Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2359801 A1 2000/08/17

(21) 2 359 801

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2000/02/09
(87) Date publication PCT/PCT Publication Date: 2000/08/17
(85) Entrée phase nationale/National Entry: 2001/07/26
(86) N° demande PCT/PCT Application No.: US 00/03345
(87) N° publication PCT/PCT Publication No.: WO 00/47724
(30) Priorité/Priority: 1999/02/09 (60/119,363) US

(51) Cl.Int.^{7}/Int.Cl.^{7} C12P 19/62, C12N 15/00

(71) Demandeur/Applicant:
BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY, US

(72) Inventeurs/Inventors:
KHOSLA, CHAITAN, US;
TSUJI, STUART Y., US;
GOKHALE, RAJESH S., IN

(74) Agent: FETHERSTONHAUGH & CO.

(54) Titre : PROCEDES DE MODULATION DE L'EFFICACITE DE MODULES DE POLYKETIDE-SYNTHASES
(54) Title: METHODS TO MEDIATE POLYKETIDE SYNTHASE MODULE EFFECTIVENESS

(57) Abrégé/Abstract:
The linking sequences which modulate cross-talk between modules of Type I polyketide synthases have been identified. Thus, arbitrarily chosen modules can be mixed and matched by supplying the appropriate linkers to obtain desired polyketide synthases and new polyketides. The modules are provided suitable linkers so that the polyketide chain is passed from one module to the other in the correct sequence.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 August 2000 (17.08.2000)

PCT

(10) International Publication Number
WO 00/47724 A3

(51) International Patent Classification[7]: C12N 15/00, C12P 19/62,

(21) International Application Number: PCT/US00/03345

(22) International Filing Date: 9 February 2000 (09.02.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/119,363 9 February 1999 (09.02.1999) US

(71) Applicant: **BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; Suite 350, 900 Welch Road, Palo Alto, CA 94304 (US).

(72) Inventors: **GOKHALE, Rajesh, S.**; 88 Sadar Apartments, Mayur Vihar Phase I (Extn), Delhi 110091 (IN). **TSUJI, Stuart, Y.**; 121 Campus Drive, Stanford, CA 94305 (US). **KHOSLA, Chaitan**; 740 La Para Avenue, Palo Alto, CA 94306 (US).

(74) Agents: **MURASHIGE, Kate, H.** et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue N.W., Washington, DC 20006-1888 (US).

(81) Designated States *(national)*: AU, CA, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

(88) Date of publication of the international search report:
7 December 2000

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/47724 A3

(54) Title: METHODS TO MEDIATE POLYKETIDE SYNTHASE MODULE EFFECTIVENESS

(57) Abstract: The linking sequences which modulate cross-talk between modules of Type I polyketide synthases have been identified. Thus, arbitrarily chosen modules can be mixed and matched by supplying the appropriate linkers to obtain desired polyketide synthases and new polyketides. The modules are provided suitable linkers so that the polyketide chain is passed from one module to the other in the correct sequence.

METHODS TO MEDIATE POLYKETIDE SYNTHASE MODULE EFFECTIVENESS

This application claims priority under 35 U.S.C. § 119(e) from provisional application No. 60/119,363 filed 9 February 1999, the contents of which are incorporated herein by reference.

The invention herein was made, at least in part, based on support by grants from the National Institutes of Health, CA-66736 and GM-22172 and from the National Science Foundation, BES-9806774. The U.S. government has certain rights in this invention.

Technical Field

The invention is directed to facilitating usage by polyketide synthase modules of nascent polyketide chains. More specifically, the invention concerns including intermodule and intramodule linkers in constructions for synthesis of desired polyketides.

Background of the Invention

Polyketides are a class of compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes. Polyketides are biologically active molecules with a wide variety of structures, and the class encompasses numerous compounds with diverse activities. Tetracycline, erythromycin, epothilone, FK-506, FK-520, narbomycin, picromycin, rapamycin, spinocyn, and tylosin are examples of polyketides. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds.

The biosynthetic diversity of polyketides is generated by repetitive condensations of simple monomers by polyketide synthase (PKS) enzymes that mimic fatty acid synthases. For instance, the deoxyerythronolide-B synthase catalyzes the chain extension of a primer with several methylmalonyl coenzyme A (MeMalCoA) extender units to produce the erythromycin core.

The cloning, analysis, and recombinant DNA technology of genes that encode PKS enzymes allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The technology also allows one to produce molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters. See, *e.g.*, PCT publication Nos. WO 93/13663; 95/08548; 96/40968; 97/02358; 98/27203; and 98/49315; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; 5,712,146; 5,830,750; and 5,843,718; and Fu, *et al.*, 1994, *Biochemistry 33*: 9321-9326; McDaniel, *et al.*, 1993, *Science 262*: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between acylthioester building blocks. The building blocks used to form complex polyketides are typically acylthioesters, such as acetyl, butyryl, propionyl, malonyl, hydroxymalonyl, methylmalonyl, and ethylmalonyl CoA. Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesized. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

The present invention concerns modular PKS. In the Type I or modular PKS enzyme group, a set of separate catalytic active sites (each active site is termed a "domain", and a set thereof is termed a "module") exists for each cycle of carbon chain elongation and modification in the polyketide synthesis pathway. The typical modular PKS is composed of several large polypeptides, which can be segregated from amino to carboxy terminii into a loading module, multiple extender modules, and a releasing (or thioesterase) domain. The PKS enzyme known as 6-deoxyerythronolide B synthase (DEBS) is a typical Type I PKS. In DEBS, there is a loading module, six extender modules, and a thioesterase (TE) domain. The loading module, six extender modules, and TE of DEBS are present on three separate proteins (designated DEBS-1, DEBS-2, and DEBS-3, with two extender modules per protein). Each of the DEBS polypeptides is encoded by a separate open reading frame (ORF) or gene; these genes are known as *eryAI*, *eryAII*, and *eryAIII*. See Figure 1. There is considerable interest in

the genetic and chemical reprogramming of modular PKSs (see, *e.g.*, Khosla, 1997, *Chem. Rev. 97*:2577-2590, and Staunton, *et al.*, 1997, *Chem. Rev.* 2611-2629, each of which is incorporated herein by reference).

Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The loading module of DEBS consists of an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain. Another type of loading module utilizes an inactivated KS, an AT, and an ACP. This inactivated KS is in some instances called KS^{Q}, where the superscript letter is the abbreviation for the amino acid, glutamine, that is present instead of the active site cysteine required for ketosynthase activity. In other PKS enzymes, including the FK-520 PKS, the loading module incorporates an unusual starter unit and is composed of a CoA ligase activity domain. In any event, the loading module recognizes a particular acyl-CoA (usually acetyl or propionyl but sometimes butyryl) and transfers it as a thiol ester to the ACP of the loading module.

The AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, *i.e.*, methylmalonyl, ethylmalonyl, and hydroxymalonyl) and transfers it to the ACP of that extender module to form a thioester. Each extender module is responsible for accepting a compound from a prior module, binding a building block, attaching the building block to the compound from the prior module, optionally performing one or more additional functions, and transferring the resulting compound to the next module. The transfer into a module is mediated by the KS domain which is upstream of the remaining catalytic domains. The additional functions are performed by enzymes which comprise a ketoreductase (KR) which reduces the carbonyl group generated from the condensation to an alcohol, a dehydratase (DH) which converts the alcohol to a double bond, and an enoyl reductase (ER) which reduces the double bond to a single bond. These catalytic domains appear to be immediately adjacent and not separated by any linking sequences. Collectively, they can be called "beta-carbonyl modifying" domains. Thus, a particular module may contain none of these activities, only KR, or KR+DH, or KR+DH+ER. Thus, the order of domains from the N-terminus of a particular module is KS,

AT, beta-carbonyl modifying domains (if present), ACP. The order, N→C of the beta-carbonyl modifying enzymes is DH ER KR.

Thus, each extender module of a modular PKS contains zero, one, two, or three enzymes that modify the beta-carbon of the growing polyketide chain downstream of the AT catalytic domain. A typical (non-loading) minimal Type I PKS extender module is exemplified by extender module 3 of DEBS, which contains only a KS domain, an AT domain, and an ACP domain. The next extender module, module 4, contains all three beta-carbonyl modifying enzymes. (The beta-carbonyl modifying enzymes effect such modification on the extender unit that has been added by the previous module.)

Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module migrates to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module one possesses an acyl-KS adjacent to a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading building block (elongation or extension).

After traversing the final extender module, the polyketide encounters a releasing domain that cleaves the polyketide from the PKS and typically cyclizes the polyketide. For example, final synthesis of 6-dEB is regulated by a TE domain located at the end of extender module six. In the synthesis of 6-dEB, the TE domain catalyzes cyclization of the macrolide ring by formation of an ester linkage. In FK-506, FK-520, rapamycin, and similar polyketides, the ester linkage formed by the TE activity is replaced by a linkage formed by incorporation of a pipecolate acid residue. The enzymatic activity that catalyzes this incorporation for the rapamycin enzyme is known as RapP, encoded by the *rapP* gene. The polyketide can be modified further by tailoring enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, *i.e.*, oxidation or reduction, on the polyketide core molecule. For example, 6-dEB is hydroxylated at C6 and C12 and glycosylated at C3 and C5 in the synthesis of erythromycin A.

In PKS polypeptides, the regions that encode enzymatic activities (domains) are separated by linker or "scaffold"-encoding regions. These scaffold regions encode amino acid

sequences that space the domains at the appropriate distances and in the correct order. Thus, the linker regions of a PKS protein collectively can be considered to encode a scaffold into which the various domains (and thus modules) are placed in a particular order and spatial arrangement. Generally, this organization permits PKS catalytic domains of different or identical substrate specificities to be substituted (usually at the DNA level) between PKS enzymes by various available methodologies. Thus, there is considerable flexibility in the design of new PKS enzymes with the result that known polyketides can be produced more effectively, and novel polyketides useful as pharmaceuticals or for other purposes can be made.

PCT publication WO 98/49315, the contents of which are incorporated herein by reference, describes an approach for modifying the enzymatic activities included within modules of a PKS by maintaining the scaffolding intact but replacing catalytic domains with different catalytic domains. U.S. Serial No. 09/346,860 filed 2 July 1999 and the corresponding PCT publication WO 00/01838, also filed on that date, and incorporated herein by reference describe alternative methods by altering the hypervariable region of the AT domains so as to alter the specificity for an extender unit and alteration of the KS domains to control stereochemistry. The present invention takes advantage of the approach of manipulating modules so that the catalytic activities of an entire module are placed in the appropriate sequence to construct a desired polyketide. The ability to utilize this approach depends on effecting an appropriate means for the module to incorporate a growing polyketide chain, which involves assuring that an appropriate linker region is included. Since the filing of the provisional application from which the present application claims priority, a related paper has been published by Ranganathan, A., *et al., Chem. & Biol.* (1999) 6:731-741. In this paper, intrapolypeptide linkages are fortuitously supplied to chimeric modules by including the KS region of the native downstream module in a chimera between the corresponding upstream module and the portions downstream of the KS domain in a heterologous module. Alternatively, the downstream module will include the ACP catalytic domain of the native upstream module fused to the remainder of a heterologous module upstream in the chimera.

Disclosure of the Invention

The invention is directed to an efficient method for constructing an arbitrarily chosen polyketide synthase, and therefore a desired polyketide, by manipulating entire modules of Type I polyketide synthases. The invention enables this approach by providing the modules with the appropriate "lead-in" or linker sequence to the ketosynthase (KS). Applicants have discovered that the appropriate linker between modules is required upstream of the relevant KS in order to permit the module to accept the nascent polyketide chain, and, in the case of intermolecular transfer, appropriate pairing of N-terminal and C-terminal regions assures the appropriate transfer. The nature of this linker varies depending on whether the module is covalently linked downstream from another module, or whether it forms the N-terminus of the polypeptide.

Thus, in one aspect, the invention is directed to a method to construct a functional polyketide synthase which method comprises providing each module contained in the desired polyketide synthase with an appropriate intrapolypeptide linker (RAL) when said module is downstream in the same polypeptide from a module derived from a different PKS and with an appropriate interpolypeptide linker (ERL) when the module is derived from a PKS where the module is the N-terminal module of a polypeptide. If the module at the N-terminus of a polypeptide is to accept a nascent polyketide chain from an upstream module, the interpolypeptide linker needs to include the appropriate amino acid sequence at the C-terminus of the module donating the nascent chain.

In describing a "module" being provided with linker(s) the term "module" refers to the functional portions extending approximately from the N-terminus of the KS catalytic region to the C-terminus of the ACP - *i.e.*, excludes the linker portions otherwise considered a portion of the module.

As further described below, any order of modules of desired specificity can be assured by providing the appropriate linkers either intermolecularly or intramolecularly. Thus, the polyketide synthase can be assembled from individual modules by providing the appropriate linkers to assure that the polyketide chain will be passed in the correct sequence from one module to the next and by assembling these modules either by directly providing the

polypeptides containing them or by co-expressing nucleotide sequences and coding them in a host cell.

In other aspects, the invention is directed to materials and compositions useful in carrying out the method, in particular to isolated DNA fragments which contain the appropriate intrapolypeptide and interpolypeptide linkers. The invention also relates to methods to construct functional polyketide synthases from libraries of modules and to polyketides prepared by supplying appropriate substrates to reconstructed polyketide synthases. The polyketides thus prepared can be "tailored" using either isolated enzymes or feeding the polyketides to an organism containing these enzymes to convert them to anti-infectives or compounds of other activities such as motolides by such post-polyketide modifications as hydroxylation and glycosylation. The ketides or ketolides or their modified forms can also be further derivatized using chemical synthetic methods.

Brief Description of the Drawings

Figure 1 is a diagram of the erythromycin PKS which forms 6-dEB, the core precursor of erythromycin. As shown, the PKS is comprised of three proteins, DEBS-1, DEBS-2 and DEBS-3 which are encoded by three genes, commonly called *eryAI, eryAII* and *eryAIII.*

Figure 2 shows the conversion of a diketide thioester to the triketides corresponding to those produced by DEBS-1 of the erythromycin PKS.

Figure 3 shows the structures of intrapolypeptide linkers and the N-terminal portions of interpolypeptide linkers derived from various Type I PKS.

Modes of Carrying Out the Invention

The invention takes advantage of the identification of the amino acid sequences for supplying an appropriate linker between modules of a Type I PKS depending on the position of the module in the synthetic scheme for the polyketide. If the module is at the N-terminus of the polypeptide in which it resides - *i.e.*, there is no additional module covalently bound upstream to it, an "interpolypeptide linker"(ERL) is placed upstream of the KS catalytic domain. Conversely, if the module resides in a polypeptide wherein there is an additional module upstream of it and covalently linked to it as a fusion protein, the two modules should be separated by an "intrapolypeptide linker" (RAL). If the module residing at the N-terminus

of a polypeptide is downstream in the synthesis process for a polyketide - *i.e.*, if it must accept a nascent polypeptide chain from a different module not on the same molecule, it may be necessary as well to supply a portion of the interpolypeptide linker at the C-terminus of the module providing the nascent polyketide chain in order to assure orderly transfer.

In the discussion that follows, polyketide synthases are discussed either at the protein level or the DNA level. As is well understood, manipulation of the sequence of amino acids in the polyketide synthase proteins is most conveniently done using recombinant techniques. Thus, for example, the appropriate linker sequences can be introduced to or modified with respect to those of an existing module by modifying the appropriate gene and expressing it in a suitable host. Interchange of linkers is also conveniently done in this manner. Further, modifications of amino acid sequences so as to obtain "variants" are effected by mutating the gene. The referent polyketide synthase should be understood to exist at both the protein level and nucleic acid level, and which form is being discussed should be apparent from the context.

Further, the action of polyketide synthases on their appropriate substrates can be effected either extracellularly by using isolated enzymes or may be effected by producing the enzymes intracellularly. By "appropriate substrate" is meant the extender units in their thioester forms that are recognized by the various modules in the PKS and "starter" units which are either thioesters of carboxylic acids or partially synthesized polyketides such as diketides. For example, as described in PCT application PCT/US96/11317, the ketosynthase domain of module 1 may conveniently be inactivated thus making more efficient the utilization of the diketide by module 2.

The linkers can be supplied by conventional recombinant DNA manipulations through the use of restriction enzymes and ligation procedures commonly practiced. The linkers in the PKS of the invention will be "isolated" from their natural environments. By "isolated," as used herein, is meant simply that the referent is found linked in association with moieties with which it is not normally associated, or in an environment in which it is not naturally found. It may be linked, if a nucleotide sequence to additional sequence with which it is not normally linked, or, if a peptide, to additional amino acid sequence with which it is not ordinarily

linked, or it may be simply detached from additional moieties with which it is usually associated.

As seen from Figure 3, the intrapeptide linkers (RAL) of the invention contain approximately 16-20 amino acid and typically contain a proline residue at approximately the middle of the sequence. On the other hand, the N-terminal upstream interpolypeptide linkers are approximately twice as long and appear to contain conserved acidic amino acid residues and basic amino acid residues at positions in the upstream half of the molecule. Thus, typical N-terminal upstream interpolypeptide linker (ERL) will contain an acidic amino acid within the first 3-10 residues, which is followed after 8-10 residues by a basic amino acid, and then after another 2-5 amino acid residues by an additional acidic amino acid. Additional acidic and basic residues may also occur in these linkers.

The intrapeptide linkers or interpeptide linkers shown in Figure 3 can be used as described below in the present invention or the corresponding amino acid sequences from native Type I PKS in general can be employed. In addition to the sequences that occur in nature, "variants" may be used. These variants are obtained by altering the amino acid sequence of the linker in minor ways that do not affect the ability of the linker to "feed" the nascent polyketide chain to the module in question. Typically, such "variants" are obtained from the native sequences by amino acid substitution, deletion or insertion; preferably the substitutions are "conservative" substitutions - *i.e.*, an acidic amino acid for a different acidic amino acid, a basic amino acid for a different basic amino acid, and the like. Preferably, the variants contain no more than three amino acid alterations, preferably only two, and more preferably only one.

For construction of polyketide synthases which contain more than one polypeptide, the appropriate sequence of transfers is assured by matching the appropriate C-terminal amino acid sequence of the donating module with the appropriate N-terminal amino acid sequence of the interpolypeptide linker of the accepting module. This can readily be done, for example, by selecting such pairs as they occur in native PKS. For example, two arbitrarily selected modules could be coupled using the C-terminal portion of module 4 of DEBS and the N-terminal of portion of the linking sequence for module 5 of DEBS.

In general, the method of the invention involves supplying to a module used in a PKS for synthesis of a desired polyketide with the appropriate N-terminal upstream portion interpolypeptide linker (N-ERL), C-terminal downstream portion of an interpolypeptide linker (C-ERL) or with an intrapolypeptide linker (RAL) at either terminus. As stated above, if the module is at the N-terminal portion of a polypeptide, an N-terminal upstream interpolypeptide linker should be appended at its N-terminus. If the module resides in a polypeptide where there is an additional module fused upstream from it, the two modules should be separated by an intrapolypeptide linker.

For ease of construction, a library of functional modules can be maintained to provide the appropriate desired module for construction of the PKS. One way to ensure the appropriate sequence of polyketide chain growth is to link the modules covalently, so that all but the first module will contain upstream intrapolypeptide linkers. Alternatively, and preferably, appropriate communication between functional modules non-covalently associated on separate polypeptide molecules can be achieved by providing appropriate matching between the C-terminal downstream portion of the interpolypeptide linker associated with the module contributing the nascent polyketide chain and the N-terminal upstream portion of the interpolypeptide linker placed upstream of the module which accepts and extends this nascent polyketide. Thus, an appropriate linker to ensure that the growing polyketide chain will be passed from module A to module B, which modules are not covalently bound, would be to couple, for example, the C-terminal scaffold portion of module 4 from erythromycin to module A and the N-terminal interpolypeptide linker (scaffold) portion from module 5 of the erythromycin PKS to the N-terminus of the KS of module B.

To design and construct the PKS, one straightforward approach is to utilize the existing linker regions of a native PKS, such as erythromycin PKS, and simply to "plug in" modules, for example from a library.

A library of modules derived from naturally occurring PKS which contains modules incorporating all alternative extender units used in native PKS combined with all variants of beta-carbonyl modification is not large. Extender units that are incorporated naturally include malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, and hydroxymalonyl-CoA. The appropriate native molecule for incorporation of each of these can readily be found.

Methylmalonyl-CoA extender units are incorporated, for example, by the modules of the erythromycin PKS. Certain modules of the picromycin PKS incorporate malonyl-CoA, while modules of the epothilone PKS incorporate ethylmalonyl-CoA or hydroxymalonyl-CoA. Modules occur naturally which contain the full spectrum of beta-carbonyl modifying activities; to the extent it is desirable to couple a particular beta-carbonyl modifying activity with a particular extender specificity, this can be accomplished by altering catalytic domains, *per se*, as described in the above-referenced PCT publication WO 98/49315. The complete combination of extender unit choices with all beta-carbonyl modification choices is thus only a total of 4x4 or 16 modules. As the KS unit determines the stereoselectivity of the module, accommodation can be made for various stereoisomeric forms of precursor by adjusting the KS domain in the module library. This expands the total number of modules necessary only to 32. An arbitrary number of modules can be included in a particular PKS construct, thus also determining the length of the polyketide chain and the size of the macrolide product. Of course, the macrolide product can be modified, if desired, by the known tailoring enzymes which convert naturally occurring macrolides to hydroxylated and/or glycosylated forms and the like. Such modification can be achieved in a variety of ways - by chemical modification, by *in vitro* treatment with appropriate enzymes, or by feeding the polyketides to a host organism which contains the appropriate tailoring enzymes, as is well understood in the art.

To construct the desired PKS, modules are selected from the library and provided the appropriate upstream intrapolypeptide or interpolypeptide linkers. Suitable linkers can be selected from the group consisting of those shown in Figure 3, the corresponding sequences from any Type I PKS, or can include variants of these depicted sequences which are conservative in nature and do not interfere with the ability of the linker to permit effective uptake of the nascent polypeptide chain. The linkers can be added by standard recombinant techniques to the modules in the library, or, the library can be composed of the collection of modules wherein each module has been further manipulated to include either an intrapolypeptide or interpolypeptide linker. It may be desirable, for example, to provide each type of extender module with an intrapolypeptide linker, including the possibility of retaining the linker that is normally associated with it. If the linker is placed at the N-terminus of the module, the module is suitable for covalent linking downstream of an additional module in a

single polypeptide. If the intrapolypeptide linker is at the C-terminus, ordinarily that module will be placed and linked covalently upstream of an additional module. In any case, a module may arbitrarily be provided with an intrapolypeptide linker (RAL) at either its N- or C-terminus depending on where it is ultimately to be placed in the PKS to be constructed or may be provided with the N-terminal upstream portion of an interpolypeptide linker (N-ERL) if it is to be placed at the N-terminus of a polypeptide in the PKS, or with a C-terminal portion of an interpolypeptide linker (C-ERL) if it is to be placed at the C-terminus of a polypeptide and is intended to transfer a nascent polyketide chain to a subsequent module.

The various modules, with appropriate linkers are then assembled into the desired polyketide synthase. As stated above, the construction of the PKS can be based on plugging in active portions of modules into an existing linker array The assembly can be performed by simply mixing the peptides containing the modules or may be generated recombinantly from expression constructs in a host cell. The cell may provide the appropriate substrates for the PKS, or the substrates may need to be provided to the reaction mixture containing the polypeptides or to the cells in which they are generated. Depending on the choice of host, provision may need to be made for providing these substrates.

In this way, the modules can be "mixed and matched" as desired to construct a polyketide product from the desired extender units and with the desired beta-carbonyl modification, choosing the linkers in accordance with the position of the module in a polypeptide, and the number of modules cam be altered as desired.

A preferred starter unit for such an assembly of modules is a diketide thioester either formed *in situ* by including a module which contains a loading domain to incorporate a starter unit along with an extender unit to attain this resultant, or the diketide may be synthesized independently and used as the substrate for the PKS. The synthesized diketide may be supplied as the thioester, such as the N-acylcysteamine thioesters. Preparation methods for these thioesters are described in the above-referenced U.S. Serial No. 09/346,860 filed 2 July 1999 and the corresponding PCT application, as well as U.S. Serial No. (Atty. docket No. 30062-20032.00) filed 27 January 2000.

Using the techniques of the invention, it is thus possible to manipulate entire modules and effect efficient cross-talk so as to assure production of the desired macrolide. Such

techniques can be used, for example, to alter the structure of macrolide anti-infectives by, for example, replacing the module 2 of the erythromycin gene cluster with module 2 of the tylosin gene cluster, or replacing the erythromycin module 6 (along with its thioester sequence) with the corresponding module 6 from narbomycin.

In addition, 14-membered macrolides could be expanded to become 16-membered macrolides by fusing modules 2-3 of the tylosin, spiramycin or niddamycin modules 2-3 between modules 1 and 3 of the erythromycin synthase or by adding any arbitrarily chosen module from other Type I PKS clusters into the synthase for production of erythromycin. Alternatively, modules 1-2 of erythromycin could be deleted and replaced by modules 1-3 of tylosin, spiramycin or niddamycin.

In addition, new substituents can be introduced into, for example, PKS erythromycin or its precursors by replacing the second module of the erythromycin PKS with module 5 from tylosin PKS where the substituted module has the enoyl reductase catalytic activity inactivated. This results in erythromycins substituted with an ethyl group at the 10-position. Alternatively, erythromycin module 5 could be replaced by the spiramycin module 6 to obtain 5-desmethyl-4-OH erythromycins.

Improved forms of FK-506 are obtained by replacing rapamycin modules 2-10 with FK-506 modules 2-6, or by replacing rapamycin modules 2-11 with FK-506 modules 2-7 or by replacing rapamycin modules 2-12 with FK-506 modules 2-8 or by replacing rapamycin modules 11-14 with FK-506 modules 7-10. Any combination or subset of the above could also be employed. Improved forms of FK-520 can be made in a similar manner. An alternative form of rapamycin is synthesized by substituting the FK-506/520 module 1 for rapamycin module 1.

The foregoing are merely exemplary of the types of manipulations that could be employed. The polyketides, obtained by supplying the appropriate substrates either *in vitro* or *in vivo*, may then be further modified if desired by hydroxylation, glycosylation and the like to obtain desired products. Further, chemical synthetic manipulations may also be employed.

Some of the resulting compounds described above could be prepared by alternative techniques previously disclosed, for example, in PCT applications PCT/US99/22886 or

PCT/US99/24483. However, the procedure described above, which manipulates entire modules, may result in better yield or more convenient synthesis.

The following examples are intended to illustrate but not to limit the invention.

Preparation A

Construction of Single Module Based Systems

Single Module Gene Constructs

Single module constructs from the DEBS gene cluster were prepared for modules 2, 3, 5 and 6 as follows. The TE domain is fused to the module to facilitate termination. The (M3+TE) gene was prepared from the tri-modular construct pKAO318 (McDaniel, R., *et al., Chem. Biol.* (1997) 4:667) having an *NheI* site engineered at the start of the DEBS-2 gene. Fusion of the TE gene at the end of ACP3 was described in connection with the construction of pCK13 in Cortes, J., *et al., Science* (1995) 268:1487; Kao, C.M., *et al., J. Am. Chem. Soc.* (1995) 117:9105 and Kao, C.M., *et al., J. Am. Chem. Soc.* (1996) 118:9184, collectively cited below as the "Cortes-Kao documents." The *NheI-EcoRI* fragment was cloned into pET 21c (Novagen) to construct pRSG34. The *EcoRI* site was used to delete the stop codon of the TE domain so that the protein could be overproduced as a carboxy terminal $(His)_6$ tagged fusion protein.

(M5+TE) was constructed by combining the engineered *NdeI* site from pJRJ10 (Jacobsen, *et al., Biochem* (19___) 37:4928) with the *EcoRI* site from pCK15 (Cortes-Kao documents). The *Nde-EcoRI* fragment was cloned in pET21c to obtain the expression plasmid pRSG46. Expression constructs for (M2+TE) and (M6+TE) were prepared similarly using an engineered *Nhe* site immediately upstream of the corresponding KS (at position 7570, 5′-GCTAGCGAGCCGATC-3′ and at position 28710, 5′-GCTAGCGACCCGATC-3′).

These constructs were expressed in *E. coli* BL21 (DE3) along with an expression system for *sfp* phosphopantetheinyl transferase from *B. subtilis*. The co-expression is described by Lambalot, R.H., *et al., Chem. Biol.* (1996) 3:923. For the construction of the *sfp* gene, the *NdeI-HindIII* fragment derived from the pUC8-*sfp* (Nakano, *et al., Mal. Gen. Genet.* (19___) 232:313) was cloned into pET28 which has a kanamycin resistance gene to give

resultant plasmid pRSG56. The resulting proteins were then isolated for use in the reaction mixtures described in the Examples below.

In more detail, the expression was induced with 1 mM isopropyl-b-D-thiogalactopyranoside, and the cells were harvested by centrifugation after 10 hours and resuspended in disruption buffer, 200 mM sodium phosphate pH 7.2, 200 mM sodium chloride, 2.5 mM dithiothreitol, 2.5 mM sodium ethylenediamine tetra-acetate (EDTA), 1.5 mM benzamidine, 2 mg/L pepstatin and leupeptin and 30% v/v glycerol. The cells were lysed by passing through a french press, and the lysate was collected after centrifugation. Nucleic acids were precipitated with polyethylenimine (0.15%) and removed via centrifugation. The supernatant was made 50% (w/v) saturated with ammonium sulfate and precipitated overnight. After centrifugation, the pellet containing protein was redissolved in buffer A (100 mM sodium phosphate pH 7.2, 2.5 mM DTT, 2 mM EDTA and 20% glycerol (v/v)) and stored at -80°C. For chromatography, the buffer was exchanged to buffer A + 1 M ammonium sulfate using a gel filtration PD10 (Pharmacia) column. The resulting sample was loaded on a Butyl Sepharose (Pharmacia) column. Fractions containing DEBS proteins were pooled and applied on an anion exchange column (Resource Q; 6 mL, Pharmacia). Purified protein fractions were pooled and concentrated using Amicon centriprep30. Typical purified protein yields were ~ 3-4 mg/liter of culture. Greater than 90% of proteins were phosphopantetheinylated *in vivo* as a result of the overexpression of *sfp* phosphopantetheinyl transferase. Although the proteins were expressed as $(His)_6$-tagged proteins, they did not bind to a Ni-column under experimental conditions. It is unclear whether this inability to bind to a Ni-agarose column is due to steric effects or if the $(His)_6$ peptide was lost during purification.

Example 1

Requirements for Cell-Free Synthesis of Triketides by Individual Modules - Identification of Linker Regions

A cell-free system, tested for the ability to convert the cysteamine thioester of 2S,3R-2-methyl-3-hydroxypentanoic acid (compound 2 in Figure 2), consisted of 0.5-10 mM concentration of the thioester, 2.5 mM ^{14}C-labeled methylmalonyl CoA and 100 pmoles of

purified protein prepared in Example 1 in a 100 µl reaction. In some cases, 1 mM of NADPH was added; this was done in assay mixtures containing (M2+TE), (M5+TE) and (M6+TE) proteins. The protein was, in each case, a single module of the DEBS PKS fused to the thioesterase (TE) termination region of module 6.

The reaction mixtures were quenched and extracted by ethyl acetate and separated by thin-layer chromatography (TLC) to discern the formation of the triketide ketolactone 3 and triketide lactone 4 (both shown in Figure 2) after 30 minutes. Results are shown as the first four entries in Table 1.

Table 1
Formation of Triketides by Single Module Constructs

	Construct	Plasmid	Triketide Formed
1.	M3+TE	pRSG34	Yes
2.	M5+TE	pRSG46	Yes
3.	M2+TE	-	No
4.	M6+TE	-	No
5.	ERL-M2+TE	pRSG64	Yes
6.	ERL-M6+TE	pRSG54	Yes
7.	M1-RAL-M6-TE	pST96	Yes
8.	M1-RAL-M3-TE	pST97	Yes
9.	*ery* M1-RAL-*rif* M5+TE	pST110	Yes
10.	(*ery* M1-RAL *rif* M5+ERL)+DEBS-2+DEBS-3	pST113	Yes (dEB6)

As seen in Table 1, although the expected triketides were formed from (M3+TE) and (M5+TE) (modules which reside at the upstream portion of their respective polypeptides), no triketides were formed from (M2+TE) or (M6+TE), (modules which reside at the C-terminal portions of their polypeptides). These latter results were unexpected since the diketide can be incorporated by module 2 when it is supplied as a part of the complete polypeptide DEBS-1. It was verified that the ACP domain was pantetheinylated in modules 2 and 6, and that for (M2+TE), the KS domain could not be acylated with radiolabeled diketide.

Example 2

Modification of Single Modules with Linker Sequences

The constructs for (M2+TE) and (M6+TE) were modified by deleting the sequences encoding the amino acids upstream of the KS catalytic domain and substituting the first 39 amino acids from (M5+TE) containing the N-terminal portion of the interpolypeptide linker (N-ERL). The relevant constructs were prepared by replacing the *BsaBI-EcoRI* fragment in pRSG46 by the corresponding fragment from pCK4 to obtain (N-ERL-M2+TE), in plasmid pRSG64, or from pJRJ10 to obtain (N-ERL-M6+TE) in plasmid pRSG54. These constructs yield modules which contain the upstream 39 amino acids from module 5. The constructs were expressed in *E. coli* and proteins obtained as described in Preparation A. These proteins were able to produce the triketide product from diketide in the cell free system of Example 1, as shown in entries 5 and 6 in Table 1.

The various constructs which are successful in converting diketide to triketide were then evaluated for the kinetic constants k_{cat} and K_M. These results are shown in Table 2. As shown in Table 2, the results are quite similar for all constructs except that the results from module 3 show a several-fold decrease in k_{cat} as compared to the other modules. This is evidently due to the absence of beta-carbonyl modifying enzymes in module 3 as verified by the fact that removal of NADPH, (which is required for the activity of such modules) from the reaction mixture of (N-ERL-M6+TE) also results in a lowering of the k_{cat}.

Table 2

Proteins	**k_{cat} X 100 (min^{-1})**	**K_M (mM)**
(N-ERL-M2+TE)	8 ± 0.6	4.6 ± 0.4
(M3+TE)	1.5 ± 0.3	4.4 ± 0.4
(M5+TE)	7.5 ± 0.7	4.7 ± 0.4
(N-ERL-M6+TE)	9.5 ± 0.6	4.3 ± 0.4
(N-ERL-M6+TE) (- NADPH)	4.5 ± 0.7	4.1 ± 0.4

It is apparent from these results that the presence of the N-terminal upstream sequence associated with modules located at the N-terminal portion of the polypeptide is essential for permitting a module in this position to incorporate the growing polyketide chain.

Example 3

Construction of (M1-RAL-M3+TE) and (M1-RAL-M6+TE)

The *BsaBI-EcoRI* fragments containing modules 3 and 6 respectively were cloned behind the M1 module which contains the intrapolypeptide linker (RAL) that natively resides between M1 and M2. The resulting M1-RAL-M3+TE and M1-RAL-M6+TE genes were then excised as *PacI-EcoRI* fragments and inserted into pCK12 resulting in plasmids pST97 and pST96 respectively. The corresponding proteins were produced by transformation into *S. coelicolor* CH999. The resulting strains of *S. coelicolor* were able to incorporate the diketide thioester into the triketide as shown by entries 7 and 8 in Table 1. (The triketide produced is the ketolactone 3 in Figure 2.)

Example 4

Additional Intrapolypeptide Mediated Transfer

A construct wherein the first module of the DEBS PKS cluster (*ery*), which contains the intrapolypeptide linker of the corresponding M1-M2 polypeptide from the erythromycin PKS, is fused to the fifth module of the rifamycin PKS (*rif*) was constructed by replacing the natural sequence at 28024 of *rif* ACP5 (5′-CGCGAC-3′) with the *SpeI* recognition sequence 5′-ACTAGT-3′. The *BsaBI-SpeI* fragment containing *rif* M5 was excised and replaced the corresponding *ery* M1-RAL- fragment in pCK12 to obtain plasmid pST110. This plasmid, containing *ery* M1-RAL-*rif*-M5+TE was transformed into *S. coelicolor* CH999 and the resulting strain was able to incorporate the diketide into the triketide lactone as shown by entry 9 in Table 1. The amount is comparable to that produced in this strain transformed with DEBS-1+TE.

Example 5

Construction of Modules for Intermolecular Transfer

The *PacI-SpeI* fragment of pST110 was inserted into a derivative of pCK7 (Kao, C.M., *et al., Science* (1994) 265:509) which had an *SpeI* site engineered at the beginning of the scaffolding sequence at the carboxy terminus of the polypeptide downstream of ACP2. The resulting pST113 construct still contains *ery* M1 linked to *rif* M5 via the natural intrapolypeptide linker between *ery* molecules 1 and 2, and also now contains *rif* M5 covalently linked to the downstream C-terminal portion of the ERL derived from *ery* M2. Thus, the complete ERL between the polypeptide generated by pST113 and the protein generated by a construct which generates DEBS-2 would correspond to the native ERL in the *ery* PKS - *i.e.*, *rif* M5 would be associated with *ery* M3 via the natural interpolypeptide linker between *ery* molecules 2 and 3. Co-transformation into *S. coelicolor* of pST113 along with constructs that produce DEBS-2 and DEBS-3 results in the production of 6-dEB, as shown by entry 10 of Table 1.

Claims

1. A method to prepare a desired polyketide synthase from individual modules which method comprises

providing successive covalently linked modules that comprise an intrapolypeptide linker (RAL) and successively non-covalently linked modules that comprise interpolypeptide linkers (ERL), so as to facilitate the transfer of a nascent polyketide chain from the a module of said PKS to a succeeding module of said PKS; and

assembling said modules.

2. The method of claim 1 wherein said assembling is by incubating the polypeptides which comprise said modules in a reaction mixture.

3. The method of claim 1 wherein said assembling is by expressing constructs which encode said modules in a host cell.

4. The method of claim 1 wherein at least some of the modules in the desired polyketide synthase are derived from a library of Type I PKS modules and wherein at least one module in said polyketide synthase is heterologous with respect to the remaining modules.

5. The method of claim 1 wherein each RAL has an amino acid sequence selected from the group consisting of those set forth in Figure 3 or a variant thereof and the N-terminal portions of each ERL has the amino acid sequence set forth in Figure 3 or a variant thereof.

6. A polyketide synthase prepared by the method of claim 4.

7. The polyketide synthase of claim 6 which contains erythromycin modules 1, and 3-6 and tylosin module 2, and wherein said polyketide chain is transferred from *ery* module 1 to *tyl* module 2 and then to *ery* modules 3-6.

8. The polyketide synthase of claim 6 which contains erythromycin modules 1-5 and narbomycin module 6, wherein said polyketide chain is passed from *ery* modules 1-5 to *nar* module 6.

9. The polyketide synthase of claim 6 which contains modules 1 and 3-6 of erythromycin and modules 2-3 of tylosin, spiramycin or niddamycin, wherein said polyketide chain is passed from *ery* module 1 to modules 2-3 of tylosin, spiramycin or niddamycin and then to *ery* modules 3-6.

10. The polyketide synthase of claim 6 which contains modules 1-3 of tylosin, spiramycin or niddamycin and modules 3-6 of erythromycin, and wherein said polyketide chain is passed from modules 1-3 of said tylosin, spiramycin or niddamycin to *ery* modules 3-6.

11. The polyketide synthase of claim 6 which contains a module of tylosin, spiramycin or niddamycin and modules 1-2 and 3-6 of erythromycin, wherein said polyketide chain is passed from *ery* modules 1-2 to the tylosin, spiramycin or niddamycin module and then to *ery* modules 3-6.

12. The polyketide synthase of claim 6 which contains modules 1 and 3-6 of erythromycin and module 5 of tylosin, spiramycin or niddamycin having the enoyl reductase catalytic activity inactivated, wherein said polyketide chain is passed from *ery* module 1 to module 5 of tylosin, spiramycin or niddamycin and then to *ery* modules 3-6.

13. The polyketide synthase of claim 6 which contains erythromycin modules 1-4 and 6 and module 6 of spiramycin or niddamycin, wherein said polyketide chain is passed from *ery* modules 1-4 to module 6 of spiramycin or niddamycin and then to *ery* module 6.

14. The polyketide synthase of claim 6 which contains module 1 of FK-506/520 and modules 2-14 of rapamycin, wherein said polyketide chain is passed from module 1 of FK-506/520 and then to modules 2-14 of rapamycin.

15. The polyketide synthase of claim 6 which contains module 1 and 11-14 of rapamycin and modules 2-6 of FK-506/520 wherein said polyketide chain is passed from module 1 of rapamycin to modules 2-6 of FK-506/520 and then to modules 11-14 of rapamycin.

16. The polyketide synthase of claim 6 which contains module 1 of rapamycin, modules 2-7 of FK-506/520 and modules 12-14 of rapamycin, wherein said polyketide chain is passed from module 1 of rapamycin to modules 2-7 of FK-506/520 and then to modules 12-14 of rapamycin.

17. The polyketide synthase of claim 6 which contains module 1 of rapamycin, modules 2-8 of FK-506/520 and modules 13-14 of rapamycin, wherein said polyketide chain is passed from module 1 of rapamycin to modules 2-8 of FK-506/520 and then to modules 13-14 of rapamycin.

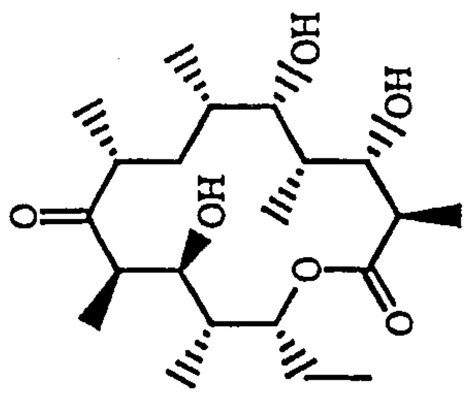
18. The polyketide synthase of claim 6 which contains modules 1-10 of rapamycin and modules 7-10 of FK-506/520, wherein said polyketide chain is passed from modules 1-10 of rapamycin to modules 7-10 of FK-506/520.

19. A method to prepare a desired polyketide which method comprises incubating required substrates with the polyketide synthase of claim 6.

20. The method of claim 19 wherein the substrates comprise a diketide thioester and thioesters of the required extender units.

21. The method of claim 20 wherein the extender units are malonyl, methylmalonyl, ethylmalonyl or hydroxymalonyl thioesters.

22. A polyketide or polyketide derivative synthesized by a method comprising the method of claim 19.



1

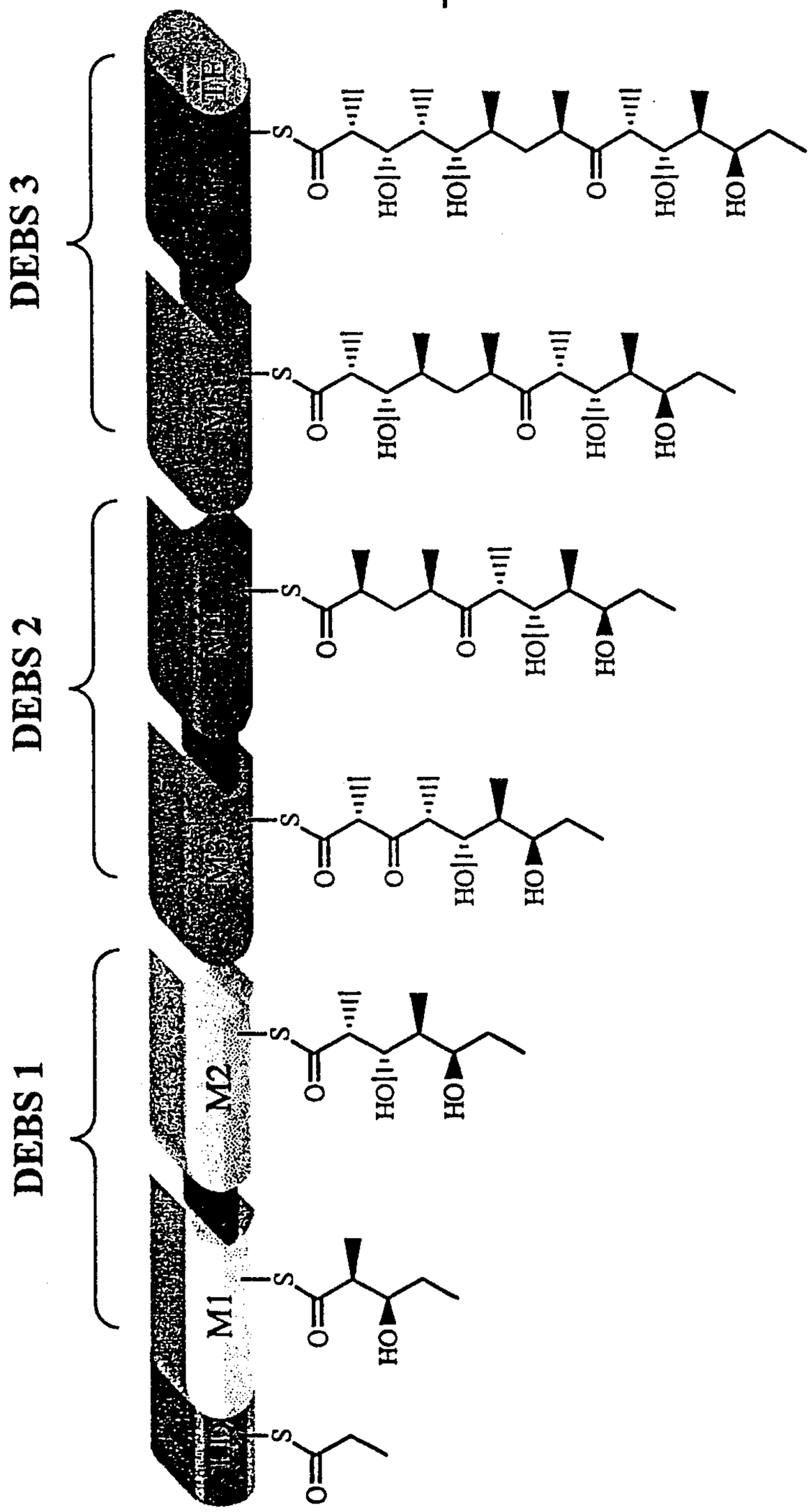


Figure 1

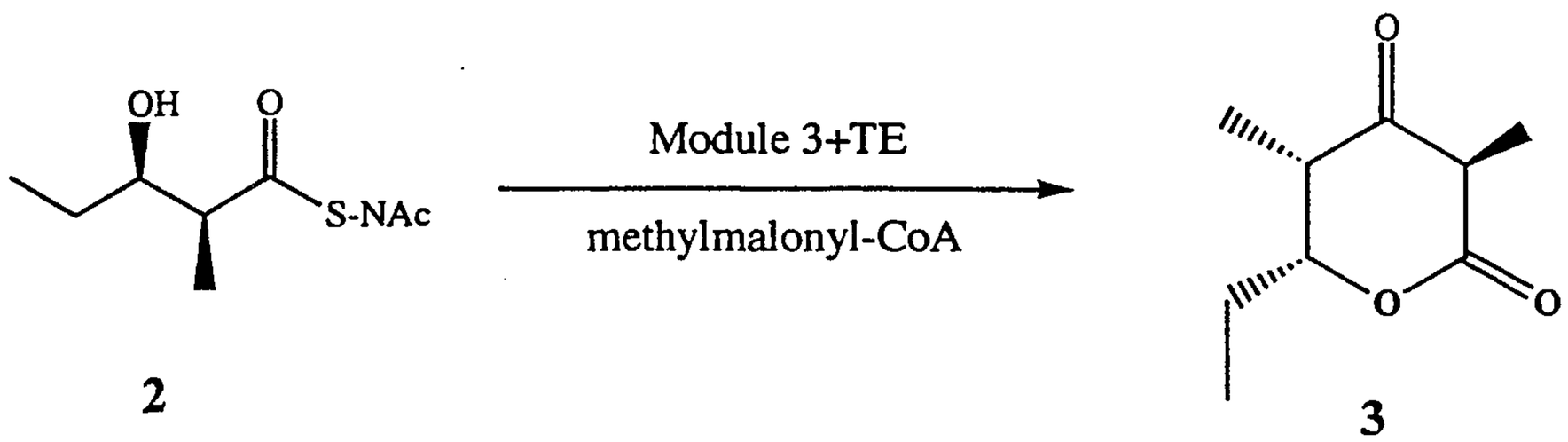

OH
O
S-NAc
Module 3+TE
methylmalonyl-CoA
O
O
O
2
3
2
Module 2+TE or Module 5+TE or Module6+TE
NADPH
NADP+
methylmalonyl-CoA
OH
O
O
4

Figure 2

(a) INTRA-POLYPEPTIDE LINKER

M2ery: GGATGAEQAAPATT..APVD

M4ery: VGDAD..QAA.VRVVGAA.DES

M6ery: VGAAEAEQA.PALVREVPKDAD

M2rif: FGSA.A.NR.PAEIGTAAAE

M3rif: LG..ER.PAAPAPVTRDVSD

M5rif: GETVAGAPATPVTTVADAG

M3rap: .ELFTGENPAPVRGPVSAVGQD

M4rap: .ELFTGENPAPVRGPVSVVGQD

M7rap: .ELFTGENPAPVRGPVSA.GQD

(b) N-TERMINAL INTER-POLYPETIDE LINKER

M3ery: .....VTD SE KVAEYLRR .ATLDLRAAR QRIRE..LES

M5ery: MSGDNGM.TE E.KLRRYLKR TVT.ELDSVT ARLRE..VEH RAG

M4rif: .....MSAPNE QIVDAL.R ASLKE....N VRLQQENSAL AAAAA

M7rif: .....VSASYE KVVEAL.R KSLEE....V GTLKKRNRQL ADAAG

M8rif: .....V.AD EGQLRDYLKR .AIADARDAR TRLRE..VEE QAR

M9rif: .....MATD E.KLLKYLKR .VTAELHS.. ..LRKQGARH .AD

M5rap: .....MR.. EDQLLDAL.R KSVKE....N ARLRKANTSL RAAMD

M11rap: .....M.PEQD KVVEYL.R WATAELHTTR AKL.....EA LAAANT

Figure 3