequences, which can be inserted into host vectors, the resulting vectors transformed or transfected into host cells, and the resulting cells plated out into individual colonies. Suitable control sequences for single cell cultures of various types of organisms are well known in the

art. Control systems for expression in suitable host cells, such as yeast and prokaryotic host cells, are widely available and are routinely used. Control elements include promoters, optionally containing operator sequences, and other elements depending on the nature of the host, such as ribosome binding sites. Particularly useful promoters for prokaryotic hosts include those from PKS gene clusters that result in the production of polyketides as secondary metabolites, including those from Type I or aromatic (Type II) PKS gene clusters. Examples are act promoters, tcm promoters, spiramycin promoters, and the like. However, other bacterial promoters, such as those derived from sugar metabolizing enzymes, such as galactose, lactose (lac) and maltose, are also useful. Additional examples include promoters derived from biosynthetic enzymes such as for tryptophan (trp), the β-lactamase (bla), bacteriophage lambda PL, and T5. In addition, synthetic promoters, such as the tac promoter (U.S. Pat. No. 4,551,433; hereby incorporated by reference), can be used.

[0060] As noted, particularly useful control sequences are those which themselves, or with suitable regulatory systems, activate expression during transition from growth to stationary phase in the vegetative mycelium. Illustrative control sequences, vectors, and host cells of these types include the modified *S. coelicolor* CH999 and vectors described in PCT publication no. WO 96/40968 and similar strains of *S. lividans*. See U.S. Pat. Nos. 5,672,491; 5,830,750; 5,843,718; and 6,177,262, each of which is hereby incorporated by reference. Other regulatory sequences may also be desirable which allow for regulation of expression of the PKS sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer sequences.

[0061] Selectable markers can also be included in the recombinant expression vectors. A variety of markers are known which are useful in selecting for transformed cell lines and generally comprise a gene whose expression confers a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium. Such markers include, for example, genes that confer antibiotic resistance or sensitivity to the plasmid.

[0062] The various PKS nucleotide sequences, or a mixture of such sequences, can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements or under the control of a single promoter. The PKS subunits or components can include

flanking restriction sites to allow for the easy deletion and insertion of other PKS subunits. The design of such restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such as site-directed mutagenesis and PCR. Methods for introducing the recombinant vectors of the present invention into suitable hosts are known to those of skill in the art and typically include the use of $CaCl_2$ or other agents, such as divalent cations, lipofection, DMSO, protoplast transformation, conjugation, and electroporation.

Host Cells comprising the PKS

[0063] The present invention provides for a host cell comprising any of the recombinant nucleic acid and/or PKS of the present invention. In some embodiments, the host cell, when cultured, is capable of producing a carboxylic acid or diacid. The host cell can be a eukaryotic or a prokaryotic cell. Suitable eukaryotic cells include yeast cells, such as from the genus *Saccharomyces* or *Schizosaccharomyces*. A suitable species from the genus *Saccharomyces* is *Saccharomyces cerevisiae*. A suitable species from the genus *Schizosaccharomyces* is *Schizosaccharomyces pombe*. Suitable prokaryotic cells include *Escherichia coli* or *Streptomyces* species.

[0064] The PKS can be in a host cell, or isolated or purified. The PKS can synthesize the carboxylic acid or diacid *in vivo* (in a host cell) or *in vitro* (in a cell extract or where all necessary chemical components or starting materials are provided). The present invention provides methods of producing the carboxylic acid or diacid using any of these *in vivo* or *in vitro* means.

Methods of Using the PKS

[0065] The present invention provides a method of producing a carboxylic acid or diacid, comprising: providing a host cell of the present invention, and culturing said host cell in a suitable culture medium such that the carboxylic acid or diacid is produced. The method can further comprise isolating said carboxylic acid or diacid from the host cell and the culture medium. The method can further comprise reacting the diacid with a diamine to produce a nylon. A suitable diamine is an alkane diamine, such as hexane-1,6-diamine. Alternatively, the method can further comprise reacting the diacid with a dialcohol to produce a polyester. A suitable dialcohol is an alkane diol, such as ethylene glycol, propane diol, or butanediol. A variety of methods for heterologous expression of PKS genes and host cells suitable for

expression of these genes and production of polyketides are described, for example, in U.S. Pat. Nos. 5,843,718; 5,830,750 and 6,262,340; WO 01/31035, WO 01/27306, and WO 02/068613; and U.S. Patent Application Pub. Nos. 2002/0192767 and 2002/0045220; hereby incorporated by reference.

[0066] The present invention provides for a composition comprising a carboxylic acid or diacid isolated from a host cell from which the carboxylic acid or diacid is produced, and trace residues and/or contaminants of the host cell. Such trace residues and/or contaminants include cellular material produced by the lysis of the host cell.

[0067] The diacids, such as adipic acid, provide for the production of "green" nylon, such as that used in Mohawk carpet fibers. Besides nylon production, the ability to manipulate the side chains of the diacids provides for the production of novel polymer precursors that would lead to polymers with a variety of properties. These products may also serve as adhesive, lubricants or precursors for pharmaceuticals or other more complicated compounds.

[0068] The present invention has one or more of the following advantages: (1) it reduces the dependence on oil for producing certain chemicals, and (2) it serves as a means of capture and sequestration of carbon from the atmosphere.

[0069] References cited:

Alini, S., Basile, F., Blasioli, S., Rinaldi, C., and Vaccari, A. (2007). Development of new catalysts for N2O-decomposition from adipic acid plant. Appl. Catal. B Environ. *70*, 323–329.

Aparicio, J.F., Caffrey, P., Marsden, A.F., Staunton, J., and Leadlay, P.F. (1994). Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. J. Biol. Chem. *269*, 8524–8528.

Clomburg, J.M., Blankschien, M.D., Vick, J.E., Chou, A., Kim, S., and Gonzalez, R. (2015). Integrated engineering of β-oxidation reversal and ω-oxidation pathways for the synthesis of medium chain ω-functionalized carboxylic acids. Metab. Eng. *28*, 202–212.

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. Science *252*, 675–679.

Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M., and Katz, L. (1993). An erythromycin analog produced by reprogramming of polyketide synthesis. Proc. Natl. Acad. Sci. *90*, 7119–7123.

Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., et al. (2014). Structure of a modular polyketide synthase. Nature *510*, 512–517.

Epstein, B.N. (1979). Tough thermoplastic nylon compositions (U.S. Patent No. 4,174,358).

Gaisser, S., Kellenberger, L., Kaja, A.L., Weston, A.J., Lill, R.E., Wirtz, G., Kendrew, S.G., Low, L., Sheridan, R.M., Wilkinson, B., et al. (2003). Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of Streptomyces avermitilis ATCC31272. Org. Biomol. Chem. *1*, 2840.

George, K.W., Chen, A., Jain, A., Batth, T.S., Baidoo, E.E.K., Wang, G., Adams, P.D., Petzold, C.J., Keasling, J.D., and Lee, T.S. (2014). Correlation analysis of targeted proteins and metabolites to assess and engineer microbial isopentenol production: Targeted Proteomics-Based Correlation Analysis. Biotechnol. Bioeng. *111*, 1648–1658.

Hagen, A., Poust, S., de Rond, T., Yuzawa, S., Katz, L., Adams, P.D., Petzold, C.J., and Keasling, J.D. (2014). *In Vitro* Analysis of Carboxyacyl Substrate Tolerance in the Loading and First Extension Modules of Borrelidin Polyketide Synthase. Biochemistry (Mosc.) *53*, 5975–5977.

Hong, H., Appleyard, A.N., Siskos, A.P., Garcia-Bernardo, J., Staunton, J., and Leadlay, P.F. (2005). Chain initiation on type I modular polyketide synthases revealed by limited proteolysis and ion-trap mass spectrometry: Limited proteolysis and MS of modular PKSs. FEBS J. *272*, 2373–2387.

Kellenberger, L., Galloway, I.S., Sauter, G., Böhm, G., Hanefeld, U., Cortés, J., Staunton, J., and Leadlay, P.F. (2008). A Polylinker Approach to Reductive Loop Swaps in Modular Polyketide Synthases. ChemBioChem *9*, 2740–2749.

Khosla, C. (2009). Structures and Mechanisms of Polyketide Synthases. J. Org. Chem. *74*, 6416–6420.

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Betlach, M., and Ashley, G. (1999). Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc. Natl. Acad. Sci. *96*, 1846–1851.

Meluzzi, D., Zheng, W.H., Hensler, M., Nizet, V., and Dorrestein, P.C. (2008). Top-down mass spectrometry on low-resolution instruments: characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. Bioorg. Med. Chem. Lett. *18*, 3107–3111.

Tang, Y., Kim, C.-Y., Mathews, I.I., Cane, D.E., and Khosla, C. (2006). The 2.7-A crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. *103*, 11124–11129.

Vergnolle, O., Hahn, F., Baerga-Ortiz, A., Leadlay, P.F., and Andexer, J.N. (2011). Stereoselectivity of Isolated Dehydratase Domains of the Borrelidin Polyketide Synthase: Implications for cis Double Bond Formation. ChemBioChem *12*, 1011–1014.

Williams, G.J. (2013). Engineering polyketide synthases and nonribosomal peptide synthetases. Curr. Opin. Struct. Biol. *23*, 603–612.

Yoon, Y.J., Beck, B.J., Kim, B.S., Kang, H.-Y., Reynolds, K.A., and Sherman, D.H. (2002). Generation of Multiple Bioactive Macrolides by Hybrid Modular Polyketide Synthases in Streptomyces venezuelae. Chem. Biol. *9*, 203–214.

Yu, J.-L., Xia, X.-X., Zhong, J.-J., and Qian, Z.-G. (2014). Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*: Adipic Acid Production From a Synthetic Pathway. Biotechnol. Bioeng. *111*, 2580–2586.

Zheng, J., Piasecki, S.K., and Keatinge-Clay, A.T. (2013). Structural Studies of an A2-Type Modular Polyketide Synthase Ketoreductase Reveal Features Controlling α-Substituent Stereochemistry. ACS Chem. Biol. *8*, 1964–1971.

[0070] The invention having been described, the following examples are offered to illustrate the subject invention by way of illustration, not by way of limitation.

EXAMPLE 1

Engineering a polyketide synthase for production of adipic acid

[0071] Polyketides have enormous structural diversity, yet polyketide synthases (PKSs) have thus far been engineered to produce only derivatives of drugs or drug candidates. Thousands of other molecules, including commodity and specialty chemicals could be synthesized using PKSs if composing hybrid PKSs from well-characterized parts derived from natural PKSs was more efficient. Here, using modern proteomics techniques, we demonstrate construction of a chimeric PKS extension module capable of producing one of the most widely used commodity chemicals, adipic acid. To accomplish this, we introduced heterologous reductive domains from various PKS clusters into the borrelidin PKS' first extension module, which we previously showed produces a 3-hydroxy-adipoyl intermediate when co-incubated with the loading module and a succinyl-CoA starter unit. Acyl-ACP intermediate analysis revealed an unexpected bottleneck at the dehydration step which was overcome by introduction of a carboxyacyl-processing dehydratase domain from the second module of the borrelidin. Adipic acid was released from the synthase after appending the erythromycin thioesterase domain to the hybrid PKS.

[0072] The results demonstrate the following:

1. Demonstration of commodity chemical production by an engineered polyketide synthase.

2. Acyl-ACP intermediate analysis is used to identify unexpected catalytic bottlenecks.

3. Identification of previously unknown dehydratase domain selectivity.

4. Construction of a broad specificity, fully-reducing PKS module.

Using proteomics-based covalent intermediate analysis, Hagen et al engineered a chimeric polyketide synthase capable of producing adipic acid. In the process they revealed unexpected selectivity in the β-carbon reduction cycle.

INTRODUCTION

[0073] Here we demonstrate engineering a PKS to produce the commodity chemical adipic acid. Current production of adipic acid results in approximately 10% of anthropogenic emissions of N_2O–a potent greenhouse gas (Alini et al., 2007)); therefore, a biological route

to adipic acid could be an important alternative.

[0074] Within the context of type I PKS-based biosynthesis, we proposed that adipic acid synthesis would most conveniently start from the four-carbon succinyl-CoA, undergo one round of extension with full reduction using a malonyl-CoA extender unit to produce the six carbon adipoyl-ACP intermediate. Adipic acid would then be released from adipoyl-ACP by the action of a thioesterase. Due to its important role in the TCA cycle, succinate/succinyl-CoA is readily available in organisms capable of aerobic respiration (e.g. common production hosts like *E. coli*, *Saccharomyces cerevisiae* and Actinobacteria), as is malonyl-CoA, which is used in fatty acid biosynthesis. Therefore production of adipic acid using a PKS and succinyl-CoA starter would be relatively host and feedstock agnostic, as minimal metabolic engineering would be necessary to ensure adequate precursor supply. Another advantage of using a PKS system is the extensibility inherent in its modular nature. For example, longer diacids could be generated by use of additional (or iterative) modules, and novel adipic acid analogs could be created with α-substitutions (e.g. methyl-, fluoro-, or allyl groups) that may yield polymers with useful attributes such as cross-linkable chemical handles (Epstein, 1979).

[0075] Previous work in our lab demonstrated that the loading and first extension modules of the borrelidin PKS (hereafter referred to as "BorLM" and "BorMod1", respectively) are capable of producing a 3-hydroxy-adipoyl-ACP intermediate in vitro using succinyl-CoA as a starter substrate and the natural extender substrate, malonyl-CoA (Hagen et al., 2014) (Figure 1B). To proceed from the 3-hydroxyadipoyl-ACP intermediate to adipic acid, additional β-carbonyl processing and hydrolytic chain release is required. We therefore sought to introduce additional reducing domains into BorMod1, and upon verification of complete reduction, append a thioesterase domain capable of releasing the linear product. "Reductive loop" swaps were among the earliest and most successful demonstrations of modularity in type I PKS systems (Donadio et al., 1993; Gaisser et al., 2003; McDaniel et al., 1999; Yoon et al., 2002). These findings along with limited proteolysis experiments and recent structural studies indicate that reductive loops function as integral units (Aparicio et al., 1994; Dutta et al., 2014; Hong et al., 2005). Despite these examples, no prescriptive rules have been developed to guide successful reductive loop swaps and the most extensive, combinatorial study of reductive loop swaps to date ultimately concluded, "no single donor [module] and no single pair of splice sites were found to be reliably optimal to effect a given alteration" (Kellenberger et al., 2008).

[0076] We selected donor reductive loops from the aureothin, indanomycin, nanchangmycin and spinosyn PKS clusters: AurB, IdmO, NanA2, SpnB, respectively, based on three criteria: (1) the loop contained the full complement of reducing domains (ketoreductase, dehydratase and enoyl reductase, hereafter referred to as "KR," "DH," and "ER," respectively), (2) the loop originated from a "standalone" module in which the open reading frame or "subunit" encodes just a single module, and (3) the module harboring the reductive loop naturally incorporates a malonate extender unit. Previous work has suggested a reduction in catalytic efficiency and relaxed stereoselectivity when KR domains are presented with an α–carbon differentially substituted than the KR's normal substrate (McDaniel et al., 1999; Zheng et al., 2013). These loops were introduced combinatorially into BorMod1 using two alternative N-terminal and a single C-terminal splice sites to generate eight chimeras to be tested for adipoyl-ACP production *in vitro* (Figure 1A).

[0077] In the absence of a thioesterase, intermediates covalently attached to the PKS could be monitored using the "PPant ejection assay" (Meluzzi et al., 2008). This system allows us to identify bottlenecks in the biosynthesis. As PKSs are complex enzymes, determining the point of failure for engineered PKSs is challenging. Most PKS engineering efforts thus far have relied on the presence of the desired final product to determine success, however this approach does not provide information as to where the enzymatic assembly line has stalled if the product is not observed. As part of our efforts to produce the commodity chemical adipic acid, we demonstrate the utility of acyl-carrier protein (ACP) intermediate analysis (via the PPant ejection assay) to "debug" PKSs. Upon satisfactory production of adipoyl-ACP after several rounds of chimeragenesis, a thioesterase was introduced to produce adipic acid.

RESULTS

Beta-carbonyl processing stalls at the dehydration step; is alleviated by provision of BorDH2 *in trans*

[0078] The initial engineered reductive loop BorMod1 library was incubated with the synthetic starter substrate succinyl-SNAC, along with malonyl-CoA and NADPH. Six out of eight constructs were catalytically active, but the primary acyl-ACP species, after introduction of the full reducing loop, remained the partially reduced 3-hydroxy-adipoyl-ACP intermediate; the 3-keto, 2,3-ene and fully reduced (adipoyl-ACP) products were not detected (see Figure 2A), indicating that reductive processing was stalled at the dehydratase step.

[0079] We hypothesized the dehydratase domains from the reductive loop variants were not competent to dehydrate 3-hydroxyadipoyl-ACP and therefore sought to test the activity of a different dehydratase domain which processes a substrate carrying a terminal carboxyl group in its natural context. Because of its proximity to a terminal carboxyl group (see Figure 1B), the first DH domain in the borrelidin cluster, BorDH2, was chosen and provided to the reductive loop library *in trans* in stoichiometric excess as previous work showed a low rate of DH activity *in vitro* (Vergnolle et al., 2011). As shown in Figure 2B, provision of BorDH2 resulted in the production of higher levels of the adipoyl-ACP intermediate when compared to the constructs without BorDH2. A particularly interesting case is the comparison between S2 and S2t, where provision of the dehydratase *in trans* (S2t), increased adipoyl-ACP production from nearly undetectable levels to the highest level amongst all variants. No significant accumulation of the 2,3-ene-ACP intermediate was observed when BorDH2 was provided (data not shown). This, along with the observed production of adipoyl-ACP in all loop variants, indicates the 2,3-ene intermediate, the immediate product of the dehydration, was readily reduced by the enoyl reductase domains present *in cis*.

[0080] Figure 6 shows the N- and C- terminal junctions for initial constructs. Arrows indicates crossover point. Figure 7 shows the junctions for DH swap constructs. Arrows indicates crossover point. Figure 8 shows the N-terminal junctions including junction 3.

BorDH2 *in cis* further increases the proportion of adipoyl-ACP

[0081] Having demonstrated that BorDH2 provided *in trans* is capable of promoting adipoyl-ACP formation, we next asked whether this was a property unique to this particular dehydratase domain or simply because BorDH2 was provided in stoichiometric excess. To determine this, BorDH2 was swapped into a subset of the most active reductive loop library members in order to replace the native DH domain. After purification, these DH swapped variants were compared to previous constructs as before *via* intermediate analysis after extension of succinyl-SNAC.

[0082] As shown in Figure 2C, DH swapped variants clearly promoted the formation of adipoyl-ACP (e.g. compare A2 (Figure 2A) to A2c (Figure 2C)) at levels comparable to where the DH was provided *in trans* at 50-fold stoichiometric excess (e.g. compare A2t (Figure 2B) to A2c (Figure 2C)). These data demonstrate that it is the unique identity of the BorDH2 domain which allows β-carbonyl processing and which is not required at

stoichiometric excess for maximum activity.

Refined chimeric junction further promotes proportion of adipoyl-ACP

[0083] Despite junction 2 PKS variants generally showing higher production of adipoyl-ACP than junction 1 variants (especially when BorDH2 was included *in cis*), further sequence and structural analysis indicated that junction 2 constructs may be truncated by approximately 15 residues (depending on how domain boundaries are annotated) at the N-terminus of the dehydratase domain (see supplemental information). These residues are distal to the active site and ACP docking interface and are clearly not essential, however their influence on the overall tertiary structure and kinetics of PKS enzymes was unclear. Therefore, a new N-terminal junction was selected intermediary to junctions 1 and 2 (junction 3). Variants were created for a subset of the reductive loop library which included the best performing AurB and SpnB loop sources both with and without the BorDH2 swap. This location immediately follows the post-AT linker region which is believed to be important for proper KS-AT domain orientation (Tang et al., 2006) and also restores the missing segment in the DH domain N-terminal truncations.

[0084] As shown in Figure 2D, junction 3 was found to be superior to junction 1 and junction 2 as gauged by total production of the adipoyl-ACP intermediate. Strikingly, the combination of the new junction with the BorDH2 swap displayed a synergistic effect as evidenced by the nearly complete intermediate conversion to adipoyl-ACP in the case of A3c and S3c constructs.

BorDH2 is necessary solely for carboxy-acyl processing

[0085] The aforementioned data suggest that dehydration of a carboxyacyl-ACP intermediate is a trait unique to BorDH2 and not shared by the four DH domains in the un-engineered reductive loops. The possibility, however, remains that the ACP in BorMod1 does not interact well with non-native DH domains, precluding the presentation of the 3-hydroxy-adipoyl-ACP intermediate, whereas the ACP more readily associates with a DH domain from the same PKS cluster. To interrogate this possibility, we incubated the isolated ACP monodomain from BorLM (which naturally presents loaded substrates to BorMod1, Fig. 1B) acylated with a variety of carboxy and descarboxy-CoA substrates with BorDH2-swapped and unswapped version of the S3 variant to determine which substrates could be processed. The CoAs employed were succinyl-CoA and its descarboxy analog propionyl-CoA as well as

the natural substrate 1,2-cyclopentanedicarboxyl-CoA (CPDA-CoA) and its respective descarboxy analog cyclopentanemonocarboxyl-CoA (CPMA-CoA). As shown in Figures 3 and 5, the descarboxy substrates propionyl- and CPMA-ACP were extended and fully reduced to their respective products by both S3 and S3c protein variants. In contrast, only the BorDH2 swapped variant converted a significant fraction of the 3-hydroxy intermediates to the fully reduced species when carboxylated substrates were provided. These results demonstrate unambiguously that the un-engineered reductive loop of SpnB is competent to perform full β-carbonyl processing of the more typical non-carboxylated intermediates, however BorDH2 is required for full β-carbonyl processing when the substrate contains a distal carboxy group.

Appending a thioesterase domain allows for production of free adipic acid

[0086] Having demonstrated the construction of a highly engineered extension module capable of producing adipoyl-ACP, we next sought to produce free adipic acid by the addition of a thioesterase (TE). The well-characterized TE domain from the erythromycin cluster was therefore appended to the best performing S3c variant in place of the C-terminal docking domain to create S3c-TE. In order to compare the activity and product profile of the engineered extension module with that of the wild type module, the TE was also appended to wild type BorMod1 to create BorMod1-TE. The proteins were purified and extension assays performed as before and titers were measured via LC-MS/MS by comparison to authentic standards (see materials and methods for synthesis of 3-hydroxy-adipic acid). Figure 4 shows that as expected, the BorMod1-TE construct produced exclusively 3-hydroxy-adipic acid whereas S3c-TE produced a mixture of the partially and fully reduced adipic acid products. While titers are modest, it is noteable that the titers for the wildtype and engineered extension modules are similar. This would suggest that despite the introduction of five chimeric junctions and utilization of domains from three different PKS clusters, the overall kinetics of the engineered extension module are comparable to wildtype.

DISCUSSION

[0087] In this study, using proteomics-based intermediate analysis to inform design iterations, we have demonstrated, for the first time, production of a commodity chemical by an engineered polyketide synthase. This was facilitated by prior identification of an extension module, BorMod1, which naturally accepts carboxyacyl substrates and extends with malonyl-

CoA. Metabolomic analysis of intermediates in solution has been utilized for bottleneck determination and subsequent improvement of engineered pathways (George et al., 2014). Here, by analysis of the covalent intermediates on the PKS assembly line, we have demonstrated the utility of this methodology to pinpoint and alleviate unexpected catalytic bottlenecks.

[0088] Initial activity tests indicated that replacing the reductive loop from BorMod1 with a library of reductive loops from fully reducing modules does not compromise the catalytic competence of the module for the extension reaction. This lends further support to the idea that the reductive loop functions as an "integral unit" apart from the core catalytic activity of the acyltransferase and ketosynthase domains in the module, and that the chimeric junctions used in this study did not perturb the module's tertiary structure such that condensation is precluded.

[0089] Interestingly, intermediate analysis showed that dehydration activity on carboxylated 3-hydroxy-ACP intermediates was poor, whereas β-carbonyl processing proceeded uninterrupted using descarboxy substrate analogs, revealing a previously unknown biochemical incompatibility between carboxylated substrates and typical dehydratase domains. In contrast, BorDH2 which in its native context processes a carboxylated substrate, appears substrate agnostic, though more kinetic data would be required to determine whether it prefers one species over the other. It is interesting to note that BorDH2 normally processes a cyclic intermediate with a sterically constrained carboxy group at the 8 position (see Figure 1B), rather than a linear 6-carboxy intermediate. Future bioinformatic and structural studies could reveal structural determinants of diacid tolerance and could enable engineering of diacid tolerance into typical reductive loops using precise amino acid substitutions of the dehydratase domain rather than chimeric domain swaps.

[0090] Addition of a thioesterase to the S3c PKS variant enabled production of free acids. Attenuating the TE activity or tuning its specificity towards the fully reduced product through mutagenesis could possibly shift the product profile further towards adipic acid. Alternatively, increasing the rate of β-carbonyl processing, perhaps through further refined chimeric boundary sampling (including at the C-terminus of the reductive loop) or selection of alternative reductive loops, would increase the proportion and possibly titer of adipic acid. Encouragingly, the overall activities of the wildtype BorMod1-TE and S3c-TE are within error indicating that despite extensive reductive loop engineering, the kinetics of the

engineered PKS module was not significantly compromised. Further engineering of hosts for improved expression of heterologous PKSs will be required to improve the productivity of these enzymes.

[0091] In recent years a number of biological routes to adipic acid have been developed, typically dependent on reversal of beta-oxidation of dicarboxylic acids (Yu et al., 2014) or omega-oxidation of fatty acids (Clomburg et al., 2015). However, as demonstrated here, the ability to engineer diacid tolerance in a PKS system sets the stage for production of other valuable commodity chemicals (e.g. the eight carbon suberic acid) as well as branched diacids that are not readily accessible through conventional synthetic chemistry or the above biosynthetic routes.

SIGNIFICANCE

[0092] Polyketide synthases have tremendous synthetic potential, yet have historically been used only for the production of drugs and their derivatives. We show PKSs can also be used for the production of commodity chemicals by engineering a PKS that produces adipic acid. In so doing, we have highlighted the utility of LC-MS/MS based acyl-intermediate analysis techniques which allowed for identification and alleviation of the dehydratase catalytic bottleneck and revealed an unexpected biochemical incompatibility between typical dehydratase domains and carboxylated intermediates. As type I PKSs are inherently modular, this work sets the stage for production of other valuable commodity chemicals such as branched diacids which are not readily accessible through conventional synthetic chemistry.

EXPERIMENTAL PROCEDURES

[0093] For details of plasmid construction, protein purifications, chemical syntheses and LC-MS/MS methods, refer to supplemental information.

Intermediate analysis of PKS variants

[0094] For extensions with succinyl-SNAC, a master mix (final concentrations: 1 mM succinyl-SNAC, 0.2 mM malonyl-CoA, 1 mM NADPH, 2.5 mM TCEP in 100 mM phosphate buffer pH 6.8) was aliquoted to separate tubes, to which 5 uM final concentration of each respective PKS variant was added. For relevant experiments, 50 μM BorDH2 was provided *in trans*. For extensions using acyl-ACP reagents, ACPs were expressed in apo form

and charged using Sfp and various acyl-CoAs as described in (Hagen et al., 2014). These were added to enzyme mixes containing either the S3 or S3c PKS variants and other reaction components at the same concentration as described above. Reactions were incubated at room temperature overnight (~16 hr). Samples were digested with 1:20 w/w porcine trypsin (Sigma-Aldrich) for 4-6 hours at 37C prior to LC-MS/MS analysis.

Product analysis of thioesterase-harboring constructs

[0095] 50 µl reactions were set up as described in intermediate analysis except the final concentration of malonyl-CoA was 0.5 mM. After incubation, samples were diluted with one volume of LC-MS grade water and filtered through 3K molecular weight cut off spin filters (Amicon) which were washed prior to use by filtration of 500 µl of LC-MS grade water. Samples were acidified by the addition of 1% formic acid prior to LC-MS/MS analysis. A dilution series of (3-hydroxy) adipic acid authentic standards was created and processed identically in parallel with samples to generate a concentration standard curve for quantification.

EXPERIMENTAL PROCEDURES

[0096] Plasmid construction. Reductive loops were codon-optimized for *E. coli* and introduced into pARH100 (Hagen et al., 2014) via scarless Gibson assembly (see below for junction boundaries). The j5 algorithm and Device Editor graphical user interface were used to design oligonucleotides and DNA assembly strategies (Hillson et al., 2012).

[0097] Purification of PKS constructs. Plasmids were introduced into *E. coli* strain BAP1 (Pfeifer, 2001) and cultures (1L) were grown at 37°C in terrific broth to an O.D. of approximately 1.0 and then 60 ng/ml anhydrotetracycline and 200 uM isopropyl-β-D-galactopyranoside (IPTG) were added to induce expression of PKS proteins and Sfp, respectively. Cultures continued incubation at 18C for 20 hours after which cells were pelleted and stored at -20C until further processing. Pellets were resuspended in lysis buffer (300 mM NaCl, 50 mM sodium phosphate, pH 6.8, 10 mM imidazole) supplemented with 0.1 mg/ml lysozyme. Suspensions were lysed by several passages through an EmulsiFlex C3 homogenizer (Avestin) and cellular debris was removed by centrifugation (15000g, 30 minutes). Cobalt resin (2-3 ml) was added to the supernatant and mixed at 4C for one hour before being applied to a fritted column. Resin was washed with lysozyme-free lysis buffer

until flow-through resulted in no color change when mixed with Bradford reagent. Proteins were eluted with several resin volumes of elution buffer (300 mM NaCl, 50 mM phosphate, pH 6.8, 200 mM imidazole) and concentrated via spin filtration (Amicon, 100 kDa MWCO). Concentrated eluate was exchanged into storage buffer (50 mM phosphate, pH 6.8, 10% glycerol) using a PD-10 column (GE Life Sciences), and then further concentrated prior to being flash frozen in liquid nitrogen and stored at -80C.

[0098] Purification of BorDH2. BorDH2 monodomain was purified as above with the exception that protein was concentrated with a 10 kDa MWCO filter and stored as a 50% glycerol solution at -20C after buffer exchange.

[0099] Reagents and Chemicals. HisPur cobalt resin was purchased from Thermo Scientific, Bradford reagent was from Bio-Rad and SDS-PAGE gels from Life Technologies.

[00100] Chemical synthesis and NMR data. Solvents (hexanes, ethyl acetate, dichloromethane and methanol) were purchased from EDH; *trans-β*-hydromuconic acid was purchased from Alfa Aesar; all other reagents were purchased from Sigma-Aldrich or as indicated.

[00101] Column chromatography was performed on a Teledyne Isco Combiflash Rf, with RediSep Rf Gold normal phase silica columns.

[00102] Gas chromatography – electron impact mass spectrometry (GC-EIMS) was performed on a Agilent5973 - HP6890 GC-MS using a 30 meter db5-ms column

[00103] ^{1}H NMR and ^{13}C NMR were obtained on a Bruker AVB 400 MHz spectrometer and a Bruker AV 500 MHz spectrometer at the UC Berkeley College of Chemistry NMR facility, funded in part by NSF grant CHE-0130862. Chemical shifts are reported in ppm relative to residual solvent signal ($\delta^{1}H$=3.31 and $\delta^{13}C$=49.0 for Methanol-d_4, $\delta^{1}H$=2.05 and $\delta^{13}C$=29.84 for Acetone-d_6).

Succinyl-SNAC

[00104] A 100 ml round-bottom flask was charged with 1 g of succinic anhydride and dissolved in a minimal volume of dichloromethane (DCM). 1 eq. *N*-acetylcysteamine (1.07 ml) was added dropwise to the stirring solution. After overnight incubation at ambient temperature with stirring, the mixture was extracted several times with saturated aqueous

sodium bicarbonate solution. The pH of this solution was lowered to approximately 6 with dropwise addition of 1 molar hydrochloric acid in order to protonate unreacted *N*-acetylcysteamine. The mixture was extracted several times with DCM to remove *N*-acetylcysteamine and then the pH of the aqueous solution was lowered to approximately 3, again with dropwise addition of 1M HCl to protonate the title compound. This was extracted several times with ethyl acetate (EtOAc), dried with the addition of sodium sulfate and filtered into a round-bottom flask. The solution was concentrated *in vacuo* to afford a fluffy white powder (0.663 g, 3.02 mmol, 30.4% yield).

Synthesis of 3-hydroxyadipic acid standard

Solvent-free synthesis of 2-(γ-butyrolactone)acetic acid

[00105] 2-(γ-butyrolactone)acetic acid (systematic name: 2-(5-oxotetrahydrofuran-2-yl)acetic acid), InChI=1S/C6H8O4/c7-5(8)3-4-1-2-6(9)10-4/h4H,1-3H2,(H,7,8)

[00106] 2 g of *trans*-β-hydromuconic acid (13.88mmol) and 4 g of silica gel (60 Å – 200 mesh) were mixed in a 50mL round-bottom flask with stir bar. The free-flowing mixture was heated to 200°C in a sand bath while gently stirring. The reaction was monitored by pipetting a few milligrams of the hot mixture into 1mL of dichloromethane (DCM), of which 40 µL was treated with 10 µL of *N*,*O*-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) and analyzed by gas chromatography-mass spectrometry (GC-MS). 3 hours into the reaction the mixture starts to turn yellow. After 24 hours, all starting material had been consumed. The mixture was cooled to room temperature and extracted with 50 mL DCM and filtered. The filtrant was extracted with another 50 mL of DCM. The light yellow filtrate was evaporated under reduced pressure and purified by flash chromatography (70:30 Ethyl acetate:Hexane) to afford the title compound as a viscous slightly yellow liquid that solidified upon standing (474mg, 3.29 mmol, 24% yield)

[00107] ^{1}H NMR (500 MHz, MeOD) δ 4.91 (p, J = 6.5 Hz, 1H), 2.73 (d, J = 6.4 Hz, 2H), 2.66 – 2.52 (m, 2H), 2.49 – 2.40 (m, 1H), 2.09 – 1.92 (m, 1H). ^{13}C NMR (126 MHz, MeOD) δ 179.68, 173.35, 78.64, 49.00, 40.55, 29.35, 28.31.

[00108] EIMS (TMS derivative): 201 (7%, $(M\text{-}Me)^{+}$), 159 (27%), 157 (54%), 117 (11%), 101 (8%), 85 (17%), 76 (7%), 75 (100%), 73 (53%), 59 (9%)

Hydrolysis of 2-(γ-butyrolactone)acetic acid to yield 3-hydroxyadipic acid

[00109] 3-hydroxyadipic acid (systematic name: 3-hydroxyhexanedioc acid) InChI=1S/C6H10O5/c7-4(3-6(10)11)1-2-5(8)9/h4,7H,1-3H2,(H,8,9)(H,10,11)

[00110] 30 mg of 2-(γ-butyrolactone)acetic (208 μmol) was dissolved in 10.4 mL 0.1 M aqueous potassium hydroxide (5 eq.), distributed among the wells of a 96-well PCR plate and heated to 99 °C for 3 hours in an Applied Biosciences Venti thermocycler with heated lid (105 °C). This was diluted into 100 mM sodium phosphate buffer (pH 6.8) to make standard curves. To acquire NMR data, the solution was consolidated and acidified to pH 3 using 6 M hydrochloric acid. The solution was flash frozen in liquid nitrogen and lyophilized to dryness (~24 h). The remaining powder was extracted with 2×2 mL acetone and filtered through a pipette filter (KCl has negligible solubility in acetone). At this point 1 μL of the solution was diluted down to a final volume of 40 μL and derivatized with 10 μL BSTFA to yield the GC-MS chromatogram below. The remaining solution was evaporated under reduced pressure at room temperature to yield the title compound as a white powder (26.7 mg, 165 μmol, 79% yield). Re-subjecting the product to GC-MS shows increasing amounts of 2-(γ-butyrolactone) acetic acid over time, suggesting that neat 3-hydroxyadipic acid spontaneously re-lactonizes at room temperature (unlike 3-hydroxyadipic acid solutions in phosphate buffer, which are stable and hence used for standard curves as described above). Hence, the NMR spectrum reported below shows 2-(γ-butyrolactone)acetic acid as an impurity.

[00111] ^{1}H NMR (500 MHz, MeOD) δ 4.02 (tdd, J = 8.6, 4.8, 3.9 Hz, 1H), 2.52 – 2.33 (m, 4H), 1.90 – 1.78 (m, 1H), 1.77 – 1.63 (m, 1H). ^{13}C NMR (126 MHz, MeOD) δ 179.68, 173.35, 78.64, 49.00, 40.55, 29.35, 28.31.

[00112] EIMS (tris(TMS) derivative): 363 (32% $(M\text{-}Me)^{+}$), 247 (26%), 233 (11%), 203 (12%), 149 (14%), 147 (55%), 133 (10%), 129 (24%), 75 (27%), 73 (100%)

Synthesis of Acyl-CoAs

[00113] Synthesis of cyclopentanecarboxyl-CoA (CPMA-CoA) and cyclopentanedicarboxyl-CoA (CPDA-CoA) was previously reported (Hagen et al., 2014)

Construct design and plasmid construction

[00114] AurB, IdmO, and SpnB DNA was generously provided by Ryan Phelan.

Codon-optimization and synthesis of AurB was performed by Genscript; codon-optimization and synthesis of IdmO, and SpnB was performed by the Joint Genome Institute (JGI). **NanA2** DNA was generously provided by Satoshi Yuzawa; codon-optimization and synthesis was performed by DNA2.0.

[00115] Amino acid sequences for various modules were aligned using the MUSCLE or Clustal Omega algorithms (Edgar, 2004; Sievers et al., 2014). All DNA pieces were amplified via PCR with either Q5 or Phusion polymerases (New England BioLabs) according to manufacturer's recommendations. Gel-extracted DNA was assembled via Gibson cloning using Gibson Assembly® master mix (New England BioLabs). In the case of construct A3 (pARH159), Gibson assembly failed and sequence was introduced by oligonucleotides using inverse PCR with pARH137 (A2) as a template. A similar strategy was used to create (A,S)3c constructs (pARH163, 164 respectively) starting from (A,S)2c constructs (pARH147, 149 respectively). A complete list of plasmids follows

BorDH2 monodomain

[00116] BorDH2 domain boundaries were selected after (Vergnolle et al., 2011) and DNA was codon-optimized and synthesized as a gBlock (Integrated DNA Technologies) and ligated into the pET28a vector (Novagen) to yield an N-terminal hexahistidine tagged construct.

[00117] Table 1. Plasmids described in Example 1 herein. Strains may be accessed and requested through the website for public-registry.jbei.org.

Strain ID	Alias	Summary
JBx_045172	pARH136	pBbS2k::6xHisMBP-BorA2eloopswap-A1
JBx_045173	pARH137	pBbS2k::6xHisMBP-BorA2eloopswap-A2
JBx_045174	pARH138	pBbS2k::6xHisMBP-BorA2eloopswap-I1
JBx_045175	pARH139	pBbS2k::6xHisMBP-BorA2eloopswap-I2
JBx_045176	pARH140	pBbS2k::6xHisMBP-BorA2eloopswap-N1
JBx_045177	pARH141	pBbS2k::6xHisMBP-BorA2eloopswap-N2
JBx_045178	pARH142	pBbS2k::6xHisMBP-BorA2eloopswap-S1
JBx_045179	pARH143	pBbS2k::6xHisMBP-BorA2eloopswap-S2
JBx_045183	pARH147	pBbS2k::6xHisMBP-BorA2eloopswap-A2c
JBx_045184	pARH148	pBbS2k::6xHisMBP-BorA2eloopswap-I2c
JBx_045202	pARH149	pBbS2k::6xHisMBP-BorA2eloopswap-S2c
JBx_045078	pARH150	pET28a::BorDH2
JBx_045189	pARH159	pBbS2k::6xHisMBP-BorA2eloopswap-A3

JBx_045190	pARH162	pBbS2k::6xHisMBP-BorA2eloopswap-S3
JBx_045191	pARH163	pBbS2k::6xHisMBP-BorA2eloopswap-A3c
JBx_045192	pARH164	pBbS2k::6xHisMBP-BorA2eloopswap-S3c
JBx_045199	pARH176	pBbS2k::6xHisMBP-BorA2eloopswap-S3c-eryTE

Intermediate analysis of Succinyl-SNAC extensions:

[00118] Samples were analyzed on an AB Sciex (Foster City, CA) 4000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. 1-2 µg of total peptide was injected onto a Sigma (St. Louis, MI) Ascentis Peptide Express C-18 column (2.1 mm x 50 mm) via an autosampler. A 20.5-minute method was used with a flow-rate of 400 ul/min. The method begins with 95% Buffer A (water, 2% acetonitrile, 0.1% formic acid) and 5% buffer B (water, 98 % acetonitrile, 0.1% formic acid) for 1.2 minutes followed by a rapid rise to 25% over 1 minute and then a very slow rise to 36% over 10 minutes. After the slow gradient step, buffer B was rapidly increased to 90%, held, and dropped back down to re-equilibrate the column as above. The peptides eluting from the column were ionized by a Turbo V Ion source (curtain gas flow: 20 l/min, temperature: 400 C, ion spray voltage: 4,800 V, ion source gas flow: 50 l/min, entrance potential: 10 V) operating in positive-ion mode.

[00119] Table 2. Mass spectrum parameters for intermediate analysis experiments.

ID	Q1	Q3	Declustering potential	Collision energy
ACP1_ctrl	680.38	846.48 (y8)	125	40
Holo-ACP1	905.76	261.12	50	44
Keto-ADA-ACP1	953.11	403.15	50	44
hydroxy-ADA-ACP1	953.77	405.16	50	44
2,3-ene-ADA-ACP1	947.78	387.16	50	44
ADA-ACP1	948.45	389.17	50	44

Intermediate analysis of various acyl-ACP_{LM} extensions

[00120] Samples were analyzed on an Agilent 6460QQQ mass spectrometer operating in MRM (SRM) mode as previously reported (Dahl et al., 2013). Briefly, 1-2 ug of total peptide was injected on a Sigma Ascentis Peptide Express C-18 column (2.1 mm x 50 mm) via an autosampler and separated at 400 ul/min. Liquid chromatography conditions used were as described above. Peptides eluting from the column were ionized using an Agilent Jet

Stream source (sheath gas flow: 11 l/min, sheath gas temperature: 350 c, nozzle voltage: 1,000 v, nebulizing pressure: 30 psi, chamber voltage: 4,500 V) operating in positive-ion mode

[00121] For all experiments, transitions were monitored using a collision cell exit potential of 10 V.

[00122] ACP1_ctrl peptide: VVESVAFGVPSLR (SEQ ID NO:3)

[00123] ACP1 peptide: AAIGPDSSFHAIGFDSLTAVELR (SEQ ID NO:4)

(site of phosphopantetheinylation underlined)

[00124] Methods for SNAC extensions were designed and data collected in Analyst 3.1 and data was quantified in MultiQuant 2.1 (AB Sciex). Methods for acyl-ACP extension were designed in Skyline (MacLean et al., 2010) and data collected in MassHunter (Agilent)

Data analysis

[00125] Raw data for each transition was normalized by dividing a transition's peak area by that of a control peptide present in BorMod1, but which does not participate in catalysis and should therefore be invariant across samples ("ACP1_ctrl") to generate values in "control peptide equivalents."

Adipic acid analytical methods

[00126] Adipic acid (commercially available) and 3-hydroxy-adipic acid were directly infused into the mass spectrometer operating in negative mode and a scan was conducted to identify product ions during adjustment of relevant acquisition parameters.

(3-hydroxy) Adipic acid production

[00127] Samples were analyzed on an AB Sciex (Foster City, CA) 4000 Q-Trap mass spectrometer operating in MRM (SRM) mode coupled to an Agilent 1100 system. 15 ul of each reaction was injected onto a Phenomenex (Torrance, CA) Kinetex XB C-18 column (3 mm x 100 mm, 1.7u) via an autosampler. A 24 minute method was used with a flow-rate of 200 ul/min and started with 97.5 % Buffer A (water, 0.1% formic acid) and 2.5% Buffer B (acetonitrile, 0.1% formic acid) for 3 minutes followed by a rise to 90% buffer B over 10

minutes where it was held for 2 minutes and then a return to 2.5% Buffer B for 9 minutes to re-equilibrate the column. Analytes eluted from column were ionized using a Turbo V Ion source (curtain gas flow: 20 l/min, temperature: 400 C, ion spray voltage: -4,500 V, ion source gas flow: 60 l/min, entrance potential: -10 V) operating in negative-ion mode.

[00128] Table 3. Mass spectrum parameters for adipic acid and 3-OH-adipic acid detection

ID	Q1	Q3	Declustering potential	Collision energy
3-hydroxy	161	99	-45	-18
Adipic	145	101	-45	-18

[00129] Further references cited:

Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D., et al. (2013). Engineering dynamic pathway regulation using stress-response promoters. Nat. Biotechnol. *31*, 1039–1046.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Hagen, A., Poust, S., de Rond, T., Yuzawa, S., Katz, L., Adams, P.D., Petzold, C.J., and Keasling, J.D. (2014). *In Vitro* Analysis of Carboxyacyl Substrate Tolerance in the Loading and First Extension Modules of Borrelidin Polyketide Synthase. Biochemistry (Mosc.) *53*, 5975–5977.

Hillson, N.J., Rosengarten, R.D., and Keasling, J.D. (2012). j5 DNA Assembly Design Automation Software. ACS Synth. Biol. *1*, 14–21.

MacLean, B., Tomazela, D.M., Abbatiello, S.E., Zhang, S., Whiteaker, J.R., Paulovich, A.G., Carr, S.A., and MacCoss, M.J. (2010). Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry. Anal. Chem. *82*, 10116–10124.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2014). Fast, scalable generation of high-quality protein

multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

Vergnolle, O., Hahn, F., Baerga-Ortiz, A., Leadlay, P.F., and Andexer, J.N. (2011). Stereoselectivity of Isolated Dehydratase Domains of the Borrelidin Polyketide Synthase: Implications for cis Double Bond Formation. ChemBioChem *12*, 1011–1014.

[00130] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

What we claim is:

1. A polyketide synthase (PKS) capable of synthesizing a carboxylic acid, said PKS comprising a synthetic module comprising the S3c variant module, or a functional variant thereof, wherein the PKS is capable of synthesizing a carboxylic acid.

2. A polyketide synthase (PKS) capable of synthesizing a carboxylic acid, said PKS comprising a hybrid module comprising a BorA2 KS domain, or functional variant thereof, a BorA2 AT domain, or functional variant thereof, a DH described in Example 1, or functional variant thereof, a heterologous KR domain, a heterologous ER domain, and a BorA2 ACP domain, or functional variant thereof, wherein the PKS is capable of synthesizing a carboxylic acid.

3. A recombinant nucleic acid encoding the polyketide synthase (PKS) of claim 1 or 2.

4. A replicon comprising the recombinant nucleic acid 3, wherein the replicon is capable of stable maintenance in a host cell.

5. The replicon of claim 4, wherein the replicon is a plasmid or vector.

6. The replicon of claim 5, wherein the vector is an expression vector.

7. A host cell comprising the recombinant nucleic acid of claim 3.

8. A host cell comprising the replicon of one of claims 4-6.

9. The host cell of claim 7 or 8, wherein the host cell when cultured produces the carboxylic acid.

10. A method of producing carboxylic acid, comprising: providing a host cell of claim 9, and culturing said host cell in a suitable culture medium such that the carboxylic acid is produced.

11. The method of claim 10, further comprising isolating the carboxylic acid.

12. The method of claim 11, wherein the carboxylic acid is a diacid, and further comprising reacting the diacid with a diamine to produce a nylon.

13. The method of claim 11, wherein the carboxylic acid is a diacid, and further comprising reacting the diacid with a dialcohol to produce a polyester.

14. A composition comprising a carboxylic acid isolated from a host cell of claim 11 from which the carboxylic acid is produced, and trace residues and/or contaminants of the host cell.

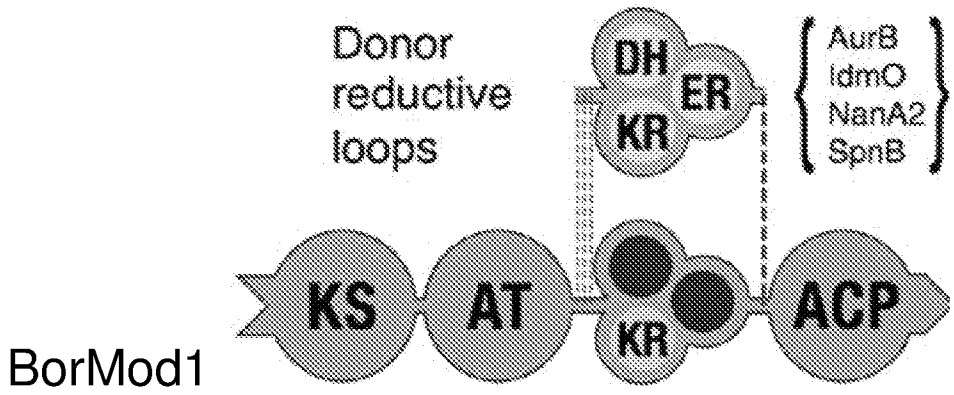

Figure 1A

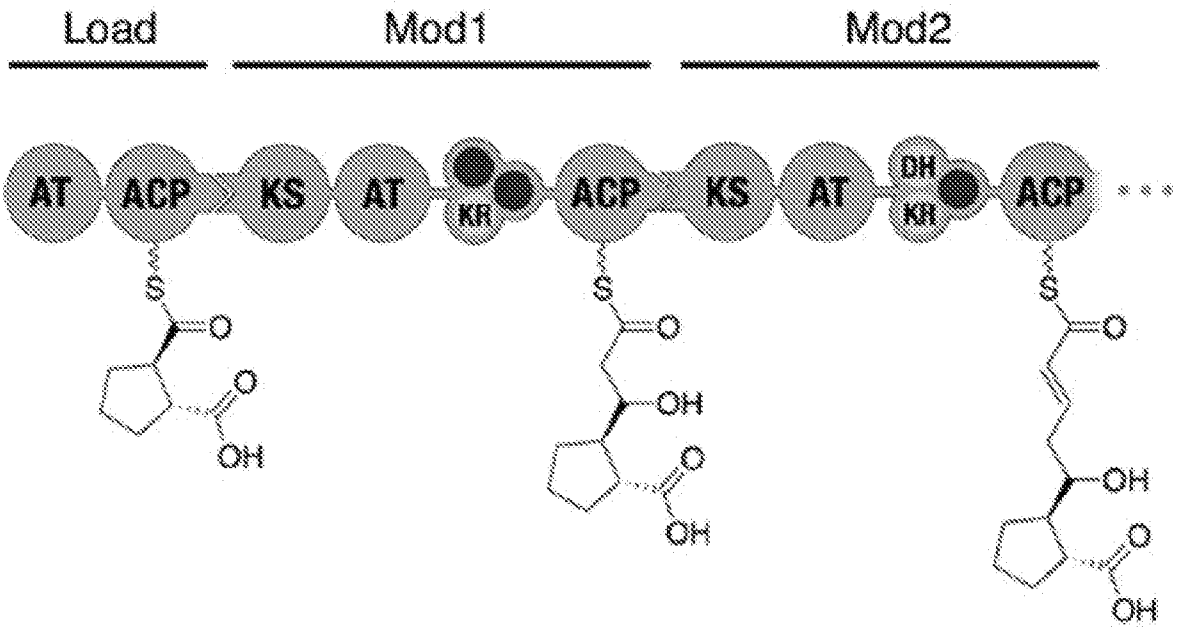

Figure 1B

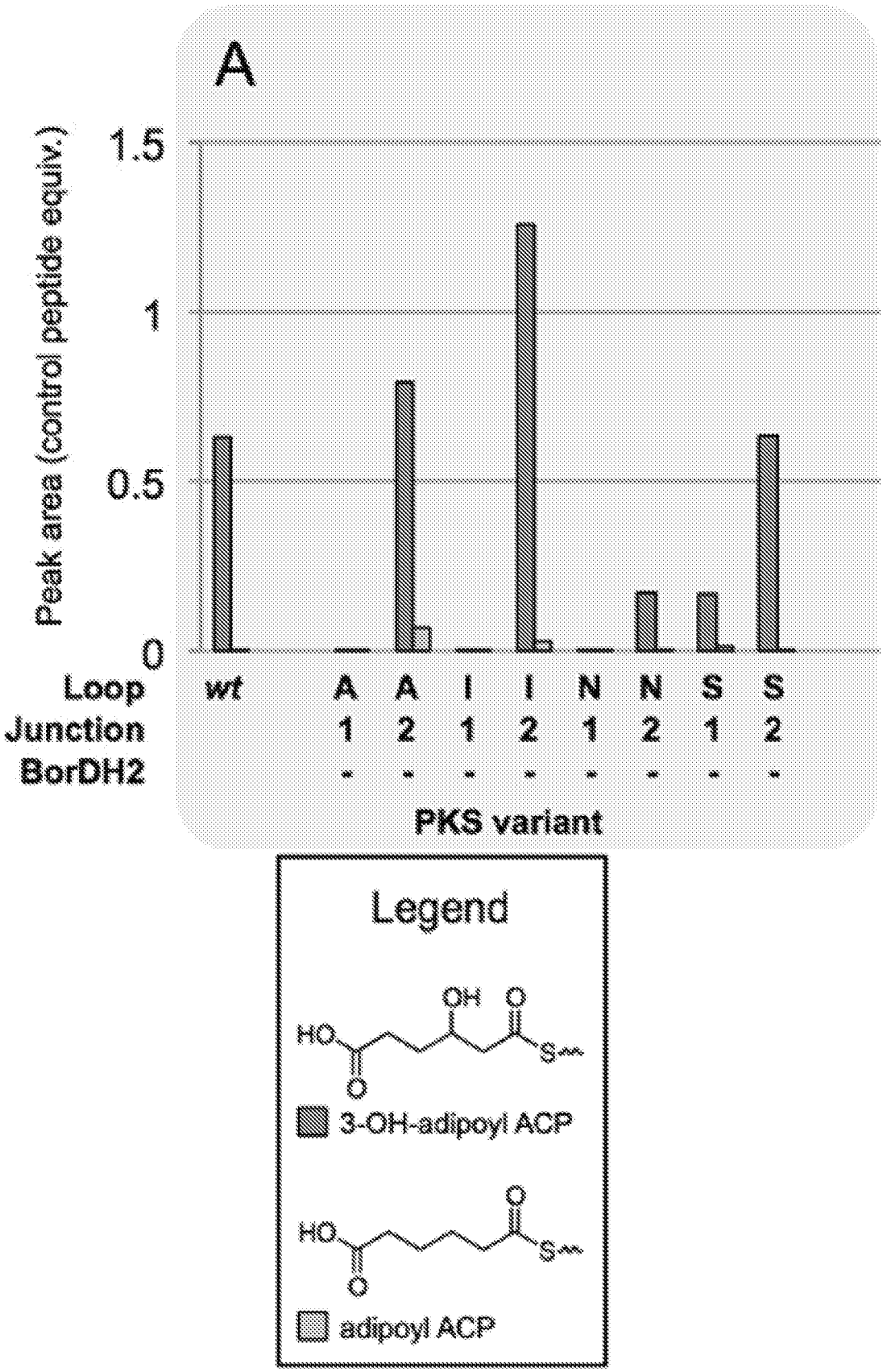
A
Peak area (control peptide equiv.)
1.5
1
0.5
0
Loop
Junction
BorDH2
wt
A1-
A2-
I1-
I2-
N1-
N2-
S1-
S2-
PKS variant
Legend
OH
O
HO
S
3-OH-adipoyl ACP
HO
S
adipoyl ACP

Figure 2A

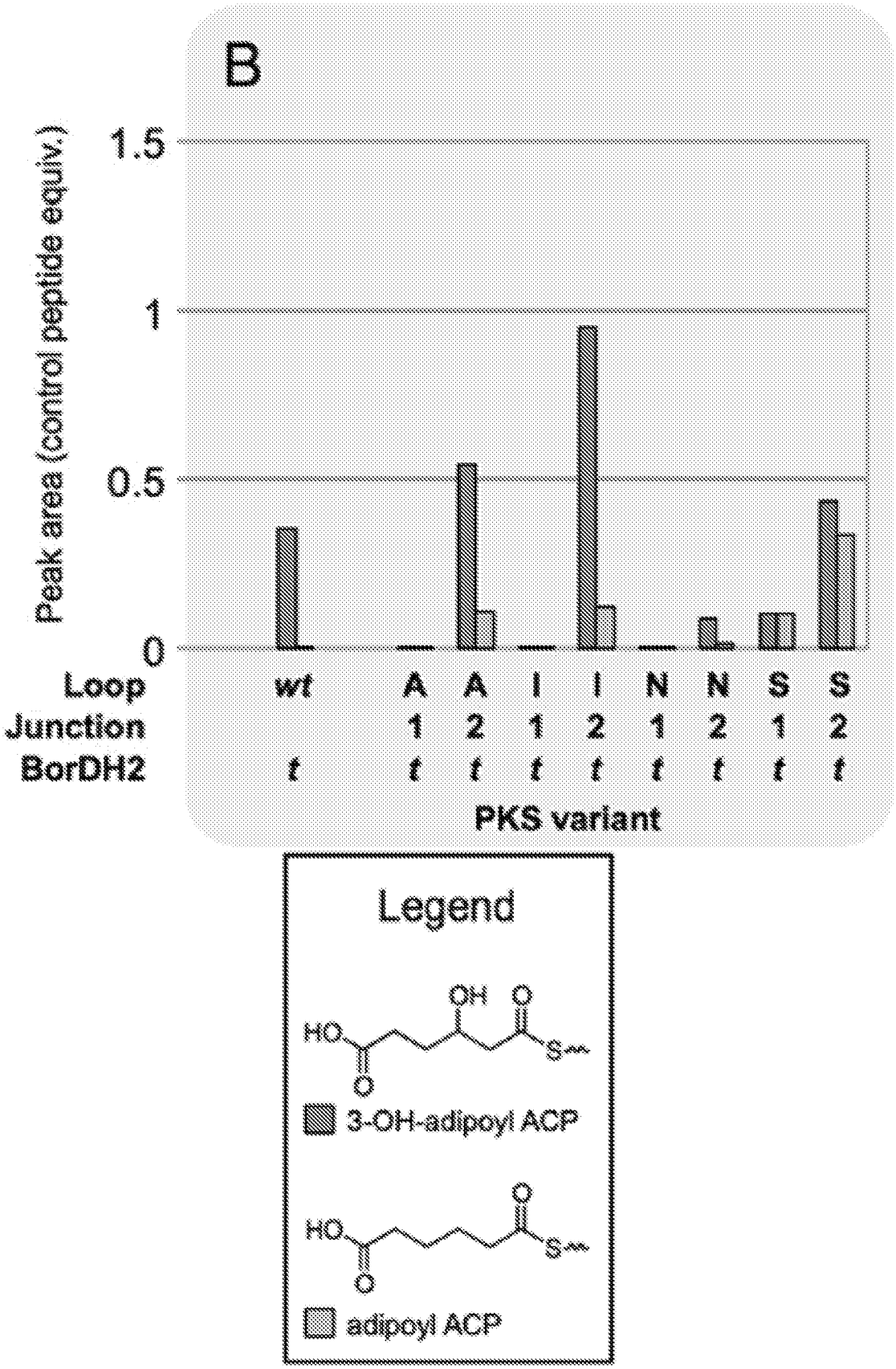
B
Peak area (control peptide equiv.)
1.5
1
0.5
0
Loop
Junction
BorDH2
wt
A 1 t
A 2 t
I 1 t
I 2 t
N 1 t
N 2 t
S 1 t
S 2 t
PKS variant
Legend
OH
HO
S
3-OH-adipoyl ACP
HO
S
adipoyl ACP

Figure 2B

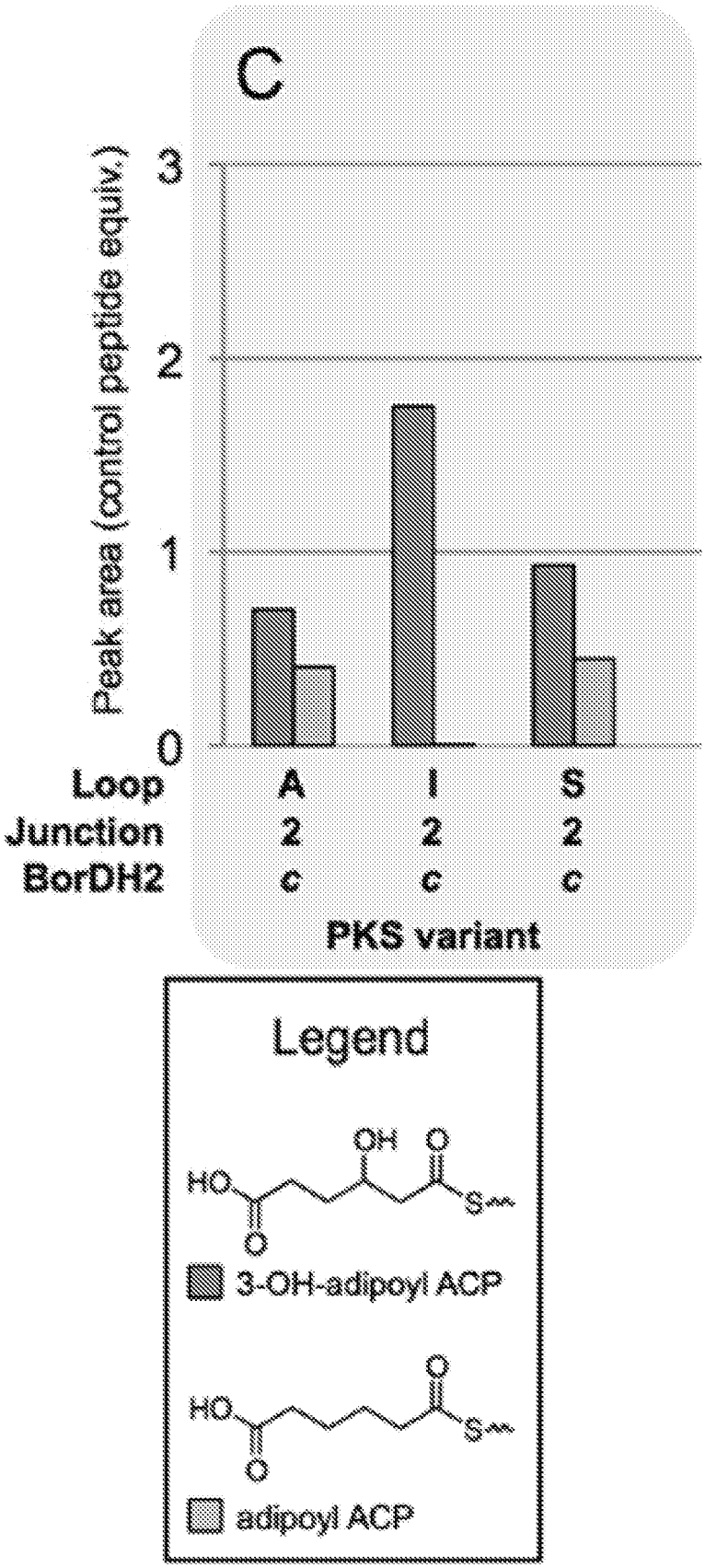
C
Peak area (control peptide equiv.)
3
2
1
0
Loop
Junction
BorDH2
A
2
c
I
2
c
S
2
c
PKS variant
Legend
OH
O
HO
S
O
3-OH-adipoyl ACP
HO
O
S
O
adipoyl ACP

Figure 2C

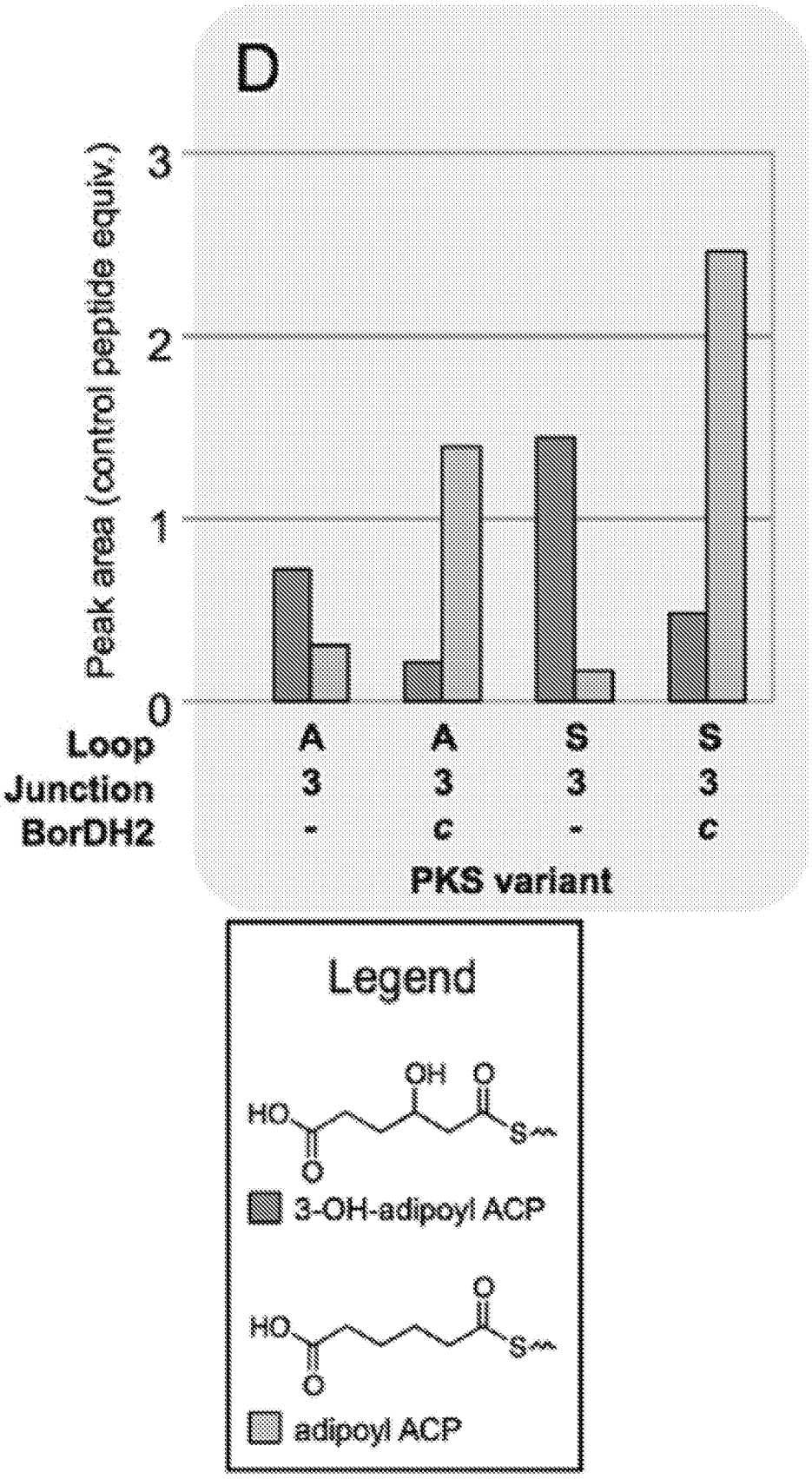
D
Peak area (control peptide equiv.)
3
2
1
0
Loop
Junction
BorDH2
A 3 -
A 3 c
S 3 -
S 3 c
PKS variant
Legend
OH
O
HO
S
O
3-OH-adipoyl ACP
HO
O
S
O
adipoyl ACP

Figure 2D

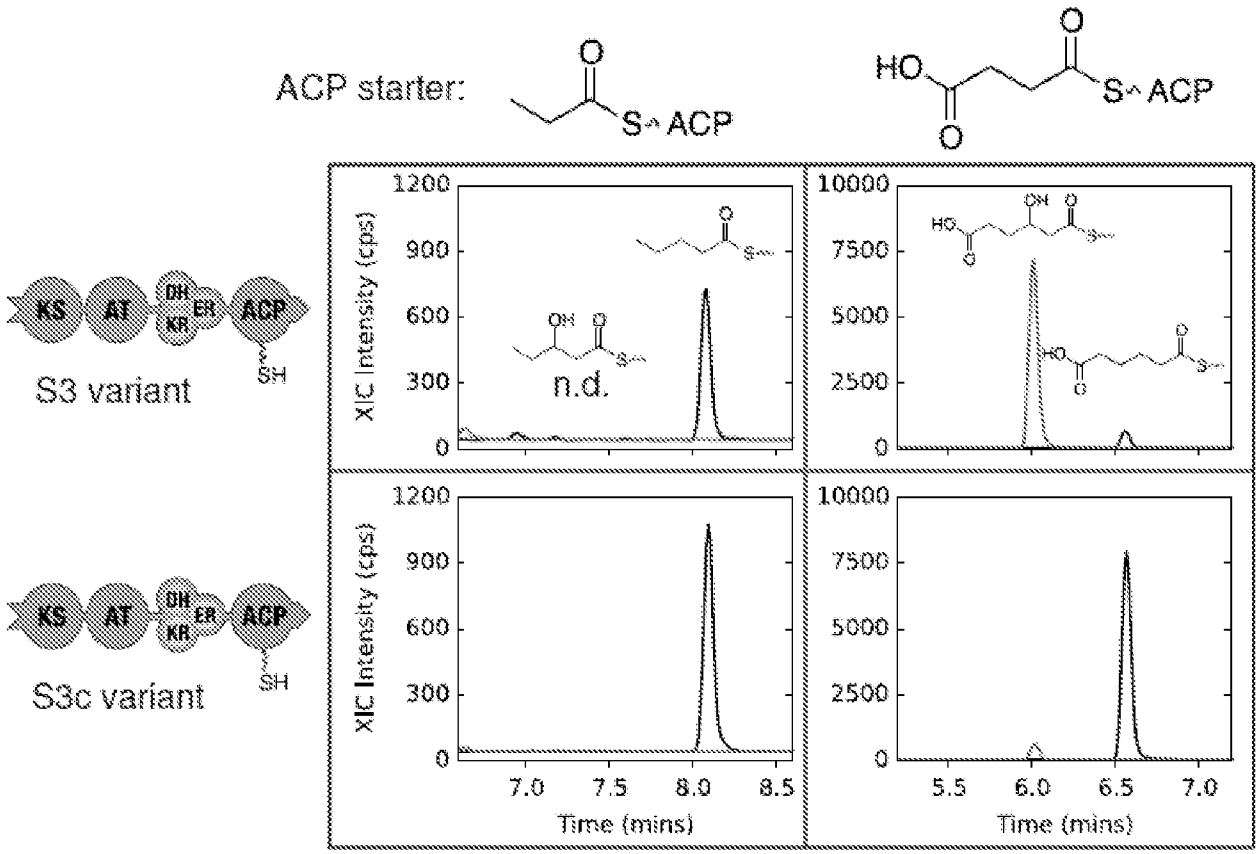
ACP starter:
S ACP
HO
S ACP
KS
AT
DH
KR
ER
ACP
SH
S3 variant
S3c variant
XIC Intensity (cps)
1200
900
600
300
0
10000
7500
5000
2500
n.d.
7.0
7.5
8.0
8.5
5.5
6.0
6.5
7.0
Time (mins)

Figure 3

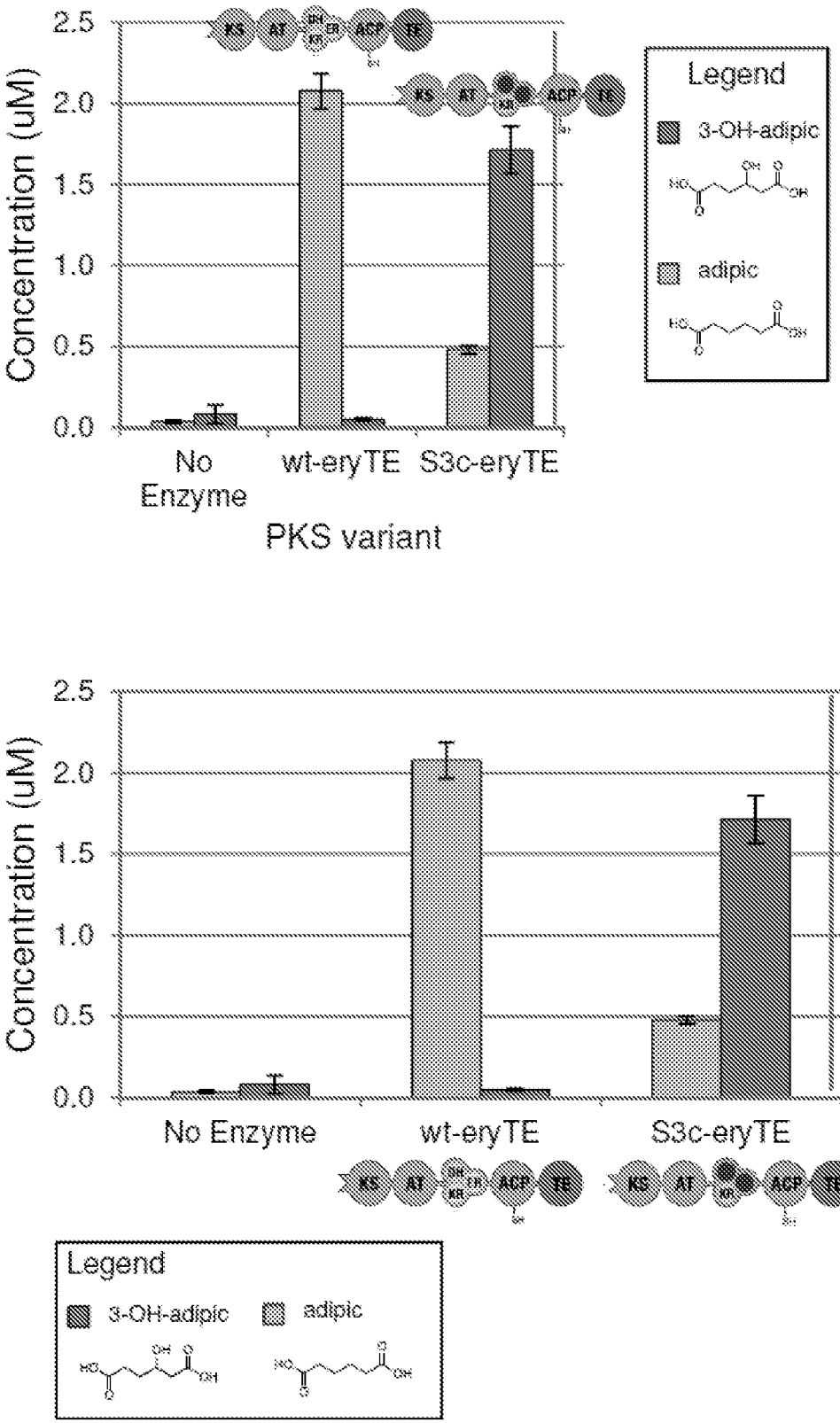
Concentration (uM)
2.5
2.0
1.5
1.0
0.5
0.0
No Enzyme
wt-eryTE
S3c-eryTE
PKS variant
Legend
3-OH-adipic
adipic
KS
AT
KR
ER
ACP
TE

Figure 4

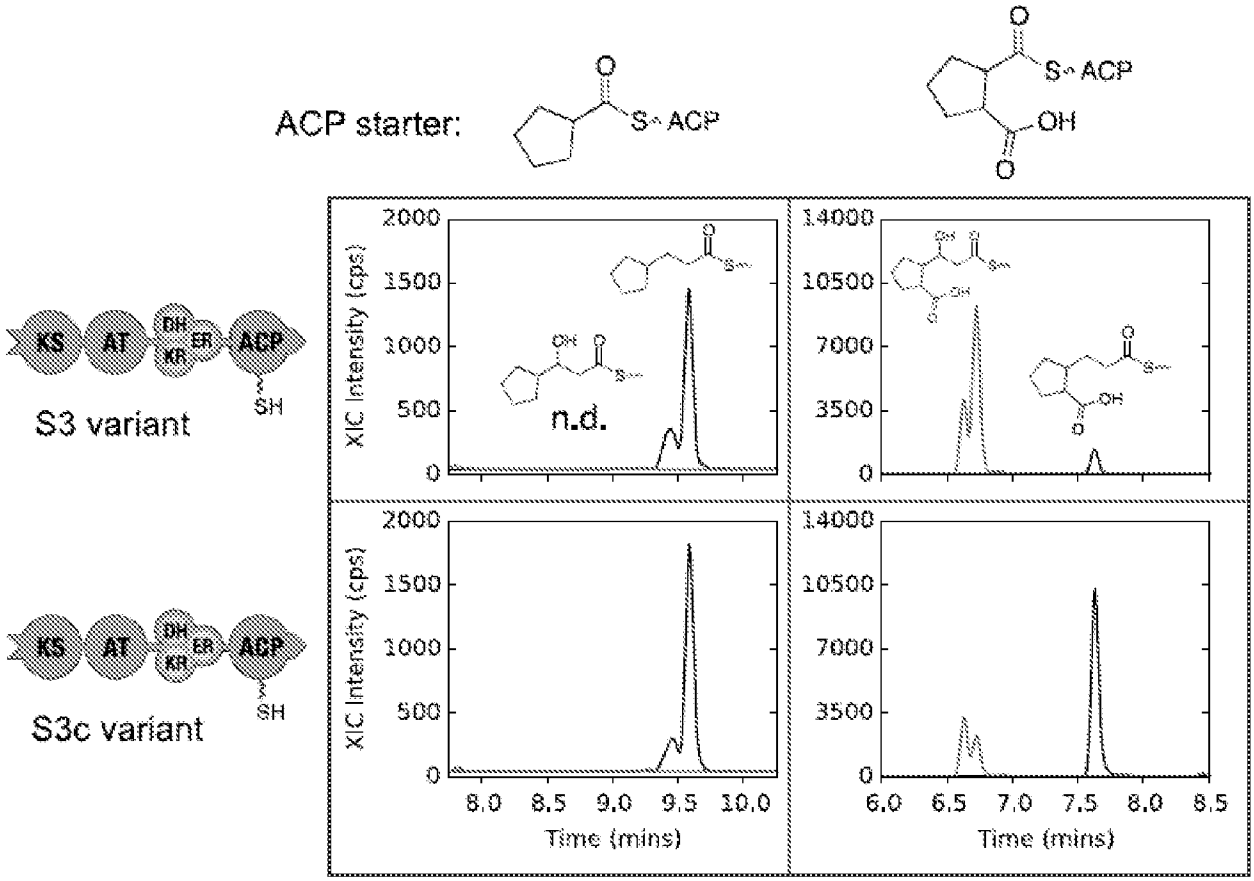

Figure 5

N-terminal junctions:

```
BorMod1   QGHPVDWAPLPF---RARTVDLPTYPFQHQRYWLDVPPL--------------------PTR   900
AurB      RGVRVDWDRFFAG-TGARRIDLPTYAFRRRSFPWIQAA-PD-ADVTSAGLAGLGHPLLGA   937
IdmO      RGVQVDWPALLP---GGRRVELPTYAFQRRRYWLEDAD-PTGGDPAALGLTAADHPLLGA   930
NanA2     AGVAVDWASWYADGPAPAAVPLPAYAFQRERYWLPACAGSGPGDVAGAGLTAVGHALLPV   950
SpnB      RGGAVDWRSFFAG-TGARQIELPTYAFQRQRYWLVPSD---SGDVTGAGLAGAEHPLLGA   945
```

1 2

C-terminal junction:

```
BorMod1   VRSLGEQREDGPGGQGEEDGLAGRLAALPEADRRRAVLDLVEELVLGVLGHETRAAIGPD  1442
AurB      ---------DAGPLP--AADALRGRLASLAPEERREAVLDVVRAQVAVVLGHGAPEGIDPQ  2035
IdmO      ---------RAASTAAAQGPSLAERLAGLPVTERERIVVELVRADLAAVLGRSSSAGIDPG  2024
NanA2     ARSGGAAAARPTAAAAEEPADWAARLRALAPAEQRRLLTDLVRRHAAPVLGHADPEAVPAD  2076
SpnB      ----------RRAATADDAESLRKRLGRLPDAEQRERILLDLVRNEVAAVLGPASSQEITAD  2024
```

Figure 6

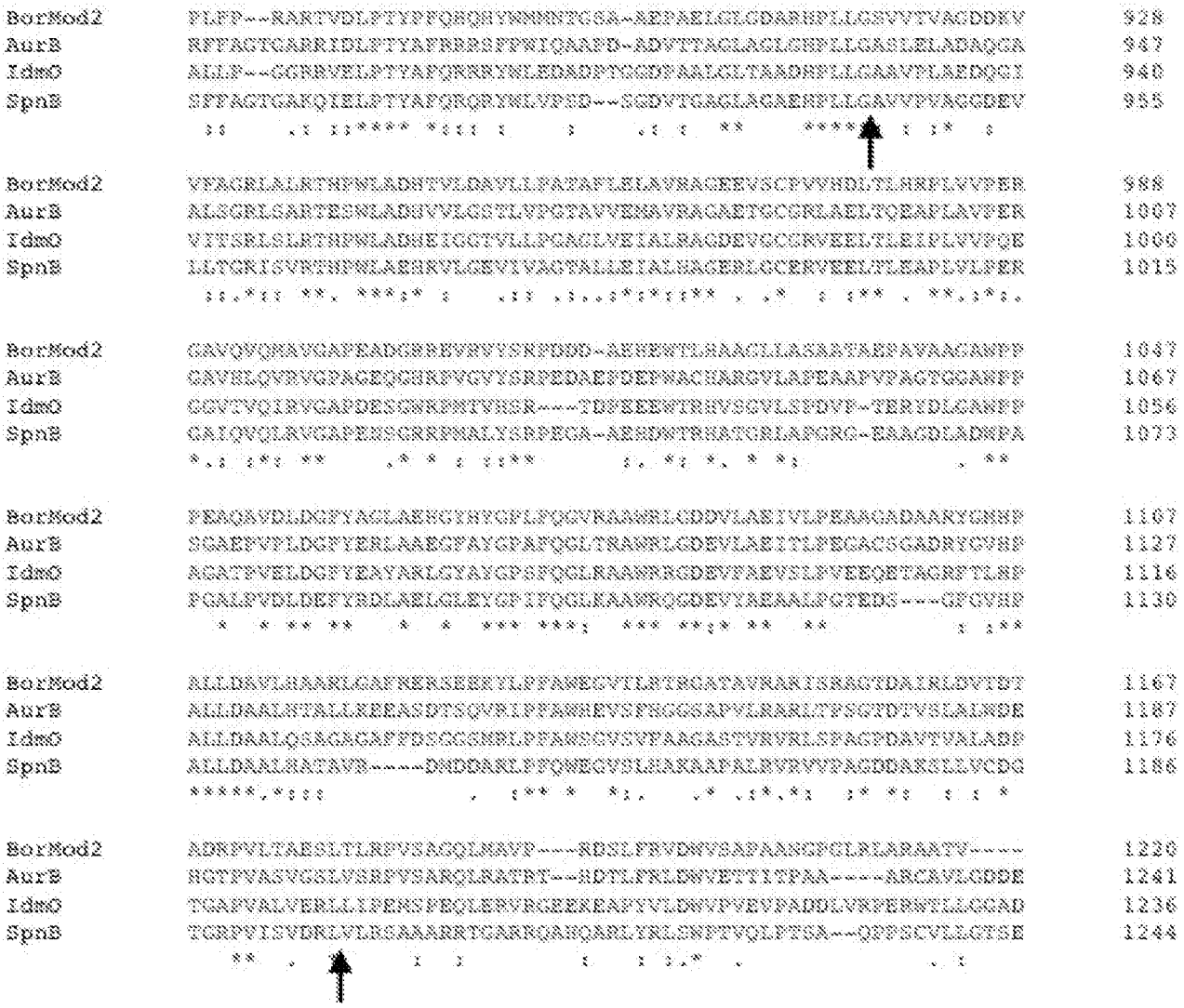

Figure 7

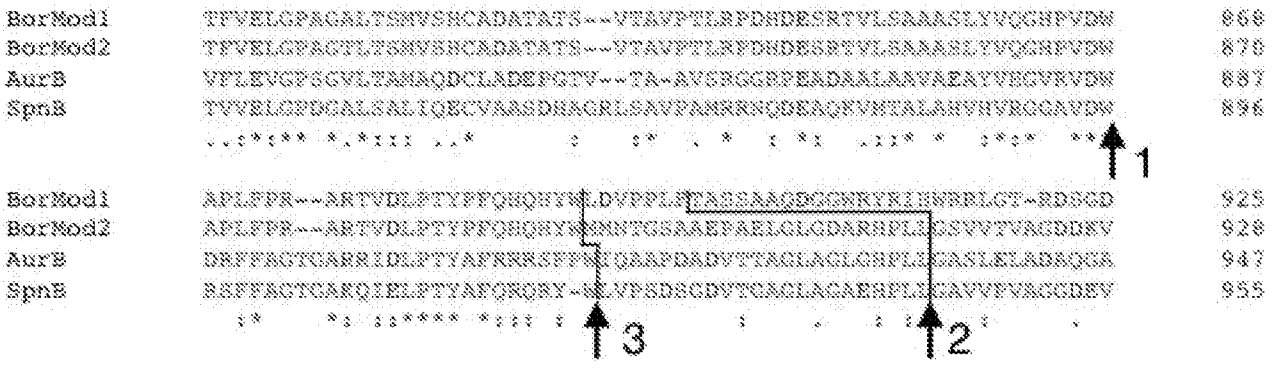

Figure 8

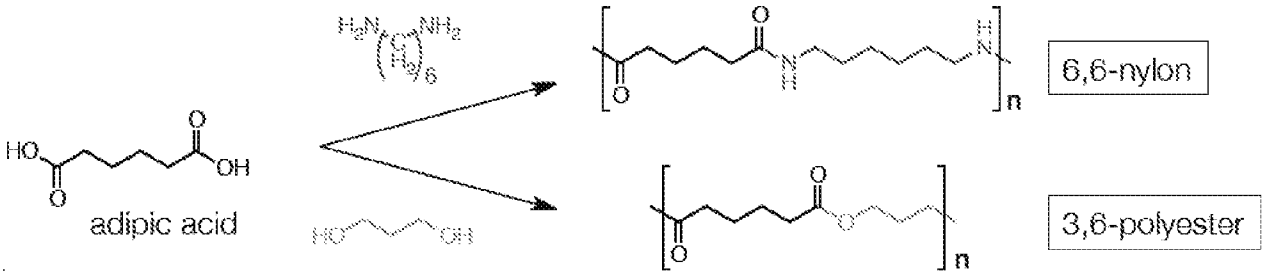

Figure 9

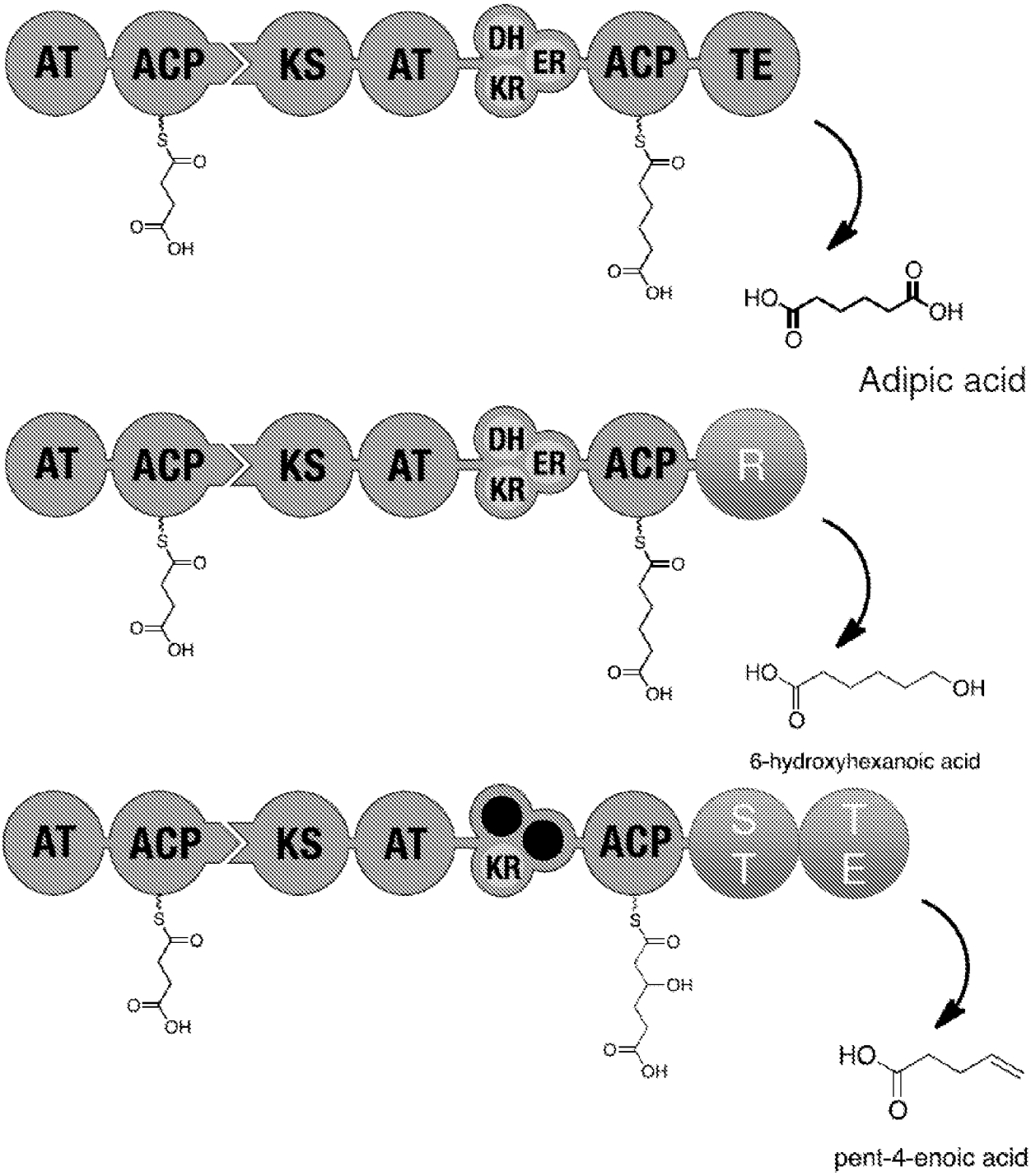
AT
ACP
KS
AT
DH
KR
ER
ACP
TE
Adipic acid
AT
ACP
KS
AT
DH
KR
ER
ACP
R
6-hydroxyhexanoic acid
AT
ACP
KS
AT
KR
ACP
S
T
T
E
pent-4-enoic acid

Figure 10

Extender Modules

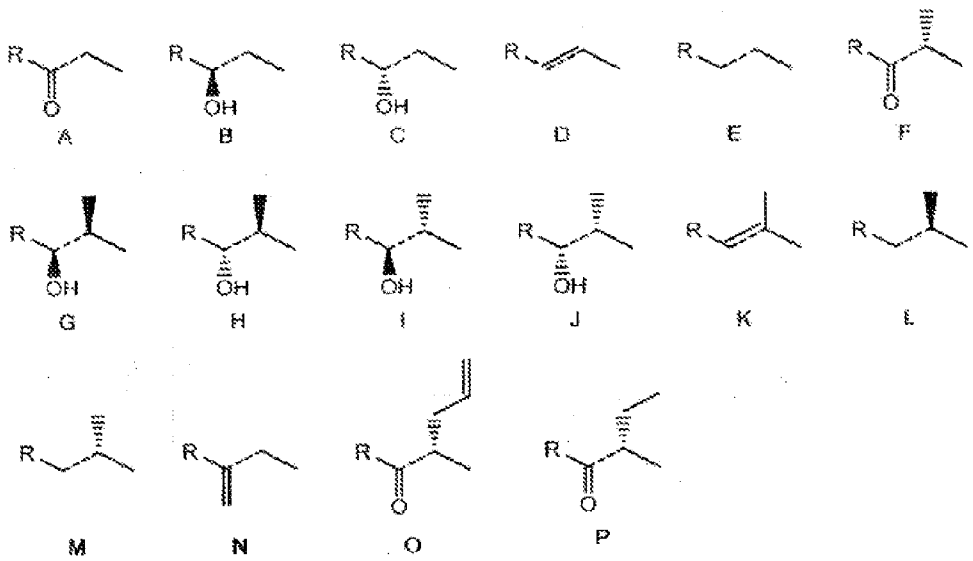

Figure 11