Molecular Engineering of Bacterial Natural Product Biosynthetic Pathways via Red/ET Recombineering

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von

Xiaoying Bian

Saarbrücken

2012

Tag des Kolloquiums: Feb 08, 2013

Dekan: Prof. Dr. Volkhard Helms

Berichterstatter: Prof. Dr. Rolf Müller

Prof. Dr. Jörg Walter

Vorsitz: Prof. Dr. Alexandra K. Kiemer

Akad. Mitarbeiter: Dr. Qingzhong Hu

ACKNOWLEDGEMENTS

It is a pleasure to thank those individuals who made this thesis possible in the last four years.

First of all, I would like to thank my supervisor Prof. Dr. Rolf Müller. I am extremely thankful to Prof. Müller who gave me the opportunity to work in the truly remarkable research environment in the Department of Pharmaceutical Biotechnology at Saarland University, supported me throughout my thesis work, and reviewed my thesis thoroughly. I learned so much from his research ability and attitude which will benefit my whole research career in my life.

I thank Prof. Dr. Jörn Walter for his kind suggestions during PhD studies.

I owe my deepest gratitude to my co-advisor Dr. Youming Zhang from Gene Bridges GmbH who took me into the fantastic world of Red/ET recombineering and tought me a lot of skills about molecular biology. His superb skills and interesting ideas/designs always enlighten me in the scientific research. He also provided so many advices and guidance for my research and career. I also would like to show my gratitude to him and Dr. Youhua Deng who gave me a lot of help for my life in Germany.

I would like to thank Prof. A Francis Stewart, and Dr. Jun Fu from BioInnovationsZentrum, Technische Universität Dresden. Over the past several years, we cooperated for several projects, and had very pleasant and successful team works. Theirs ambition for techniques always inspires me and I also thank theirs recombineering reagents and suggestions.

I am grateful to Dr. Alberto Plaza from Helmholtz Institute for Pharmaceutical Research Saarland who directed me in the isolation and structural elucidation of natural products. I also thank his excellent craft of NMR measurement.

I extremely thankful to Dr. Silke C. Wenzel, Liujie Huo, Hilda Sucipto, Ying Tang in our group, and Sarah Ongley from University of New South Wales, Sydney, Australia, for their kind and careful proof-reading of the manuscript of this thesis.

I am indebted to all my colleagues who support me, discuss with me, and help me during the last four years in Prof. Müller's group. I specially wish to thank Dr. Silke C. Wenzel for her extremely useful suggestions of my research, Thomas Hoffmann, Thorsten Klefisch, and Eva Luxenburger for LC-MS measurement, Jennifer Herrmann for bioactive activity test, Liujie Huo for protein expression and translation of summary, Dr. Daniel Krug for LC/MS analysis

and computer assistance, and secretaries Birgitta Lelarge, Natja Mellendorf and Claudia Thiele for their many kind helps. I also own thanks to my Chinese colleagues Dr. Yi Chai, Dr. Yanyan Li, Dr. Wei Ding, Dr. Shengbiao Hu, Dr. Yixin Ou, Dr. Shiping Shan, Dr. Chengzhang Fu, Liujie Huo, Qiang Tu and Ying Tang for all their support in various ways as both scientific collaborators and friends.

I want to acknowledge Chinese Scholarship Council (CSC) and Northwest A&F University (NWAF, China) for the PhD scholarship. I also thank Prof. Wenjun Wu, Prof. Zhaonong Hu and Mr. Zhenming Wang from NWAF for their supports and encouragements.

Last but not least, my heartfelt thanks to my parents and my sisters for their support and sacrifices throughout all the years of this thesis. I specially thank my girlfriend Xiangmei Ren for her indefatigable support and suggestions for my thesis.

Xiaoying Bian

Saarbrücken, Oct, 2012

Vorveröffentlichungen der Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultät III, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung:

Fu, J.*; **Bian, X.***; Hu S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F.; Zhang, Y. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.*, **2012,** *30*, 440-446.
(* Authors contributed equally to this work.)

Bian, X.; Plaza, A.; Zhang, Y.; Müller, R. Luminmycins A-C, cryptic natural products from *Photorhabdus luminescens* identified by heterologous expression in *Escherichia coli. J. Nat. Prod.*, **2012**, *75*, 1652-1655.

Wang, H.; **Bian, X.**; Xia, L.; Ding, X.; Müller, R.; Zhang, Y.; Fu, J.; Stewart, A. F. Efficient and seamless DNA recombineering using *ccdB* for counterselection in *Escherichia coli. Nucleic. Acids Res.*, **2012**, *submitted.*

Bian, X.; Wenzel, S. C.; Plaza, A.; Zhang, Y.; Müller, R. Efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives *in vivo. ChemBioChem*, **2012**, *to be submitted.*

Bian, X.; Huang, F.; Stewart, A. F.; Xia, L.; Zhang, Y.; Müller, R. Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering. *ChemBioChem*, **2012**, *13*, 1946-1952.

Summary

Bacterial genome-sequencing projects have revealed that a large number of natural product biosynthetic pathways presented in genomes are cryptic. Cloning, engineering and expression of a cryptic biosynthetic pathway in a well-characterized heterologous host to discover and optimize its product is an effective alternate to traditional approaches.

We developed RecE/RecT mediated linear-linear homologous recombination (LLHR) for direct cloning of ten unknown NRPS/PKS gene clusters from the genomic DNA of *Photorhabdus luminescens* into expression vectors in *E. coli*. Coupled with robust heterologous expression in *E. coli*, the products of *plu3263* and *plu1881-plu1877*, luminmides A-G and luminmycins A-C, were determined, respectively.

By site-directed modification of the third adenylation domain of NRPS Plu3263, we efficiently created mutants using *ccdB* counterselection recombineering, which led to a significant alteration of relative yields of luminmides A and B.

Partial syringolin biosynthetic pathway (*sylCDE*) from *Pseudomonas syringae* was cloned by LLHR and expressed in *E. coli* resulting in the identification of two new syringolins. Intact syringolin pathway was reassembled by addition of *sylAB* and engineering of promoter via Red/ET recombineering. The varying production distribution of syringolins showed the different efficiencies of native and synthetic promoters in *E. coli*.

Zusammenfassung

Bakterielle Genom-Sequenzierungsprojekte zeigten, dass zahlreiche im Genom vorhandene Sekundärmetabolit Biosynthesewege kryptisch sind. Um die Produkte zu entdecken und optimieren, ist Klonierung, Engineering und Expression von diesen Biosynthesewegen in einem gut charakterisierten heterologen System eine effektive Alternative zu traditionellen Ansätzen.

Wir entwickelten die von RecE/RecT vermittelten linear-linear homologen Rekombination (LLHR) für die direkte Klonierung von zehn unbekannten NRPS/PKS-Genen aus Genom von *Photorhabdus luminescens* in *E. coli*-Expressionsvektoren. Mittels robuster heterologer Expression in *E. coli* wurden die Produkte von *plu3263* und *plu1881-plu1877*, Luminmide A-G und Luminmycin A-C, jeweils identifiziert.

Die "site-directed" Modifikation der dritten Adenylierungsdomaine von NRPS Plu3263 wurde effizient durch *ccdB* Conter-selektion Rekombination erstellt. Das führte dazu, dass sich die relative Ausbeute von Luminmides A und B stark veränderten.

Der partielle Syringolin Biosyntheseweg (*sylCDE*) aus *Pseudomonas syringae* wurde durch LLHR kloniert und die Expression in *E. coli* versursachte die Identifizierung von zwei neuen Syringolinen. Der intakte Syringolin Biosynthesesweg wurde durch Zugabe von *sylAB* und Promotor-Engineering via Red/ET-Rekombination wieder zusammengesetzt. Die variierende Verteilung der Produktion aller Syringolin-derivate zeigte die unterschiedliche Effizienzen zwischen nativen und synthetischen Promotoren in *E. coli*.

Table of contents

A.Introduction

1 The impact of natural products on drug discovery

Nature has evolved over a long time to generate a huge number of natural products with surprising novelty, diversity and bioactivity.[1-3] These molecules have played an important role across the world in treating and preventing human diseases and agricultural pests for thousands of years.[4,5] Since the middle of last century, natural products have formed a central pillar of the modern pharmaceutical industry.[6] From 1981 to 2010, 1355 new chemical entities (small-molecule drugs) were introduced into clinical use, and 64% of them are of natural-product origin or inspired derivatives, with the number rising to 68.4% and 74.9% when limited to drugs as anti-infective and anticancer molecules, respectively.[7] The discovery and development of antibiotics (e.g., penicillin, tetracycline, erythromycin, vancomycin), antiparasitics (e.g., avermectin), antimalarials (e.g., quinine, artemisinin), lipid control agents (e.g., lovastatin and analogs), immunosuppressants for organ transplants (e.g., cyclosporine, rapamycins) and anticancer drugs (e.g., taxol, doxorubicin, epothilone) based on natural products revolutionized medicine, and extended people's life expectancy (Figure A-1).[8] In addition, natural product-based environment friendly pesticides, such as fungicides (e.g., strobilurins, kasugamycin), insecticides and miticides (e.g., avermectin, spinosyns, pyrethroids, neonicotinoids), and herbicides (e.g., glufosinate, sulcotrione), reinforced fungal, insect, and weed management in agriculture and forest, and improved people's life quality.[5]

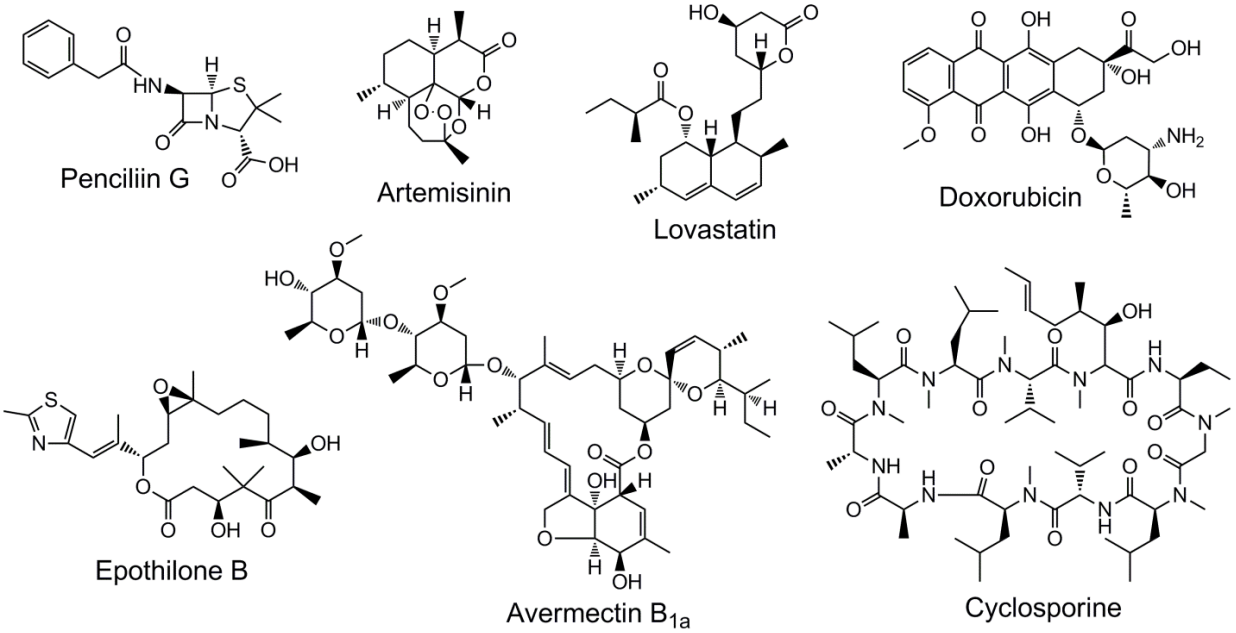


Figure A-1. Examples for important natural compounds in clinical and agricultural applications.

This fact, allied with the lack of marketed or late phase clinical candidates from combinatorial chemistry, increases the confidence that natural products will continue to fulfill their role as a major source of new antibiotics or anticancer drugs for the predictable future.[9,10]

Compared to plants and animals, microorganisms are the dominant sources of natural products in clinical and agricultural use because they are often renewable, thus opening up the possibility to scale-up production.[11] In addition, the increased understanding of microbial secondary metabolite biosynthesis and regulation coupled with advances in molecular genetics drive us to focus on the microbial natural products′ discovery and optimization at the post-genomic era.

2 Biosynthetic logics of microbial natural products

Microbial natural products are chemical compounds isolated from microbes. These compounds may derive from microbial primary or rather secondary metabolism. The primary metabolites (polysaccharides, proteins, nucleic and fatty acids) are common in all biological systems. While secondary metabolites are low molecular (MW<3000), they represent chemically and taxonomically extremely diverse compounds and their biosynthesis starts from some primary metabolites or from intermediates of the primary metabolism. Microbial secondary metabolites are mainly classified as polyketides, peptides, oligosaccharides, terpenoids and alkaloids.[12] But the majority of bacterial secondary metabolites isolated so far are polyketides and nonribosomally biosynthesized peptides, or hybrids thereof, which are biosynthesized by large molecular assembly lines composed of polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) multienzymes.[13-15] It should, however, be noted that many bacterial natural products are derived from other pathways, such as isoprenoid or shikimate pathways.[16-18]

PKSs are multifunctional and multimodular enzymes; each module contains a set of distinct, non-iteratively acting activities responsible for several biosynthetic steps during the catalysis of one cycle of the polyketide chain elongation.[19] The basic domain equipment of a typical PKS module includes: the C–C bond-forming ketosynthase (KS) domain; the acyltransferase (AT) domain, which introduces malonyl or methylmalonyl units during each cycle of chain elongation; and the acyl carrier protein (ACP) domain, also known as a thiolation (T) domain where the acyl chain is assembled and elongated. Furthermore, a variety of optional domains may be present for modification of the polyketide intermediate, including ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains. The last

module of the PKS assembly line usually harbors a thioesterase (TE) domain, which in charge of the release of the biosynthetic intermediate in linear, cyclic or branched cyclic forms (Figure A-2a).[14]

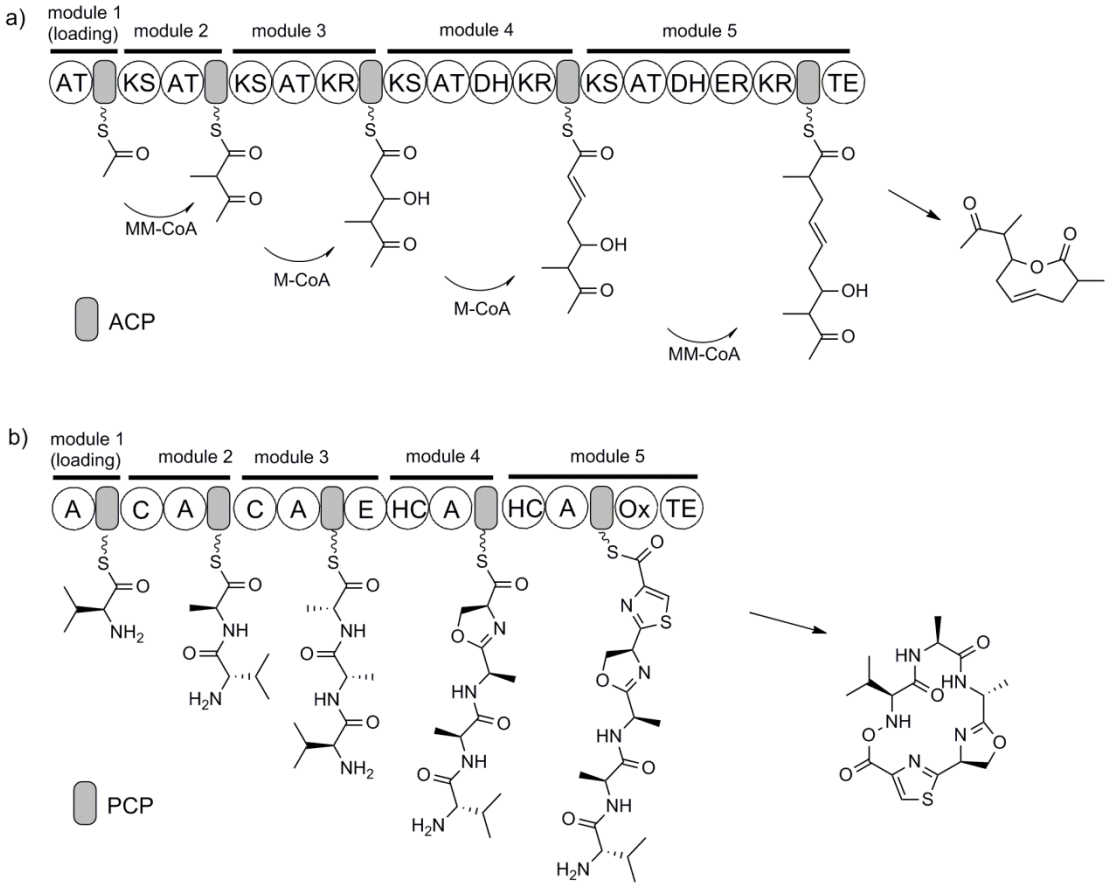


Figure A-2. Biosynthetic logics of bacterial PKS and NRPSs. a) Generalized example of a modular PKS consisting of non-iteratively acting domains; acyltransferase (AT), acyl carrier protein (ACP), ketosynthase (KS), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), thioesterase(TE); the specificity of the AT domain for malonyl-CoA (M-CoA) or methylmalonyl-CoA (MM-CoA) is indicated. b) Generalized example of NRPS biosynthesis; adenylation (A), peptidyl carrier protein (T), condensation (C), epimerization (E), heterocyclization (HC), oxidation (Ox). The substrates for NRPS modules are Val, Ala, Ala, Thr, and Cys, respectively.

NRPSs are similar to the PKSs; they are also multifunctional enzymes that are organized into modules for the incorporation of one amino acid per module.[20] NRPS elongation minimal module consists of an adenylation (A) domain, a condensation (C) domain and a peptidyl carrier protein (PCP), for incorporation of one amino acid into the polypeptide intermediate. The A domains select and activate the amino acids as substrates and thereby act as gatekeepers for specificity of NRPS modules. On the basis of the crystal structure of PheA,[21] a Phe-activating A-domain from gramicidin S synthetase, combined with phylogenetic analysis and biochemical studies, 10 active-site residues, designed as 'specificity-conferring code' or 'non-ribosomal code', that are responsible for binding the amino acid substrates were proposed.[22-24] This code allows us to predict the substrate

specificity of NRPS from their primary protein sequences.[25-27] Furthermore, by changing as few as one of these residues, it has been shown possible to switch the specificity from one amino acid to another *in vitro* and *in vivo*, even to a non-natural amino acid.[22,28-35] The C domains is similar to KS domains in PKS which is the chain-elongating catalyst via formation of peptide bonds. The function of PCP domain is equivalent to the role of ACP in PKS assemble lines. Optional tailoring domains include epimerization (E), methyltransferase (MT), heterocyclization (HC) and oxidase (Ox) domains (Figure A-2b).[14] The termination modules of NRPS assembly lines usually have a C-A-T-TE organization, and thioesterase (TE) domain usually catalyzes the hydrolysis or cyclization for nascent products to form linear or cyclic peptides.

In addition, both PKS and NRPS require a post-translational modification by a phosphopantetheinyl transferase (PPTase).[36] The PPTase catalyses the covalent attachment of a 4'-phosphopantetheine moiety, derived from coenzyme A, to the peptidyl (NRPS), or acyl (PKS) carrier proteins, which converts the enzyme from the inactive *apo*- to the active *holo*-form.

3 The impact of bacterial genomics on the natural product research

Genes encoding these PKSs and NRPSs in bacteria are often clustered in one region of the chromosomes, thus the PKS/NRPS-derived biosynthetic pathways are also known as biosynthetic gene clusters. These megasynthetases are characterized by having repeat motifs in the same functional domains (e.g., KS and AT domains in PKS, C and A domains in NRPS) which make them easier to be annotated from the genomic sequences *in silico*.[37,38] With the progress in the genome sequencing, a large number of bacterial genomes have been sequenced. These genomic data have unveiled that the genetic potentials of secondary metabolite biosynthesis in the bacteria producing natural products via PKS and NRPS are remarkably higher than the amount of known compounds identified from these bacteria (Figure A-3).[11,37,39] For examples, *Streptomyces coelicolor* and *S. avermitilis* which were known to produce 4 and 3 secondary metabolites at the time of genome sequencing, but actually possess 20 and 25 PKS/NRPS clusters for secondary metabolite biosynthesis, respectively.[40,41] Even more unexpectedly, many microorganisms that were not known as natural product producers also contain the typical PKS/NRPS clusters for the production of hypothetical or unidentified secondary metabolites.[11] These gene clusters that encode biosynthetic pathways whose products have not yet been detected are referred as 'cryptic' or 'orphan' biosynthetic pathways.[42,43] The following efforts to decode these cryptic gene

clusters supported a renaissance in natural product research.[44] With continuously increasing numbers of genome sequences and metagenomics data, it becomes more and more apparent that the compounds identified by classical phenotypic screening and bioassay-guided fractionation only represented the tip of the iceberg,[39] because either their hosts cannot be cultivated or if they can, the low expression level of pathways in laboratory conditions resulting in low amount or nonproduction of natural products that often precludes them from being discovered during conventional screening.[45] Even if a compound is detected, low yield and unstable production may restrain subsequent purification, structure elucidation and biological activity test. Thus there is a wealth of natural product chemistry that remains to be uncovered.

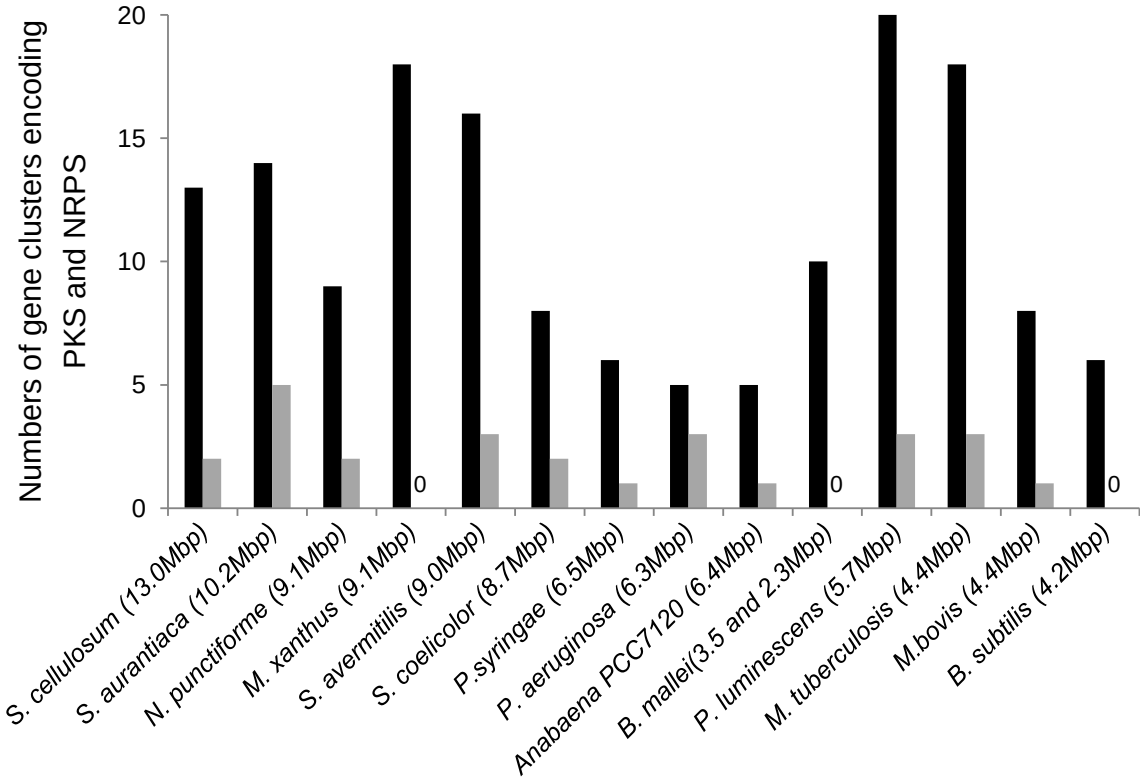


Figure A-3. Biosynthetic potentials of selected fully sequenced prokaryotic natural product producers. Identified PKS and/or NRPS biosynthetic gene clusters are shown as black bars, isolated polyketides (PKs), nonribosomal peptides (NRPs) and PKS/NRPS hybrid compounds known at the time of the genome sequencing are shown as grey bars (The figure adapted from Ref. 11 and 37).

Strategies to exploit the biosynthetic potential of bacteria like genome mining[46] and microbial metagenomic approaches[47] have been developed and proven particularly powerful. Genome mining is normally based on genome sequence information and aims to correlate putative biosynthetic pathways with products. A number of genome mining examples for the discovery of 'hidden' secondary metabolites, which were not detected by classical screening approaches, have been reported in the last decade since the genome sequence projects of

several typical natural product producers.[40,41,48,49] Several recent reviews have systematically summarized the current available approaches and strategies of genome mining to discover cryptic natural products and their biosynthetic pathways,[43,46,50-54] such as OSMAC (One strain-many compounds),[55] bioinformatics prediction/subsequent screening for predicted properties,[42,56] gene inactivation studies and metabolite profiling,[57-59] heterologous gene expression and comparative metabolic profiling approach,[60,61] the genomisotopic approach,[62] *in vitro* reconstitution of biosynthetic pathways,[63] activation of silent cryptic biosynthetic gene clusters by manipulation of regulatory genes and chromatin remodeling,[64-66] ribosome engineering,[67,68] 'protein-first' PrISM (Proteomic Investigation of Secondary Metabolism) [69,70] and mass spectrometry-guided natural product peptidogenomics (NPP).[71]

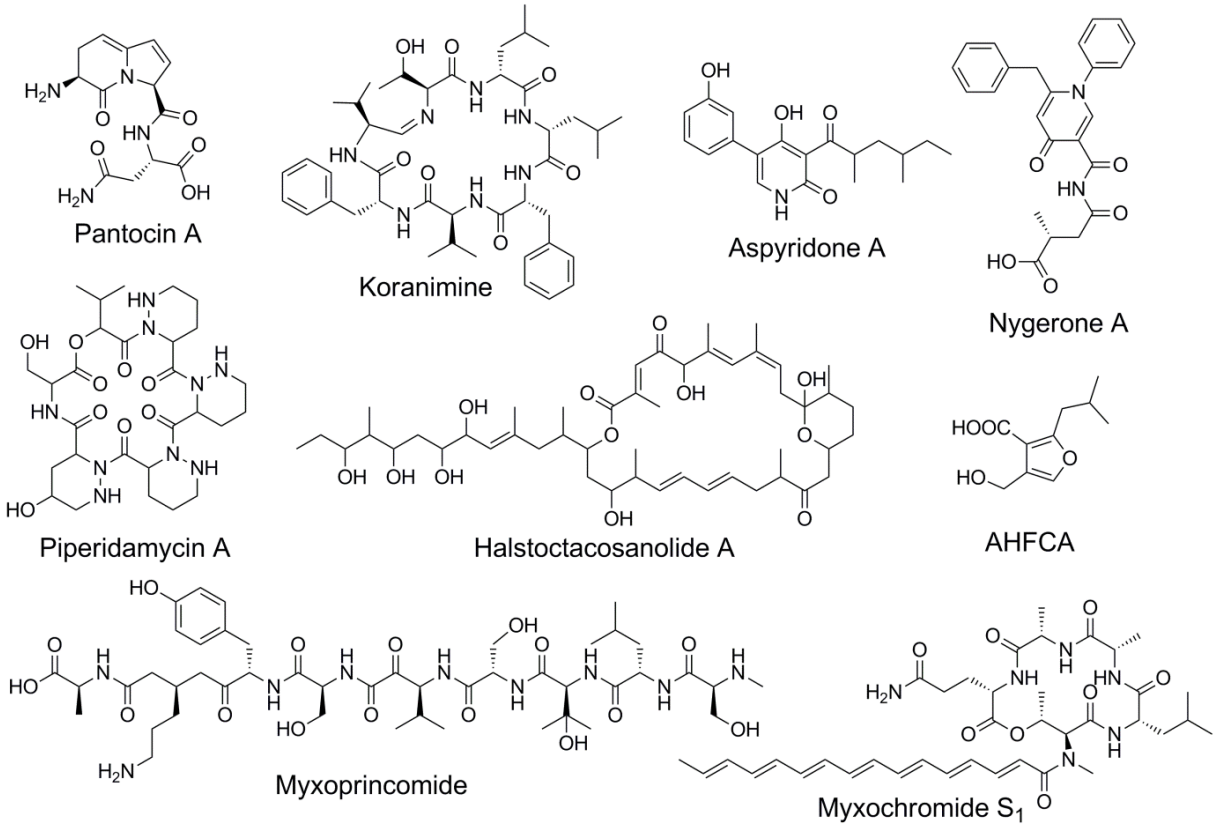


Figure A-3 Selected natural products discovered by genome mining. Pantocin A from *Pantoea agglomerans* and 2-alkyl-4-hydroxymethylfuran-3-carboxylic acid (AHFCA) from *Streptomyces coelicolor* A3(2) by heterologous expression;[60,61] Koranimine from *Bacillus* sp. by PrISM;[70] Nygerone A from *Aspergillus niger* by chemical epigentics methodology;[75] Aspyridone A from *A.nidulans* by activation of regulatory gene;[66] Piperidamycin A from *Streptomyces* sp. 631689 by ribosome engineering;[67] Halstoctacosanolide A from *S. halstedii* by bioinfromatic analysis;[76,77] myxoprincomide from *M. xanthus* and myxochromide S_1 from *Stigmatella aurantiaca* by knockout strategy.[57,59]

Most genome mining methods depend on the comprehensive comparative analysis of secondary metabolomes between wild type strains and mutants, but the low abundance of the corresponding compounds under standard laboratory conditions may have prevented their detection by frequently tedious, and lowly sensitive manual comparison of LC-MS data. The

M*ü*ller group implemented several statistical tools, like principal component analysis (PCA), for metabolome mining based on LC-HRMS data.[59,72-74] Combined targeted gene-inactivation studies with a downstream holistic comparison of secondary metabolomes by PCA led to the identification of a novel NRPS/PKS natural product, myxoprincomide (Figure A-3), and the discovery of two additional new classes of secondary metabolites from myxobacterium model strain *Myxococcus xanthus*.[59]

In addition, most methods for investigating cryptic gene clusters rely on their functional expression in the native hosts. The OSMAC method requires the change of cultivation parameters, e.g., medium composition and aeration, to identify conditions that privilege expression of secondary metabolites. Targeted gene inactivation and overexpression of pathway-specific regulators are also workable only in genetically tractable hosts. The use of epigenetic factors to modulate secondary metabolite regulation is often difficult to predict the conditions but also, if any, only cause the expression of a particular cryptic gene cluster. Furthermore, some natural product producers are difficult to handle and even most of them are uncultured. Thus, heterologous expression of cryptic gene clusters in a well-established host or a plug-and-play[78] system using advanced DNA engineering techniques becomes a more doable approach, especially for the hard-to-handle native host and uncultured microbes, which ultimately leads to discovery of new types of active compounds.[45,79] The further structural diversification of bioactive natural products can be achieved by combinatorial biosynthetic engineering in this amenable host. However, the application of heterologous expression strategy was stunted by some restrictions, such as difficulties in cloning and engineering large gene clusters and lack of suitable heterologous hosts (Section 4).[51]

4 Heterologous expression of microbial natural products

The clustered pattern of genes responsible for natural product biosynthesis, regulation and resistance in the microbial genomes possibly allows us to clone a gene cluster into a physical entity and transfer it into a well-established heterologous host for the production of natural products in higher quantities or for the discovery of products. We can also take full advantage of the recombinant DNA technologies and tools to generate novel natural product analogs by combinatorial biosynthesis methods, circumventing the need to develop a genetic system for every microbe producing an interesting compound.[80] The requirements and challenges associated with this strategy are also encountered and mainly include: 1) cloning and engineering of biosynthetic gene cluster; 2) selection of suitable heterologous hosts (e.g., precursor supply, post-translational protein modification, self-resistance, codon usage).

4.1 Cloning and engineering of biosynthetic gene clusters

The cloning and sequencing of bacterial secondary metabolite biosynthetic gene clusters is now routine but time and labor consuming, involving the construction and screening of cosmid or BAC genomic DNA libraries by random cloning.[10,81] Generally, relatively small biosynthetic pathways (<30kb) can often be cloned within a single cosmid and isolated by screening the genomic cosmid libraries of the natural producer strains, as the average insert size of a cosmid vector is estimated to be approximately 40 kb. In this case, the biosynthetic gene cluster located on one vector can be directly transferred into closely related host, e.g., from *Streptomyces* to *Streptomyces*, in which the promoter and regulatory elements are similar.[82] However, many natural product biosynthetic pathways are much larger in size than the average insert of a cosmid, and some are even larger than fragments commonly found in BAC vectors. Thus, the coexpression of several plasmids each harboring parts of biosynthetic genes is demanded to achieve heterologous expression of such ample biosynthetic gene clusters. Reassembly of a large natural product biosynthetic pathway on one construct before transferring into the heterologous host is an alternative to the tedious coexpression procedures. By using Red/ET recombineering (section 5.4.1),[83-86] the complete biosynthetic pathways can be stitched together from several cosmids into one construct, followed by the insertion of transfer genes (e.g., genes for conjugation), and promoter(s) which is required for heterologous expression in unrelated hosts. This is a time-saving method, especially for biosynthetic pathways from slow-growing bacteria that can be expressed in a bacterium which grows relatively quickly. Several complete biosynthetic pathways, such as 43 kb *Stigmatella aurantiaca* DW4/3-1 myxochromide S gene cluster,[87] 57 kb *S. aurantiaca* DW4/3-1 myxothiazol gene cluster,[88] the 42 kb *Streptomyces* sp. Tü6071 phenalinolactone gene cluster,[89] the 32.5 kb *S. refuineus* anthramycin gene cluster,[90] the 36.8 kb *S. rishiriensis* DSM40489 coumermycin A_1 gene cluster,[91] the 58 kb *Sorangium cellulosum* epothilone gene cluster,[92] the 40 kb *Cystobacter* sp. SBCb004 tubulysin gene cluster,[93] 35 kb *S. ansochromogenes* nikkomycin gene cluster,[94] have been reassembled by construction and screening of cosmid libraries and subsequent Red/ET recombineering mediated stitching and modification. A modified version of gene cluster stitching, involving linearization of both vectors prior to transformation of recombineering proficient *Escherichia coli*, was used to reconstitute the 38.5 kb *Streptomyces nooses* subsp. *asukaensis* asukamycin gene cluster.[95]

In addition, the large gene cluster can be directly and specifically cloned from sequenced genome using transformation-associated recombination (TAR) cloning which uses the native recombination potential of *Saccharomyces cerevisiae*.[96-98] Brady's group created pTARa, a

BAC-based *S. cerevisiae*/*E. coli*/*Streptomyces* shuttle capture vector.[99] This vector contains elements that allow pathways to be assembled in *S. cerevisiae*, characterized and maintained in *E. coli*, and conjugatively transferred into a wide range of *Streptomyces* for heterologous expression studies. Using pTARa, they directly cloned the 56 kb colibactin gene cluster directly from genomic DNA of *Citrobacter koseri*.[99] This method provides a rapid alternative to traditionally genomic library for cloning intact natural product biosynthetic gene clusters from sequenced microorganisms. To access large, functionally intact, natural product gene clusters (e.g., PKSs, NRPSs) from the environment, partial gene clusters need to be reconstituted or stitched from metagenomic environmental DNA (eDNA) library clones. The same group also used TAR to reconstruct or assemble large gene cluster from several overlapping eDNA cosmid clones, resulting in the identification of several new compounds.[100]

An alternative approach is to assemble gene clusters by recently developed synthetic biology techniques, such as the Gibson assembly method and DNA assembler,[101-103] from several smaller fragments which can be prepared via standard PCR or direct chemical DNA synthesis. But these methods required multiple PCR amplification steps that would introduce unexpected mutations, or synthesized DNA that still expensive, and the assembly process also causes mutations, insertions, or deletions due to incorrect pairing of fragments.[79]

The whole or partial natural product biosynthetic gene cluster is usually cloned into an *E. coli* vectors firstly by genomic DNA library or direct cloning method. The subsequent manipulation should be determined by strategies of entrance and maintenance in heterologous hosts. If the intact gene cluster is already contained in a single plasmid/BAC, the next step is to delete non-required gene fragments, to introduce genes for transfer, or to exchange native promoters. If the whole gene cluster is found scattered in several plasmids, we can adopt the stitching strategy to put all fragments in a single plasmid/BAC, or use multi-plasmid based method. The former should be followed by stitching into a BAC or a low copy-plasmid through Red/ET recombineering or TAR cloning, while the latter need to change the origin of replication and selection marker of plasmids to make them compatible in one cell.

The challenges of genetic engineering biosynthetic pathways have been generally overcome by Red/ET recombineering.[10] It has streamlined the genetic modification of complex biosynthetic pathways by eliminating many of the time-consuming steps in traditional digestion-ligation genetic engineering (section 5.4).

A promoter is a region of DNA that initiates transcription of a particular gene. The native promoters present in a specific gene cluster might not be recognized by the heterologous

hosts. This problem could mostly be circumvented by chosen of closely related heterologous hosts maintaining the native promoters for the genes.[80,82] However, when the gene cluster expressed in a phylogenetically distant host, the native promoter has to be replaced by an available promoter in heterologous host to promote the transcription of foreign gene cluster, in particular the promoter with inducibility (e.g., the thiostrepton-inducible *tipA* promoter and tetracycline-inducible promoter for streptomycetes,[104-106] the toluic acid-inducible Pm promoter for *Pseudomonas* strains[87,107,108] and L-arabinose-inducible P_{BAD} promoter and T7 promoter for *E. coli*[109-111]). Inducible expression is desirable particularly in gene clusters whose encoded metabolites may be cytotoxic to the hosts. Exchange of the native promoter with an inducible promoter not only solves the problem of inefficient transcription in heterologous host, but also protects the host from the toxicity of products. If the products are not toxic to host or the resistance gene is present in the host, a much stronger or constitutive promoter, such as *Tn5* promoter,[92] P_{aphII} promoter,[112] *ErmE* promoter,[113] *act* promoters[114,115] or a synthetic promoter, coupled with optimized ribosomal binding sites, could be used to improve the production of secondary metabolites. In addition, the native TTG or GTG start codon exchanged for an ATG codon can aid translational efficiency in heterologous host and improve production, e.g., the native TTG of *tubC* was changed to ATG lead to a five-fold increase of pretubulysin A production in *P. putida*.[93]

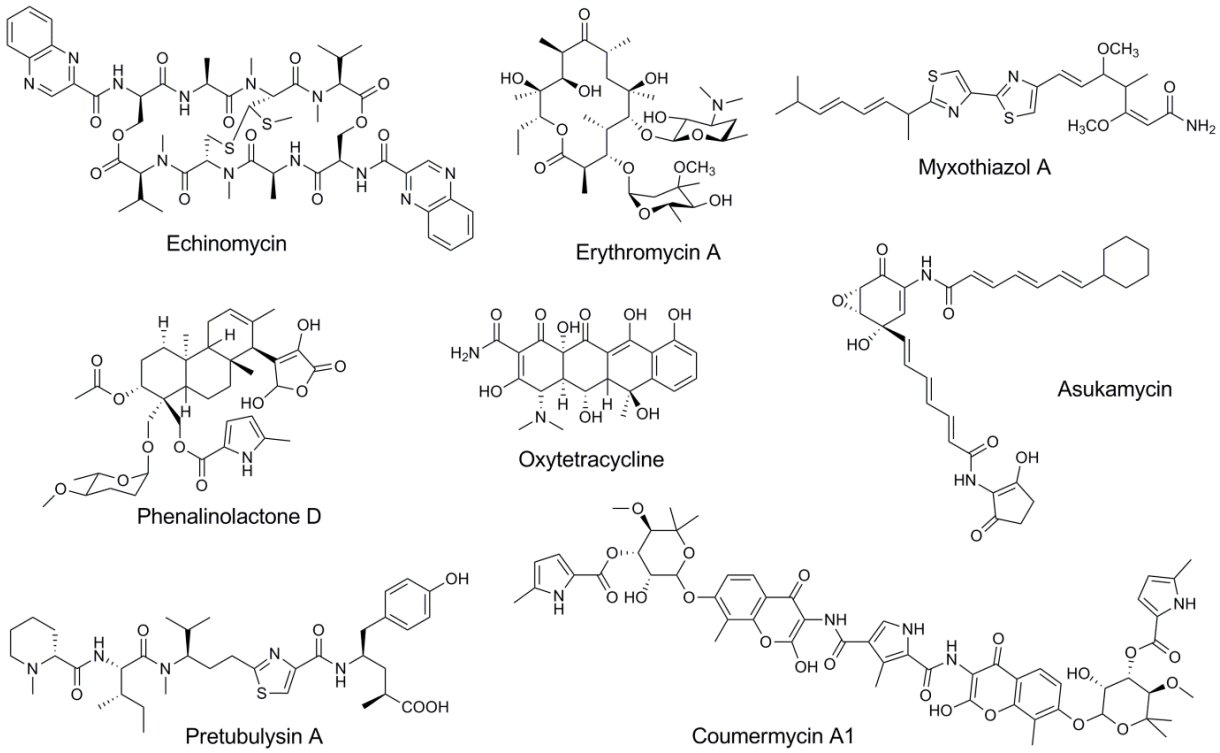


Figure A-4. Selected natural products that have been successfully expressed in a heterologous host.

4.2 Selection of suitable heterologous hosts

It is important to choose a suitable host for heterologous expression of the desired product, after the successful reconstitution of biosynthetic gene cluster into a mobile vector system (a cosmid or BAC, replicative or integrative, singular or multi-plasmid based) in *E. coli*. An ideal heterologous host should grow fast, be genetically tractable, ensure functional expression of the required proteins, provide all necessary precursors and have a low background for native secondary metabolites.[80] Although such an ideal host may not exist for all natural products, these criteria can often be met by strains closely related to the native producer but with better growth and genetic characteristics. For example, secondary metabolite biosynthetic gene clusters from actinomycetes species usually can be successfully expressed in some streptomycetes (e.g., *S albus, S. lividans, S. coelicolor*), all of which are suitable for genetic manipulation. However, some phylogenetically distant strains (unrelated hosts) provide better results although obstacles, such as codon usage bias, promoter recognition, precursor availability and post-translational protein modification, have to be overcome by genetic engineering (introduction of an exogenous pathways for supplement of precursors and broad substrate specificity PPTases) or precursor feeding. While many different bacteria hosts have been used for expression of heterologous genes, *E. coli*, *Pseudomonas putida*, *Myxococcus xanthus*, and several *Streptomyces* species have been shown to express large clusters at reasonable initial titers.[116,117]

E. coli is a model microorganism for genetic and metabolic engineering with a rapid growth rate and its use for biological product overproduction (primarily recombinant proteins) are also well established. However, *E. coli* lacks the intracellular machinery (e.g., post-transcriptional modification, intracellular metabolism, or biosynthetic enzymes) to natively produce natural products. Nonetheless, *E. coli* does have the ability to produce secondary metabolites, such as enterobactin (NRP),[118,119] and the structurally unknown colibactin (NRP/PK).[120,121] The rapid growth rate, easy and cheap scale-up cultivation, various genetic manipulation tools, and well-studied genetic system allured people to engineer *E. coli* for producing exogenous natural products. To express 6-deoxyerythronolide B (6dEB), the precursor of erythromycin, a phosphopantetheine transferase gene (*sfp*) from *Bacillus subtilis* was inserted into the *E. coli* chromosomal *prp* operon responsible for propionate catabolism to delete the *prpRBCD* genes but leaving the *prpE* gene under an inducible T7 promoter. PrpE converts propionate into propionyl-CoA, one of the precursors needed for 6dEB. The propionyl-CoA carboxylase (PCC) from *S. coelicolor*, was also introduced to convert propionyl-CoA to (2*S*)-methylmalonyl-CoA (MM-CoA), the second precursor required for

6dEB formation.[110] In the heterologous epothilone production, the methylmalonyl-CoA decarboxylase gene (*ygfG*) was replaced by the *pcc* genes to accumulation of MM-CoA.[109] The engineering enables the successful expression of large PKS/NRPS gene clusters, such as erythromycin (PKS, feeding of propionate),[122] yersiniabactin (NRPS/PKS, feeding of salicylate),[123] epothilone (NRPS/PKS, optimized codon usage and feeding of intermediate),[109] echinomycin (NRPS),[111] in *E. coli*. But heterologous protein folding, foreign codon usage and substrate availability must be addressed for further heterologous expression of PKSs and NRPSs in *E. coli*.

As heterologous hosts for natural products, pseudomonads have several advantages; they grow rapidly and are genetically tractable. The high GC genomic content, codon preference as well as the possession of PPTases with promiscuous substrate specificity makes them more suitable than *E. coli* for the expression of biosynthetic gene clustes.[124,125] Müller group has successfully expressed several myxobacterial biosynthetic gene clusters (e.g., myxochromide S, pretubulysin) in *P. putida*.[87,93] However, *P. putida* cannot provide MM-CoA in a detectable level, which precludes the expression of many polyketides. To address this issue, methylmalonyl-CoA epimerase, methylmalonyl-CoA mutase, and MeaB (an enzyme complex-stabilizing factor) from the myxobacterium *S. cellulosum* So ce56 were introduced into *P. putida* resulting in a strain producing MM-CoA.[126] Afterwards, myxothiazol, an MM-CoA-dependent myxobacterial secondary metabolite, was heterologous produced in this engineered strain with a final yield of 600 μg/L.[126] An unknown PKS gene from *S. cellulosum* So ce56 was expressed in *P. putida* yielded flaviolin, a product formerly unknown from myxobacteria.[127] In addition, *Pseudomonas* strains have also been successfully used to confirm the identity of the safracin biosynthetic gene cluster from *P. fluorescens* A2-2 by heterologous expression.[128]

M. xanthus is the best characterized myxobacterium and possesses more genetic engineering tools than all other myxobacteria. It also produces a number of PKS/NRPS natural products that showed the presences of PPTases, MM-CoAs, and self-resistance to bioactive compounds. Furthermore, *M. xanthus* exhibits a much shorter doubling time (5 h) than *S. cellulosum* (16h), the anticancer epothilone producer. Thus, Julien and colleages introduced the 65.4 kb epothilone gene cluster into *M. xanthus* that led to production of this hybrid PKS/NRPS product.[129] The Müller group also accomplished the heterologous production of myxobacterial secondary metabolites myxochromide S, myxothiazol and pretubulysin in *M. xanthus*.[88,92,93] In addition to myxobacterial biosynthetic pathways, *M. xanthus* can heterologously express streptomycete-derived oxytetracycline biosynthetic

pathway in a high titer 10 mg/L.[130] The above facts make *M. xanthus* an attractive potential candidate for a 'universal' host for facilitating heterologous expression of PKS/NRPS biosynthetic pathways from myxobacteria and actinomycetes.

The *Streptomyces* or relative actinomycetes have also been harnessed as heterologous hosts for the expression of diverse natural products. First, the genus has the ability to produce a wide variety of secondary metabolites, which implies that they have sufficient precursors and PPTases, and provide resistance mechanisms to sequester themselves from their own bioactive compounds. Second, several *Streptomyces* species, such as *S. coelicolor*,[40] *Saccharopolyspora erythraea*,[48] and *S. avermitilis*,[41] have been sequenced; and the tools for genetic manipulation have been established. These characteristics, as well as a nonpathogenic nature and established fermentation technology, made *Streptomyces* an evident option for producing natural products, especially for the polyketides.[117] Secondary metabolites in actinomycetes can usually be heterologously expressed in these *Streptomyces* hosts as reviewed by Baltz.[131]

Moreover, some streptomycetes are also engineered to improve the production of foreign products, or to decrease the background metabolite profile. The complete deletion of the gene clusters for actinorhodin, undecylprodigiosin, CDA and the *cpk* cluster[132] in the heterologous host *S. coelicolor* (M1146) resulted in the 7.5 times increased production of coumermycin A_1 when compared to that in *S. coelicolor* M512.[133] The introduction of point mutation into *rpoB* and *rpsL* in *S. coelicolor* M1146 dramatically increased the levels of heterologous production of chloramphenicol and congocidine and relatively simplified extracellular metabolite profiles.[134] A region of more than 1.4 Mb was deleted stepwise from the 9.02 Mb *S. avermitilis* chromosome to generate a series of defined deletion mutants that did not produce any of the major native secondary metabolites found in the original strain.[135] Exogenous streptomycin and cephamycin C biosynthetic pathways were efficiently expressed at levels higher than those of the native-producing species. This feature will markedly facilitate the discovery of new compounds by heterologous expression of cryptic gene clusters that are prevalent within actinomycete genome sequences.

Considering the growth rate, precursor supply, and self-resistance, the safe selection of heterologous host was proposed. Generally, the actinomycete-derived gene clusters are firstly considered to the engineered streptomycetes, like *S. coelicolor* M1146 or M1152, engineered *S. avermitilis*, and then *M. xanthus*. *M. xanthus* is also the optimal choice for the expression of myxobacterial gene clusters, especially for the polyketide gene clusters. The NRPS gene

clusters can be attempted to express in *P. putida*, and *E. coli*, but promoter exchange is ordinarily essential for successful expression in a distantly related host (Section 4.1).

4.3 others

The methods for transfer of biosynthetic gene clusters into heterologous hosts are dependent on the choice of hosts (Section 4.2). *E. coli* is the model genetic engineering microorganism and electroporation is the most frequently-used and efficient method to transfer of large DNA into most *E. coli* strain now. Large gene clusters can maintain in *E. coli* by replicative plasmids,[110,111,123] only few examples by integration of gene clusters into the chromosome.[136] The routine method to introduce foreign DNA into *P. putida*, and streptomycetes is conjugation. The large DNA fragments enter *M. xanthus* always by electroporation. Because no stable plasmid replicon is available for this strain, all of the heterologous gene clusters should be integrated into the chromosome. After access into cells, the stable maintenance of the foreign biosynthetic gene cluster in a heterologous host is the groundwork for successful heterologous expression. The plasmid-based system and chromosomal integration are the two prevalent patterns, which have been well-established.[10] Biosynthetic gene clusters can be expressed from self-replicating plasmids; either from a single plasmid or from multiple plasmids individually expressing modules *in cis.* Plasmid-based expression has the benefit of increasing the number of copies of the biosynthetic gene cluster which can provide an enhanced titer of compound, but unstability is the main drawback of this strategy. The chromosomal integration is more stable and it includes homologous recombination by endogenous recombinases in the heterologous host, transposition for random integration[92] and Phage derived integration which mediated by a large serine recombinase that catalyzes unidirectional recombination between the bacteriophage *attP* and chromosomal *attB* sites.[137]

Once the biosynthetic gene cluster for the desired compound is cloned, transferred and stably maintained within a chosen heterologous host, the selection of suitable fermentation conditions serves as the last variable to activate or improve metabolite production. Based on the many successful examples reported, in most cases the original fermentation conditions used with the native producer of the desired compounds can also be applied to the experiments within the heterologous host.[80] The conditions are usually optimized for good production of the respective natural product and, at least for host related to the native producer; these conditions should provide all the nutrients needed to ensure sufficient growth of the respective recombinant strain. For the unrelated host, the condition is difficult to predict, but generally, the low temperature incubation and coexpression of chaperones are

beneficial for large protein expression which can facilitate the production of desired products, especially for the expression of large PKS and NRPS in *E. coli*. [110,122]

Among these, the biggest challenge is the cloning of big size biosynthetic gene clusters, though advanced recombinant DNA techniques (i.e., Red/ET recombineering, TAR cloning) slightly simplified the procedure for reconstruction of large gene cluster, and promoted the genetic engineering of gene clusters. Besides, the efforts to improve the production of heterologous products are must be strengthened.

5 Molecular engineering of microbial natural product pathways

The genome sequences of natural product producers set the stage for genetic engineering of microbial natural product biosynthetic pathways to characterize gene function, generate novel derivatives, or improve production yields.

5.1 General strategies for engineering of natural product pathways

The biosynthesis of PKs and NRPs that condenses monomeric building blocks into a natural product scaffold, uses a process which can be divided into three stages: provision of precursors (acyl-CoAs for PKS, amino/aryl acids for NRPS) to feed into the assembly lines; PKS/NRPS catalyzes chain initiation, elongation, and termination during assembly; post-assembly-line tailoring of nascent released product to form mature (bioactive) natural products.[138] Based on the biosynthetic pathway, the modifications to yield new 'unnatural' natural products should include three levels: Precursor-directed biosynthesis/mutagenesis which aims to incorporate modified starter/extender units or biosynthetic intermediates into the assembly process by feeding experiments, engineering of the PKS/NRPS assembly line, and modification of post-assembly line enzymes (Figure A-5).

Precursor-directed biosynthesis (PDB) and mutasynthesis utilize nature's machinery to produce complex natural product analogs, which take advantage of the natural flexibility of biosynthetic pathways toward the acceptance of unnatural precursor analogs.[139] The alternative biosynthetic building blocks are designed, synthesized and fed to the wild type (PDB) or engineered (mutasynthesis) microorganisms to generate the natural products of interest.[140,141] Due to competition with endogenous precursors, the yields of new molecules by PDB are usually low. This problem is circumvented in mutasynthesis approaches where the native pathway is blocked at specific positions by genetic manipulations (mutagenesis) resulting in an abolishment of product formation. The biosynthesis can then be restored by feeding appropriate unnatural precursors to the mutant strains to trigger a 'jump-start' of the assembly process. However, incorporation of the altered precursors requires a broad substrate

tolerance of all catalytic activities downstream from the blocked step and, in practice, only a subset of analogs will be accepted.[141]

The modification of the assembly line is generally based on the manipulation, such as deletion, insertion, exchange, fusion, and site-directed mutagenesis, of catalytic domains, complete modules to cause the alteration of the 'gatekeeper' domains, A domains for NRPS and AT domains for PKS. By rational manipulation of domains/modules, or site-directed mutagenesis, it has been possible to alter the selection of building blocks by domain/modular exchange,[142-148] to increase or decrease chain length by module insertion,[149] fusion[150] or deletion,[151,152] to modify the β-keto reduction for PKS by inactivation of active sites or domain exchange,[153-156] or to inactivate tailoring domains in NRPS by point mutations.[157] The recombination of whole modules or domain is a rather drastic intervention in PKS/NRPS biosynthesis that usually results in reduced catalytic efficiency and product yields. Manipulating the A domain's specificity through point mutations of substrate-coordinating amino acid residues according to the 'non-ribosomal code' represent a more conservative strategy for NRPS engineering.[158,159]

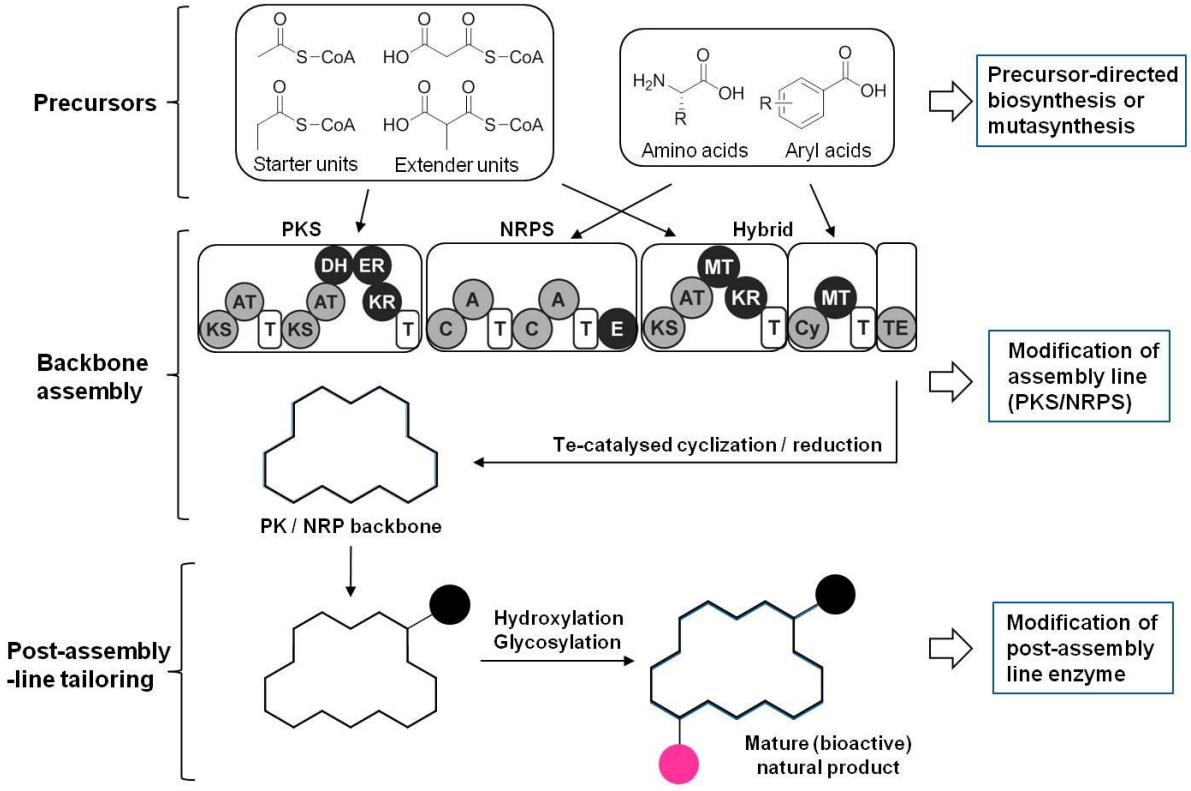


Figure A-5. Strategies for molecular engineering based on the biosynthetic stages of PKS and NRPS.

The post-assembly-line tailoring enzymes include glycosyltransferases, methyltransferases, halogenases, prenyltransferases, carbamoyltransferases, and oxygenases, which can mature

the bioactive natural products. It is possible to generate novel 'unnatural' natural products by alteration, deletion or insertion of those post-assembly-line tailoring enzymes.

5.2 Molecular engineering in native producers

There are two basic ways for engineering of bacterial natural product biosynthetic pathways, one is the genetic manipulation in the chromosome of native producers, the second being transfer of entire biosynthetic gene cluster into a more amenable heterolgous host followed by modification and engineering.[37] The former way usually requires double crossover to introduce, or delete genes, to the best, in a markerless pattern. However, some natural product producers, like some myxobacteria and streptomycetes, are growing slowly and they are refractory to accept exogenous DNA for elaborate genetic manipulation.

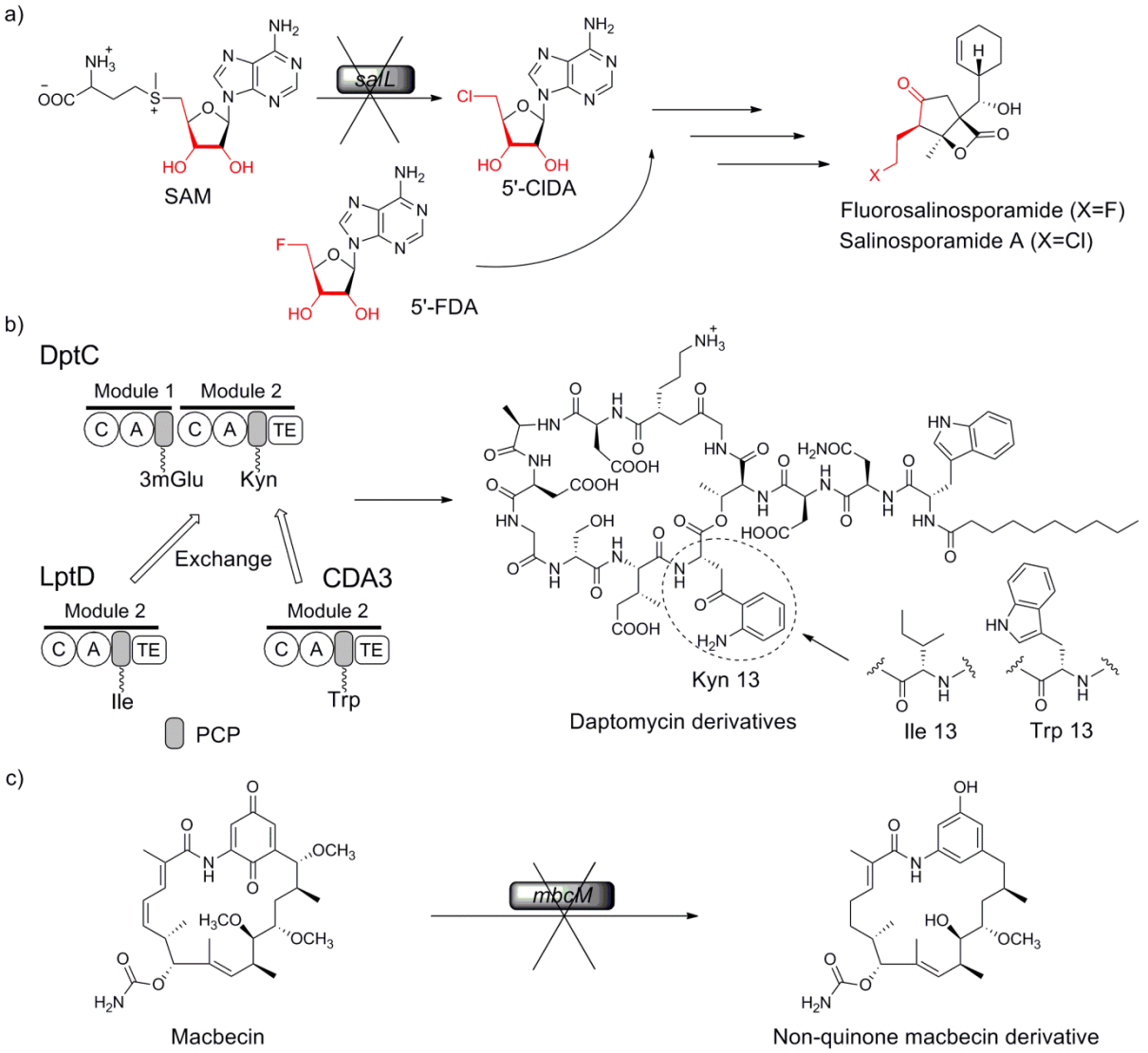


Figure A-6. Examples of molecular engineering of natural product biosynthetic pathaways for generating new derivatives. a) Disruptants in *salL*, a gene encoding the unusual chlorinase can be complemented with 5'- FDA in order to generate fluorosalinosporamide. b) Module exchange was applied to produce daptomycin derivatives with Trp13 and Ile13. c) Targeted disruption of an oxygenase (MbcM), resulting in production of a non-quinone macbecin derivative.

In the naive producer, mutasynthesis is the more convenient method to generate new derivatives. Eustáquio and Moore disrupted *salL*, a gene encoding the unusual chlorinase for the biosynthetic chlorination of *S*-adenosylmethionine (SAM), in salinosporamide producing strain *Salinispora tropica*, thereby blocking the halogenation. Complementation of 5′-fluoro-5′-deoxyadeninoside (5′-FDA) yielded the fluoro derivative fluorosalinosporamide (Figure A-6a).[160] Disruptants in *rhiA*, a gene encoding the heterocyclization in the loading module in the rhizoxin biosynthetic pathway coupled with feeding of *N*-acetylcysteamine derivative led to formation of thiarhizoxin in *Burkholderia rhizoxinica*.[161] This PDB/mutagenesis method have been widely applied in generation of new derivatives for many natural products, such as balhimycin,[162,163] rapamycin,[164] pacidamycin,[165] CDA (calcium-dependent antibiotic),[166,167] myxalamid,[168] geldanamycin,[169-172] erythromycin,[173] ansamitocin,[174-178] cryptophycin,[179,180] aureothin,[181] cinnabaramide,[182] elansolid.[183]

The modification of PKS and NRPS backbones in the native producer are usually performed by gene complementation strategy. Firstly the targeted subunit was cleanly deleted by double crossover, and then complemented with a replicative or integrative plasmid containing the engineered fragment. An elegant study on the genetic manipulation of the PKS assembly line was performed by Kosan Biosciences, Inc.[142] Genetic engineering of the geldanamycin gene cluster (GdmPKS) in native producer *S. hygroscopicus* resulted in new geldanamycin derivatives that selectively lack methyl or methoxy groups in the ansa chain by replacement of AT domains in six different GdmPKS modules that commonly accept MM-CoA or methoxymalonyl-CoA with malonyl AT domains from the rapamycin PKS. Four of these manipulations led to the production of 2-desmethyl, 6-desmethoxy-, 8-desmethyl-, and 14-desmethylgeldanamycin derivatives, the γ,δ-saturated, and the hybridization products. Another successful example is the combinatorial biosynthesis of novel antibiotics related to daptomycin and A54154. The scientists in Cubist Pharmaceutical used module exchange (Figure A-6b), NRPS subunit exchanges, inactivation of the tailoring enzymes, and natural variations of the lipid tail based on the daptomycin gene clusters to generate a library of novel lipopeptides, some of which show improved properties.[145-147,184-186] They employed Red/ET recombineering to exchange new domains or modules at specific interdomain linker sites (Section 5.4.2). Direct double crossover in the native producer was also workable. Marahiel group made many efforts in the combinatorial biosynthesis base on this strategy by domains/modules exchanges, in-frame module deletion, module fusion, and site-directed mutagenesis in the surfactin biosynthetic gene cluster in *Bacillus subtilis*.[31,144,148,151,187] It was shown possible to switch the specificity of A domain in NRPS from one amino acid to

another to change the polypeptide backbone. A second specificity change in module 5 from Asp to Asn yielded the expected surfactin derivative *in vivo*.[31] The saturation mutagenesis of nonribosomal codes on NRPS AdmK allowed generating new andrimid derivatives in original producer *Pantoea agglomerans*.[32] The deletion or insertion of whole module, or domain fusion also was utilized to change the size of NRPS backbones. The directly manipulation of biosynthetic gene cluster in the native chromosome is usually difficult and tedious, thus ectopic *trans*-complementation system is an efficient alternative to accomplish this goal.

The inactivation of post-modifying genes can be achieved by simple single crossover to produce reduced derivatives, meanwhile to identify the function of inactive genes. For example, targeted disruption of an oxagenase involved in post-PKS quinone formation during the biosynthesis of an Hsp90 inhibitor macbecin led to the production of a non-quinone analog which has an enhanced binding affinity and significantly reduced toxicity compared with the parent molecule (Figure A-6c).[188]

5.3 Molecular engineering in heterologous hosts

As the mutagenesis of secondary metabolite pathways in the chromosome of natual producers is often very difficult except for some relatively genetically tractable strains (section 5.2), the robust heterolgous expression of biosynthetic gene cluster in an amenable host provides an important basis for the molecular engineering, because the engineering of gene cluster in *E. coli* is usually well-established. The successful expression of echinomycin gene cluster in *E. coli* enabled to produce an unnatural (synthetic) quinomycin antibiotic TANDEM by deletion of the gene *ecm18* encoding thioacetal formation enzyme and inactivation of both MT domains by point mutations in the *ecm7* NRPS.[111,157] The same group also substituted an echinomycin biosynthetic gene *ecm7*, with a gene *swb17*, encoding a bimoduler NRPS that contains the A domain capable of accepting (+)-(1*S*,2*S*)-norcoronamic acid ((+)-NCA) as extender unit, from SW-163 biosynthetic gene cluster. This rational engineering of plasmid-based echinomycin biosynthetic pathway and feeding of a chemically synthesized precursor (+)-NCA caused L-Val in echinomycin was replaced with (+)-NCA and further production of a novel bioactive compound ecolimycin C in *E. coli*.[189] This showed that *E. coli*-based plasmid-borne heterologous natural product biosynthetic expression can not only facilitate the reconstruction of various biosynthetic gene clusters but also afford an applicable platform for rational redesigning of biosynthetic pathways through modular reassembling of the genes from another biosynthetic cluster. In addition, the loading module of DEBS1 was replaced with the loading module from the rifamycin PKS and this

change led to the production of an analog of erythromycin D with a designed benzyl unit in *E. coli* at a titer of 0.5 mg/L. Similarly, the second module of DEBS3 was altered to accept a malonyl-CoA extender unit, which resulted in the lack of a methyl group at the C2 position of erythromycins A-D at titers between 0.05 and 0.2 mg/L.[122] The reengineering of megaenzymes is usually performed in *E. coli*, but expression is not always functional in *E. coli*, and the limiting factor is the lower or no product titers.

The large gene cluster can also be modified in *E. coli* using Red/ET recombineering and then transferred into the heterologous host for robust gene expression. Heterologous expression of the reconstituted gene cluster in *S. coelicolor* M512 resulted in the production of coumermycin A_1. Inactivation of the MT gene *couO* by Red/ET recombineering and heterologous expression of the modified cluster led to an accumulation of a *C*-8-unsubstituted coumermycin A_1 derivative. Subsequent expression of the halogenase gene *clo-hal* from the clorobiocin gene cluster in the heterologous producer strain resulted in the formation of two new hybrid antibiotics.[91] The same group also produced 8-halogenated and 8-unsubstituted novobiocin derivatives in genetically engineered *S. coelicolor* strains.[190] Similarly, deletion of a glycosyltransferase–encoding gene *plaA6* from the phenalinolactone expression construct by Red/ET recombineering gave rise to a novel, unglycosylated phenalinolactone derivative in the heterologous host *S. coelicolor* M512.[89] Inactivation of *p*-aminobenzoate synthase gene *aurG* involved in the formation of the natural aureothin starter unit (*p*-nitrobenzoyl-CoA) via Red/ET recombineering, and expression of the modified pathway in *S. lividans*, resulted in mutants lacking aureothin production. Addition to the non-natural substrate *p*-cyano benzoate to this mutant resulted in the production of a new nitrile analog from aureothin, aureonitrile, which shows significantly enhanced biological activity.[181] This can be considered as the mutasynthesis in the heterologous host, analogously, the precursor-directed biosynthesis of epothilone was also achieved in a heterologous host *E. coli*.[191]

The biosynthetic gene cluster can also be reconstructed from PCR-amplified multiple DNA fragments by DNA assembler (Section 4.1) via *in vivo* homologous recombination in *S. cerevisiae*. This method is flexible for genetic manipulations, such as site-directed mutagenesis, scar-less gene deletion and insertion, of the target pathway. The Zhao group created point mutations onto the core motif of DH domain in AurB in the aureothin biosynthetic pathway using DNA assembler and expressed the resulting mutants in heterologous host *S lividans* led to expected production of a new hydroxyl aureothin together with the original compound.[103]

A couple of new derivatives have been successfully generated by the combinatorial biosynthesis in the relatively amenable hosts, such *as E. coli*, streptomycetes, but the difficult in cloning of large gene cluster and low titers of heterologous expression (lack of appropriate heterologous host) restricted the application of this strategy. In addition, design of modification of biosynthetic pathways needs the detailed understanding of protein-protein interaction in the gene cluster.[82] Amongst them, the cloning of large gene cluster into *E .coli* is the primary issues and the prerequisite for prosperous heterologous expression and molecular engineering to exploit the natural products. Thus more convenient and advanced DNA cloning techniques are required to efficiently clone and manipulate large DNA fragments.

5.4 Red/ET recombineering facilitates molecular engineering

5.4.1 Red/ET recombineering

Red/ET recombineering, also known as 'ET cloning',[83] 'ET recombination',[84,85] 'λ Red recombination'[86,192] or 'recombineering',[193,194] is an *in vivo* homologous recombination-based genetic engineering method employed primarily in *E. coli* by using short homology arms, which is based on the expression of either *redα/redβ* from the Red operon of λ phage[86] or the analogous *recE/recT* from Rac prophage[83] that are located in the *E. coli* K12 chromosome. In the interests of clarity, Stewart referred to the homologous recombination in *E. coli*, mediated by either RecE/RecT or Redα/Redβ, as Red/ET recombination.[195] In this thesis, we prefer to use Red/ET recombineering for all the techniques involving RecE/RecT or Redα/Redβ.

The Rac operon and the λ Red phage are functionally equivalent.[196] RecE and Redα are 5′ to 3′ exonucleases, RecT and Redβ are single strand DNA-binding proteins (SSAP) with annealing and strand exchange activity, Redγ is a inhibitor of nuclease RedBCD and prevent the degradation of linear DNA fragment.[197] The Red/ET recombineering is started with the RecE or Redα protein binds to linear DNA and degrades the linear DNA in a 5′ to 3′ direction, thereby exposing a 3′ single strand DNA (ssDNA) overhang. Then, RecT or Redβ binds to the ssDNA, forming a recombinogenic proteonucleic filament which is used in recombination, either by single strand annealing or by strand invasion (Figure A-7).[198-200] In the annealing model, the protein-ssDNA filament anneals to a complementary single-stranded region that has arisen from either a similarly prepared double strand break (DSB) or from a DSB produced by DNA replication. In the strand invasion, the protein-ssDNA filament establishes a D-loop in an unbroken DNA region.[200] The resulting recombinants can be acquired after the repair and replication.

In comparison to Redα, RecE is a much larger protein. The C terminus starting at residue 588 (RecE588) of which only constitute the exonuclease domain, together with RecT, was used for recombineering and they are better than the full length RecE/RecT.[83] But RecE588/RecT pair was less efficient than the Redα/Redβ pair for homologous recombination, thus the Red system was extensively applied for precise insertion, deletion and insertion of sequence of interest.[85,92,192,201-203] This kind of Red/ET recombineering approaches relied on homologous recombination occurring between linear (generally PCR products) and replicating circular DNA molecules, termed it LCHR (linear plus circular homologous recombination). It is proficient with homology arms as short as 40-50bp and permits the use of polymerase chain reaction (PCR)-amplified fragments generated from oligonucleotides synthesized to introduce the short homology arms.[192,204] As Red/ET recombineering is based on homologous recombination, it allows precise modification bacterial chromosomes or any episomal DNA in *E. coli* and is independent on the location of restriction sites.

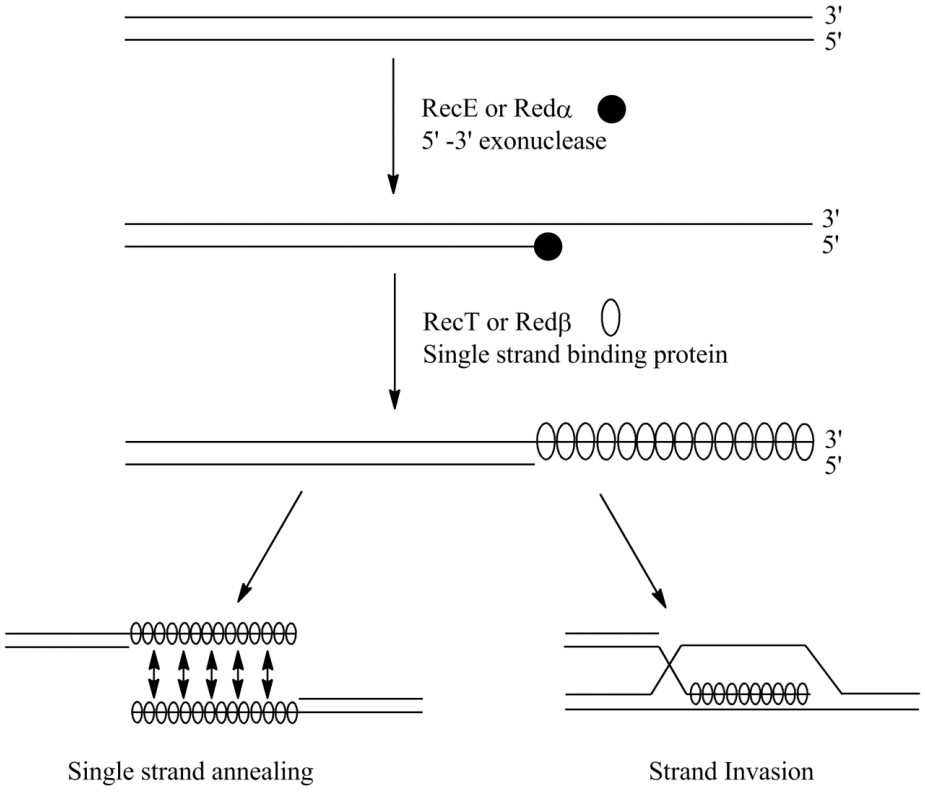


Figure A-7. Two classical models for Red/ET recombineering using double-stranded DNA as a donor. (The figure adapted from Ref.196)

Besides linear double-stranded DNA (usually in the pattern of PCR products), synthetic oligonucleotides or single-strand DNA (ssDNA) can also be substrate of recombineering, named as 'single stranded oligo repairing (ssOR)', or oligonucleotide-mediated

recombineering, which can be used to create single base changes, inserts and substitute short DNA sequences, and generate deletions.[195,205] The repair by oligos that hybridized with the lagging strand template (lagging) is consistently higher than repair by oligos that hybridize with the leading strand template (leading) (Figure A-8).[195,205] The phage annealing proteins, RecT or Redβ, mediated ssOR-directed mutagenesis allow rapidly and efficient introducing point mutations in BACs or chromosomes by a counterselection-based recombineering method.[195,206,207] A major problem in counterselection is intramolecular recombination; hence some improvements have been made to reduce these unwanted events by expressing Redβ alone at crucial step, and enhance the efficiency by using phosphothioated oligonucleotides.[206]

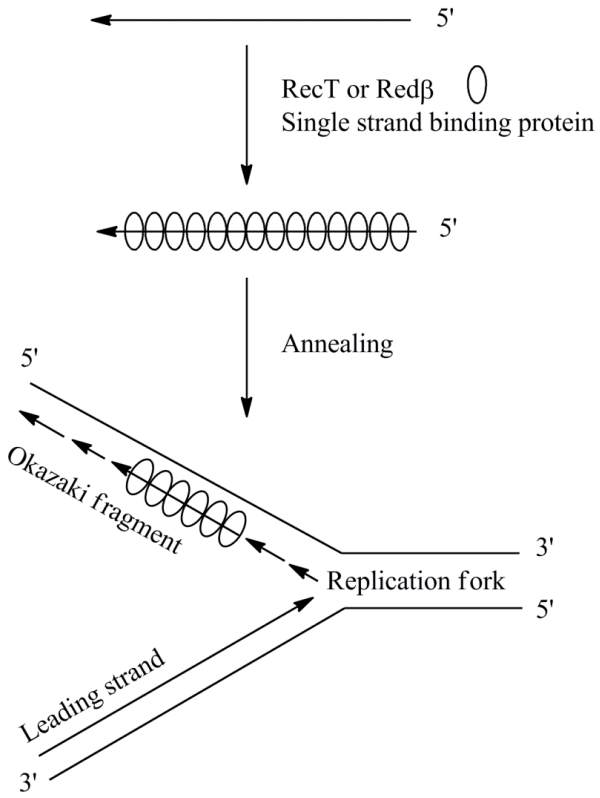


Figure A-8. Current model for oligonucleotide-mediated recombineering. RecT or Redβ promotes annealing of an oligonucleotide to ssDNA of the lagging strand template of a replication fork. Following elongation by DNA polymerase I and ligation to neighboring Okazaki fragments, the oligonucleotide is incorporated into the DNA. (The figure adapted from Ref.208)

Owing to its short homology requirements and high efficiency, Red/ET recombineering can be extensively utilized in DNA engineering.[208] It can be used to accurately insert selectable or nonselectable markers in plasmids, bacterial chromosomal DNA or BACs without size limits, to generate gene-targeting constructs for making reporter-tagged and conditional alleles in embryonic stem cells.[209] It can also be used to generate subtle alterations in BACs or bacterial chromosomal DNA by counterselection recombineering

without site-specific recombination system.[194] Moreover, Red/ET recombineering can be applied to clone and subclone certain DNA regions of interest from a DNA source (e.g., plasmid, BAC, genomic DNA), into a plasmid. [84]

5.4.2 Red/ET recombineering mediated engineering of biosynthetic gene clusters

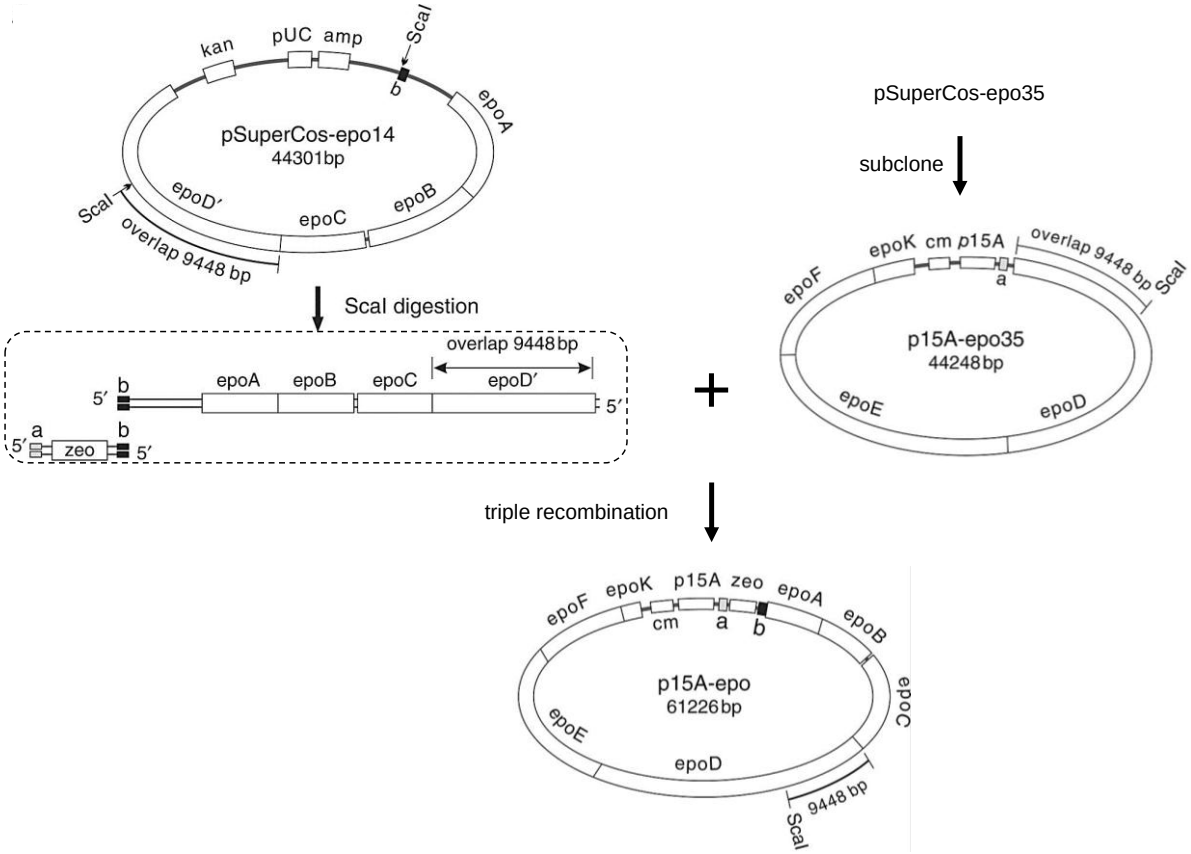


Figure A-9 Diagram for stitching of the epothilone gene cluster. pSuperCos-epo35 was retrofitted with the p15A origin and the chloramphenicol resistance gene to remove the pSuperCos backbone and introduce the short homology arm 'a' to create p15A-epo35. pSuperCos-epo14 was digested with *Sca*I and the *epoA-D* genes were recombined into p15A-epo35 by triple recombination using a bridging zeocin (*zeo*) resistance gene with a 5′ end homology arm to p15A-cm-epo35 ('a') and a 3′ end homology arm to a region in front of *epoA* in pSuperCos-epo14 ('b'). (The figure adapted from Ref.92)

Using Red/ET recombineering, the large gene cluster spanning in several cosmids can be stitched into a single plasmid or BAC, which is simple and convenient for the further modification and delivery. Then genetic elements required for the transfer and integration into the chromosome were introduced, if necessary, the original promoter was exchanged with a promoter which is available in the unrelated host. Figure A-9 shows the stitching of the 61 kb epothilone gene cluster from two cosmids by triple recombination.[92] The stitching was based on two rounds of recombineering and the generation of two intermediates (*zeo* PCR product and linear *epoA-D* fragment) (Figure A-9). The subsequent insertions of genetic elements for gene transfer, stable integration, and regulated expression can be completed by further Red/ET recombineering. The most important advantage of this method is that the

whole secondary metabolite pathway is located on one construct, which can be transferred into the heterologous host strain once. As recombineering is based on homologous recombination, one might have expected difficulties using the described method in the assembly of repetitive sequences. Nevertheless, in related report employing gene clusters up to 80 kb in size demonstrated that highly repetitive sequences can be recombined efficiently.[87]

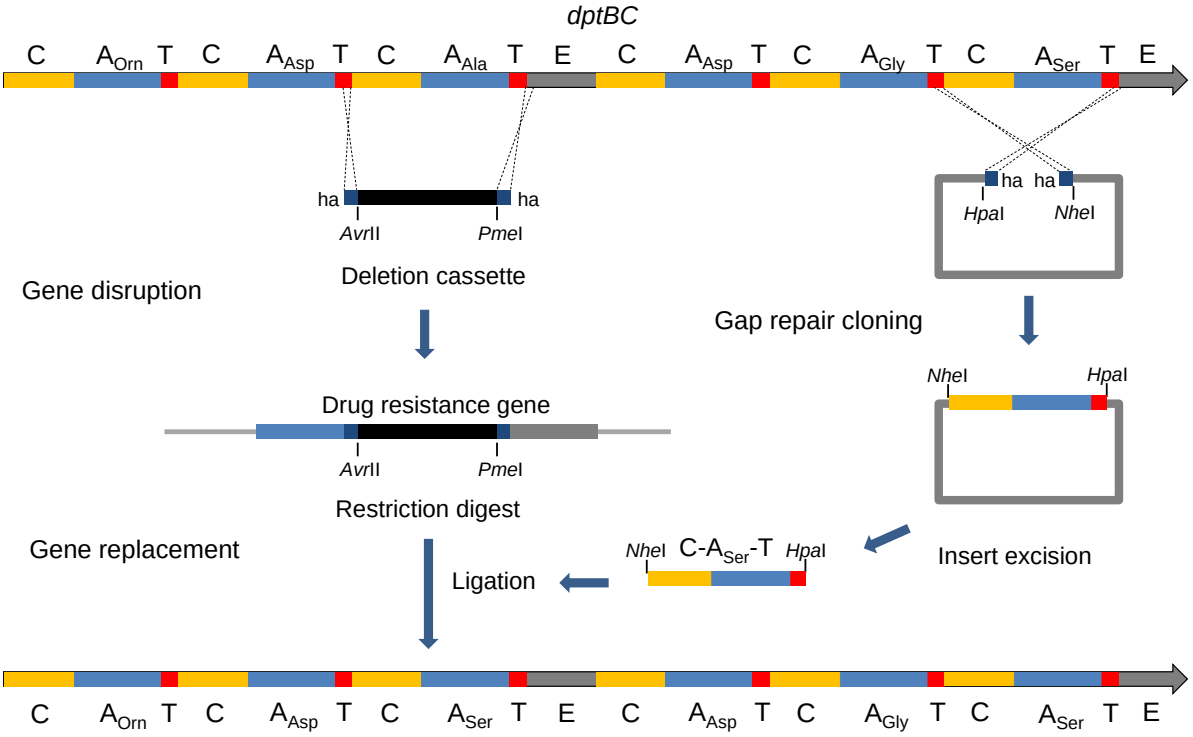


Figure A-10. Red/ET recombineering mediated module exchange for construction of hybrid *dptBC in daptomycin biosynthetic pathway*. ha means 45-bp homology arm. (The figure adapted from Ref.145)

The other important application of Red/ET recombineering is the engineering of large gene clusters. The seamless and precise recombination of a sequence of interest flanked by homologous 'arms' to any targeted location on a replicating DNA molecule is extremely important for engineering of PKS/NRPS assembly lines, because in some case even the alteration of one base pair would lead to complete loss of product. The modular/domain exchange, deletion, or insertion, and precise point mutations can be accomplished by Red/ET recombineering mediated strategy. For example, in the module exchange of daptomycin biosynthetic gene cluster (Figure A-10),[145] the targeted module was initially replaced with a durg resistance gene flanked with rare restriction sites *Avr*II or *Pme*I by Red/ET recombineering (gene disruption). The cassette was then excised, and the linearized plasmid was used for ligation. In the meanwhile, the DNA fragment coding for another CAT domains was cloned by the 'gap-repair cloning' method using Red/ET recombineering, the suitable

restriction sites, *Hpa*I and *Nhe*I, were also added in both termini. The gap repair cloned fragment was cut off and used in ligation with the linearized targeted plasmid by the compatible restriction sites for gene replacement. This strategy leads to the remains of splicing sites in the T-C linker, the T-E linker or the E-C linker and further incorporation of amino acid residues in these regions. But these subtle alterations seemingly have no big effects on the final expression.[145] Using this method for module/domain substitution combined with deletion of tailoring genes, 120 novel lipopeptides were generated based on daptomycin and A54154, and over 40 novel lipopeptide antibiotics were produced in sufficient quantities for facile isolation, confirmation of structure, and analysis of antibacterial activities.[186] Importantly, some of the compounds displayed the desired properties of antibacterial activities similar to those of daptomycin, minimal or no inhibition by surfactant, and low mammalian toxicity.

For the deletion of post-assembly-line tailoring enzymes, one round of Red/ET recombination is harnessed to replace a target gene with a antibiotic resistance gene flanked by site-specific recombination sites (i.e., LoxP, FRT)[89,210] or rare restriction sites,[91] and subsequent site-specific recombination (Cre-LoxP or Flp-FRT) (Figure A-11) or religation to remove selection marker for in-frame deletion, but this resulted in a short 'scar' sequence remaining.

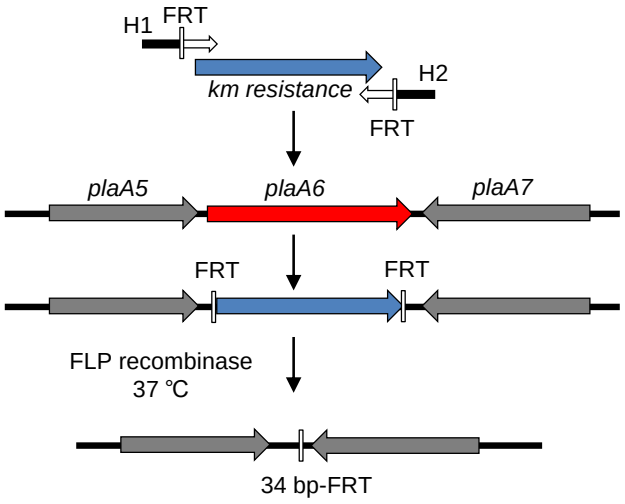


Figure A-11. Deletion of the glycosyltransferase encoding gene *plaA6* by using a kanamycin-resistance cassette that was flanked by FRT recognition sites. Red/ET recombineering was used to replace *plaA6* with the kanamycin-resistance cassette. Incubation with FLP recombinase at 37 °C led to a markerless deletion of *plaA6* in the *pla* cluster. (The figure adapted from Ref. 89)

The emergence of Red/ET recombineering extremely promotes the engineering of large natural product biosynthetic gene cluster in *E. coli*. Coupled to efficient heterologous expression, it will facilitate the natural product bioprospecting at this post-genomic era.

6 Outline of my dissertation

The aim of this thesis was to develop technology to improve discovery and modification of bacterial natural products based on the increasing numbers of sequenced bacterial genomes. Together with the progress of DNA engineering techniques, such as Red/ET recombineering, the large secondary metabolite biosynthetic gene clusters can now be rapidly cloned and engineered. When these steps are followed by robust expression in a suitable heterologous system, the products encoded by cryptic biosynthetic gene clusters can be identified, and even optimized by combinatorial biosynthetic engineering. The present thesis involves rapid cloning, genetic engineering, and heterologous expression of several bacterial secondary biosynthetic gene clusters via Red/ET recombineering. In particular, luminmide and luminmycin biosynthetic pathways from an entomopathogenic bacterium *Photorhabdus luminescens* subspecies *laumondii* strain TT01 (DSM15139), and syringolin biosynthetic pathway from *Pseudomonas syringae* pv. *syringae* (DSM1242) were investigated. In addition, an improved technique for modification of biosynthetic pathway by *ccdB* counterselection recombineering was introduced.

Laborious and time-consuming DNA library construction and screening are required for recovering large genomic sequences of interest, e.g. secondary metabolite biosynthetic pathways. In **Section B Chapter 1**, we describe direct DNA cloning based for the first time to discover that the full-length Rac prophage protein RecE (866 amino acids) and its partner RecT mediated highly efficient homologous recombination between two linear DNA substrates, which is mechanistically distinct from conventional replication-dependent recombineering meditated by Redαβ from lambda phage or truncated version of RecET. This finding had immediate implication for cloning of linear DNA constructs from bacterial artificial chromosomes, cDNA and genomic DNA into linear vectors in *E. coli*. The capability of direct cloning was demonstrated by successfully cloning ten cryptic natural product biosynthetic gene clusters (each 10-52 kb in length) from digested genomic DNA of *P. luminescens* into expression vectors. When these cloned PKS/NRPS gene clusters were transformed into *E. coli* GB05-MtaA and Nissle 1917, two of them, *plu3263* and *plu1881-plu1877*, produced new groups of compounds at an immediate detectable level (>0.1mg/L). They were designated as luminmides and luminmycins, respectively.

The luminmycin biosynthetic pathway shows high similarity with that of glidobactin in *Burkholderia* DSM2029 except for the absence of a homologue of the biosynthetic gene *glbH*. This gene encodes a dioxygenase which has been proposed to add the -OH group to the lysine

in the12-membered ring. Addition of *glbH* into the luminmycin gene cluster *plu1881-plu1877* leads to partial conversion from luminmycin A to glidobactin A which shows higher cytoxic and antifungal activity. Luminmycins A-C were isolated from the heterologous host and their structures were elucidated using 2D NMR and HRESIMS in **Section B Chapter 2** (Figure A-12). Luminmycin A is a dehydroxy derivative of the previously reported glidobactin A, while luminmycins B and C most likely represent its acyclic biosynthetic intermediates. Luminmycin A showed cytotoxicity against human colon carcinoma HCT-116 cell line with an IC_{50} value of 91.8nM while acyclic luminmycin B was inactive at concentrations as high as 100 μg/mL.

Plu3263 is encoded by a 15.6 kb NRPS gene and it contains five NRPS modules and produces a group of cyclic peptides, luminmides A-G, when it expressed in *E. coli* GB05-MtaA. These luminmides were identified by HRMS, direct and inverse feeding experiments and 2D-NMR in **Section B Chapter 1** and **Chapter 4** (Figure A-12). The third module (C/E-A_3-T) can activate and incorporate five different amino acids, Phe, Leu, Try, Met, and Ala, into the final cyclic pentapeptides, which suggest the C/E and A domain in this module exhibits broad substrate specificity and the exnyme system also shows downstream toleration for altered substrates in subsequent steps. **Section B Chapter 3** describes the efficient point mutation of a multi-copy plasmid and BACs by using *ccdB* counterselection recombineering. In **Section B chapter 4**, we employed *ccdB* counterselection recombineering to create single or double mutations at the 'specificity-conferring codes' of A domain in a complex NRPS gene *plu3263* efficiently. Followed by subsequent robust expression of mutants in *E. coli*, the minor product luminmide B was changed to be prominent in several mutants without significant reduction of product yields, even higher (A301G) than the wild type, while the dominating luminmide A in the wild type Plu3263 was decreased to different extent in all the mutants. This fact showed the feasibility of alteration of the specificity of A3 domain by site-directed mutagenesis.

The syringolin biosynthetic gene cluster (*sylCDE*, 19 kb) from a digested genomic DNA mixture of *P. syringae* was directly cloned into a plasmid in which *sylCDE* is under the control of an inducible promoter by one step linear plus linear homologous recombination (LLHR) in *E. coli* in **Section B chapter 5**. After expression in *E. coli* GB05-MtaA, two new syringolin derivatives (sryingolin G and H) were discovered (Figure A-12). The complete syringolin gene cluster was assembled by addition of *sylAB* and exchange of a synthetic bidirectional promoter against the native promoter to drive *sylB* and *sylC* expression by using

Red/ET recombineering. The varied production distribution of syringolin derivatives showed the different efficiencies of native and synthetic promoters in *E. coli*.

The direct cloning and engineering of natural product biosynthetic pathways by Red/ET recombineering coupled with suitable heterologous expression hosts represents an effective approach to investigate or optimize known and unknown biosynthetic pathways discovered by bacterial genome sequencing projects, especially from slow growing bacteria and such for which genetics tools are only poorly established.

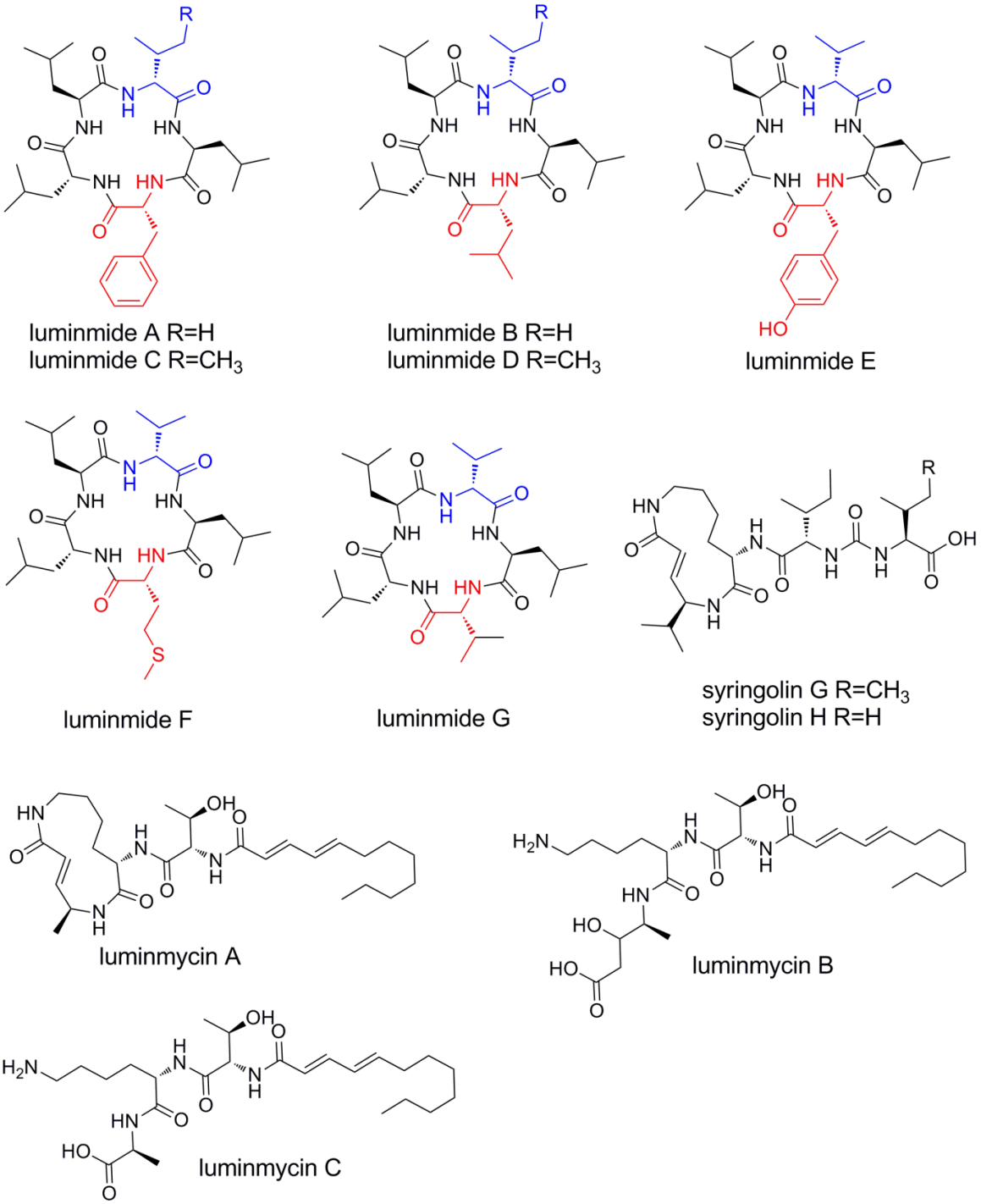


Figure A-12. Structures of newly identified natural products in this thesis.

B. Publications

Chapter 1

Full-length RecE Enhances Linear-linear Homologous Recombination and Facilitates Direct Cloning for Bioprospecting

Jun Fu*, Xiaoying Bian*, Shengbaio Hu, Hailong Wang, Fan Huang, Philipp M Seibert, Alberto Plaza, Liqiu Xia, Rolf Müller, A Francis Stewart, and Youming Zhang
(* These authors contributed equally to this work.)

Nature Biotechnology, 30, 440–446, **2012**

This article is available online at:

http://www.nature.com/nbt/journal/v30/n5/pdf/nbt.2183.pdf

Note: Supplementary information is available on the Nature Biotechnology website.
http://www.nature.com/nbt/journal/v30/n5/extref/nbt.2183-S1.pdf

Chapter 2

Luminmycins A-C, Cryptic Natural Products from *Photorhabdus luminescens* Identified by Heterologous Expression in *Escherichia coli*

Xiaoying Bian, Alberto Plaza, Youming Zhang, and Rolf Müller

Journal of Natural Products, 75, 1652-1655, **2012**

This article is available online at:

http://pubs.acs.org/doi/pdf/10.1021/np300444e

Supporting information is available free of charge via the Internet at
http://pubs.acs.org/doi/suppl/10.1021/np300444e/suppl_file/np300444e_si_001.pdf.

Chapter 3

Efficient and Seamless DNA Recombineering using *ccdB* for counterselection in *Escherichia coli*

Hailong Wang, Xiaoying Bian, Liqiu Xia, Xuezhi Ding, Rolf Müller, Youming Zhang, Jun Fu, and A. Francis Stewart

Nucleic Acids Res., submitted, **2012**

ABSTRACT

Recombineering, which is the use of homologous recombination for DNA engineering in *Escherichia coli*, usually employs antibiotic selection to identify the intended recombinant. When combined in a second step with counterselection, seamless products can be obtained. Most counterselection options employ drug selection. Here, we report the advantages of a non-drug strategy using *ccdB* as the counterselectable agent. Expression of CcdB is toxic to *E. coli* in the absence of the CcdA antidote so the counterselection step is initiated by the removal of CcdA expression. Because counterselection strategies necessarily differ according to the copy number of the target, we describe two variations. For multi-copy targets, we employ two *E. coli* hosts so that counterselection is exerted by the transformation step that is needed to separate the recombined and unrecombined plasmids. For single copy targets, we put the *ccdA* gene onto the temperature-sensitive, pSC101 Red expression plasmid so that counterselection is exerted by the standard temperature shift to remove the expression plasmid. Hence this drug-free counterselection method is less work than existing methods and will be useful for genetic engineering not only BACs and plasmids, but also the *E. coli* chromosome.

INTRODUCTION

DNA engineering methods are central to molecular biology. However, the original, revolutionary, 'cut and paste' methods based on restriction enzymes and DNA ligations are limited to engineering exercises for small DNA molecules. Recombineering using phage protein-mediated homologous recombination in *E. coli* (1-10) was developed to engineer cloned DNA molecules of all sizes. Recombination occurs through homology regions, which are stretches of identical DNA sequence shared by the two molecules that recombine. Because the homology regions can be chosen freely, recombineering is not dependent on the location of restriction sites and any position on a target molecule can be specifically altered.

Recombineering is mediated by transient expression of the phage recombinase pairs, either RecE/RecT from the Rac prophage (1,11,12) or Redα/Redβ from λ phage (2-5,11-13). RecE and Redα are 5′-3′ exonucleases (14,15), and RecT and Redβ are DNA annealing proteins (16-18). Linear DNAs, either double-stranded, usually in the form of PCR products, or single-stranded synthetic oligonucleotides (19,20) are introduced by electroporation and provide the substrates to introduce genetic change adjacent to the region of homologous recombination. Interaction between RecE and RecT, or Redα and Redβ, facilitates double stranded homologous recombination (11), however, only the annealing protein is required for

recombination promoted by single stranded oligonucleotides (19,20). Recombineering is convenient because efficient recombination can be achieved with short lengths of perfect sequence identity, typically 50 bp or less, and the adjacent regions of non-homology can range from 1bp to >50kb (1,12). Hence the homology regions can be easily incorporated into PCR primers by oligonucleotide synthesis and virtually any mutation or insertion can be achieved. Recombineering is now widely employed to engineer cloned DNA in all commonly used vectors (e.g. BACs, fosmids, and plasmids) and several prokaryotic chromosomes. Applications include subcloning of precisely defined sections by gap repair (3), oligonucleotide-directed mutagenesis (19,20), BAC engineering for gene targeting (21-23), high-throughput DNA engineering (24-27), and a variety of other precise applications.

Seamless mutagenesis refers to site-directed mutagenesis without any other nearby change, such as the presence of the selectable gene used to introduce the mutation. Mutagenesis in a protein coding region is a clear example of the need for seamless DNA engineering because any extraneous sequence introduced during the mutagenic step could interfere with expression. Seamless mutagenesis has been achieved using a two-step selection/ counterselection strategy, which first involves the replacement of the target site with a selectable cassette such as an antibiotic resistance gene accompanied by a counterselectable gene. The cassette is then subsequently replaced seamlessly with the desired sequence by selecting against the counterselectable gene usually involving the administration of a small molecule, such as streptomycin or a sugar (1,28-37). Popular options of counterselectable markers include *sacB* (1,28), *rpsL* (29-32), as well as markers that can, in the right host background, both be selected for and against including *galK* (33) , *thyA* (34), and *tolC* (35).

Bacterial toxin-antitoxin (TA) systems are based on closely linked genes that together encode a protein poison and an antidote. The best characterized TAs operate to maintain plasmids. Typically the toxin is stable and the antitoxin is unstable so if the plasmid encoding the TA pair is lost, the host will die (38-40). The *ccdA* and *ccdB* TA pair was found in the *ccd* (control of cell death) operon of the *E. coli* F plasmid where it serves to maintain this low copy plasmid (41,42), by a mechanism known as post-segregational killing (38-40). Binding of CcdB to the GyrA subunit of DNA gyrase blocks the passage of DNA and RNA polymerases and leads to double-strand DNA breakage and cell death (43-47). The CcdA antidote (72 amino acids) prevents the CcdB toxin (101 amino acids) from inhibiting gyrase by forming a tight CcdA-CcdB complex. Upon losing the F plasmid, CcdA is quickly degraded by Lon protease so that the concentration of CcdA decreases faster than CcdB, leaving free CcdB to bind gyrase and kill the cells (48-50).

TA systems have been implemented in several biotechnology applications. Most directly, they have been used to maintain plasmids in industrial culture (51,52) and to report ligation efficiency through loss of function. The *ccdB* gene has been used as a counterselectable marker in a number of commonly used applications (53-57).

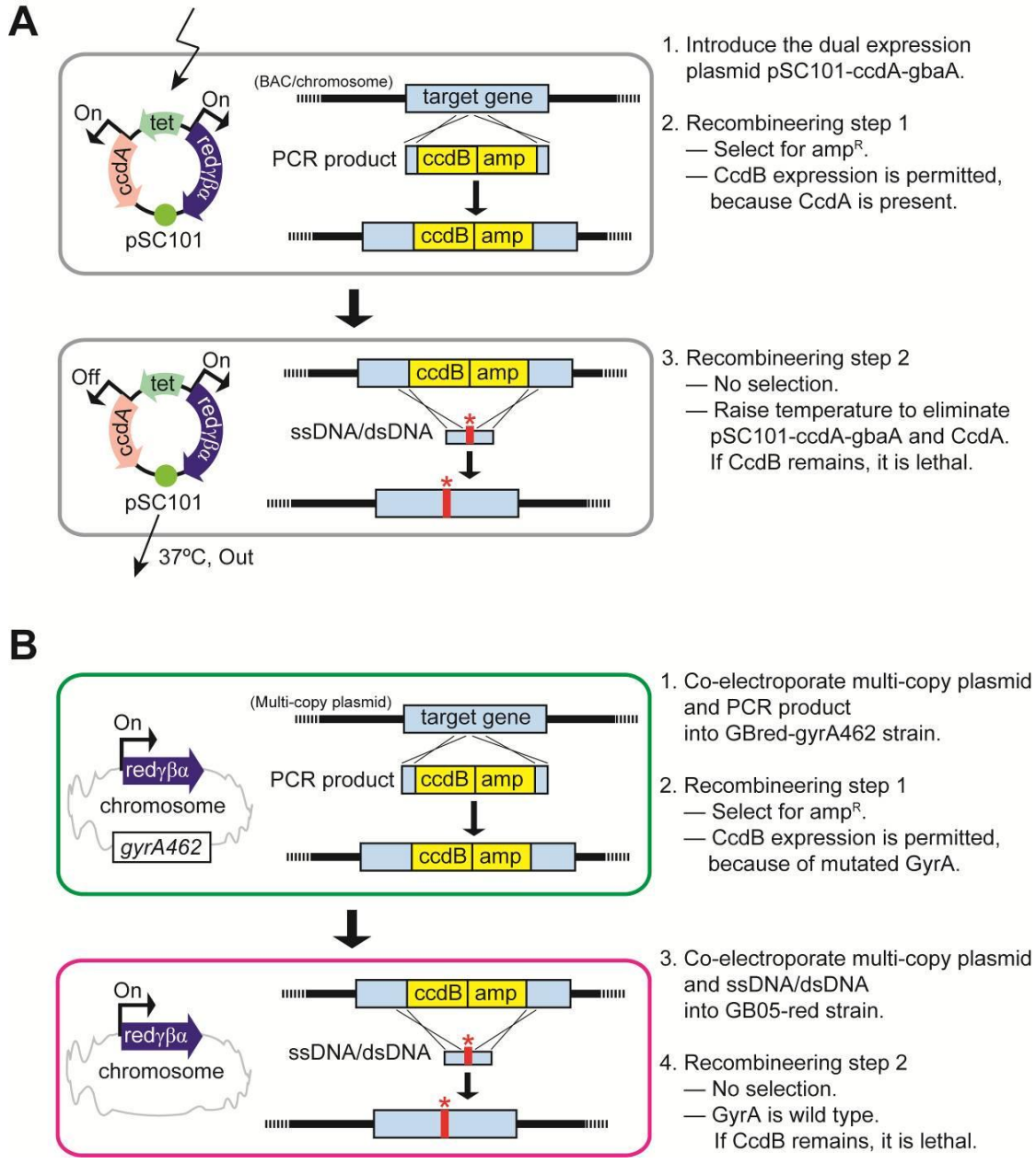


Figure 1. Overview of *ccdB* counterselection strategy. **(A)** Strategy for BAC or chromosome recombineering using *ccdB* counterselection. First, pSC101-ccdA-gbaA was transformed into *E. coli* cells. Subsequently, Redα/Redβ and CcdA expression were induced by L-rhamnose and by L-arabinose respectively. Electro-competent cells were prepared and electroporated with the linear targeting molecule containing *ccdB-sm* (here *amp*). The *ccdB-amp* PCR product has two 40 bp homology arms to the target molecule. Correct recombinants were obtained after plating on LB plates containing ampicillin and L-arabinose, and incubated at 30°C. Keep the expression of CcdA at 30°C until the *ccdB* gene was removed. *E. coli* cells harboring correct *ccdB-amp* recombinants and pSC101-ccdA-gbaA were grown in LB medium in the presence of ampicillin and L-arabinose at 30°C, and induced with L-rhamnose for Redα/Redβ, RecA and Redγ expression. Then, electro-competent

cells were prepared and electroporated with another linear targeting molecule or oligo, containing the DNA of interest and two same 40 bp homology arms as *ccdB-amp* PCR product. Stop the expression of CcdA and Redα/Redβ (no inducers) at 37°C. Correct recombinants were obtained after plating on LB plates (for BAC modification, appropriate antibiotics need to be added, while not for chromosome modification) at 37°C. **(B)** Strategy for multi-copy plasmid recombineering using *ccdB* counterselection. Redα/Redβ expression in GBred-gyrA462 was induced by L-arabinose. Electro-competent cells were prepared and electroporated with the target multi-copy plasmid together with the linear targeting molecule containing *ccdB-amp*. The *ccdB-amp* PCR product has two 40 bp homology arms to the target plasmid. Correct recombinants were obtained after plating on LB plates containing ampicillin. Redα/Redβ expression in GB05-red was induced by L-arabinose. Electro-competent cells were prepared and electroporated with the recombinant plasmid together with another linear targeting molecule or oligo containing the DNA of interest and two same 40 bp homology arms as *ccdB-amp* PCR product. Correct recombinants were obtained after plating on LB plates containing appropriate antibiotics.

In this paper, we describe the use of *ccdB* counterselection in combination with recombineering for efficient and seamless DNA engineering in *E. coli*. This two-step recombineering procedure first involves the replacement of the target site with a *ccdB*-antibiotic resistance gene cassette by selection for antibiotic resistance. The first step needs to be performed either with CcdA expression or in a CcdB-resistant *E. coli* strain, which has the $gyrA462_{Arg\rightarrow Cys}$ mutation. Because counterselection protocols must be different for single or multiple copy situations, we describe two versions of the strategy. For single copy targets such as BACs or the *E. coli* chromosome, we put the *ccdA* gene on our standard recombineering expression plasmid. For multi-copy targets we employ stable *E. coli* strains.

MATERIALS AND METHODS

Bacterial strains, plasmids and culturing conditions

E. coli strains and plasmids used in this work are listed in Supplementary Table 1. *E. coli* strains were maintained in LB medium at 30°C or 37°C, and selected with appropriate antibiotics (chloramphenicol [*cm*], 15 μg/ml; ampicillin [*amp*], 100 μg/ml; gentamycin [*genta*], 3 μg/ml and tetracycline [*tet*], 5 μg/ml). The concentration of L-arabinose and L-rhamnose used for induction was 1.4 mg/ml.

Recombineering

Recombineering was performed as previously described by Fu *et al*. (58). The selection of recombinants was carried out depending on the selection marker in the cassette. PCRs were performed with Phusion polymerase (New England Biolabs, GmbH, Frankfurt am Main, Germany) according to the manufacturer's protocol. Only lagging oligos were used in the counterselection step to achieve higher efficiencies of single strand oligonucleotide repair (ssOR) (19,20).

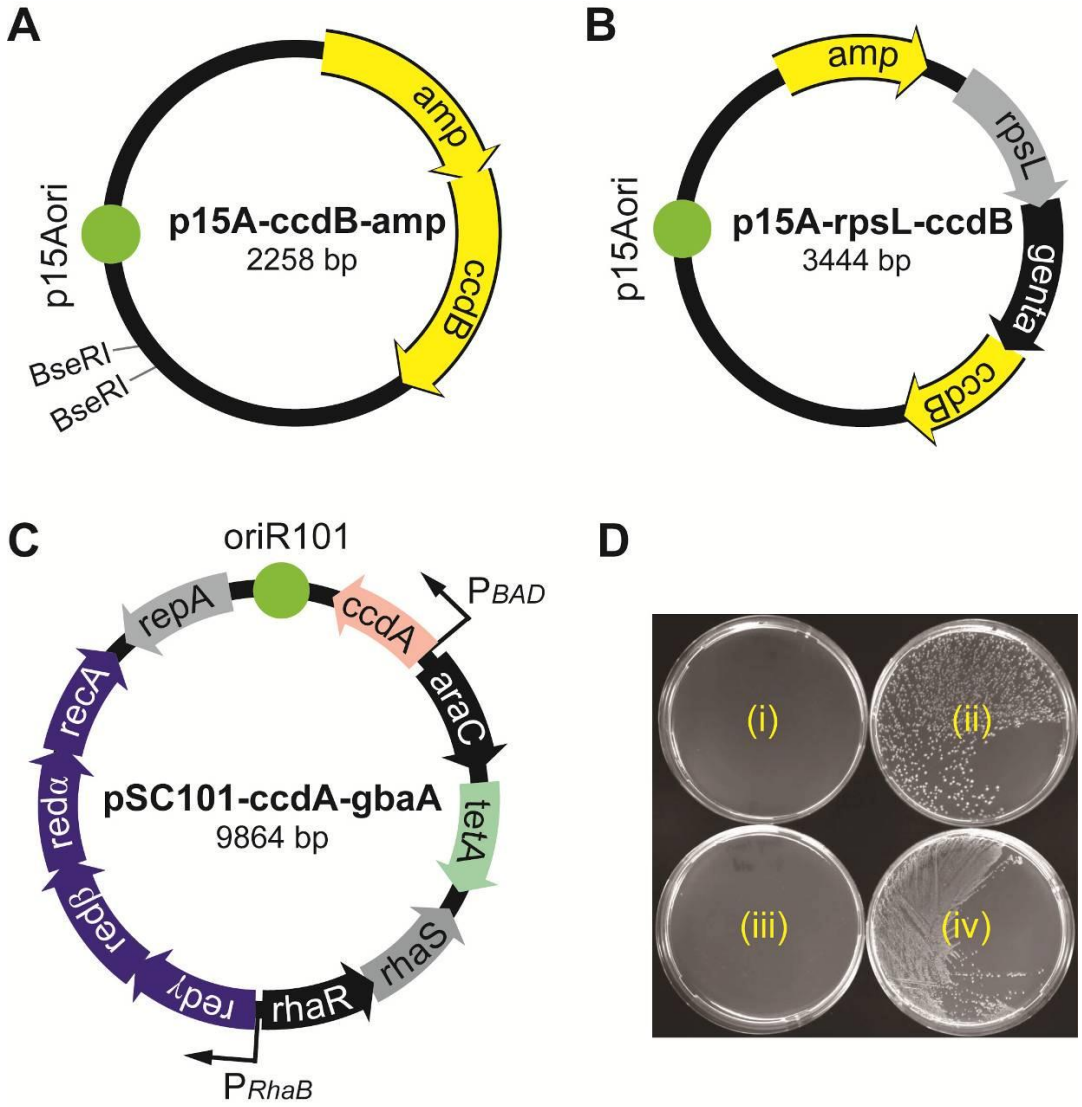


Figure 2. Schematic presentation of the plasmids used. **(A)** p15A-ccdB-amp contains the *ccdB-amp* fusion gene. **(B)** p15A-rpsL-ccdB contains ampicillin resistance gene (*amp*), *rpsL*, gentamycin resistance gene (*genta*) and *ccdB*. Both A and B carry the *p15A* replication origin. **(C)** pSC101-ccdA-gbaA was constructed from pRedFlp (39). *ccdA* is expressed from the L-arabinose-inducible BAD promoter (P_{BAD}) and *redγβαA* is expressed from the L-rhamnose-inducible RhaB promoter (P_{RhaB}). **(D)** **(i)** and **(ii)** are function test of the cassette of *rhaS-rhaR*-P_{RhaB}-*redγβαA* in pSC101-ccdA-gbaA. The function test was performed by co-electroporating a linear PCR fragment and a plasmid into L-rhamnose induced (ii) or uninduced (i) GB2005 cells harboring pSC101-ccdA-gbaA. The linear PCR fragment has two 50 bp homology arms to the plasmid. After recovery, 100 µl of culture were plated on LB plates supplemented with appropriate antibiotics. The results showed that the function of *redγβαA* is fine and L-rhamnose-inducible RhaB promoter is stringent. (**iii)** and (**iv)** are function test of the cassette of *araC*-P_{BAD}-*ccdA* in pSC101-ccdA-gbaA. The function test was performed by electroporating p15A-ccdB-amp into L-arabinose induced (iv) or uninduced (iii) GB2005 cells harboring pSC101-ccdA-gbaA. After recovery, 10 µl of culture were plated on LB plates containing ampicillin plus L-arabinose (iv) or not (iii) respectively. The results showed that the function of *ccdA* is fine and L-arabinose-inducible BAD promoter is stringent.

Construction of dual inducible expression plasmid pSC101-ccdA-gbaA

pSC101-ccdA-gbaA (Figure 1A and Figure 2C) carries the *ccdA* gene under the control of the arabinose-inducible P_{BAD} promoter, and *redα, redβ* and *redγ* genes of the λ phage together

with the *recA* gene (*redγβαA*) in a polycistronic operon under the control of the rhamnose-inducible P_{RhaB} promoter. In the presence of L-rhamnose, Redα/Redβ expression in *E. coli* cells allows efficient homologous recombination between linear DNA fragments and chromosome or BAC. While, in the presence of L-arabinose, CcdA is expressed, conferring CcdB resistance to *E. coli* cells (Figure 1A bottom and Figure 2D).

pSC101-ccdA-gbaA is a derivative of pSC101 replicon, which is a low copy number (~5 per cell) and thermo sensitive plasmid. It replicates at 30°C, but not at 37°C (59). Consequently, it can be easily eliminated from the host by temperature shift in the absence of selection. Since pSC101-ccdA-gbaA is based on pSC101 replicon, it can be propagated in *E. coli* together with BACs and most colE1 origin-derived plasmids.

Construction of p15A-ccdB-amp (cm) and p15A-rpsL-ccdB

p15A-ccdB-amp (Figure 2A) and p15A-ccdB-cm were constructed by recombineering. They are derivatives of p15A replicon, which contain different resistance genes followed by *ccdB*. They can be used as the template to amplify the ccdB-antibiotic resistance gene cassette by PCR reactions. p15A-rpsL-ccdB (Figure 2B) was derived from p15A-ccdB-amp by inserting *rpsL-genta* cassette between *ccdB* and *amp*.

Generation of CcdB-resistant *E. coli* GBred-gyrA462 strain

50 pmol mixture of synthetic single-stranded lagging and leading oligos, containing the mutation of $gyrA_{Arg462Cys}$ (Figure 3A) was electroporated into recombinase induced *E. coli* GB05-red (58). The cells were incubated at 37°C overnight after electroporation. Electro-competent cells were prepared and electroporated with a test plasmid, p15A-rpsL-ccdB (Figure 2B), to fish out the CcdB-resistant *E. coli* mutant GBred-gyrA462. Finally, the test plasmid was cured by adding 500 μg/ml streptomycin into the medium. The recombineering efficiency of GBred-gyrA462 was tested using our standard protocol and compared with GB05-red and DB3.1 harboring pSC101-BAD-gbaA (60).

Oligos used for generating GBred-gyrA462 strain are listed below. The homology arm sequence is in lowercase. The mutation site is in uppercase and bold, and another two bases in underlined uppercase are synonymous mutations designed to reduce mismatch repair.

gryA1 5'-gtcgagcagtttttcgtgctcaagaccggtcagtttctgcaG**GCA**caAatccagaatcgcctgagcttgctgttc ggtcaggtagtac-3'

gyrA2 5'-gtactacctgaccgaacagcaagctcaggcgattctggatTtg**TGC**Ctgcagaaactgaccggtcttgagcac gaaaaactgctcgac-3'

Using *ccdB* counterselection to modify pBeloBAC11

The general strategy for BAC or chromosome recombineering using *ccdB* counterselection

is depicted in Figure 1A. Briefly, the dual inducible expression plasmid pSC101-ccdA-gbaA was electroporated into *E. coli* DH10B harboring pBeloBAC11. The linear targeting molecule containing *ccdB-amp* was amplified from BseRI digested p15A-ccdB-amp by PCR using oligonucleotides BACccdB-amp1 (5'-aatacgactcactatagggcgaattcgagctcggtacccgTTTG TTTATTTTTCTAAATAC-3') and BACccdB-amp2 (5'-ctcaagcttgcatgcctgcaggtcgactctagag gatcccAGCCCCATACGATATAAGTTG-3').

As shown in Figure 4A, in the first round of recombineering, the correct recombinant, pBeloBAC11-ccdB-amp, was obtained after plating on LB plates containing Amp and L-arabinose. The resulting recombinants were analyzed by restriction digestion of EcoRI.

In the second round of recombineering, 50 pmol of synthetic lagging oligos (5'-ctcaagcttgcatgcctgcaggtcgactctagaggatccccgggtaccgagctcgaattcgccctatagtgagtcgtatt-3') was electroporated. After recovery, the cultures were diluted 100 times and 100 µl was plated on LB plates supplemented with *Cm*, 0.1 mM IPTG and 40 µg/ml X-Gal. The plates were incubated at 37°C overnight and the number of colonies was counted.

Making a single base pair substitution in *plu3263* in pBR322 vector

Two rounds of recombineering were used to introduce a point mutation in *plu3263* in a pBR322 vector (Figure 5A). The general strategy for multi-copy plasmid recombineering using *ccdB* counterselection is depicted in Figure 1B. Briefly, the target plasmid pGB-plu3263 together with the PCR products of *ccdB-cm*, amplified from the BseRI digested p15A-ccdB-cm, were coelectroporated into recombinase induced GBred-gyrA462 competent cells, which are resistant to CcdB and have the artificial arabinose inducible Red operon on the chromosome. The recombinant plasmid pGB-plu3263-PCP3-ccdB-cm was selected on LB plates containing *Cm* and *Amp*. The correct clones were identified by restriction analysis, and transformed into GBred-gyrA462 again to obtain pure population of the recombinants. The function of the *ccdB* gene of these clones was tested by electroporation into the gyrase wide type GB2005 strain.

Oligos used for PCR amplification of the *ccdB-cm* cassette are listed below. Sequence as homology arm for recombineering is in lowercase.

3263PCP3ccdBcm5 5'-tttgaacaggttggccgacatgacagtttctttgccttgggcggtcacTGTGACGGAA GATCACTTCGCAG-3'

3263PCP3ccdBcm3 5'-ccaatcctatacgccgtaaacgttcgatcatcctgactgccaacagcgTTTGTTCAAA AAAAAGCCCGCTC-3'

In the second round of recombineering, 50 pmol of synthetic single-stranded lagging oligos, 3263SPCP3A-B (5'-caggcccaatcctatacgccgtaaacgttcgatcatcctgactgccaacag**CGC**gtgac

cgcccaaggcaaagaaactgtcatgtcggccaacctgttcaaca-3', the mutation site is in uppercase and bold) and the pGB-plu3263-PCP3-ccdB-cm plasmid DNA were co-electroporated into the recombinase induced GB05-Red competent cells. The recombinant plasmid pGB-plu3263M was screened on LB plates containing *Amp*. The correct recombinants were identified by restriction analysis, following verification by sequencing. Three correct pGB-plu3263M clones were subsequently electroporated into *E. coli* GB05-MtaA (12,61) to check the production of luminmides (the products of Plu3263).

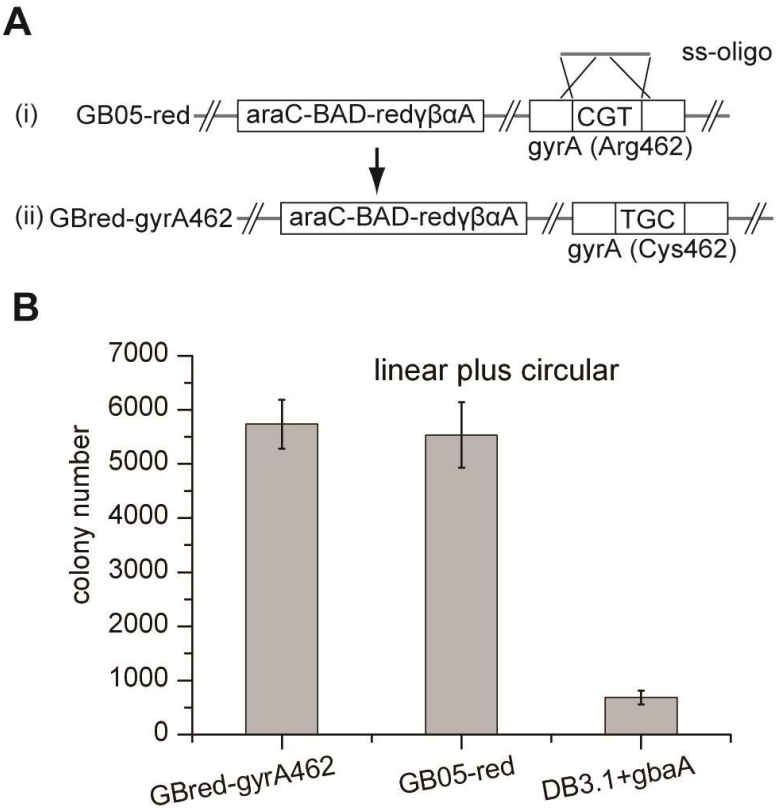


Figure 3. Generation of *E. coli* GBred-gyrA462. **(A)** Schematic presentation of generation of CcdB-resistant *E. coli* GBred-gyrA462. The Arg462-coding codon CGT was changed into Cys-coding codon TGC. **(B)** Comparing the recombineering efficiency of GBred-gyrA462 with GB05-red and DB3.1 harboring pSC101-BAD-gbaA. The recombineering was performed by co-electroporating a linear PCR fragment and a plasmid into recombinase induced *E. coli* cells. The linear PCR fragment has two 50 bp homology arms to the plasmid. The recombineering efficiency was indicated by the colony number on LB plates supplemented with appropriate antibiotics. Results are from three independent experiments. Error bars represent the standard deviation.

Cultivation, extraction and HPLC/MS analysis

E. coli GB05-MtaA containing pGB-plu3263 or pGB-plu3263M was inoculated from overnight cultures (2%) into 5 ml LB medium with *Amp* in 15 ml glass tubes. The expression of Plu3263 was induced with *Tet* (0.5 μg/ml) at 4 hours after inoculation, the cultures were incubated for another 4 hours and 2% of XAD-16 absorber resin was added. The cultures

were continued to be incubated at 30°C for 24 hours. The cells and the resin were harvested by centrifugation and extracted with acetone and methanol. Solvents were removed *in vacuo* by Genevac, and the residue was dissolved in 500 µl methanol. An aliquot of 5 µl was analyzed by HPLC-MS; an Agilent 1100 series solvent delivery system coupled to Bruker HCTplus ion trap mass spectrometer. The chromatographic conditions were: RP column Nucleodur C18, 125 by 2 mm, 2.5 µm particle size, and precolumn C18, 8×3 mm, 5 µm. Solvent gradient (with solvents A [water and 0.1% formic acid] and B [acetonitrile and 0.1% formic acid]) from 5% B to 95% B within 20 min, followed by 3 min with 95% B at a flow rate of 0.4 ml/min. Detection was carried out in positive and negative ion models.

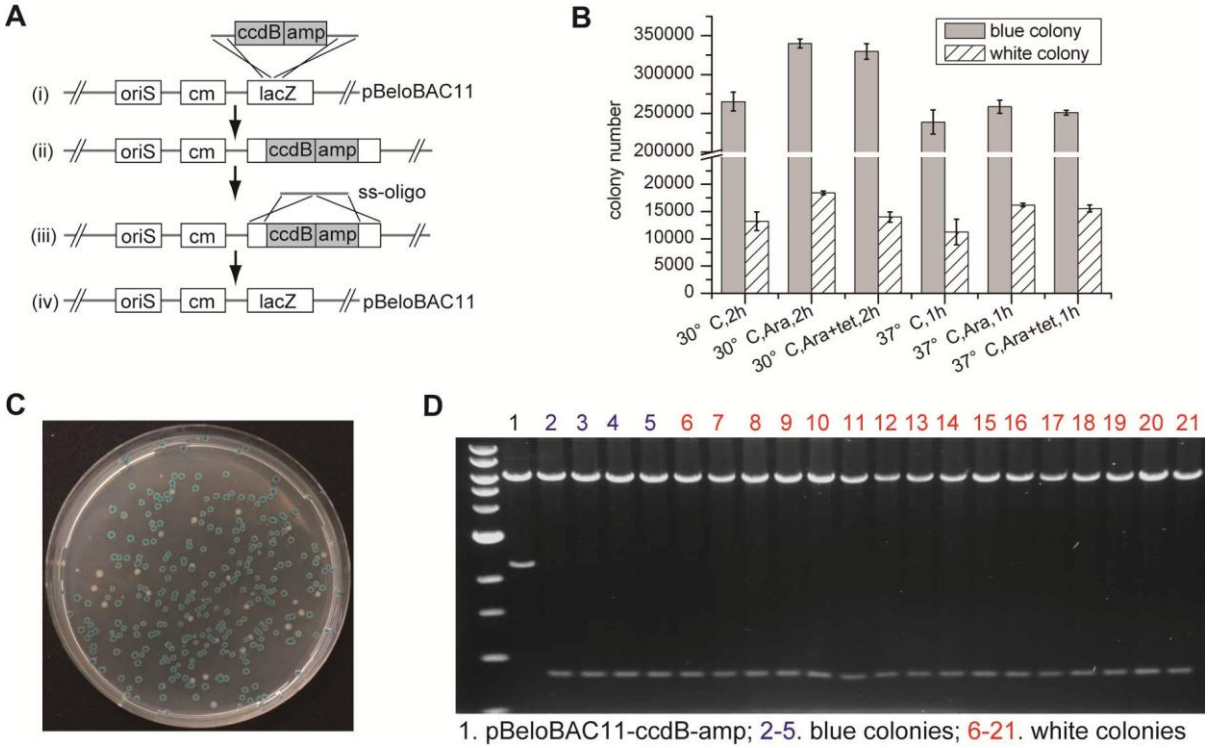


Figure 4. Point mutation generated in BAC using *ccdB* counterselection**. (A)** Schematic presentation of using *ccdB* counterselection to modify pBeloBAC11. **(B)** The recombineering efficiency of different recovery methods in the *ccdB* counterselection step. The efficiency is indicated by the number of blue colonies and white colonies on LB plates. Results are from three independent experiments. Error bars represent the standard deviation. **(C)** A picture of the colonies on the LB plate supplemented with *Cm*, IPTG and X-Gal after *ccdB* counterselection. About 95% colonies are blue and about 5% are white. **(D)** EcoRI restriction analysis of BAC DNAs from counter-selected blue colonies and white colonies. The BAC DNAs from white colonies have the same digestion pattern as those from blue colonies.

RESULTS AND DISCUSSION

Overall strategy

The principles of recombineering with *ccd* counterselection are illustrated in Figure 1. For

single copy targets, the *ccdA* gene is expressed from the same expression plasmid as the Red genes for recombineering. In pSC101-ccdA-gbaA, *ccdA* is expressed from the arabinose--inducible P_{BAD} promoter and the Redγβα*A* operon from the rhamnose-inducible P_{RhaB} promoter. The plasmid is introduced into a host containing the target followed by arabinose and rhamnose induction and electroporation with a *ccdB*-antibiotic resistance gene cassette (here ampicillin) flanked by homology arms for recombineering. After selection for *Amp* resistance, the culture is induced only with rhamnose, electroporated with DNA containing the seamless mutation (either double stranded DNA or a single stranded oligonucleotide) and maintained at 37°C to eliminate the expression plasmid. Cells that have eliminated the *ccdB* gene will survive, whereas cells that have not will die from CcdB toxicity.

For multi-copy targets, there is an implicit limitation with counterselection. If a host cell contains both the intended product without the counterselectable gene and the parental plasmid with the counterselectable gene, it will die under counterselection pressure. In other words, the presence of the counterselectable gene is dominant and will occlude recovery of the intended product. Hence counterselection with plasmids is best exerted by including a transformation step so that each host cell can only harbor a single type of plasmid. To incorporate the required transformation step in an optimized counterselection strategy, we made a gyrA462 mutation in the recombineering strain, GB05-red, which has the arabinose--inducible P_{BAD}-Redγβα*A* operon integrated into the chromosome (58). Shortly before making cells for electroporation, GBred-gyrA462 cells are induced with arabinose, then co-electroporated with the target plasmid and the *ccdB-cm* cassette flanked by homology arms for recombineering. After selection for *Cm* and plasmid-mediated resistance, plasmids are isolated and co-electroporated with DNA containing the seamless mutation (either double stranded DNA or a single stranded oligonucleotide) into arabinose induced GB05-red and selected for plasmid-mediated resistance. Cells harboring plasmids that have eliminated the *ccdB* gene will survive.

To achieve methods based on *ccdB* counterselection, we built pSC101-ccdA-gbaA and GBred-gyrA462. The performance of these reagents was verified by functional tests (Figure 2 and 3). Then we applied the ccdB counterselection strategies to (a) evaluate efficiencies in a BAC-based reporter assay; and (b) make a point mutation in a plasmid that we wanted for our other research goals.

Generation of CcdB-resistant *E. coli* GBred-gyrA462

We found that recombineering in the gyrA462 strain, *E. coli* DB3.1 (Invitrogen GmbH, Karlsruhe, Germany) was not so efficient. Hence, we introduced gyrA462 into our

recombineering-proficient *E. coli* strain GB05-red by oligonucleotide-directed mutagenesis (Figure 3A). The Arg462-coding codon CGT of gyrA was changed into Cys-coding codon TGC to create GBred-gyrA462. The recombineering efficiency of GBred-gyrA462 is comparable with GB05-red and higher than DB3.1 harboring pSC101-BAD-gbaA (Figure 3B).

Using *ccdB* counterselection to modify pBeloBAC11

A functional test based on β-galactosidase expression from pBeloBAC11 was used to evaluate the efficiency and fidelity of *ccdB* counterselection for BAC recombineering (Figure 4A). The *lacZ* gene of pBeloBAC11 was first disrupted by the *ccdB-amp* cassette. Then, in the second round of recombineering, the *ccdB-amp* cassette was replaced with an 80nt oligonucleotide by counterselection against CcdB. CcdA expression was terminated by removing arabinose and culturing at 37°C to eliminate pSC101 replication.

After counterselection, restoration of the *lacZ* gene was scored by blue versus white colonies on LB plates supplemented with IPTG and X-Gal (Figure 4C). We used this test to evaluate the relationship between CcdA removal and CcdB toxicity by employing several recovery methods with different parameters (Figure 4B). All variations worked well indicating robustness of the method. Although the recombineering efficiency after recovery at 30°C for 2 hours was a little higher than 37°C for 1 hour, the latter is more convenient.

After counterselection, about 95% of the colonies on LB plates supplemented with IPTG and X-Gal were blue, and only about 5% were white (Figure 4C). We picked 16 white colonies and 4 blue colonies for restriction analysis (Figure 4D). Unexpectedly, all 16 white clones had the same digestion pattern as the 4 blue clones suggesting that the white clones were also successfully repaired but other mutations had been introduced by the synthetic oligos.

Mutation of a single base pair in *plu3263* cloned in pBR322

The luminmide biosynthetic gene *plu3263* was identified in the genome sequence of *Photorhabdus luminescens* TT01 DSM15139 (62). It encodes a NRPS (Non-ribosomal peptide synthetase) composed of 5 modules. *plu3263* was directly cloned into pASK3 vector (pBR322 origin, IAB, Gottigen) by using linear plus linear homologous recombination (12) to form pGB-plu3263, in which *plu3263* is under the control of a tetracycline inducible promoter. The luminmides can be produced in *E. coli* GB05-MtaA (12,61). The invariant serine2742 residue (OH group) in the signature sequence (GGD/HSL) of PCP3 domain in Plu3263 is the active site for attachment of the 4'-phosphopantetheine cofactor (63), which catalyzes the conversion of *apo*-PCP to *holo*-PCP. If this serine residue is mutated, the

function of PCP3 domain should be completely lost and no luminmides will be produced. The serine-coding codon 2742, TCG, was changed to GCG generating S2742A (Figure 5A). In the first round of recombineering, a 1 bp deletion was introduced into codon 2742 in addition to insertion of the *ccdB-cm* cassette. In this step, 10 out of 12 random checked colonies on the selection plates were shown to be mixtures of pGB-plu3263 and pGB-plu3263-PCP3-ccdB-cm as expected (data not shown). The recombinant and unmutated plasmids were separated by retransformation into *E.coli* GBred-gyrA462. First we checked 12 random colonies to find 6 clones of pGB-plu3263-PCP3-ccdB-cm. These were electroporated into gyrase wild type *E. coli* GB2005 to check the function of *ccdB* gene and all 6 clones killed the host indicating that CcdB was expressed as expected.

pGB-plu3263-PCP3-ccdB-cm was then co-electroporated into GB05-red with an oligo to replace the *ccdB-cm* cassette. We checked 24 colonies and found 15 to be correct (data not shown). Like most NRPS genes, *plu3263* contains many repeated sequences. The remaining 9 clones were all due to intramolecular recombination between the repeated sequences. This background is inherent to counterselection, because any mutation that ablates expression of the counterselectable gene will be selected. Sequencing showed that all 6 pGB-plu3263M candidates examined were correctly mutated (Figure 5B). Three correct pGB-plu3263M clones were subsequently electroporated into *E. coli* GB05-MtaA to check for the production of luminmides (the products of Plu3263). HPLC-MS analysis showed that the mutations abolished luminmide production (Figure 5C).

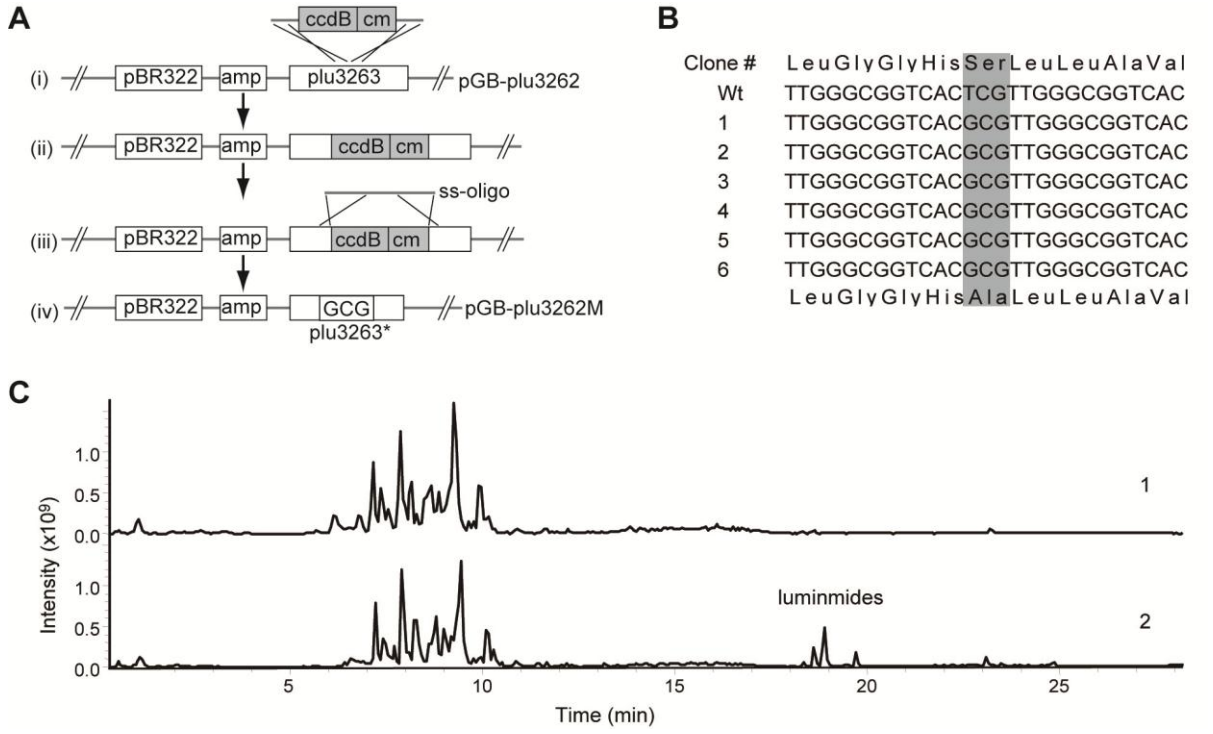


Figure 5. Mutation and expression of luminmide biosynthetic gene *plu3263*. **(A)** Schematic presentation of making a single base pair substitution in *plu3263*. The Ser2742-coding codon TCG was changed into Ala-coding codon GCG. **(B)** Sequence analysis of the point-mutation region. All clones had the intended substitution (shadow). **(C)** HPLC-MS analysis (base peak chromatogram (BPC), *m/z* = 200 to 1100 + All MS) of *E. coli* GB05-MtaA containing pGB-plu3263M (line 1) and pGB-plu3263 (line 2). The products of Plu3263, luminmides (Retention Time=18-20 min) were completely abolished.

CONCLUSION

The development of recombineering has enabled a wide variety of DNA engineering applications. Amongst these, methods based on consecutive steps of insertion by selection and replacement by counterselection are prominent because they permit seamless mutagenesis. Seamless mutagenesis is an ideal DNA engineering goal, so work to develop better counterselection applications continue to be developed. We first used *sacB* for counterselection in recombineering (1,28), which is based on addition of a high concentration of sucrose. However, sucrose counterselection is often inefficient with false positives commonly surviving. Furthermore the *sacB* gene is present in the vector backbones of pBACe3.6 and pTARBAC series, making it unsuitable for use with many BAC constructs. To find a better counterselection system we employed *rpsL*, which conveys sensitivity to streptomycin (20). In our hands this is more useful than *sacB*/sucrose counterselection. Other labs have employed other counterselection systems including *galK*/2-deoxy-galactose+glycerol (33), *thyA*/trimethoprim+thymine (34), and *tolC*/colicin E1 (35). Our experience with these counterselection systems indicates that they are not better than *rpsL*/streptomycin and in some cases involve certain inconveniences such as a need for special media and/or unusual selection reagents. Use of *galK*, *thyA* or *tolC* requires additional knockout steps to generate *ΔgalK* (33), *ΔthyA* (34), *ΔtolC* (35) mutant strains respectively. Consequently these counterselection systems are not portable and BACs must be transformed into the mutant strains. All these counterselection systems often require experiment-to-experiment titration to find the optimal concentration of the conditional toxin for the counterselection window, which can be quite small.

Here we explored the merits of a genetic counterselection system based on the *ccdA*/*ccdB* toxin/antitoxin system. The system is more convenient than a chemical based system and as shown in Figure 4, appears to be effective across a wide range of conditions thus reducing the need to optimize the counterselection step for each different application. We also developed different reagents and protocols for single and multiple copy applications, which may also reduce the need to adapt the counterselection window for each experiment. Consequently we are confident that counterselection with *ccdA*/*ccdB* adds to the recombineering repertoire and

will be a helpful improvement. Recently we described an improved method for counterselection with oligonucleotides based on the reduction of intramolecular recombination at the counterselection step by expression of only Redβ without Redα (32). Here again, as with many counterselection applications, we found that intramolecular recombination was a prominent source of background (Figure 5). To combine the advantages of that method with the *ccdA*/*ccdB* method, plasmids and strains are under construction. The combination should further improve the general utility of seamless mutagenesis via recombineering.

SUPPLEMENTARY DATA

A table showing the *E. coli* strains used is available at NAR Online:

ACKNOWLEDGEMENTS

The authors would like to thank Professor Laurence Van Melderen for sharing the plasmid pKT-ccdA-ccdB. X.B. received a fellowship of China Scholarship Council (CSC).

FUNDING

This work was supported by European Union's 7th Framework Program EUCOMMTOOLS [261492 to A.F.S.]; Research in the laboratory of R.M. was funded by the DFG and BMBF. Y.Z. was funded by the BMBF (MiPro).

REFERENCES

1. Zhang, Y., Buchholz, F., Muyrers, J.P.P. and Stewart, A.F. (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.*, **20**, 123-128.

2. Muyrers, J.P.P., Zhang, Y., Testa, G. and Stewart, A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.*, **27**, 1555-1557.

3. Zhang, Y., Muyrers, J.P.P., Testa, G. and Stewart, A.F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.*, **18**, 1314-1317.

4. Yu, D., Ellis, H.M. and Lee, E. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **97**, 5978-5983.

5. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K12 using PCR products. *Proc. Natl. Acad. Sci. USA*, **97**, 6640-6645.

6. Muyrers, J.P.P., Zhang, Y. and Stewart, A.F. (2001) Techniques: recombinogenic engineering-new options for cloning and manipulating DNA. *Trends Biochem. Sci.*, **26**, 325-331.

7. Copeland, N.G. and Jenkins, N.A. (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.*, **2**, 769-779.

8. Court, D.L., Sawitzke, J.A. and Thomason, L.C. (2002) Genetic engineering using homologous recombination. *Annu. Rev. Genet.*, **36**, 361-388.

9. Sawitzke, J.A., Thomason, L.C., Costantino, N., Bubunenko, M. and Datta, S. (2007) Recombineering: in vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol.*, **421**, 171-199.

10. Sharan, S.K., Thomason, L.C. and Kuznetsov, S.G. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc*., **4**, 206-223.

11. Muyrers, J.P.P., Zhang, Y., Buchholz, F. and Stewart, A.F. (2000) RecE/RecT and Redα/Redβ initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev*., **14**, 1971-1982.

12. Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F. et al. (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol*., **30**, 440-446.

13. Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y. and Stewart, A.F. (2010) Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol. Biol.*, **11**, 54.

14. Kovall, R. and Matthews, B.W. (1997) Toroidal structure of lambdaexonuclease. *Science*, **277**, 1824-1827.

15. Zhang, J., Xing, X., Herr, A.B. and Bell, C.E. (2009) Crystal structure of *E. coli* RecE protein reveals a toroidal tetramer for processing double-stranded DNA breaks. *Structure*, **17**, 690-702.

16. Thresher, R.J., Makhov, A.M., Hall, S.D., Kolodner, R. and Griffith, J.D. (1995) Electron microscopic visualization of RecT protein and its complexes with DNA. *J. Mol. Biol*., **254**, 364-371.

17. Passy, S.I., Yu, X., Li, Z., Radding, C.M. and Egelman, E.H. (1999) Rings and filaments of β protein from bacteriophage λ suggest a superfamily of recombination proteins. *Proc. Natl. Acad. Sci. USA*, **96**, 4279-4284.

18. Erler, A., Wegmann, S., Elie-Caille, C., Bradshaw, C.R., Maresca, M., Seidel, R., Habermann, B., Muller, D.J. and Stewart, A.F. (2009) Conformational adaptability of Redβ during DNA annealing and implications for its structural relationship with Rad52. *J. Mol. Biol.*, **391**, 586-598.

19. Ellis, H.M., Yu, D., DiTizio, T and Court, D.L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA*, **98**, 6742-6746.

20. Zhang, Y., Muyrers, J., Rientjes, J. and Stewart, A.F. (2003) Phage annealing proteins promote oligonucleotide-directed mutagenesis in *Escherichia coli* and mouse ES cells. *BMC Mol. Biol.*, **4**, 1.

21. Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J. et al. (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat. Biotechnol*., **21**, 652-659.

22. Yang, Y. and Seed, B. (2003) Site-specific gene targeting in mouse embryonic stem cells with intact bacterial artificial chromosomes. *Nat. Biotechnol*., **21**, 447-451.

23. Testa, G., Zhang, Y., Vintersten, K., Benes, V., Pijnappel, W.W., Chambers, I., Smith, A.J., Smith, A.G. and Stewart, A.F. (2003) Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat. Biotechnol*., **21**, 443-447.

24. Sarov, M., Schneider, S., Pozniakovski, A., Roguev, A., Ernst, S., Zhang, Y., Hyman, A.A. and Stewart, A.F. (2006) A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nat. Methods*, **3**, 839-844.

25. Poser, I., Sarov, M., Hutchins, J.R.A., Hériché, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B. and Bird, A.W. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods*, **5**, 409-415.

26. Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T. et al. (2011) A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*, **474**, 337-342.

27. Sarov, M., Murray, J.I., Schanze, K., Pozniakovski, A., Niu, W., Angermann, K., Hasse, S., Rupprecht, M., Vinis, E., Tinney, M. et al. (2012) A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. *Cell*, **150**, 855-866.

28. Muyrers, J.P.P., Zhang, Y., Benes, V., Testa, G., Ansorge, W. and Stewart, A.F. (2000) Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO reports*, **1**, 239-243.

29. Imam, A.M., Patrinos, G.P., de Krom, M., Bottardi, S., Janssens, R.J., Katsantoni, E., Wai, A.W., Sherratt, D.J. and Grosveld, F.G. (2000) Modification of human β-globin locus PAC clones by homologous recombination in *Escherichia coli*. *Nucleic Acids Res*., **28**, e65.

30. Wang, S., Zhao, Y., Leiby, M. and Zhu, J. (2009) A new positive/negative selection scheme for precise BAC recombineering. *Mol. Biotechnol*., **42**, 110-116.

31. Westenberg, M., Soedling, H.M., Mann, D.A., Nicholson, L.J. and Dolphin, C.T. (2010) Counter-selection recombineering of the baculovirus genome: a strategy for seamless modification of repeat-containing BACs. *Nucleic Acids Res*., **38**, e166.

32. Bird, A.W., Erler, A., Fu, J., Heriche, J.K., Maresca, M., Zhang, Y., Hyman, A.A. and Stewart, A.F. (2012) High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. *Nat. Methods*, **9**, 103-109.

33. Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. and Copeland, N.G. (2005) Simple and highly efficient BAC recombineering using *galK* selection. *Nucleic Acids Res.*, **33**, e36.

34. Wong, Q.N.Y., Ng, V.C.W., Lin, M., Kung, H., Chan, D. and Huang, J.D. (2005) Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in *Escherichia coli*. *Nucleic Acids Res*., **33**, e59.

35. DeVito, J.A. (2008) Recombineering with *tolC* as a Selectable/Counter-selectable Marker: remodeling the rRNA Operons of *Escherichia coli*. *Nucleic Acids Res*., **36**, e4.

36. Jamsai, D., Orford, M., Nefedov, M., Fucharoen, S., Williamson, R. and Ioannou, P.A. (2003) Targeted modification of a human β-globin locus BAC clone using GET Recombination and an I-SceI counterselection cassette. *Genomics*, **82**, 68-77.

37. Tischer, B.K., von Einem, J., Kaufer, B. and Osterrieder, N. (2006) Two-step Red-mediated recombination for versatile high-efficiency markerless DNA manipulation in *Escherichia coli*. *BioTechniques*, **40**, 191-197.

38. Engelberg-Kulka, H. and Glaser, G. (1999) Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol*., **53**, 43-70.

39. Gerdes, K. (2000) Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J. Bacteriol*., **182**, 561-572.

40. Hayes, F. (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*, **301**, 1496-1499.

41. Ogura, T. and Hiraga, S. (1983) Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. USA*, **80**, 4784-4788.

42. Jaffe, A., Ogura, T. and Hiraga, S. (1985) Effects of the *ccd* function of the F plasmid on bacterial growth. *J. Bacteriol*., **163**, 841-849.

43. Bernard, P. and Couturier, M. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol*., **226**, 735-745.

44. Couturier, M., Bahassi, E.M. and Van Melderen, L. (1998) Bacterial death by DNA gyrase poisoning. *Trends Microbiol*., **6**, 269-275.

45. Van Melderen, L. (2002) Molecular interactions of the CcdB poison with its bacterial target, the DNA gyrase. *Int. J. Med. Microbiol*., **291**, 537-544.

46. Dao-Thi, M.H., Van Melderen, L., De Genst, E., Afif, H., Buts, L., Wyns, L. and Loris, R. (2005) Molecular basis of gyrase poisoning by the addiction toxin CcdB. *J. Mol. Biol*., **348**, 1091-1102.

47. De Jonge, N., GarciaPino, A., Buts, L., Haesaerts, S., Charlier, D., Zangger, K., Wyns, L., De Greve, H. and Loris, R. (2009) Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Mol. Cell*, **35**, 154-163.

48. Van Melderen, L., Bernard, P. and Couturier, M. (1994) Lon dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid free segregant bacteria. *Mol. Microbiol.*, **11**, 1151-1157.

49. Van Melderen, L., Thi, M.H.D., Lecchi, P., Gottesman, S., Couturier, M. and Maurizi, M.R. (1996) ATP-dependent degradation of CcdA by Lon protease. *J. Biol. Chem.*, **271**, 27730-27738.

50. Afif, H., Allali, N., Couturier, M. and Van Melderen, L. (2001) The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison–antidote system. *Mol. Microbiol*., **41**, 73-82.

51. Szpirer, C.Y. and Milinkovitch, M.C. (2005) Separate-component-stabilization system for protein and DNA production without the use of antibiotics. *BioTechniques*, **38**, 775-781.

52. Peubez, I., Chaudet, N., Mignon, C., Hild, G., Husson, S., Courtois, V., De Luca, K., Speck, D. and Sodoyer, R. (2010) Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. *Microb. Cell Fact.*, **9**, 65.

53. Bernard, P., Gabarit, P., Bahassi, E.M. and Couturier, M. (1994) Positive-selection vectors using the F plasmid ccdB killer gene. *Gene*, **148**, 71-74.

54. Betton, J.M. (2004) Cloning vectors for expression PCR products. *BioTechniques*, **37**, 346-347.

55. Le Roux, F., Binesse, J., Saulnier, D. and Mazel, D. (2007) Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene vsm by use of a novel counterselectable suicide vector. *Appl. Environ. Microbiol*., **73**, 777-784.

56. Chen, S., Songkumarn, P., Liu, J. and Wang, G.L. (2009) A versatile zero background T-vector system for gene cloning and functional genomics. *Plant Physiol*., **150**, 1111-1121.

57. Mandi, N., Kotwal, P. and Padmanabhan, S. (2009) Construction of a novel zero background prokaryotic expression vector: potential advantages. *Biotechnol. Lett*., **31**, 1905-1910.

58. Fu, J., Teucher, M., Anastassiadis, K., Skarnes, W. and Stewart, A.F. (2010) A recombineering pipeline to make conditional targeting constructs. *Methods Enzymol*., **477**, 125-144.

59. Hashimoto-Gotoh, T. and Sekiguchi, M. (1977) Mutations of temperature sensitivity in R plasmid pSC101. *J. Bacteriol*., **131**, 405-412.

60. Wang, J., Sarov, M., Rientjes, J., Hu, J., Hollak, H., Kranz, H., Xie, Y., Stewart, A.F. and Zhang, Y. (2006) An improved recombineering approach by adding RecA to λ Red recombination. *Mol. Biotechnol*., **32**, 43-53.

61. Bian, X., Huang, F., Stewart, F.A., Xia, L., Zhang, Y. and Müller, R. (2012) Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering. *ChemBioChem*, **13**, 1946-1952.

62. Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M. and Charles, J.F. *et al* (2003) The genome sequence of the entomopathogenic bacterium Photorhabdus luminescens. *Nat. Biotechnol*., **21**, 1307-1313.

63. Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M., Kruft, V. and Wittmann-Liebold, B. (1991) An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem*., **266**, 23135-23141.

Supplementary data

Supplementary Table S1. Strains and plasmids used in this work

Strains and plasmids	**Genotype or relevant features**
E.coli strains	
GB2005	(HS996, ΔrecET, ΔybcC). The endogenous recET locus and the DLP12 prophage ybcC, which encodes a putative exonuclease similar to the Redα, were deleted
DB3.1	(*gyrA462*$_{Arg\rightarrow Cys}$, Invitrogen GmbH, Karlsruhe, Germany). Used for propagating plasmids containing *ccdB*
GB05-red	(GB2005, *araC*-BAD-αβγA) lambda *red* operon and *recA* under BAD promoter were inserted at the ybcC locus
GBred-gyrA462	(GB05-red, *gyrA462*) GyrA mutation of Arg462Cys
GB05-MtaA	(GB2005, *mtaA-genta*) a pPant transferase MtaA from myxobacterium *Stigmatella aurantiaca* DW4/3-1 was randomly transposed into the chromosome
Plasmids	
p15A-ccdB-amp	p15A replicon. PCR templates to amplify *ccdB-amp* cassette (plasmid DNA digested with BseRI)
p15A-rpsL-ccdB	p15A replicon. The test plasmid used to generate *E.coli* GBred-gyrA462
pSC101-BAD-gbaA	pSC101 replicon. Tetracycline resistance, lambda *red* operon and *recA* (*redγβαA*) under BAD promoter
pSC101-ccdA-gbaA	pSC101 replicon. Tetracycline resistance, *ccdA* under BAD promoter and *redγβαA* under RhaB promoter

Chapter 4

Efficient Site-directed Mutagenesis of Adenylation Domain Alters Relative Yields of Luminmide Derivatives *in vivo*

Xiaoying Bian, Silke C. Wenzel, Alberto Plaza, Youming Zhang, and Rolf Müller

ChemBioChem, to be submitted, **2012**

Abstract

Cloning and engineering a natural product biosynthetic pathway combined with robust heterologous expression in a tractable host is a proven approach for combinatorial biosynthesis of microbial secondary metabolites. Herein, we employ *ccdB* counterselection combined with oligonucleotide-mediated recombineering to efficiently create point mutations in a complex NRPS gene *plu3236* located on a multi-copy plasmid in *E. coli*. Targeting the active site residues of NRPS A domains, single and double mutations were rationally constructed within the Plu3263 A3 domain which is part of a nonribosomal peptide synthetase producing luminmides. The mutations lead to an alteration of the predominant product from luminmide A in the wild type pathway to luminmide B in several mutants as a result of specifically altering the specificity code of the domain. These results demonstrate the suitability of counterselection recombineering for site-directed mutagenesis of complex NRPS biosynthetic pathways in *E. coli* and to modify the gatekeeper role of A domains resulting in specific alterations of pathway products.

Introduction

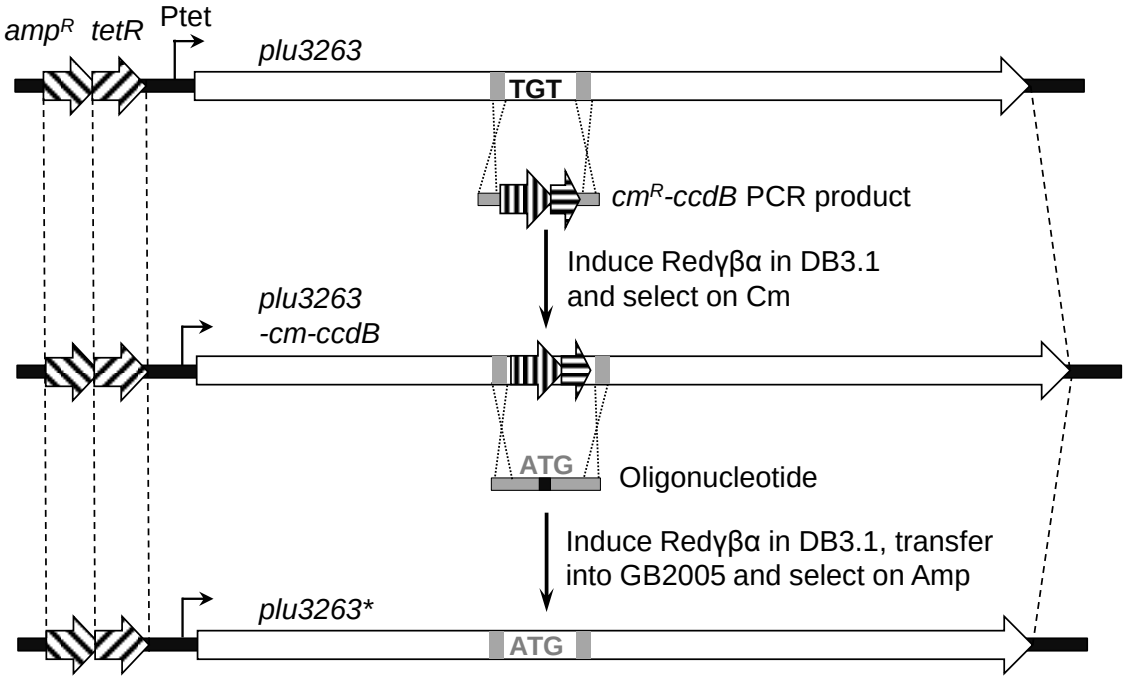
Nonribosomal peptide synthetases (NRPSs) are large multimodular and multifunctional enzymes that specifically bind, activate and condense amino acids to form peptide bonds in a template-directed fashion[1]. These megaenzymes are responsible for the biosynthesis of polypeptides exhibiting a variety of biological activities, such as antibiotics[2, 3], antitumor drugs[4], immunosuppressives[5], siderophores[6], and toxins[7]. Each NRPS module represents a functional building unit required for the recognition and activation of one amino acid monomer as well as incorporation into the growing peptide chain through condensation of a C-N bond[8]. The primary method for combinatorial biosynthesis involves the alteration of enzymatic NRPS modules which can lead to a corresponding change in the structural features of peptides[9]. A NRPS module consists of a minimal set of domains, the peptide bond-forming condensation (C) domain and the amino acid adenylation (A) domain, as well as a phosphopantetheine primed thiolation (T) domain (also known as a peptidyl carrier protein, PCP). The adenylation (A) domain that selects and activates specific amino acid from a large pool of substrates is the ‘gatekeepers’ of the modular ‘assemble-lines’[8]. Therefore, the alteration of an A domain in the assembly line by genetic engineering is a feasible method to manipulate the peptide backbone. The crystal structure of a Phe-activating A-domain, PheA[10], from gramicidin S synthetase, combined with phylogenetic analysis and *in vitro* biochemical studies classified the 10 active site residues that are responsible for recognition

and binding the amino acid substrates as 'specificity-conferring codes' or 'nonribosomal code'[11-13]. These codes were derived from characterized A domains, and they also allow us to predict the substrate of unknown A domains from their translated nucleotide sequences[14-16]. Moreover, by changing as few as one or two of these active site residues, it has been shown that it is possible to switch the specificity of A domains from one amino acid to another *in vitro* (e.g. from L-Phe to L-Leu[11,17], to L-Tyr[18] for GrsA, and from L-Phe to L-Ala for TycA[19]) and *in vivo* (e.g. from L-Glu to L-Gln for SrfA-A1[20], from L-Asp to L-Asn for SrfA-B5[20], from L-Val to L-Leu, L-IIe, L-Ala, or to L-Phe for AmdK[21], and (2*S*,3*R*)-3-methyl glutamic acid to non-natural (2*S*,3*R*)-3-methyl glutamine for CdaPS3[22]). The A domain was generally engineered in *E. coli*, and subsequently transferred into the native producer to replace the wild type A domain, which is time-consuming and inefficient for further genetic manipulation, especially for some slow-growing and genetically intractable producers.

For the genetic engineering of natural product biosynthesis genes, it is important to introduce seamless and precise mutations in the sequence of interest, and is particularly important for the modification of PKS and NRPS assembly lines, because in some cases even the alteration of one base pair could lead to complete loss of products. Considering this, subtle modification without any unwanted additional sequence is preferred and can be achieved by the selection/counterselection strategy *in vivo* to remove unexpected mutations which may be caused by *in vitro* site-directed mutagenesis. Counterselection, such as *rpsL* or *sacB*, in combination with oligonucleotide-mediated recombineering is a high efficient approach to introduce single point mutation into plasmids, large bacterial artificial chromosomes (BACs) and *E. coli* chromosomes[23-26]. Recently, we reported the employment of *ccdB* as a counter-selection marker for seamless DNA engineering in *E. coli*. By using this *ccdB* counterselection reombineering strategy, point mutations in BAC and multi-copy pBR322 plasmid were successfully created (Chapter 3).

The luminmide biosynthetic pathway is encoded by *plu3263* located in the chromosome of entomopathogenic bacterium *Photorhabdus luminescens* subspecies *laumondii* strain TT01[27]. The gene *plu3263* was directly cloned by the newly established linear plus linear homologous recombination (LLHR) from a genomic DNA and successfully heterologous expressed in *E. coli* GB05-MtaA to elucidate the products luminmides A and B[28] (Table 1). Almost in parallel, the same products were also identified by MS and stable isotope labeling from the native host and named GameXPeptides A-D[29]. Plu3263 is composed of five NRPS modules. The first, third, and fourth C domains have been assigned by phylogenetic inference as dual

E/C domains. These domains catalyze both epimerization of upstream amino-acyl-*S*-pantetheinyl donor and condensation resulting in the presence of the corresponding D-amino acids residues in the final products independent of any additional epimerization domains. The second standard $^{L}C_{L}$ domain catalyzes the condensation of two L-amino acids giving rise to a L-amino acid at the coresponding position within the luminmide backbone. This phylogenetic assignment of C domains into sub-classes[30] in general enables the determination of the configuration of amino acid residues[28,31,32]. This approach should be used, however, with caution and should be complemented by traditional chemical methods, such as advanced Marfey's method[33,34], for the determination of the absolute configurations where two or more identical amino acids are present in a single cyclic peptide backbone. In the identified luminmides/GameXpeptides, the third module (E/C-A_3-T) of Plu3263 is responsible for activation and incorporation of two different amino acids, Phe and Leu. This suggests broad substrate tolerance within both the E/C and A_3 domains in this module, as well as, a level of substrate tolerance in downstream modules. Guided by specificity-conferring codes, it is possible to alter the NRPS substrate specificity and promiscuity, e.g., by changing one or two amino residues in the amino acid binding pocket of the A3 domain using counterselection recombineering. This approach could result in removing undesirable products or in increasing the amount of favored side product.



Scheme 1. Methodfor *ccdB* counterselection recombineering to create point mutation on *plu3263* construct. A *cm*^{R}*-ccdB* counterselection cassette is introduced to remove the target nucleotides by induction of Redγβα in *E. coli* DB3.1 and selection on chloramphenicol. Then, the counterselection cassette is replaced by oligonucleotides containing a mutation of interest which involved induction of

Redγβα in *E. coli* DB3.1 followed by transformation of plasmids into *E. coli* GB2005 for counterselection of correct mutants (see method).

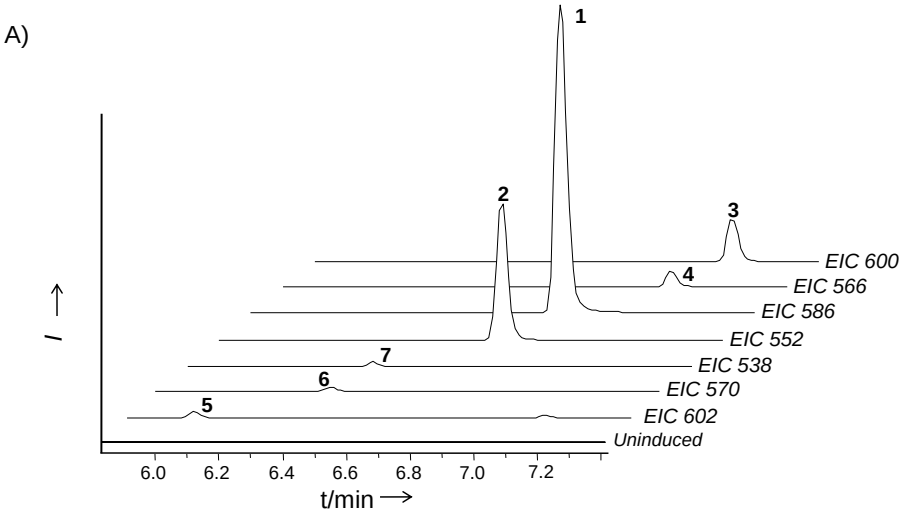
Herein, five new luminmides were identified in an *E. coli* heterologous host by using HRESIMS, direct and inverse amino acids feeding experiments as well as 2D-NMR. We then efficiently created point mutations of the NRPS gene *plu3263* in a multi-copy plasmid by using *ccdB* counterselection recombineering (Scheme 1). By subsequent expression of Plu3263 mutants in *E. coli*, we investigated the specificity of A_3 domain mutants by analyzing the corresponding products of luminmide pathway. The minor product of th wild type enzyme luminmide B, was altered to become more prominent in several mutants without significant reduction of product yields. One mutant even gave better product titers than the wild type. Moreover, the production of main product luminmide A could be deliberately and significantly reduced in all the mutants.

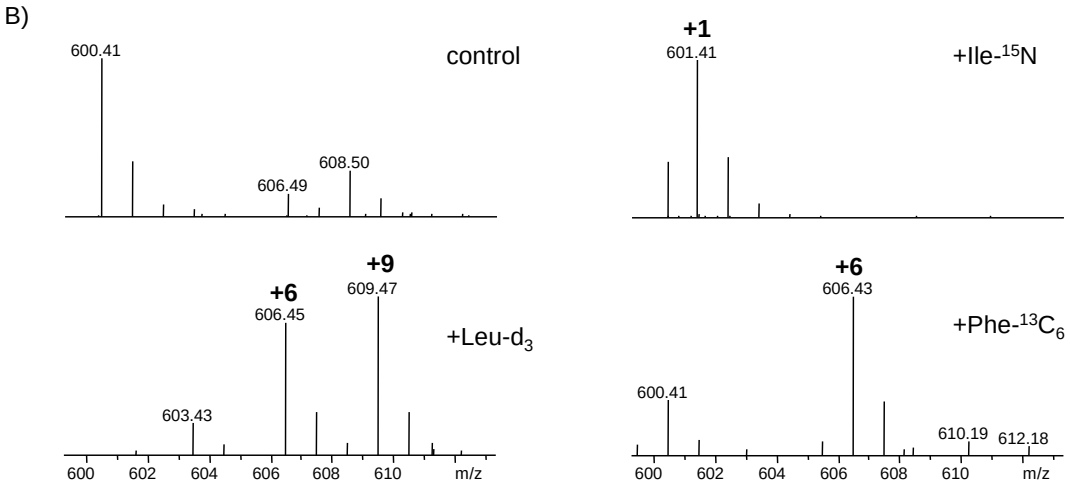
Results and discussions

Characterization of luminmide analogs

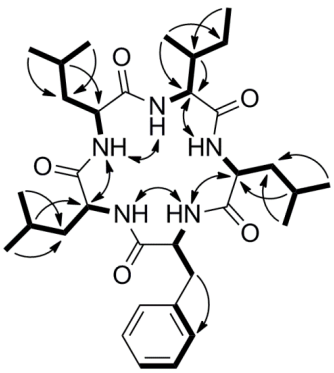
To obtain further insight into the substrate promiscuity of module 3 *in vivo*, we aimed to identify more luminmide analogs, which may have escaped detection in earlier experiments due to low production levels, by using MS fragmentation analysis and isotopes feeding experiments. Using detailed comparative LC-HRMS profile analysis between the *E. coli* GB05-MtaA with or without Plu3263, luminmides A (**1**), B (**2**) and five minor compounds (**3**-**7**) with pseudomolecular ion at *m/z* 602.3911 $[M+H]^+$ had the molecular formula $C_{32}H_{51}O_6N_5$ (**5**, calculated $C_{32}H_{52}O_6N_5$, 602.3912), *m/z* 570.3679 $[M+H]^+$ had the molecular formula $C_{28}H_{51}O_5N_5S$ (**6**, calculated $C_{28}H_{52}O_5N_5S$, 570.3684), *m/z* 538.3961 $[M+H]^+$ had the molecular formula $C_{28}H_{51}O_5N_5$ (**7**, calculated $C_{28}H_{52}O_5N_5$, 538.3963), *m/z* 566.4274 $[M+H]^+$ had the molecular formula $C_{30}H_{55}O_5N_5$ (**4**, calculated $C_{30}H_{56}O_5N_5$, 566.4276) and *m/z* 600.4117 $[M+H]^+$ had the molecular formula $C_{33}H_{53}O_5N_5$ (**3**, calculated for $C_{33}H_{54}O_5N_5$, 600.4119) produced only in the Plu3263-containing culture but lacking in the empty expression host (Figure 1A). Compounds **3**-**7** also belonged to luminmides and are named as luminmides C-G, respectively. In the known luminmides or GameXpeptides, the modules 2, 4, 5 only activate and incorporate Leu into the peptide chain as also supported by the same specificity-conferring codes of A domains[28] that shows rigid selectivity. However, module 3 demonstrates substrate tolerance and can activate and incorporate Phe (**1**) and Leu (**2**), respectively into final products (Figure 1D, Table 1). Furthermore, the A_1 domain is predicted to select two amino acids Val and Ile into the cyclic peptide[28] but Val and Leu were reported

as incorporated in the structure of GameXpeptides A-D in the native host[29]. Thus, the structural variation of the compound should be present in the amino acid residue(s) I and/or III of the pentaptide (Figure 1D).





C)



D)

	RT(min)	Amino acid residues (cyclo-)				
		I	II	III	IV	V
1	6.80	D-Val	L-Leu	D-Phe	D-Leu	L-Leu
2	6.71	D-Val	L-Leu	D-Leu	D-Leu	L-Leu
3	7.14	D-Ile	L-Leu	D-Phe	D-Leu	L-Leu
4	7.04	D-Ile	L-Leu	D-Leu	D-Leu	L-Leu
5	6.04	D-Val	L-Leu	D-Tyr	D-Leu	L-Leu
6	6.38	D-Val	L-Leu	D-Met	D-Leu	L-Leu
7	6.41	D-Val	L-Leu	D-Val	D-Leu	L-Leu

Figure 1. Characterization of luminmide derivatives. A) Combined extracted ion chromatograms (EICs) in positive mode ionization mode of luminmides (**1**-**7**) from extracts of *E. coli* GB05-MtaA containing pGB-Ptet-plu3263. B) The direct feeding of labeled L-Leu, L-Ile and L-Phe to confirm their

presence in luminmide C (**3**). C) Key correlations of TOCSY (bold lines), HMBC (single-headed arrows), and ROESY (double headed arrow) of luminmide C (**3**). D) The structures of luminmides A-G (**1-7**).

The structures of **3-7** were then determined by HRMS, feeding experiments and 2D-NMR. The feeding experiments of labeled Leu, Ile and Phe showed they were present in **3** (Figure 1B, Figure S1) but these findings contradicts with the proposed structure of GameXpeptide B[29] which does not include any Ile. Compound **3** was isolated from *E. coli* GB05-MtaA harboring pGB-Ptet-plu3263. The ^{1}H-NMR, and 2D-NMR data (HSQC, HMBC, and TOCSY) of **3** indicated that it is composed of one Ile (Figure 1C, Table S1), three Leu and one Phe. The mass data and long-range HMBC and ROESY correlations confirmed the positioning and connection of residues (Figure 1C). However, our limited material restrained us to assign confidently the configuration of C-3 of Ile by chemical methods. In our experiments, it is clear that it is Ile rather than Leu that was incorporated by the first module into **3** which is also true for compound **4**. Here, however, but third amino acid of **4** is Leu which is similar to **2**. Thus, the first module of Plu3263 can activate and incorporate Val or Ile into final products. Compound **5** differs from **1** by 16 atomic mass units which is in agreement with the addition of a hydroxy group (account for +O) and the retention time of those two compounds in HPLC chromatograms show that, **5** is more polar than **1** under reversed phase separation conditions (Figure 1A, Figure 1D). The inverse labeling experiment of [U-^{12}C]tyrosine, [U-^{12}C]valine, and [U-^{12}C]leucine in a fully labeled ^{13}C medium confirmed the existence of one Tyr, one Val, and three Leu into **5** (Figure S3). Combined with comparative MS fragmentation pattern, the third amino acid residue Phe in **1** was replaced with Tyr in **5** (Figure 1D). Similar analysis was employed to determine the amino acid composition of **6** and **7**, and the Phe in **1** is substituted with Met in **6,** and with Val in **7**, respectively (Figure 1D, Figure S4 and S5). The incorporation of Met into **7** was also corroborated by a direct feeding experiment of ^{13}C-labeled and deuterated L-Met (Figure S4). Unsurprisingly, Phe, Leu, Tyr, Val and Met can be activated and incorporated into the final product by module 3 but with different efficiencies. To the best of our knowledge, the activation and incorporation of Met is unprecedented in the elucidated NRPS biosynthetic pathways. The absolute configurations of α-position of amino acid residues in these luminmides are identical to those of corresponding positions in luminmide A because they were biosynthesized by the same NRPS Plu3263 (Figure 1).

Efficient mutation of a complex NRPS gene by site-directed mutagenesis

Table 1. Comparison of the specificity-conferring code of the NRPS A domains activating Phe/Tyr/Leu/Val with consensus sequences.[a]

		235	236	239	278	299	301	322	330	331	517
GrsA	Phe	D	A	W	T	I	A	A	I	C	K
TycA	Phe	D	A	W	T	I	A	A	I	C	K
BacC-M2	Phe	D	A	F	T	V	A	A	V	C	K
CepA	Tyr	D	A	S	T	V	A	A	V	C	K
TycC	Tyr	D	A	L	T	T	G	E	V	V	K
GrsA*[b]	Leu	D	A	W	M	I	G	A	I	C	K
SrfAA-M2	Leu	D	A	F	M	M	G	M	V	F	K
CssA-M9	Val	D	A	W	M	F	A	A	V	L	K
GrsB	Val	D	A	F	W	I	G	G	T	F	K
Plu3263-M2/4/5	Leu	D	A	W	C	I	G	A	V	C	K
Plu3263-M3	Phe/Leu	D	A	W	C	I	A	A	V	C	K

[a] Residue numbering according to the GrsA L-phenylalanine activating A domain[10].
[b] The mutated GrsA which is a L-Leucine activating A domain[11].

Our next goal was to rapidly switch the active site residues in the building pocket of an A domain and monitor the alteration of end products generated by the mutants *in vivo*. We engaged counterselection-based recombineering to introduce seamless site-directed mutations. Generally, two selections after consecutive homologous recombination were required for accomplishment of seamless plasmid mutagenesis: the positive selection for antibiotic resistance to replace the target sequences by insertion of a dual selection-counterselection cassette, and the counterselection of the toxicity marker inferred by the counterselectable gene for substitution of this cassette with sequences of interest. Here we use a chloramphenicol resistance gene (cm^R) as positive selectable marker and *ccdB* as a counterselectable gene that is only propagated in the gyrA462 *E. coli* strain DB3.1[35-37] but cannot propagated in our standard recombineering *E. coli* strain GB2005[38, 39]. Thus, the first step for replacement of target sequences with *ccdB-cm* cassette is performed in DB3.1 containing Redγβα expression plasmid pSC101-P_{BAD}-gba[40] for an efficient recombination reaction (Scheme 1). In the second step, the oligonucleotide-mediated recombineering is accomplished in DB3.1 containing Redγβα to form the mixture of repaired (minority) and unrepaired constructs in one cell because of the presence of pBR322 derived multi-copy plasmids. *ccdB* counterselection was achieved by transferring the resultant plasmid mixture from DB3.1 into GB2005 to separate the mixture and remove the unrepaired *ccdB*-containing plasmids. The correct mutants should yield viable GB2005, due to the absence of *ccdB*. However, this did not assure the correctness of mutants due to the possibility of intramolecular recombinations that removed *ccdB* and enabled GB2005 to survive. Using this

procedure of counterselection through changing host, even one copy of the correct repaired plasmid can be selected and therfore proves an effective means for generating point mutations in a multi-copy plasmid.

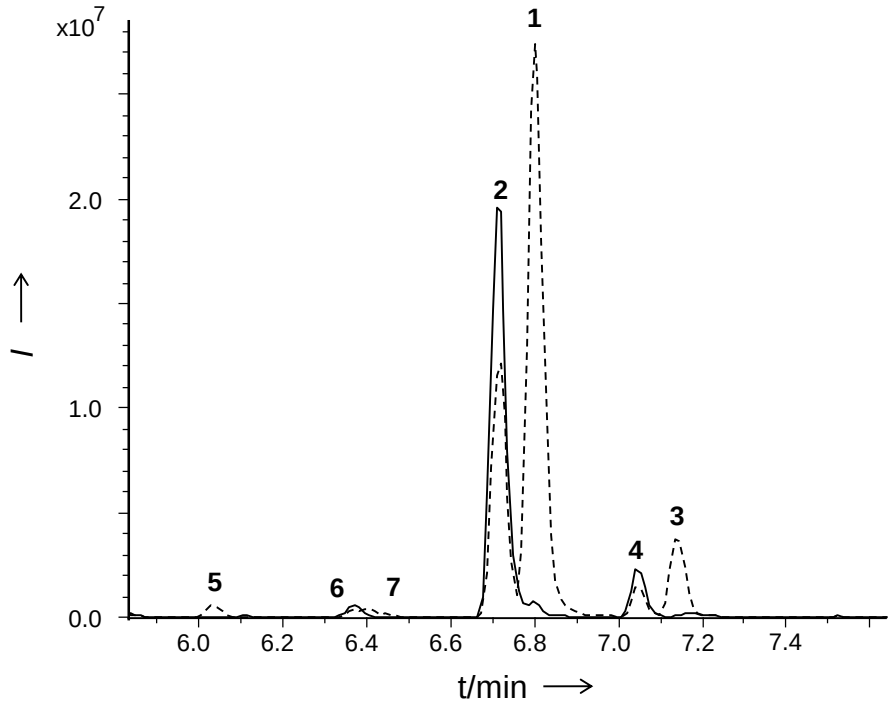


Figure 2. HPLC analysis (base peak chromatogram (BPC) from 530-620 Da in positive ionization mode) of the luminmides in *E. coli* GB05-MtaA containing Plu3263 WT (dashed line) and mutant A301G (solid line).The luminmides A-G (**1**-**7**) are shown.

In order to select residues for mutagenesis, we compared the 'specificity-conferring code' of A_3 domain with A domains known to activate Leu, Phe, Val by amino acid alignment (Table 1). The nonribosomal code at position 278 is show almost exclusively Met in Leu/Val-activating A domains while it is Thr when Phe is exclusively recognized. In the Plu3263 A_3 domain, the residue at 278 is Cys. The initial idea was to change Cys278 to Thr, or Met to increase the yields of luminmide A or B, respectively. At the residue 301 of the Leu-activating domain is Gly, but in Plu3263 A3 domain is Ala. It is proposed to increase the activation efficiency of Leu or decrease that of Phe by switching Ala301 to Gly. Furthermore, most of the residues in position 299 for activation of Val/Leu/Phe are Phe or Ile, but in Plu3263 A3 domain Ile is found, thus the mutation I229F was performed in one round with A301G in a single recombineering exercise which was possible due to the vicinity of 299 and 301 amino acid residues. We speculated that these single and/or double mutants would improve the catalytic efficiency of A_3 domain towards Leu or Phe which might lead to variation of luminmides. Single mutations C278M, C278T, I299F, A301G, and double mutation I299F/A301G were constructed in one round of selection and counterselection recombineering exercises. We then

used the C278M construct to generate double mutation C278M/A301G using a second round of recombineering (Table S1). Seamless site-directed mutagenesis of a complex NRPS biosynthetic pathway in a multi-copy plasmid using *ccdB* counterselection recombineering was achieved. The final efficiency of obtaining the correct recombinant was on average 15% (Table S2), which is reasonable considering the iterative nature of modules/domains in large NRPS biosynthetic genes which explains the presence of numerous highly homologous stretches of DNA. The mutation position in the middle of the repetitive NRPS genes would thus frequently induce intramolecular recombination that can result in the deletion of the *ccdB* cassette and the surrounding DNA to generate incorrect recombinants and hinder the efficiency of counterselection[23]. The incorrect mutants in this investigation are the result of intramolecular recombination between the upstream regions (module 1 and 2) and downstream areas (module 4 and 5) of module 3 as supported by restriction digestion examination.

Table 2. The yields of luminmides in WT and mutated Plu3263 [a]

Mutants	Yields (μg L^{-1})						
	1	2	3	4	5	6	7
WT	694.5	303.9	87.5	30.1	11.6	8.1	8.9
C278M	2.7	261.2		23.5			22.7
C278T	229.9	138.9	26.1	11.3	7.1	1.6	1.6
I299F	105.4	198.9	8.3	16.3		1.7	7.4
A301G	17.2	515.2	0.68	53.9		14.8	0.42
I299F/A301G	4.7	417.2		39.9		11.4	0.25
C278M/A301G		386.9		38.5		3.4	

[a] The yields of luminmides were average of two or three independent clones (Table S2) in the same cultivation conditions and were determined by UPLC-HRMS, and yield less than 0.1 μg L^{-1} was not shown in the table.

Alteration of luminmide derivatives

The WT and mutated Plu3263 were expressed in the heterologous host *E. coli* GB05-MtaA and the yields of luminmide derivatives were determined by UPLC-HRMS (Table 2). The total yields of luminmides in WT Plu3263 were higher than in all of the modified mutants, which showed that all the point mutations lower the efficiency of module 3 to activate and incorporate amino acid residues. This finding is quite normal for complex NRPS biosynthetic pathways that have been genetically manipulated for combinatorial biosynthesis. However, the yield of luminmide B showed a significant increase from 304 μg L^{-1} in WT Plu3263 to 515 μg L^{-1} in mutant A301G (Figure 3, Table 2, Figure S6), while the original dominant product luminmide A decreased to different extents in all the given mutants, and even disappeared in double mutant C278M/A301G. Thus, we can specifically increase the

yields of expected compounds by simple single or double mutations in the active site residues of the A domain. Besides, the productions of several minor derivatives are also increased by 1~2 fold (Table 2). For example, the increase of **7** from 9 μg L^{-1} to 23 μg L^{-1} in the mutant C278M is consistent with our anticipation that Met in position 278 is advantageous for recognition of Val. But C278T did not show any increase of luminmide A as expected because the Cys in this position maybe more efficient in this NRPS although in position 278 Thr is found in all the other Phe-activating A domains. The production of **6** is also increased in the mutant A301G (from 8 μg L^{-1} to 15 μg L^{-1}) and I299F/A301G (to 11.4 μg L^{-1}) and shows same trend with luminmide B. This fact implies that the 'specificity-conferring code' of a Met-activating A domain (**6**) would be similar to a Leu-activating A domain. Conpound **3** and **4** changed simultaneously with luminmide A (**1**) and B (**2**), respectively, because third amino acid residues are identical, Phe in **1** and **3** and Leu in **2** and **4**. The differences between the pairs **1**, **2**, and **3**, **4**, are found in the first amino acid residue, Val in **1** and **2**, but Ile in **3** and **4** (Table 1).

Table 3. The specificity of Plu3263 A3 domain *in vivo* [a]

Mutants	Proportion (%)[b]				
	Phe	Leu	Tyr	Met	Val
WT	**68.3**	29.2	1.01	0.71	0.78
C278M	0.88	91.8			**7.32**
C278T	61.5	36.0	**1.71**	0.39	0.39
I299F	33.6	63.7		0.51	2.19
A301G	2.96	94.5		**2.46**	0.07
I299F/A301G	0.99	96.6		2.4	0.05
C278M/A301G		**99.2**		0.8	

[a] The values show percentages of defined amino acids incorporated, with the sum of all final luminmides as 100%.
[b] Proportions of less than 0.05% are not shown in this table.

The reduced overall efficiency of Plu3263 with mutated A domain was complemented by the improvement of the specificity and selection for activating and processing specific amino acid residues. We calculated the percentage of certain amino acids (at the third residue) in comparison to total yields of luminmides to represent the 'specificity of A3 domain *in vivo*' regardless of the efficiency (Table 3). Because **1** and **2** as well as **3** and **4**, contain the same amino acid in the third position, Leu and Phe, respectively, the yields of Leu- and Phe-luminmides should be the sum of each pair. The WT A_3 domain in the heterologous host preferentially 'activates' (as measured by final product thus representing activation and incorporation) Phe (68.2%), and Leu (29.2%). The C278T mutant showed a similar

specificity to the WT, but the remaining mutants are favorable to 'activation' of Leu. The effectiveness of Leu activation in the I299F mutant is 64%, 92% in C278M, 94% in A301G, 96.5% in the double mutant I299F/A301G, and 99.2% in C278M/A301G (Table 3). In the latter double mutant, the Phe-luminmides were completely abolished and only a trace amount of **6** was produced in which the third amino acid is Met. In the C278M mutant, the A3 domain preferentially activates Val rather than Phe. Our strategy also represents a feasible way to remove the predominant WT compound and/or improve the yield of a minor product, which can simplify the downstream purification procedure of low-level secondary metabolites. In this case, the luminmides have a tendency to precipitate out of the solutions when the crude extract and fractions are concentrated. It is consequently difficult to purify the minor product **2** from the WT Plu3263 extracts. By using the mutant culture, the pure compound can be wasily obtained for use in further studies[28].

Luminmide A only shows weak cytotoxicity against HCT-116 human colon carcinoma (IC_{50}=27.5 μg/mL) while luminmide B has no effect on cell viability at concentration as high as 100 μg/mL. The planar structure of luminmide A is identical to sansalvamide A peptide (against HCT-116, IC_{50}=0.98 μg/mL)[41] which was regarded as an excellent lead for treating multiple cancers[42-53] and it was a chemically synthesized derivative of fungal secondary metabolite sansalvamide A (against HCT-116, IC_{50}=9.8 μg/mL)[54] whose the residue V is a leucic acid. However, the sansalvamide A peptide contains five L-configured amino acids and the cytotoxicity is much higher than luminmide A which contains mixed L- and D-configured amino acids (Figure 1D).

Conclusion

Combinatorial biosynthesis in a well-established heterologous host is in theory a powerful technique. However, still many challenges remain for generating new derivatives of complex NRP or PKs natural products[55]. We have been able to create single or double point mutations efficiently in a complex NRPS in *E. coli in vivo* by using *ccdB* counterselection recombineering. This method yielded much less unexpected mutations than may be obtained by PCR or other *in vitro* method for performing site-directed mutagenesis. As a result of the point mutations in the specificity-conferring code of the NRPS A domains, the WT minor product luminmide B (29%) became more abundant (more than 90%) in several mutants without significant reduction of product yields, while one mutant (A301G) produced higher yields than wild type. The successful *in vivo* site-directed mutagenesis of an exogenous NRPS, encoded on a multi-copy plasmid, combined with robust heterologous expression

provides a rational way to generate expected or predictable derivatives based on structural biology and bioinformatics.

Experimental section

Feeding experiments: *E. coli* GB05-MtaA/pGB-Ptet-plu3263[28] was cultivated in Luria-Bertani (LB) broth supplemented with 50 μg mL^{-1} ampicillin and 1 μg mL^{-1} gentamicin. For direct feeding experiments, L-valine-d_8 (Deutero GmbH), L-leucine-5,5,5-d_3 (Deutero GmbH), L-phenylalanine-ring-$^{13}C_6$ (Cambridge Isotope Laboratories, Inc.), L-isoleucine-^{15}N (Sigma-Aldrich), L-methionine-(*methyl*-^{13}C) (Cambridge Isotope Laboratories, Inc.) and L-methionine (*methyl*-d_3) (Sigma-Aldrich) (final concentration 1 mg mL^{-1}) were dissolved into LB medium supplemented with suitable antibiotics and sterilized by syringe filter (0.2 μm), respectively. The media was inoculated with 2% overnight culture of *E. coli* GB05-MtaA/pGB-Ptet-plu3263 and cultivated at 30 °C for 4 h, induced by tetracycline (0.5 μg mL^{-1}) and 2% adsorber resin Amberlite XAD-16 was added 8 h later. L-phenylalanine, L-tyrosine, and L-methionine (Synopharm GmbH & Co. KG, final concentration 4 mg mL^{-1}) were inversely fed into *E. coli* cultures grown in [U-^{13}C] medium and filter sterilized. Fifty milliliters of [U-^{13}C] medium was prepared from 0.5 g ISOGRO®-^{13}C powder (Sigma-Aldrich) and supplemented 0.9 mL of 100 g L^{-1} K_2HPO_4, 1.4 mL of 50 g L^{-1} KH_2PO_4, 1.0 mL of 50 g L^{-1} $MgSO_4$ and 15.0 μL of 37 g L^{-1} $CaCl_2$ solutions according to the manufacturer's protocol. The *E. coli* overnight preculture was washed twice with [U-^{13}C] medium prior to inoculation into [U-^{13}C] medium. Extraction and HPLC/MS analysis of luminmides are same as quantitative analysis method.

Isolation of luminmide C (3): The overnight *E. coli* GB05-MtaA/pGB-Ptet-plu3263[28] culture was inoculated in LB medium in the presence of 50 μg mL^{-1} ampicillin and 2% XAD-16 adsorber resin (added 8 h later) at 30 °C for 30 hours in shaking flasks. The XAD was separated from the supernatant by sieving, washed with dH_2O twice and extracted stepwise with ethyl acetate, acetone and methanol. The resulting extracts were analyzed by LC/MS and most of the compounds were present in the methanol extract. Then, this extract was partitioned with equal volume ethyl acetate and dH_2O to obtain 0.85 g extract from ethyl acetate phase. The extract was loaded onto a Sephadex LH-20 column with methanol as mobile phase. Fractions containing significant amounts of **3** were combined and evaporated to dryness. Further purification was performed on repeated semi-preparative HPLC

instrument equipped with UV detector (220 nm, and 280 nm). A RP-C_{12} column (Jupiter, Proteo 90A, 250 × 10 mm, 4 μm) was used for further purification with a solvent system consisting of water (A) containing 0.05% TFA and acetonitrile (B). The gradient was as follows: 0-5 min 50% B, 5-15 min 50-70% B, 15-40 min 70% B. Around 0.3 mg luminmide C (**3**) was afforded at the retention time of 32.2–33.2 min as a white solid. The luminmide C was dissolved into DSMO-d_6 for NMR recording. NMR spectra were recorded on a Bruker Ascend 700 MHz spectrometer equipped with a CryoProbe system. The samples were measured in DMSO-d_6, and the solvent peak was used for spectra calibration (δ_C 39.51, δ_H 2.50 ppm). Data acquisition, processing, and spectral analysis were performed with TopSpin 3.0 (Bruker BioSpin).

***ccdB* counterselection recombineering**

Introduction of counterselection cassette *cm^R-ccdB* into site of interest: By electroporation, Red/ET recombination expression plasmid pSC101-P_{BAD}-gba-tet[40] combined with targeted construct pGB-Ptet-plu3263 were cotransformed into *E. coli* DB3.1 (Invitrogen) competent cells, and 500 μL recovery culture was spread on LB agar containing 5 μg mL^{-1} tetracycline and 100 μg mL^{-1} ampicillin. Restriction analysis was performed to confirm both of them. The PCR products mixture of *ccdB* gene amplified from p15A-Amp-ccdB and chloramphenicol resistance gene were cotransformed into L-arabinose induced DB3.1 cells bearing pSC101-P_{BAD}-gba-tet and pGB-Ptet-plu3263. All the cells were spread on LB plates containing 30 μg mL^{-1}chloramphenicol and 100 μg mL^{-1} ampicillin. The resulting constructs were checked by restriction analysis. Generally, all of clones were a mixture of the original plasmid and the plasmid containing *cm^R-ccdB* cassette. One more transformation into DB3.1 was required to separate the plasmids and obtain a pure construct pGB-Ptet-plu3263-ccdB-Cm construct. Then, four independent plasmids were transformed into GB2005[38, 39] and selected on ampicillin LB plates to check the function of *ccdB* gene. If a few colonies (generally less than 10) on the LB plate, this indicates that CcdB works well. It is unusual to see no colonies on the plate due to spontaneous intermolecular recombination or mutation may be happened to make the *ccdB* deletion or malfunction.

Recombination of an oligonucleotide containing a mutation of interest to replace counterselection cassette: The DB3.1 strain carrying pure and functional constructs pGB-Ptet-plu3263-ccdB-Cm were transformed with pSC101-P_{BAD}-gba-tet plasmids. Initial

restriction analysis was performed to check both of constructs. Then the single strand oligonucleotide (Table 4) with a mutation of interest were transformed into the L-arabinose induced DB3.1 electracompetent cells harboring pSC101-P_{BAD}-gba-tet[40] and pGB-Ptet-plu3263-ccdB-Cm. After 1h at 37 °C shaking for recovery, each 1 mL recovery culture was divided into two parts and the same volume LB medium containing 100 μg mL^{-1} ampicillin was added and returned to shake for another 3 h. The plasmid DNA was extracted, resuspended into 5 μL ddH_2O, and 2 μL solution was transformed into GB2005 electrocompetent cells. All the cells were plated on the LB plate containing 100 μg mL^{-1} ampicillin. The plasmid with *ccdB* can kill the normal *E. coli* strain GB2005 by binding to the gyrase[56]. Restriction analysis and sequencing were used to verify the mutated regions and recombination arms. The correct mutants were transformed into *E. coli* GB05-MtaA[28] to check the yields of luminmides. A list of primers used for mutagenesis is presented in Table 4.

Table 4. Oligonucleotides used for mutagenesis in this work.[a]

Primer names	Sequence of primers(5'→3')
cm^R-ccdB cassette	
A3-278cm5	TCCACAGCGCCTTTGGCGTTACCTGGAAGAGCAGGCGATAACCCATGCCTGTGAC GGAAGATCACTTCG
3263ccd5	GCGTAAAGGAGGTCAGATTATGCAGTTTAAGGTTTACACC
3263ccd3	TATCGCCGGTAAATCGGTGCCGTCATGAAACATAGCCGGTGTCAATTTGTTCAAA AAAAAGCCCGCTC
A3-299Cm5	TGACGGCACCGATTTACCGGCGATAGCGATAAAACCAACCTTATGTGACGGAAGA TCACTTCG
A3-301ccdB3	CGGCTGCACAGTGCCTGAAATAGCGCGGGACTCGGCGCTTCTCCTTTGTTCAAAA AAAAGCCCGCTC
Repair oligos	
C278ToligB	TATCGCTATCGCCGGTAAATCGGTGCCGTCATGAAACATAGCCGGTGTCAA**GGT**G GCATGGGTTATCGCCTGCTCTTCCAGGTAACGCCAAAGGCGCTGTGGA
C278MoligB	TATCGCTATCGCCGGTAAATCGGTGCCGTCATGAAACATAGCCGGTGTCAA**CAT**G GCATGGGTTATCGCCTGCTCTTCCAGGTAACGCCAAAGGCGCTGTGGA
I299F-B	CGGCTGCACAGTGCCTGAAATAGCGCGGGACTCGGCGCTTCTCCTGCAAA**GAA**TA AGGTTGGTTTTATCGCTATCGCCGGTAAATCGGTGCCGTCATGAAACA
A301G-B	TCGGCCCGGCTGCACAGTGCCTGAAATAGCGCGGGACTCGGCGCTTCTCC**ACC**AA ATATTAAGGTTGGTTTTATCGCTATCGCCGGTAAATCGGTGCCGTCATGA
299-301Mut	ATCGGCCCGGCTGCACAGTGCCTGAAATAGCGCGGGACTCGGCGCTTCTCC**ACC**A AA**GAA**TAAGGTTGGTTTTATCGCTATCGCCGGTAAATCGGTGCCGTCATGAAACA
Sequencing	
3263A3seq	TGACTCCTCCGGTATAATCG

[a] Underlined sequences are homology arms for Red/ET recombination. The bold bases represent mutated amino acid.

Quantitative analysis of luminmides: The *E. coli* strains harboring the wild type and mutated luminmide biosynthetic genes (*plu3263*) were incubated in 15 mL glass tubes containing 5 mL LB plus ampicillin and gentamicin. The culture was inoculated from an overnight culture (1:50) and incubated for 4 hours at 30 °C, shaking at 150rpm. After induction with tetracycline (final concentration 0.5 µg mL^{-1}, water as control), the culture was incubated for a further 8 hours and 2% of adsorber resin Amberlite XAD-16 was added, then cultivated for 1 day. The resin and biomass were harvested by centrifugation and then extracted with 0.37 mL acetone and 1 mL methanol. 1 mL extracts were evaporated to dryness and dissolved in 350 µl methanol to form 10× crude extracts. High-resolution mass spectrometry (HRMS) was performed on an Accela UPLC-system (Thermo-Fisher) coupled to a linear trap-FT-Orbitrap combination (LTQ-Orbitrap), operating in positive ionization mode. The 10× crude extracts were separated on a Waters BEH RP-C_{18} column (50 × 2 mm; 1.7 mm particle diameter; flow rate 0.6 mL min^{-1}), with a mobile phase of water/acetonitrile each containing 0.1% formic acid, using a gradient from 5%–95% acetonitrile over 9 min. The UPLC-system was coupled to the LTQ-Orbitrap by a Triversa Nanomate (Advion), a chip-based nano-ESI interface. A standard curve was established from serial dilutions of luminmide A down to 0.15 µg mL^{-1}. The yield of luminmide A was determined from the peak area (base peak chromatogram) by reference to the standard curve and estimates of the yields of all other derivatives were obtained by comparison of the peak areas to those of luminmide A.

Cytotoxicity assay: Human HCT-116 colon carcinoma cells (DSMZ, ACC 581) were seeded at 6×10^3 cells per well in 96-well plates in 180 μL complete medium and directly treated with varying concentrations of the compounds. Each sample was tested in duplicate along with an internal methanol control. After 5 d incubation, 20 μL of 5 mg/mL MTT (Thiazolyl blue tetrazolium bromide) in PBS was added per well and it was further incubated for 2 h at 37°C. The medium was then discarded and cells were washed with 100 μL PBS before adding 100 μL 2-propanol/10 N HCl (250 : 1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader (EL808, Bio-Tek Instruments Inc.), and cell viability was expressed as percentage relative to the respective methanol control. IC_{50} values were obtained by sigmoidal curve fitting on logarithmic scale.

Acknowledgement

We thank J. Herrmann and T. Hoffmann from R.M.'s laboratory for bioactivity tests and high-resolution MS measurements, respectively. Research in the laboratory of R.M. was funded by the DFG and BMBF. Y.Z. from Gene Bridges was partially funded by the BMBF (MiPro). X.B. thanks China Scholarship Council (CSC) for a PhD fellowship.

Keywords: luminmide · point mutation · Red/ET recombineering · biosynthesis · natural products

References

[1] M. A. Marahiel, T. Stachelhaus, H. D. Mootz. *Chem. Rev.* **1997**, *97*(7), 2651-2674.

[2] A. M. A. van Wageningen, P. N. Kirkpatrick, D. H. Williams, B. R. Harris, J. K. Kershaw, N. J. Lennard, M. Jones, S. J. M. Jones, P. J. Solenberg. *Chem. Biol.* **1998**, *5*, 155-162.

[3] V. Miao, M. F. Coeffet-LeGal, P. Brian, R. Brost, J. Penn, A. Whiting, S. Martin, R. Ford, I. Parr, M. Bouchard, C. J. Silva, S. K. Wrigley, R. H. Baltz. *Microbiology* **2005**, *151*(Pt 5), 1507-1523.

[4] L. Du, C. Sanchez, M. T. Chen, D. J. Edwards, B. Shen. *Chem. Biol.* **2000**, *7*, 623-642.

[5] G. Weber, E. Leitner. *Curr. Genet.* **1994**, *26*(5-6), 461-467.

[6] C. Pelludat, A. Rakin, C. A. Jacobi, S. Schubert, J. Heesemann. *J. Bacteriol.* **1998**, *180*(3), 538-546.

[7] J. D. Walton. *Phytochemistry* **2006**, *67*(14), 1406-1413.

[8] M. A. Fischbach, C. T. Walsh. *Chem. Rev.* **2006**, *106*(8), 3468-3496.

[9] S. A. Sieber, M. A. Marahiel. *Chem. Rev.* **2005**, *105*(2), 715-738.

[10] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick. *EMBO J.* **1997**, *16*(14), 4174-4183.

[11] T. Stachelhaus, H. D. Mootz, M. A. Marahiel. *Chem. Biol.* **1999**, *6*, 493-505.

[12] G. L. Challis, J. Ravel, C. A. Townsend. *Chem. Biol.* **2000**, *7*(3), 211-224.

[13] H. von Döhren, R. Dieckmann, M. Pavela-Vrancic. *Chem. Biol.* **1999**, *6*(10), R273-R279.

[14] M. Z. Ansari, G. Yadav, R. S. Gokhale, D. Mohanty. *Nucleic Acids Res.* **2004**, *32*(Web Server issue), W405-W413.

[15] B. O. Bachmann, J. Ravel. *Meth. Enymol.* **2009**, *458*, 181-217.

[16] C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben, D. H. Huson. *Nucleic Acids Res.* **2005**, *33*(18), 5799-5808.

[17] C. Y. Chen, I. Georgiev, A. C. Anderson, B. R. Donald. *Proc. Natl. Acad. Sci. USA* **2009**, *106*(10), 3764-3769.

[18] B. W. Stevens, R. H. Lilien, I. Georgiev, B. R. Donald, A. C. Anderson. *Biochemistry-US* **2006**, *45*(51), 15495-15504.

[19] B. Villiers, F. Hollfelder. *Chem. Biol.* **2011**, *18*(10), 1290-1299.

[20] K. Eppelmann, T. Stachelhaus, M. A. Marahiel. *Biochemistry-US* **2002**, *41*(30), 9718-9726.
[21] B. S. Evans, Y. Chen, W. W. Metcalf, H. Zhao, N. L. Kelleher. *Chem. Biol.* **2011**, *18*(5), 601-607.
[22] J. Thirlway, R. Lewis, L. Nunns, N. M. Al, M. Styles, A. W. Struck, C. P. Smith, J. Micklefield. *Angew. Chem. Int. Ed.* **2012**, *51*(29), 7181-7184.
[23] A. W. Bird, A. Erler, J. Fu, J. K. Heriche, M. Maresca, Y. Zhang, A. A. Hyman, A. F. Stewart. *Nat. Methods* **2011**, *9*(1), 103-109.
[24] R. Heermann, T. Zeppenfeld, K. Jung. *Microb. Cell Fact.* **2008**, *7*.
[25] Y. Zhang, J. P. Muyrers, J. Rientjes, A. F. Stewart. *BMC Mol. Biol.* **2003**, *4*(1), 1.
[26] Y. Zhang, F. Buchholz, J. P. Muyrers, F. A. Stewart. *Nat. Genet.* **1998**, *20*, 123-128.
[27] E. Duchaud, C. Rusniok, L. Frangeul, C. Buchrieser, A. Givaudan, S. Taourit, S. Bocs, C. Boursaux-Eude, M. Chandler, J. F. Charles, E. Dassa, R. Derose, S. Derzelle, G. Freyssinet, S. Gaudriault, C. Medigue, A. Lanois, K. Powell, P. Siguier, R. Vincent, V. Wingate, M. Zouine, P. Glaser, N. Boemare, A. Danchin, F. Kunst. *Nat. Biotechnol.* **2003**, *21*(11), 1307-1313.
[28] J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Müller, A. F. Stewart, Y. Zhang. *Nat. Biotechnol.* **2012**, *30*(5), 440-446.
[29] H. B. Bode, D. Reimer, S. W. Fuchs, F. Kirchner, C. Dauth, C. Kegler, W. Lorenzen, A. O. Brachmann, P. Grün. *Chem. Eur. J.* **2012**, *18*(8), 2342-2348.
[30] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson. *BMC Evol. Biol.* **2007**, *7*, 78-92.
[31] J. M. Crawford, C. Portmann, R. Kontnik, C. T. Walsh, J. Clardy. *Org. Let.* **2011**, *13*(19), 5144-5147.
[32] S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas, H. B. Bode. *Org. Biomol. Chem.* **2011**, *9*(9), 3130-3132.
[33] K. Fujii, Y. Ikai, T. Mayumi, H. Oka, M. Suzuki, K. Harada. *Anal. Chem.* **1997**, *69*(16), 3346-3352.
[34] K. Fujii, Y. Ikai, H. Oka, M. Suzuki, K. Harada. *Anal. Chem.* **1997**, *69*(24), 5146-5151.
[35] P. Bernard, M. Couturier. *J. Mol. Biol.* **1992**, *226*(3), 735-745.
[36] P. Bernard, K. E. Kezdy, L. Vanmelderen, J. Steyaert, L. Wyns, M. L. Pato, P. N. Higgins, M. Couturier. *J. Mol. Biol.* **1993**, *234*(3), 534-541.
[37] T. Miki, J. A. Park, K. Nagao, N. Murayama, T. Horiuchi. *J. Mol. Biol.* **1992**, *225*(1), 39-52.
[38] J. Fu, S. C. Wenzel, O. Perlova, J. Wang, F. Gross, Z. Tang, Y. Yin, A. F. Stewart, R. Müller, Y. Zhang. *Nucleic Acids Res.* **2008**, *36*(17), e113.
[39] J. Fu, M. Teucher, K. Anastassiadis, W. Skarnes, A. F. Stewart. *Methods Enzymol.* **2010**, *477*, 125-144.

[40] J. Wang, M. Sarov, J. Rientjes, J. Fu, H. Hollak, H. Kranz, W. Xie, A. F. Stewart, Y. Zhang. *Mol. Biotechnol.* **2006**, *32*(1), 43-53.

[41] W. Gu, S. S. Liu, R. B. Silverman. *Org. Lett.* **2002**, *4*(23), 4171-4174.

[42] K. Otrubova, T. J. Styers, P. S. Pan, R. Rodriguez, K. L. McGuire, S. R. McAlpine. *Chem. Commun.* **2006**, (9), 1033-1034.

[43] K. Otrubova, G. Lushington, D. V. Velde, K. L. McGuire, S. R. McAlpine. *J. Med. Chem.* **2008**, *51*(3), 530-544.

[44] P. S. Pan, R. C. Vasko, S. A. Lapera, V. A. Johnson, R. P. Sellers, C. C. Lin, C. M. Pan, M. R. Davis, V. C. Ardi, S. R. McAlpine. *Bioorg. Med. Chem.* **2009**, *17*(16), 5806-5825.

[45] R. C. Vasko, R. A. Rodriguez, C. N. Cunningham, V. C. Ardi, D. A. Agard, S. R. McAlpine. *ACS Med. Chem. Lett.* **2010**, *1*(1), 4-8.

[46] R. P. Sellers, L. D. Alexander, V. A. Johnson, C. C. Lin, J. Savage, R. Corral, J. Moss, T. S. Slugocki, E. K. Singh, M. R. Davis, S. Ravula, J. E. Spicer, J. L. Oelrich, A. Thornquist, C. M. Pan, S. R. McAlpine. *Bioorg. Med. Chem.* **2010**, *18*(18), 6822-6856.

[47] L. D. Alexander, R. P. Sellers, M. R. Davis, V. C. Ardi, V. A. Johnson, R. C. Vasko, S. R. McAlpine. *J. Med. Chem.* **2009**, *52*(24), 7927-7930.

[48] C. L. Carroll, J. V. Johnston, A. Kekec, J. D. Brown, E. Parry, J. Cajica, I. Medina, K. M. Cook, R. Corral, P. S. Pan, S. R. McAlpine. *Org. Lett.* **2005**, *7*(16), 3481-3484.

[49] M. R. Davis, E. K. Singh, H. Wahyudi, L. D. Alexander, J. B. Kunicki, L. A. Nazarova, K. A. Fairweather, A. M. Giltrap, K. A. Jolliffe, S. R. McAlpine. *Tetrahedron* **2012**, *68*(4), 1029-1051.

[50] J. B. Kunicki, M. N. Petersen, L. D. Alexander, V. C. Ardi, J. R. McConnell, S. R. McAlpine. *Bioorg. Med. Chem. Lett.* **2011**, *21*(16), 4716-4719.

[51] P. S. Pan, K. L. McGuire, S. R. McAlpine. *Bioorg. Med. Chem. Lett.* **2007**, *17*(18), 5072-5077.

[52] R. A. Rodriguez, P. S. Pan, C. M. Pan, S. Ravula, S. Lapera, E. K. Singh, T. J. Styers, J. D. Brown, J. Cajica, E. Parry, K. Otrubova, S. R. McAlpine. *J. Org. Chem.* **2007**, *72*(6), 1980-2002.

[53] T. J. Styers, A. Kekec, R. Rodriguez, J. D. Brown, J. Cajica, P. S. Pan, E. Parry, C. L. Carroll, I. Medina, R. Corral, S. Lapera, K. Otrubova, C. M. Pan, K. L. McGuire, S. R. McAlpine. *Bioorg. Med. Chem.* **2006**, *14*(16), 5625-5631.

[54] G. N. Belofsky, P. R. Jensen, W. Fenical. *Tetrahedron Lett.* **1999**, *40*(15), 2913-2916.

[55] R. H. Baltz. *Nat. Biotechnol.* **2006**, *24*(12), 1533-1540.

[56] L. Van Melderen. *I. J. Med. Microbiol.* **2002**, *291*(6-7), 537-544.

Supporting information

Table of content

Table S1 NMR spectral data for luminmide C (**3**) in DMSO-d_6

unit	position	δ_C, type [a]	δ_H (J in Hz) [b]
Ile	1	172.02, qC	
	2	54.9, CH	4.32 (m)
	3	37.1,CH_2	1.60 (m)
	4	25.8, CH	1.06 (m), 1.265 (m)
	5	11.3,CH_3	0.85 (m)
	6	14.0, CH_3	0.76 (d, 6.82)
	NH		7.50 (d, 9.31)
Leu-1	1	171.87, qC	
	2	52.1, CH	4.03 (m)
	3	38.6, CH_2	1.12 (m), 1.33 (m)
	4	23.6, CH	0.83 (m)
	5	21.6, CH_3	0.712 (d, 6.51)
	6	22.6, CH_3	0.623 (d, 6.64)
	NH		8.56 (d, 6.33)
Phe-2	1	170.57, qC	
	2	55.4, CH	4.23 (m)
	3	36.5, CH_2	2.75 (m), 3.12 (m)
	4	138.1, qC	
	5	128.4, CH	7.25 (m)
	6	125.9, CH	7.19 (m)
	7	123.4, CH	7.18 (m)
	NH		8.86 (d, 8.03)
Leu-2	1	171.56,qC	
	2	51.8, CH	4.30 (m)
	3	40.0, CH_2	1.40 (m), 1.67 (m)
	4	24.8, CH	1.39 (m)
	5	23.0, CH_3	0.853 (m)
	6	22.1, CH_3	0.89 (m)
	NH		7.39 (d, 6.67)
Leu-3	1	170.45, qC	
	2	49.8, CH	4.34 (m)
	3	36.2, CH_2	1.43 (m), 1.52 (m)
	4	24.0, CH	1.56 (m)
	5	22.6, CH_3	0.87 (m)
	6	21.3, CH_3	0.787 (d, 6.41)
	NH		8.88 (d, 7.82)

[a]Recorded at 175 MHz; referenced to residual DMSO-d_6 at δ 39.51 ppm.

[b]Recorded at 700 MHz; referenced to residual DMSO-d_6 at δ 2.50 ppm.

Table S2. Point mutation in NRPS gene by Red/ET recombination in combination with *ccdB* counter selection				
position	*plu3263*	*plu3263**	RE digestion correct	Sequencing correct
C278M	TGT (Cys)	ATG (Met)	3/24	3/3
C278T	TGT (Cys)	ACC(Thr)	3/18	3/3
I299F	ATA(Ile)	TTC(Phe)	6/22	6/6
A301G	GCA(Ala)	GGT(Gly)	6/24	2/6
I299F/A301G	ATA(Ile)/GCA(Ala)	TTC(Phe)/GGT(Gly)	5/28	3/5
C278M/A301G	TGT (Cys)/GCA(Ala)	ATG (Met)/GGT(Gly)	2/16	2/2
S2742A (PCP3)	TCG(Ser)	GCG(Ala)	15/24	6/6
			30/156 19.2%	25/31 80.6% Total.15.5%
RE means restriction endonuclease				

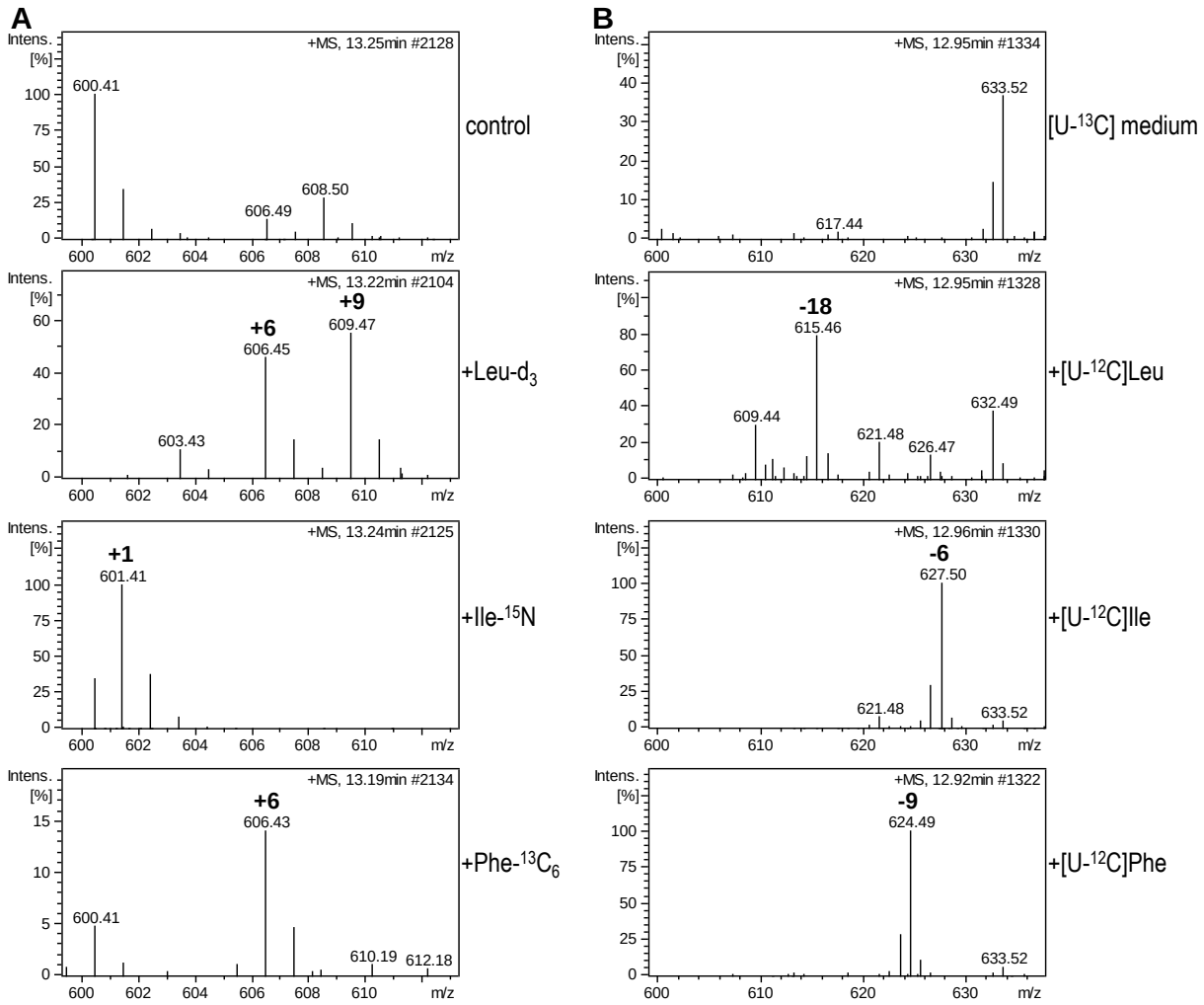


Figure S1. Direct and inverse feeding experiment of luminmide C (**3**). A) Direct feeding. B) Inverse feeding.

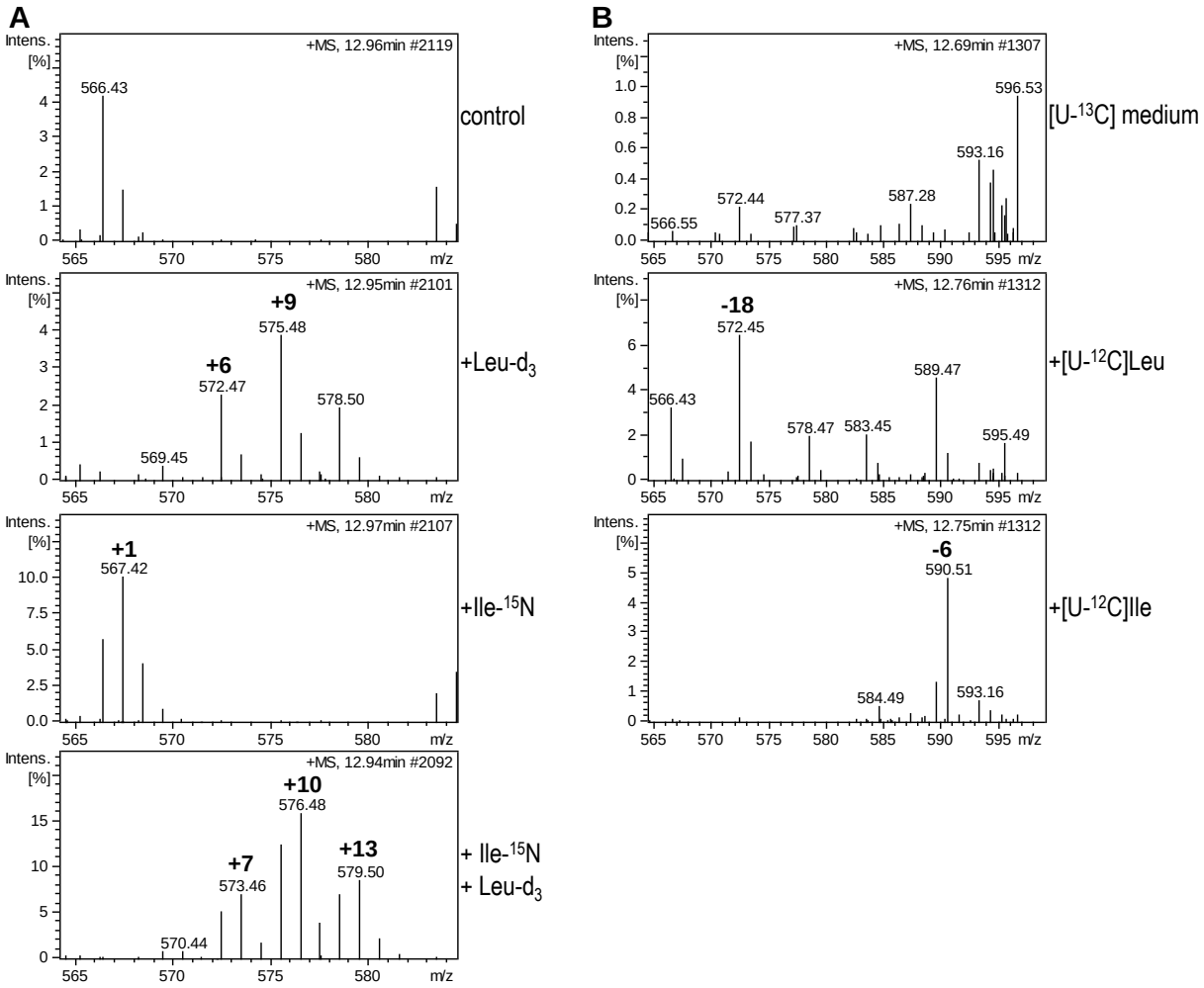


Figure S2. Direct and inverse feeding experiment of luminmide D (**4**). A) Direct feeding. B) Inverse feeding

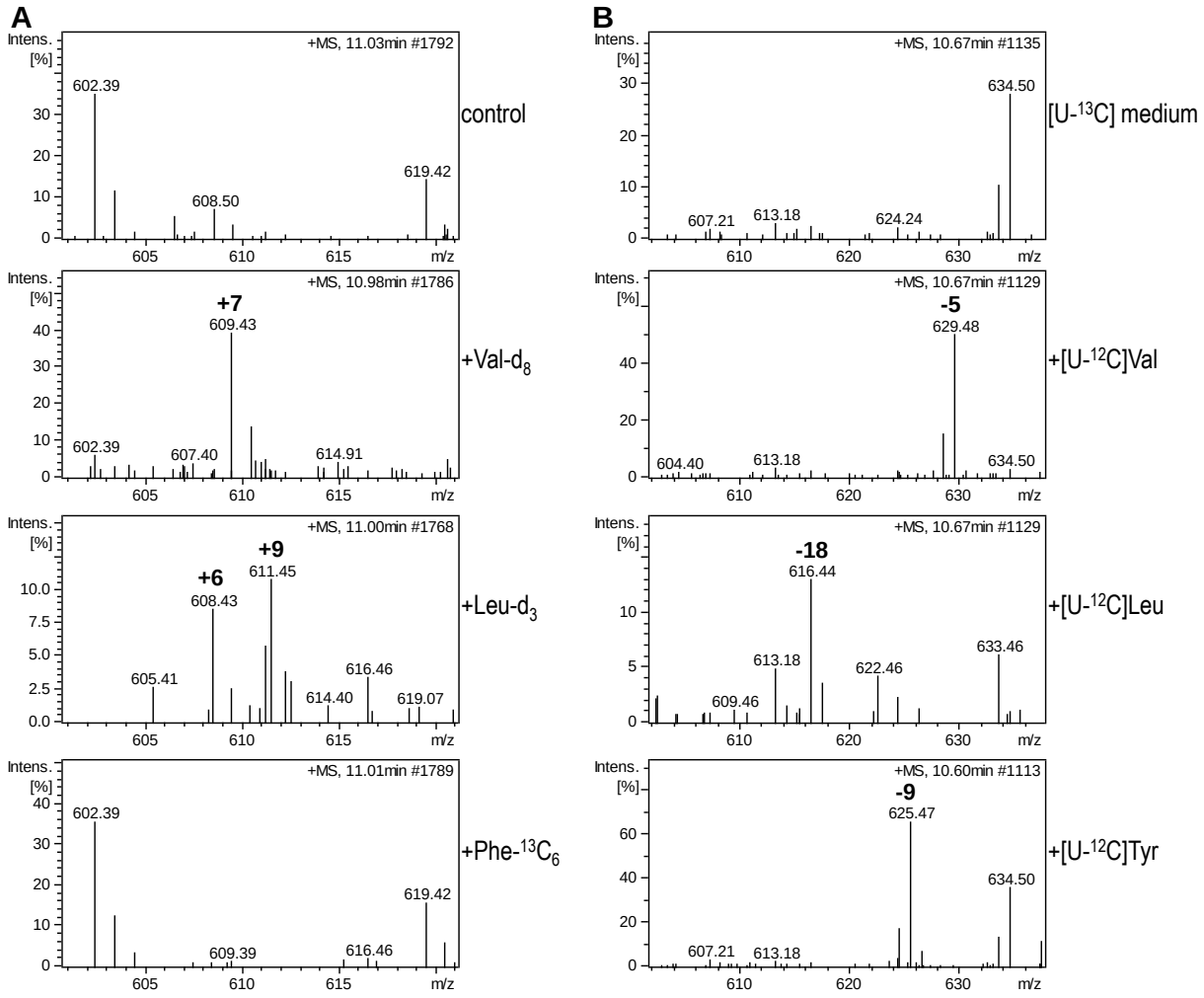


Figure S3. Direct and inverse feeding experiment of luminmide E (**5**). A) Direct feeding. B) Inverse feeding

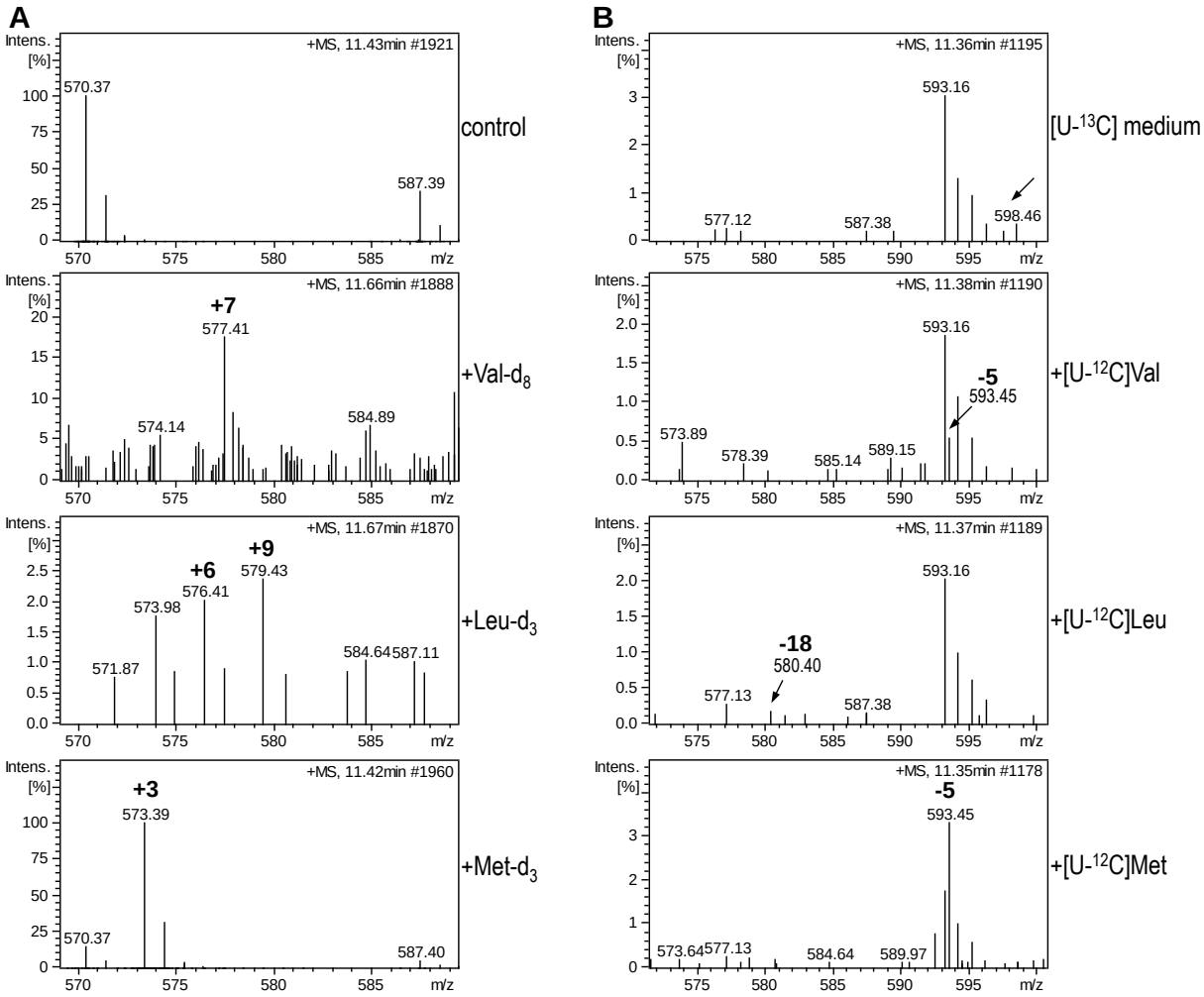


Figure S4. Direct and inverse feeding experiment of luminmide F (**6**). A) Direct feeding. B) Inverse feeding

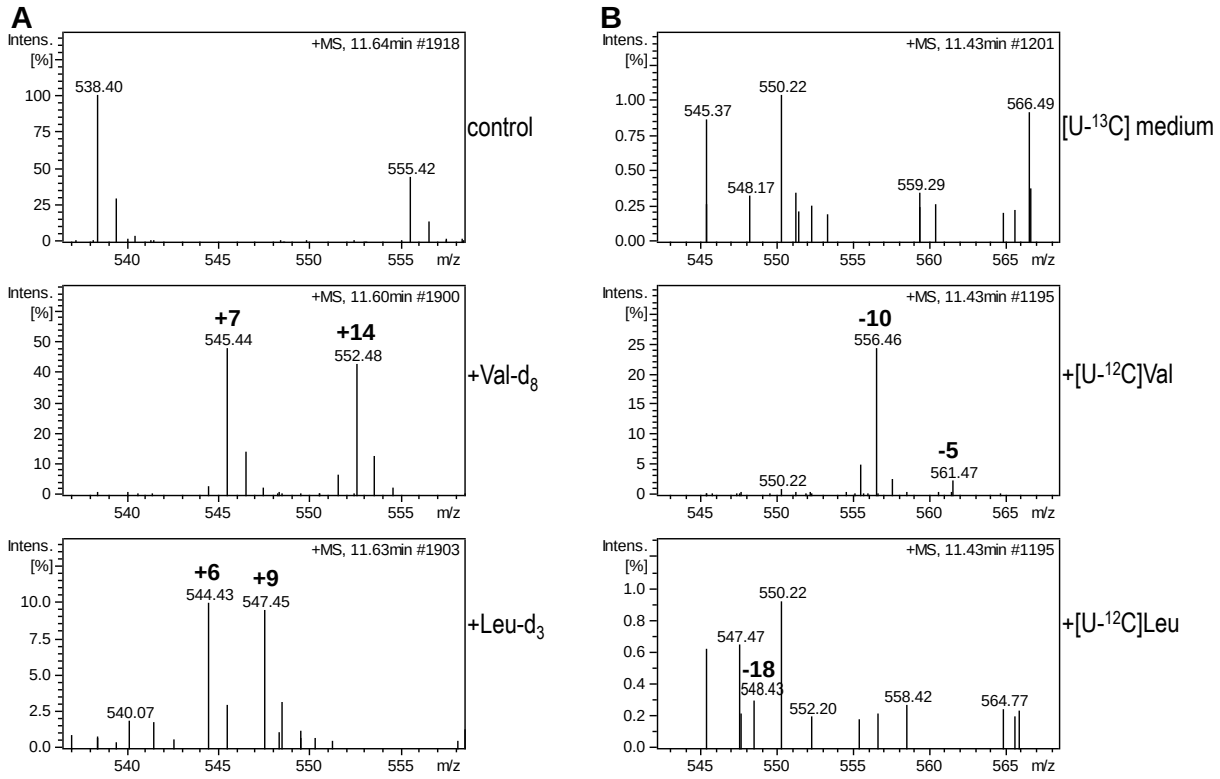


Figure S5. Direct and inverse feeding experiment of luminmide G (**7**). A) Direct feeding. B) Inverse feeding.

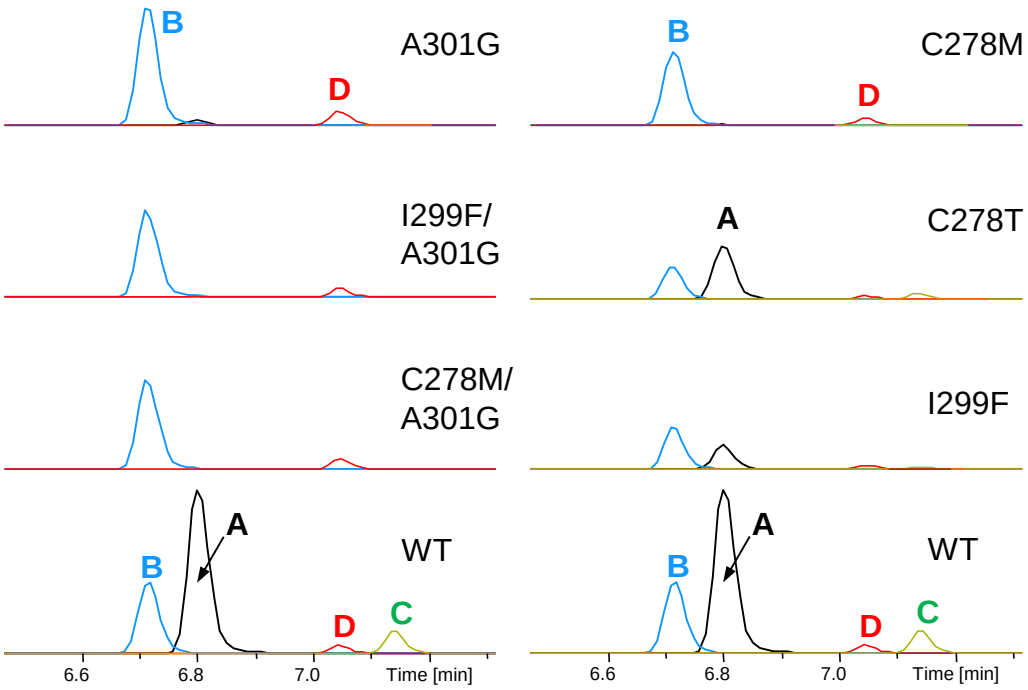


Figure S6. Comparative HPLC/MS chromatograms of WT and mutated Plu3263 expressed in *E. coli* GB05-MtaA. Luminmide A, EIC, *m/z* 586 $[M+H]^+$; Luminmide B, EIC, *m/z* 552 $[M+H]^+$; Luminmide C, EIC, *m/z* 600 $[M+H]^+$; Luminmide D, EIC, *m/z* 566 $[M+H]^+$.

Chapter 5

Direct Cloning, Genetic Engineering and Heterologous Expression of the Syringolin Biosynthetic Gene Cluster in *E. coli* through Red/ET Recombineering

Xiaoying Bian, Fan Huang, Francis A. Stewart, Liqiu Xia, Youming Zhang,

and Rolf Müller

ChemBioChem. 13, 1946-1952, **2012**.

This article is available online at:

http://onlinelibrary.wiley.com/doi/10.1002/cbic.201200310/pdf

Supporting information for this article is available on the WWW under
http://dx.doi.org/10.1002/cbic.201200310

C. Final discussion

1 RecE/RecT mediated linear-linear homologous recombination facilitates direct cloning for bioprospecting

The routine sequencing of bacterial genomes and recent progress in bioinformatics (e.g., antiSMASH which can quickly identify biosynthetic loci and predict types of compounds),[38] allow for the rapid detection of large numbers of cryptic biosynthetic gene clusters that potentially govern the biosynthesis of novel biologically active compounds. However, the low expression level of biosynthetic genes in laboratory conditions results in low amount or nonproduction of natural products which often precludes them from being mined during conventional screening.[45] Even if a compound is detected, low yield may restrain following purification, structure elucidation and biological activity test. Many genomic-based approaches that are currently employed for investigating silent gene clusters rely on provoking their expression in the native host through the optimizations of culture condition or engineering of regulatory genes. An alternative and potentially better approach to excavate potential novel products is the heterologous expression of silent clusters in a well-characterized host, e.g., *E. coli* and several *Streptomyces* species, using sophisticated DNA manipulation techniques.[79] However, some challenges, such as difficulties in cloning and engineering large gene clusters and lack of suitable heterologous hosts, hinder the application of this strategy.

The efficient cloning of cryptic gene clusters, which usually span 20-100 kb of DNA, remains a restriction for heterologous cluster expression besides the obvious problem of defining suitable/optimized host. The construction of gene libraries, screening them for target clusters, and stitching clusters together from regions spanning more than one clone have become prevalent. In addition, synthetic biology methods for the assembly of DNA (gene, biosynthetic pathway and even entire chromosome) from small synthetic DNA fragments have been developed. Although it allows codon usage and regulatory elements to be host-optimized for optimal performance, it remains rather expensive and it has the potential to introduce annoying random mutation.

The direct cloning method mediated by full length RecE and RecT presented in this thesis relies on the functional distinction and replication-independent recombination - LLHR. The LLHR refers to the homologous recombination between a linear *E. coli* vector flanked by two homologous arms which defines the limits of the target insert, and the linear target DNA

segment which is present in a digested genomic DNA mixture. The full length RecE/RecT which catalyzes homologous recombination between two linear DNA molecules with more than 20-folder greater efficiency compared to both truncated versions of RecE/RecT and Redα/Redβ pairs which are more superior for LCHR. We immediately implemented this new technique for the direct cloning of bacterial secondary metabolite biosynthetic pathways from digested genomic DNA of *P. luminescens*. This lead to discovery of two groups of compounds, luminmycins and luminmides, by heterologous expression of the pathways in *E. coli* strains Nissle 1917, and GB05-MtaA, respectively. In this case, we screened and engineered *E. coli* strain GB2005 and GB2006 to minimize the endogenous recombination occurring between the highly repetitive DNA regions of PKS and NRPS clusters. This issue enables these large heterologous DNA to be stable in *E. coli* and avails for the next step engineering and expression. Furthermore, in our cloning strategy, a tetracycline inducible promoter (P_{tetO}) was included in the linear cloning vector. After LLHR, the native promoter and ribosome binding site (RBS) at the upstream of first open reading frame of an operon in the target cluster were changed to an inducible promoter, sometime the uncommon/alternate start codons like GTG and TTG were altered to ATG at the same time. This exchange was performed based on two considerations: 1), the native promoter or regulatory genes may not function in the unrelated or phylogenetically distant host, thus the exchange of the promoter would enable expression in the heterologous host. 2), in our experience, foreign PKS/NRPS clusters are often toxic to the heterologous host when constitutively expressed. Consequently establishing the gene cluster under the control of an inducible promoter is the best choice for both plasmid maintenance and gene expression.

Although NRPS gene clusters up to 52 kb in size have been directly cloned from purified *P. luminescens* genomic DNA (5.7 Mb) with high efficiency (6/21) using a two-step strategy combining both LLHR and LCHR,[211] it is still a challenge to clone large gene cluster (>60kb) from even bigger chromosomes. Because the cloning efficiency decreases with an increasing size of the target gene cluster, we generally used a more complex strategy combined with LLHR to clone them. For example, the ~81 kb rhizoxin biosynthetic gene cluster from *Pseudomonas fluorescens* Pf-5 was separately (approximately 48 kb for upstream part, 36 kb for downstream part) cloned into p15A vectors by LLHR from digested genomic DNA. In the stitching step, the upstream part was initially subcloned into a BAC vector by Red/ET recombineering, then the downstream part was digested and ligated into the BAC construct containing upstream part to reconstitute the complete rhizoxin biosynthetic pathways(~ 91 kb).[212] The direct cloning of various PKS/NRPS gene clusters by LLHR from the genomes of

Gram-negative bacteria (*E. coli* Nissle, *Photorhabdus luminescens*, *Xenorhabdus* sp., *Pseudomonas syringae* and *P. fluorescens, Burkholderia DSM7029 and Burkholderia glumae*, myxobacteria *Myxococcus xanthus* and *Sorangium cellulosum*) and Gram-positive bacteria *Streptomyces coelicolor*, *Paenibacillus larvae* and *Bacillus thuringiensis* have in the meantime been accomplished by our group and our partners (unpublished data), which demonstrates the powerful capability and wide adaptability of direct cloning for potential bioprospecting.

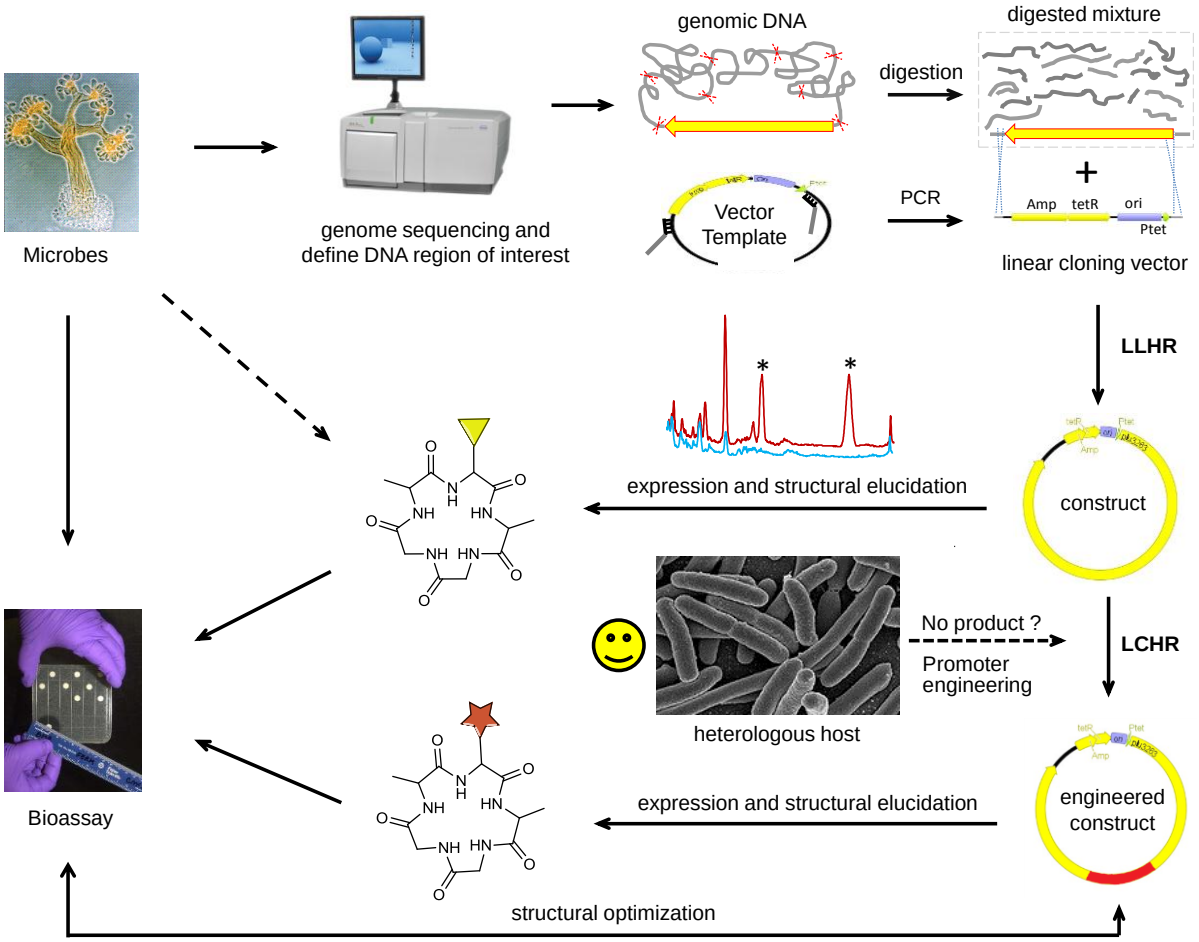


Figure C-1. Strategy of Red/ET recombineering mediated genomic-based bioprospecting of microbes for new bioactive natural product.

Using Red/ET recombineering, including LLHR by RecET and LCHR by Redαβ, cloning and mutagenesis of secondary metabolite pathways can be rapidly performed in *E. coli*. Coupled with heterologous gene cluster expression in a versatile host, a genomic-based bioprospecting of microbes for new bioactive natural products can be proposed (Figure C-1). The genome of microbial isolates can be sequenced rapidly and then analyzed by bioinformatics software such as antiSMASH to define the gene clusters of interest. Genomic DNA can be digested using suitable enzymes to release the chosen gene clusters onto single

DNA fragments and they can be direct cloned into an *E. coli* expression vector by LLHR for expression in an appropriate host to discover their products. The resulting constructs can also be modified by LCHR to exchange internal promoter(s), inactivation of repressors, or overexpression of positive regulators, to increase titers of products for bioactivity assay. The biosynthetic gene cluster encoding potential lead compounds can be further manipulated (e.g. site-directed mutagenesis, or deletion, insertion, fusion, swapping of domains or modules, or pathway fusion) by LCHR and other tools (e.g., digestion and ligation) for combinatorial biosynthesis to generate more derivatives for lead optimization. A more important issue is the choice of or engineering an ideal heterologous host for a specific gene cluster, such as addition of precursor biosynthetic pathway (MM-CoA for biosynthetis of most polyketides) and PPTase (Sfp, MtaA) into *E. coli*, or deletion of completing biosynthetic pathways. A couple of cryptic secondary metabolites were discovered by direct cloning following by heterologous expression in *E. coli*, *P. putida*, and *Burkholderia* species. Along with the successful genetic engineering of luminmide, luminmycin and syringolin biosynthetic pathways to change metabolite profile or generate new derivatives, prove Red/ET recombineering is a powerful tool for cloning and engineering of cryptic bacterial secondary metabolite biosynthetic pathways and promotes genomic-based natural product bioprospecting in a post-genomics era.

Admittedly, LLHR also has several limitations. It is contingent on the presence of unique restriction enzyme sites both upstream and downstream of the interested gene cluster to release it into single DNA segment, as the distance between the location of the restriction sites and chosen homology arm greatly affects the recombination efficiency. Otherwise, more rounds of Red/ET recombineering would be required to engineer the promoter, add necessary genes, or delete redundant sequences. Another problem which may be caused by LLHR is the mutations at 5′ ends of two cloned gene clusters, preventing functional expression. These results suggested they are toxic to *E. coli* cloning host even under the control of inducible promoter but without induction. More tightly regulated promoters are required in the future. At least, LLHR simplify the process for cloning of large gene clusters.

However, the successful expression of an unknown gene cluster is usually unpredictable owing to the differences between native and heterologous host, which may not produce all necessary precursors or low-production escaped from detection due to the inappropriate choice of heterologous host. It is safe to select the heterologous host which is related or close to native producer, increasing the possibility for successful expression. In our case, we cloned 10 gene clusters from *P. luminescens*, and seven were tried to express in the phylogenetically

close *E. coli* strain, but only two were successfully expressed to identify their products. In addition, successful expression may require several attempts with different heterologous hosts, but we only used the poor heterologous host *E. coli* for expression of unknown natural products. The further attempts of the other heterologous host would lead to identification of more compounds. In the other hand, the problem of pathways aspect includes the possibility of an internal promoter(s) or regulator(s) which may make the gene cluster 'silent' in a heterologous host. Guided by increasing understanding of biosynthesis and regulation of secondary metabolites, it can be readily addressed by Red/ET recombineering mediated gene targeting once a clone has been obtained. Hence, we have to search more 'universe' heterologous hosts and rationally manipulate cloned biosynthetic pathways in future work for robust heterologous expression and application of this strategy. To sum up, the development of Red/ET recombineering followed by suitable heterologous expression would promote natural product discovery and optimization.

2 A site-directed mutagenesis approach for NRPS: *ccdB* counterselection recombineering

Site-directed mutagenesis, also called site-specific mutagenesis or oligonucleotide-directed mutagenesis, is a molecular biology technique often used in bioengineering in which a mutation is created at a defined site in a DNA molecule. There are a number of methods for achieving this, e.g., overlapping PCR. Ideally, site-directed mutations should be seamless so that the final product does not include any unnecessary sequence that was acquired during mutagenesis, and it would best be performed *in vivo* so that unexpected mutations at the non-target sequence, which may be introduced by PCR methods, can ultimately be avoided, especially for the large constructs. These potential mutations are particularly important for mutagenesis within protein-coding regions of bacterial secondary metabolite biosynthetic megasynthetases where extraneous subtle changes would cause significant alteration of final production or even completely abolish it. The *in vivo* seamless mutagenesis can be efficiently achieved using counterselection recombineering strategy which utilizes counterselectable marker and Red/ET recombining. We report a method for performring site-directed mutagenesis, independent of traditional medium supplementation that utilizes *ccdB* as the counterselective marker.

We employed the *ccdB* as a counterselectable marker combined with Red/ET recombineering to specifically alter a few bases without introducing any unwanted sequence in the NRPS gene *plu3263* presented in a multi-copy plasmid pBR322 with a rational

efficiency (~15%). *ccdB* encodes a lethal protein that targets DNA gyrase. It can be propagated in *E. coli* strain DB3.1 which harbors a specific mutation in the gyrase rendering it resistant. In the first step of recombineering, the target sequences is replaced with a *ccdB-Cm* cassette including chloramphenicol resistant gene and counterselectable marker *ccdB* which is performed in DB3.1 strain containing *redγβα* expression plasmid. In the second step, the oligonucleotide-mediated recombineering was also performed in the DB3.1 strain containing Redγβα to form the plasmid mixture of repaired and unrepaired constructs in one cell. Because pBR322-derived plasmid is multi-copy, it is possible that only several copies can be repaired by single-stranded oligonucleotide but most copies will still contain *ccdB*. In order to select the repaired plasmid efficiently, the plasmids were extracted immediately from DB3.1 and retransformed into the *ccdB*-sensitive *E. coli* strain GB2005, thus only the mutant without *ccdB* can grow on the plate. However, not all the viable clones were the correct mutants; where possible intramolecular recombinants were generated removing the *ccdB* to allow the survival of *E. coli* strain GB2005, which was the most probable reason for the observed background of incorrect recombinants. Using this counterselection pattern through host change, even one copy of repaired plasmid can be selected and thus is a robust method for introducing point mutations into a multi-copy plasmid. One limitation of this counterselection method is that success is extremely dependent on the transformation efficiency, especially for large plasmid mutagenesis, but this is addressed by using the strain with the highest efficiency for the transformation of large plasmids - GB2005 (unpublished data). As a result of performing the site-directed mutagenesis of *plu3263,* only 15% of the recombinants were correct. While this fraction is low, the method is still relatively efficient, considering the iterative features of modules/domains in large NRPS biosynthetic genes and the mutation position in the middle of repetitive NRPS gene that would frequently lead to intramolecular recombination. The incorrect mutants in this investigation are the intramolecular recombinants between the regions upstream (module 1 and 2) and downstream (module 4 and 5) of module 3 (where the target sequence is located) as supported by restriction digestion examination. Several measures have been applied in the *rpsL* counterselection-based recombineering method to reduce these unwanted events, such as expressing Redβ alone at crucial step or using phosphothioated oligonucleotides, to improve the efficiency of site-directed mutagenesis in BAC.[206] These measures may also be implemented into *ccdB* counterselection recombineering to enhance the likelihood of obtaining the correct point mutations in multi-copy plasmid.

The efficient *ccdB* counterselction recombineering enriches the genetic engineer's tool kit and provides a powerful alternative for seamless modification of PKS and NRPS, which include not only site-directed mutagenesis using single stranded DNA as a substrate, but also domain/module swapping, and deletion using double stranded DNA (usually PCR product).

3 Examples: luminmides, luminmycins, syringolins

We used those two developed techniques, direct cloning and site-directed mutagenesis for cloning and engineering of cryptic luminmide and luminmycin biosynthetic pathways from *P. luminescens*, and for the known syringolin gene cluster from *P. syringae.*

3.1 Luminmide

The gene *plu3263* was directly cloned from genomic DNA of *P. luminescens* by LLHR and heterologously expressed in *E. coli* GB05-MtaA to uncover its previously secret seven products, Luminmides A-G. The diversity of luminmides stems from the NRPS substrate promiscuity. From the structural, genetic and bioinformatics analysis, the initiation module (A_1-T) can select two amino acids Val or Ile, and the third module (E/C-A_3-T) can activate five amino acids Phe, Leu, Val, Tyr or Met to give a diversified product scaffold. The NRPS substrate promiscuity is not uncommon and also present in other pathways such as tyrocidine[213] and xenematide.[214]

To further exploit this chemical diversity, we enforced *ccdB* counterselection recombineering to change several active residues followed by robust heterologous expression to compare the metabolic profiles of WT and mutated Plu3263. Six mutants were created and indeed great variation was found in the abundance of luminmides A-D due to the alteration of substrate specificity of A_3 domain. In C278T mutant, the production of luminmide A is still higher than other analogs. But in all the remaining mutants, the WT minor product luminmide B became the dominant product, and the yield of luminmide A decreased to diffident extents and was completely abolished in the double mutant C278M/A301G. Thus our results from altering the specificity of A_3 domain by site-directed mutagenesis of the 'specificity-conferring code' are highly encouraging as basically the domain could be altered in a way that enabled almost complete reprogramming. Gly301 mutaion is more efficient than Ala301, and Met278 is more efficient than Cys278, for activating of Leu in A_3 domain, while the Thr278 is less efficient than native Cys278 for selection and incorporation of Phe into luminmides. The latter conflicts with the fact that Thr278 presents in most of Phe-activating A domains but Cys278 is rare. In addition, the increase of luminmide G in the mutant C278M is consistent with our anticipation that Met in position 278 is also beneficial for recognition

of Val. The production of luminmide F and B are also increased in the mutant A301G, which implies that the 'specificity-conferring code' of Met-activating A domain may be similar to that of Leu, because a Met-activating A domain is rare in NRPS. Moreover, this site-directed mutagenesis is a feasible way to remove the predominant compound and/or improve yield of subordinate product, which can simplify purification procedure of low-level secondary metabolites, such as luminmide B here (Chapters 1 and 4).

3.2 Luminmycin

After the genome sequence of *P. luminescens* was published, several hidden PKS/NRPS biosynthetic gene clusters were found and structures of their products were predicted based on the NRPS rules for substrate specificity of A domains in the website NPbiogene (www.npbiogene.com, unavailable now).[215] A NRPS/PKS biosynthetic pathway was predicted from *plu1864-plu1886* genes and the corresponding putative compound was named luminmycin in this webpage (Figure C-2).[215] Blast analysis showed that this gene cluster (*plu1881-plu1877*) is highly similar to the glidobactin gene cluster from *Burkholderia* K481-B101 except for the absence of *glbH* and *glbE* homologs.[216] The genes *plu1864* to *plu1876* organized and transcribed in a different direction to *plu1881-plu1877* and therefore it is unlikely that they belong to one operon. The genes *plu1883-plu1885* (*bkdA-bkdC*) encoding enzymes for production of iso-fatty acids are involved in stilbene biosynthesis.[217] The gene *plu1882* encodes a glycoporin and is not a secondary metabolite biosynthesis-related protein. Thus we speculated that the *plu1881-plu1877* would encode an intact biosynthetic pathway and may generate glidobactin-like compounds. But no glidobactin analog was detected from the extract of *P. luminescens* WT strain using HPLC/MS analysis.

The *plu1881-plu1877* was directly cloned by LLHR and expressed in *E. coli* strain Nissle 1917 from which glidobactin is successfully heterologous expressed. There luminmycins were identified by comparative metabolome analysis and luminmycin A was found to be a 10-deoxy derivative of the previously reported glidobactin A, while luminmycins B and C most likely represent its acyclic biosynthetic intermediates (Figure A-12). Introduction of *glbH* into *plu1881-plu1877* leads to the partial conversion from luminmycin A to glidobactin A which demonstrates that GlbH involves in the formation of 4-hydroxylysine. Luminmycin A showed slightly weaker cytotoxicity than glidobactin indicating that the C-10 hydroxy group is a positive factor for cytotoxicity, while acyclic luminmycin B exhibited no activity at concentrations as high as 100 μg/mL.[218] The cryptic *plu1881-plu1877* was elucidated by heterologous expression in *E. coli*. Thus, the Red/ET recombineering mediated activation and

expression of cryptic biosynthetic gene clusters in heterologous hosts is shown to be a powerful tool for bioprospecting.

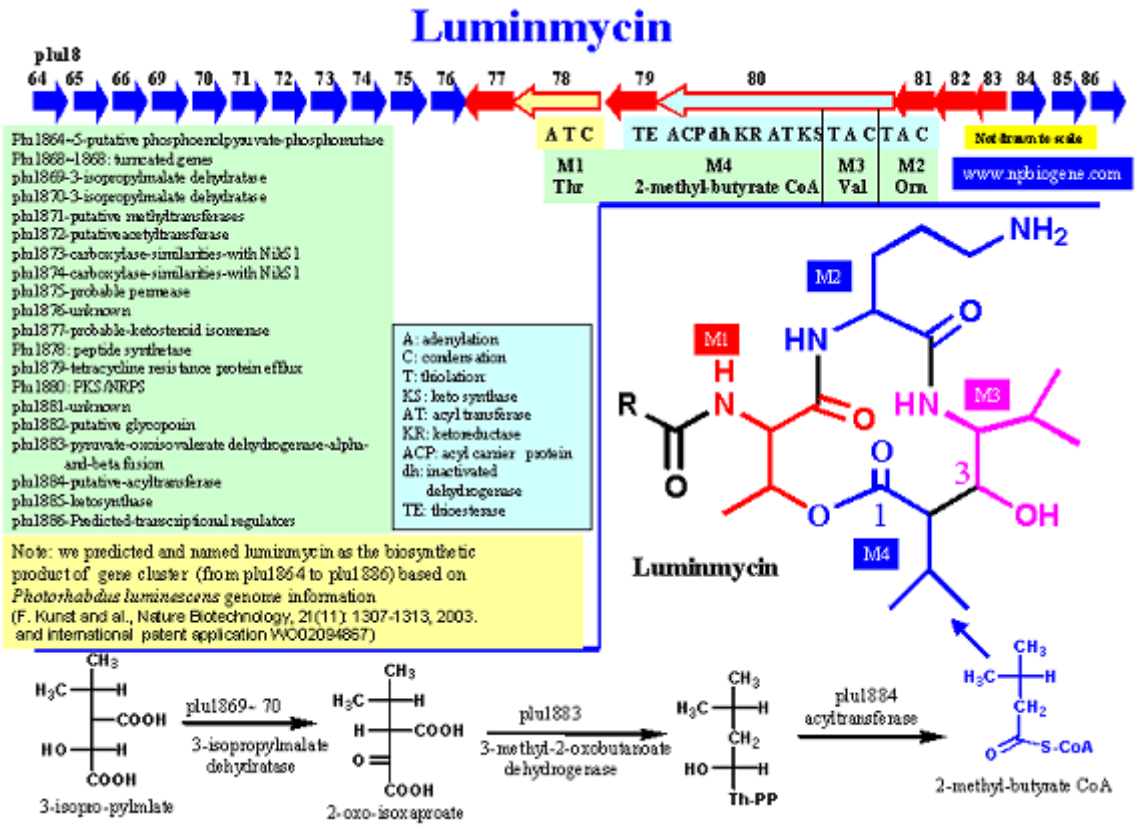


Figure C-2. The proposed luminmycin biosynthetic pathway from *plu1864* to *plu1886* in NPbiogene website.

The biosynthetic pathway for luminmycin was proposed based on comparative bioinformatics, genetic and structural analysis with glidobactin. The tripeptide part of luminmycin is formed by the NRPS/PKS megasynthetase, Plu1878 and Plu1880. Based on the signature sequence of the NRPS A domain, Plu1878, a homologue of GlbC, is predicted to activate Thr which is the *N*-terminal amino acid and it is *N*-acylated with 2(*E*), 4(*E*) dodecadienoic acid (Dda). This module contains a condensation domain prior to the first amino acid activation domain and this starter C domain would utilize fatty acyl-CoA to acylate the Thr amino group and generate an intermediate for subsequent chain elongation.[219] The Plu1880, a homologue of GlbF, is a NRPS/PKS hybrid protein including two NRPS modules and one PKS module. Its first and second NRPS modules are predicted to activate Lys which differs from glidobactin, 4-hydroxylysine, and Ala which is the same as glidobactin, respectively. If the hydrolysis happens in the second thiolation domain of Plu1880 it would result in the release of luminmycin C where one PKS extender unit is lacking. The PKS module would load malonyl-CoA and condense it into peptidyl chain and then undergo β-carbon reduction by KR and DH domains. Luminmycin B may be the

hydrolysis product of *β*-hydroxyacyl-*S*-T prior to dehydration in *β*-carbon processing. Then, the last cyclization should be achieved by the TE domain to generate luminmycin A.

The gene *plu1881* and *glbB* encodes a hypothetical protein containing a conserved domain BsmA which is a biofilm formation and stress response factor.[220] Plu1881 also shows weak similarity (Identities 26%, Positives 46%) to L-isoleucine-4-hydroxylase from *Bacillus thuringiensis*.[221] Because *glbB* is essential for glidobactin biosynthesis,[216] we reasoned Plu1881 may also be involved in the hydroxylation of lysine residue together with GlbH which is a dioxygenase-related protein. The fact that *glbH* mutant of *Burkholderia* K481-B101 can still produce trace glidobactin A shows one or more another gene products (e.g. Plu1881/GlbB) also responsible for the hydroxylation of lysine because the lack of the corresponding proteins may have impaired the function for hydroxylation. This could be testified by *in vitro* biochemical characterization, or deletion of *glbB* which caused the completely abolishment of glidobactin A but not loss of luminmycin A. Another homolog absent from luminmycin gene cluster is *glbE* encoding an MbtH-like protein which activates the A domains of NRPS GlbF,[219,222] but is not essential for biosynthesis of luminmycin. Addition of *glbF* into *plu1881-plu1877* does not improve the production of luminmycin. The *plu1877* and *glbG* encode ketosteroid isomerase-related proteins which may be involved in the biosynthesis of 2(*E*), 4(*E*) dodecadienoyl-CoA from unsaturated fatty acids.[223] The putative exporter Plu1879, a homologue of *glbD* and *sylE*, the exporters of glidobactin[216] and syringolin,[224] respectively, is hypothesized to also mediate luminmycin A export.

Luminmycins can also be one member of syrbactin beside syringolin and glidobactin,[225] which stands for a new class of potent proteasome inhibitors acting by the covalent attachment between the α,β-unsaturated carbonyl group of the 12-membered macrocylic core and the active site Thr1 of the 20S eukaryotic proteasome through a Michael-type 1,4-addition.[226,227] The high cytotoxicity of luminmycin A and the inactivity of open-ring luminmycin B are correlated well with the mechanism of proteasome inhibition.

3.3 Syringolin

We used LLHR to directly clone the partial biosynthetic pathway (*sylCDE*) from genomic DNA and then successfully expressed in a heterologous host *E. coli* GB05-MtaA, which resulted in the identification of two new members of the syringolins, syringolins G and H (Figure A-12). Both of compounds are Lys-saturated products of syrongolins F and C as a result of seizure of SylB.[228]

The gene *sylB* encodes a desaturase that would act specifically on Lys to form a non-natural derivative 3,4-dehydrolysine, the substrate for the first NRPS module of SylD.

Presumably due to toxicity of 3,4-dehydrolysine to *E. coli*, the heterologous overexpression of recombinant SylB was unsuccessful, hence the function of SylB was unable to be characterized *in vitro*.[229] But the desaturation occurred prior to activation and elongation by the SylD assembly line which was proven by the A domain preference for the activation of 3,4-dehydrolysine to lysine. Thus, heterologous expression of the remainder of the biosynthetic pathway combined with *in vivo* identification of Lys-saturated syringolins is an alternative way to identify the function of SylB *in vivo*. In the deleted *sylB* mutants, all the Lys-desaturated syringolins, syringolins A, C, D, and F, were completely abolished but Lys-saturated syringolins B and E, as well as newly discovered G and H were still present. This fact confirms that SylB catalyzes the desaturation of Lys to form 3,4-dehydrolysine *in vivo*.

The whole syringolin gene cluster was also reconstructed by addition of *sylB* and promoter exchange of a synthetic bidirectional promoter in place of the native promoter region to drive *sylB* and *sylC*. The genes *sylAB* were inserted into *sylCDE* to complete the syringolin gene cluster in which has an identical genetic organization, promoter and regulatory gene to the native producer. This construct also produced all eight syringolins when expressed in *E. coli* GB05-MtaA, but the production of Lys-desaturated syringolins was lower than that of Lys-saturated syringolins which may be partly due to the native promoter and regulatory gene being less active in *E. coli*. So we deleted the putative regulatory gene *sylA* and introduced a synthetic bidirectional promoter between *sylB* and *sylC* because the two genes are arranged in different transcription direction. The results showed an increase in the production of Lys-desaturated syringolins due to the stronger function of a synthetic promoter than native promoter in *E. coli*. Certainly, changing the native promoter to a more potent promoter can influence bacterial secondary metabolite production yields in both heterologous and native host. The Tn5-derived *npt* promotor or T7A1 promoter were inserted into the upstream of NRPS/PKS genes of thuggacin or myxoprincomide biosynthetic pathways in their native producers, *Chondromyces crocatus* and *M. xanthus*, respectively, which greatly improved their yields.[59, 230] Such an approach may ultimately prove fruitful for up-regulating biosynthesis from some of the many natural product gene clusters in native host, for which no product is currently known. Besides, the heterologous expression of biosynthetic gene clusters in an unrelated host generally requires the native promoter to be changed. For example, replacement of the native promoter of myxochromide S biosynthetic pathway from *S. aurantiaca* to toluic acid inducible P_m promoter led to improved production in *P. puptida*.[87] The heterologous expressions of biosynthetic pathways from a *Paenibacillus* species or a cyanobacterium were also achieved in *E. coli* GB05-MtaA by exchanging the native

promoter with L-arabinose inducible P_{BAD} or tetracycline inducible promoter (unpublished results). The promoter exchange together with heterologous expression in *E. coli* also awakens the silent luminmycin biosynthetic pathway in *P. luminescens.* Promoter engineering is an important factor to be considered for effective heterologous expression of bacterial secondary metabolites biosynthetic pathways.

4 Summary

Bacterial genome-sequencing projects have revealed that the majority of natural product biosynthetic pathways cannot currently be correlated to compounds and thus offers the opportunity for the discovery of novel bioactivies. Cloning, engineering and expression of a biosynthetic pathway of unknown function in a well-characterized heterologous host is an effective approach to discover and optimize its products, especially if the genes are derived from slow growing bacteria or such one for which genetics are only poorly or not at all established. The efforts towards genome sequencing, stitching of large secondary metabolite biosynthetic pathways by Red/ET recombineering, and efficient transfer of large gene clusters into heterologous hosts by transposition established a pipeline for the heterologous expression of large natural product biosynthetic pathways. In the current study, LLHR mediated direct cloning together with well-established genetic engineering of large gene clusters from genomic DNA, followed by robust heterologous expression promoted identification of some of their products and generation of new derivatives. The Red/ET recombineering mediated direct cloning and genetic engineering is expected to play an important role in facilitating genomic-based bioprospecting of microbes for new bioactive natural products (leads) in the future. Based on this technology, now gene cloning is not limiting genome mining anymore. Mostly, choice of heterologous host and biosynthetic engineering in the respective species now represents the most important limitation of heterologous expression. These issues require to be addressed in the future work.

D. References

1. Baker,D.D. & Alvi,K.A. Small-molecule natural products: new structures, new activities. *Curr. Opin. Biotechnol.* **15**, 576-583 (2004).
2. Chin,Y.W., Balunas,M.J., Chai,H.B., & Kinghorn,A.D. Drug discovery from natural sources. *AAPS Journal* **8**, E239-E253 (2006).
3. Firn,R.D. & Jones,C.G. Natural products - a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382-391 (2003).
4. Ganesan,A. The impact of natural products upon modern drug discovery. *Curr. Opin. Chem. Biol.* **12**, 306-317 (2008).
5. Cantrell,C.L., Dayan,F.E., & Duke,S.O. Natural Products As Sources for New Pesticides. *J. Nat. Prod.*(2012).
6. Baker,D.D., Chu,M., Oza,U., & Rajgarhia,V. The value of natural products to future pharmaceutical discovery. *Nat. Prod. Rep.* **24**, 1225-1244 (2007).
7. Newman,D.J. & Cragg,G.M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **75**, 311-335 (2012).
8. Li,J.W. & Vederas,J.C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161-165 (2009).
9. Alekshun,M.N. New advances in antibiotic development and discovery. *Expet Opin. Investig. Drugs* **14**, 117-134 (2005).
10. Baltz,R.H. Molecular engineering approaches to peptide, polyketide and other antibiotics. *Nat. Biotechnol.* **24**, 1533-1540 (2006).
11. Bode,H.B. & Müller,R. The impact of bacterial genomics on natural product research. *Angew. Chem. Int. Ed.* **44**, 6828-6846 (2005).
12. Walsh,C.T. & Fischbach,M.A. Natural Products Version 2.0: Connecting Genes to Molecules. *J. Am. Chem. Soc.* **132**, 2469-2493 (2010).
13. Hertweck,C. The biosynthetic logic of polyketide diversity. *Angew. Chem Int. Ed Engl.* **48**, 4688-4716 (2009).
14. Fischbach,M.A. & Walsh,C.T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468-3496 (2006).
15. Marahiel,M.A. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J. Pept. Sci.* **15**, 799-807 (2009).
16. Dewick,P.M. The biosynthesis of C5-C25 terpenoid compounds. *Nat. Prod. Rep.* **19**, 181-222 (2002).
17. Knaggs,A.R. The biosynthesis of shikimate metabolites. *Nat. Prod. Rep.* **18**, 334-355 (2001).
18. Knaggs,A.R. The biosynthesis of shikimate metabolites. *Nat. Prod. Rep.* **20**, 119-136 (2003).
19. Staunton,J. & Weissman,K.J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380-416 (2001).
20. Walsh,C.T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805-1810 (2004).
21. Conti,E., Stachelhaus,T., Marahiel,M.A., & Brick,P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174-4183 (1997).
22. Stachelhaus,T., Mootz,H.D., & Marahiel,M.A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493-505 (1999).
23. Challis,G.L., Ravel,J., & Townsend,C.A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211-224 (2000).
24. von Döhren,H., Dieckmann,R., & Pavela-Vrancic,M. The nonribosomal code. *Chem. Biol.* **6**, R273-R279 (1999).
25. Ansari,M.Z., Yadav,G., Gokhale,R.S., & Mohanty,D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32**, W405-W413 (2004).
26. Bachmann,B.O. & Ravel,J. Methods for *in Silico* Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. *Methods. Enymol.* **458**, 181-217 (2009).

27. Rausch,C., Weber,T., Kohlbacher,O., Wohlleben,W., & Huson,D.H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799-5808 (2005).
28. Chen,C.Y., Georgiev,I., Anderson,A.C., & Donald,B.R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. USA* **106**, 3764-3769 (2009).
29. Stevens,B.W., Lilien,R.H., Georgiev,I., Donald,B.R., & Anderson,A.C. Redesigning the PheA domain of gramicidin synthetase leads to a new understanding of the enzyme's mechanism and selectivity. *Biochemistry-US* **45**, 15495-15504 (2006).
30. Villiers,B. & Hollfelder,F. Directed evolution of a gatekeeper domain in nonribosomal Peptide synthesis. *Chem. Biol.* **18**, 1290-1299 (2011).
31. Eppelmann,K., Stachelhaus,T., & Marahiel,M.A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry-US* **41**, 9718-9726 (2002).
32. Evans,B.S., Chen,Y., Metcalf,W.W., Zhao,H., & Kelleher,N.L. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives *in vivo*. *Chem. Biol.* **18**, 601-607 (2011).
33. Uguru,G.C. *et al.* Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. *J. Am. Chem. Soc.* **126**, 5032-5033 (2004).
34. Thirlway,J. *et al.* Introduction of a non-natural amino Acid into a nonribosomal Peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed.* **51**, 7181-7184 (2012).
35. Han,J.W. *et al.* Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnol. Lett.* **34**, 1327-1334 (2012).
36. Lambalot,R.H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923-936 (1996).
37. Wenzel,S.C. & Müller,R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **26**, 1385-1407 (2009).
38. Medema,M.H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339-W346 (2011).
39. Wilkinson,B. & Micklefield,J. Mining and engineering natural-product biosynthetic pathways. *Nat. Chem. Biol.* **3**, 379-386 (2007).
40. Bentley,S.D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).
41. Ikeda,H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526-531 (2003).
42. Zazopoulos,E. *et al.* A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat. Biotechnol.* **21**, 187-190 (2003).
43. Gross,H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl. Microbiol. Biotechnol.* **75**, 267-277 (2007).
44. Paterson,I. & Anderson,E.A. The renaissance of natural products as drug candidates. *Science* **310**, 451-453 (2005).
45. Zotchev,S.B., Sekurova,O.N., & Katz,L. Genome-based bioprospecting of microbes for new therapeutics. *Curr. Opin. Biotechnol.* http://dx.doi.org/10.1016/j.copbio.2012.04.002 (2012).
46. Challis,G.L. Genome mining for novel natural product discovery. *J. Med. Chem.* **51**, 2618-2628 (2008).
47. Brady,S.F., Simmons,L., Kim,J.H., & Schmidt,E.W. Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat. Prod. Rep.* **26**, 1488-1503 (2009).
48. Oliynyk,M. *et al.* Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **25**, 447-453 (2007).
49. Schneiker,S. *et al.* Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **25**, 1281-1289 (2007).
50. Challis,G.L. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology-(UK)* **154**, 1555-1569 (2008).

51. Scherlach,K. & Hertweck,C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753-1760 (2009).
52. Zerikly,M. & Challis,G.L. Strategies for the discovery of new natural products by genome mining. *ChemBioChem* **10**, 625-632 (2009).
53. Nikolouli,K. & Mossialos,D. Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnol. Lett.* 34, 1393-1403 (2012).
54. Wu,M.C., Law,B., Wilkinson,B., & Micklefield,J. Bioengineering natural product biosynthetic pathways for therapeutic applications. *Curr. Opin. Biotechnol.* doi:10.1016/j.copbio.2012.03.008 (2012).
55. Bode,H.B., Bethe,B., Höfs,R., & Zeeck,A. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* **3**, 619-627 (2002).
56. Udwary,D.W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. USA* **104**, 10376-10381 (2007).
57. Wenzel,S.C. *et al.* Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: Evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *ChemBioChem* **6**, 375-385 (2005).
58. Lautru,S., Deeth,R.J., Bailey,L.M., & Challis,G.L. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol.* **1**, 265-269 (2005).
59. Cortina,N.S., Krug,D., Plaza,A., Revermann,O., & Müller,R. Myxoprincomide: A Natural Product from *Myxococcus xanthus* Discovered by Comprehensive Analysis of the Secondary Metabolome. *Angew. Chem. Int. Ed.* **51**, 811-816 (2012).
60. Corre,C., Song,L., O'Rourke,S., Chater,K.F., & Challis,G.L. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proc. Natl. Acad. Sci. USA* **105**, 17510-17515 (2008).
61. Jin,M., Liu,L., Wright,S.A.I., Beer,S.V., & Clardy,J. Structural and functional analysis of pantocin A: An antibiotic from *Pantoea agglomerans* discovered by heterologous expression of cloned genes. *Angew. Chem. Int. Ed.* **42**, 2898-2901 (2003).
62. Gross,H. *et al.* The genomisotopic approach: A systematic method to isolate products of orphan biosynthetic gene clusters. *Chem. Biol.* **14**, 53-63 (2007).
63. McClerren,A.L. *et al.* Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc. Natl. Acad. Sci. USA* **103**, 17243-17248 (2006).
64. Bok,J.W. *et al.* Genomic mining for Aspergillus natural products. *Chem. Biol.* **13**, 31-37 (2006).
65. Bok,J.W. *et al.* Chromatin-level regulation of biosynthetic gene clusters. *Nat. Chem Biol*(2009).
66. Bergmann,S. *et al.* Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat. Chem. Biol.* **3**, 213-217 (2007).
67. Hosaka,T. *et al.* Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **27**, 462-464 (2009).
68. Ochi,K. *et al.* Ribosome engineering and secondary metabolite production. *Adv. Appl. Microbiol.* **56**, 155-184 (2004).
69. Bumpus,S.B., Evans,B.S., Thomas,P.M., Ntai,I., & Kelleher,N.L. A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* 27, 951-956 (2009).
70. Evans,B.S., Ntai,I., Chen,Y.Q., Robinson,S.J., & Kelleher,N.L. Proteomics-based discovery of koranimine, a cyclic imine natural product. *J. Am. Chem. Soc.* **133**, 7316-7319 (2011).
71. Kersten,R.D. *et al.* A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **7**, 794-802 (2011).
72. Krug,D. *et al.* Discovering the hidden secondary metabolome of *Myxococcus xanthus*: a study of intraspecific diversity. *Appl. Environ. Microbiol.* **74**, 3058-3068 (2008).
73. Krug,D., Zurek,G., Schneider,B., Garcia,R., & Müller,R. Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionization-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* **624**, 97-106 (2008).
74. Pistorius,D. & Müller,R. Discovery of the Rhizopodin Biosynthetic Gene Cluster in *Stigmatella aurantiaca* Sg a15 by Genome Mining. *ChemBioChem* **13**, 416-426 (2012).

75. Henrikson,J.C., Hoover,A.R., Joyner,P.M., & Cichewicz,R.H. A chemical epigenetics approach for engineering the in situ biosynthesis of a cryptic natural product from *Aspergillus niger*. *Org. Biomol. Chem.* **7**, 435-438 (2009).
76. Tohyama,S. *et al.* Genome-inspired search for new antibiotics. Isolation and structure determination of new 28-membered polyketide macrolactones, halstoctacosanolides A and B, from *Streptomyces halstedii* HC34. *Tetrahedron* **60**, 3999-4005 (2004).
77. Tohyama,S., Kakinuma,K., & Eguchi,T. The complete biosynthetic gene cluster of the 28-membered polyketide macrolactones, halstoctacosanolides, from *Streptomyces halstedii* HC34. *J. Antibiot.* **59**, 44-52 (2006).
78. Medema,M.H., Breitling,R., Bovenberg,R., & Takano,E. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat. Rev. Microbiol.* **9**, 131-137 (2011).
79. Cobb,R.E. & Zhao,H. Direct cloning of large genomic sequences. *Nat. Biotechnol.* **30**, 405-406 (2012).
80. Galm,U. & Shen,B. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expet Opin. Drug Discov.* **1**, 409-437 (2006).
81. Gust,B. Cloning and Analysis of Natural Product Pathways. *Methods Enzymol.* **458**, 159-180 (2009).
82. Wenzel,S.C. & Müller,R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **16**, 594-606 (2005).
83. Zhang,Y., Buchholz,F., Muyrers,J.P., & Stewart,F.A. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123-128 (1998).
84. Zhang,Y., Muyrers,J.P.P., Testa,G., & Stewart,A.F. DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.* **18**, 1314-1317 (2000).
85. Muyrers,J.P.P., Zhang,Y., Testa,G., & Stewart,A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* **27**, 1555-1557 (1999).
86. Yu,D.G. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**, 5978-5983 (2000).
87. Wenzel,S.C. *et al.* Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via Red/ET recombineering. *Chem. Biol.* **12**, 349-356 (2005).
88. Perlova,O. *et al.* Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **72**, 7485-7494 (2006).
89. Binz,T.M., Wenzel,S.C., Schnell,H.J., Bechthold,A., & Müller,R. Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster by using Red/ET recombineering. *ChemBioChem* **9**, 447-454 (2008).
90. Hu,Y.F., Phelan,V.V., Farnet,C.M., Zazopoulos,E., & Bachmann,B.O. Reassembly of anthramycin biosynthetic gene cluster by using recombinogenic cassettes. *ChemBioChem* **9**, 1603-1608 (2008).
91. Wolpert,M., Heide,L., Kammerer,B., & Gust,B. Assembly and heterologous expression of the coumermycin A(1) gene cluster and production of new derivatives by genetic engineering. *ChemBioChem* **9**, 603-612 (2008).
92. Fu,J. *et al.* Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **36**, e113 (2008).
93. Chai,Y. *et al.* Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. *Chem. Biol.* **19**, 361-371 (2012).
94. Liao,G.J. *et al.* Cloning, reassembling and integration of the entire nikkomycin biosynthetic gene cluster into *Streptomyces ansochromogenes* lead to an improved nikkomycin production. *Microb. Cell Fact.* **9**, 6 (2010).
95. Rui,Z. *et al.* Biochemical and Genetic Insights into Asukamycin Biosynthesis. *J. Biol. Chem.* **285**, 24915-24924 (2010).
96. Kouprina,N. & Larionov,V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protocals.* **3**, 371-377 (2008).

97. Noskov,V.N. *et al.* A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics* **4**, (2003).
98. Kouprina,N. & Larionov,V. Innovation - TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat. Rev. Genet.* **7**, 805-812 (2006).
99. Kim,J.H. *et al.* Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers* **93**, 833-844 (2010).
100. Feng,Z.Y., Kim,J.H., & Brady,S.F. Fluostatins Produced by the Heterologous Expression of a TAR Reassembled Environmental DNA Derived Type II PKS Gene Cluster. *J. Am. Chem. Soc.* **132**, 11902-11903 (2010).
101. Gibson,D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-U41 (2009).
102. Shao,Z.Y., Zhao,H., & Zhao,H.M. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, (2009).
103. Shao,Z.Y., Luo,Y.Z., & Zhao,H.M. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* **7**, 1056-1059 (2011).
104. Otsuka,M., Ichinose,K., Fujii,I., & Ebizuka,Y. Cloning, sequencing, and functional analysis of an iterative type I polyketide synthase gene cluster for biosynthesis of the antitumor chlorinated polyenone neocarzilin in "*Streptomyces carzinostaticus*". *Antimicrob. Agents Chemother.* **48**, 3468-3476 (2004).
105. Dangel,V., Westrich,L., Smith,M.C.M., Heide,L., & Gust,B. Use of an inducible promoter for antibiotic production in a heterologous host. *Appl. Microbiol. Biotechnol.* **87**, 261-269 (2010).
106. Rodriguez-Garcia,A., Combes,P., Perez-Redondo,R., Smith,M.C.A., & Smith,M.C.M. Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*. *Nucleic Acids Res.* **33**, (2005).
107. Marques,S., Gallegos,M.T., & Ramos,J.L. Role of sigma S in transcription from the positively controlled Pm promoter of the TOL plasmid of *Pseudomonas putida*. *Mol. Microbiol.* **18**, 851-857 (1995).
108. Mermod,N., Lehrbach,P.R., Reineke,W., & Timmis,K.N. Transcription of the TOL plasmid toluate catabolic pathway operon of *Pseudomonas putida* is determined by a pair of co-ordinately and positively regulated overlapping promoters. *EMBO J.* **3**, 2461-2466 (1984).
109. Mutka,S.C., Carney,J.R., Liu,Y., & Kennedy,J. Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry-US* **45**, 1321-1330 (2006).
110. Pfeifer,B.A., Admiraal,S.J., Gramajo,H., Cane,D.E., & Khosla,C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790-1792 (2001).
111. Watanabe,K. *et al.* Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat. Chem. Biol.* **2**, 423-428 (2006).
112. Perlova,O., Gerth,K., Kuhlmann,S., Zhang,Y., & Müller,R. Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermopilic isolate *Corallococcus macrosporus* GT-2. *Microb. Cell Fact.* **8**, 1-11 (2009).
113. Sanchez,C. *et al.* The biosynthetic gene cluster for the antitumor rebeccamycin: Characterization and generation of indolocarbazole derivatives. *Chem. Biol.* **9**, 519-531 (2002).
114. Kwon,H.J., Smith,W.C., Xiang,L., & Shen,B. Cloning and heterologous expression of the macrotetrolide biosynthetic gene cluster revealed a novel polyketide synthase that lacks an acyl carrier protein. *J. Am. Chem. Soc.* **123**, 3385-3386 (2001).
115. Tang,L. *et al.* Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640-642 (2000).
116. Zhang,H, Boghigian,B.A., Armando,J., & Pfeifer,B.A. Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.* **28**, 125-151 (2011).
117. Zhang,H., Wang,Y., & Pfeifer,B.A. Bacterial hosts for natural product production. *Mol. Pharmaceut.* **5**, 212-225 (2008).
118. Walsh,C.T., Liu,J., Rusnak,R., & Sakaitani,M. Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chem. Rev.* **90**, 1105-1129 (1990).
119. Gehring,A.M., Mori,I., & Walsh,C.T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry-US* **37**, 2648-2659 (1998).

120. Nougayrede,J.P. *et al. Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848-851 (2006).
121. Homburg,S., Oswald,E., Hacker,J., & Dobrindt,U. Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*. *FEMS Microbiol. Lett.* **275**, 255-262 (2007).
122. Zhang,H., Wang,Y., Wu,J., Skalina,K., & Pfeifer,B.A. Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem Biol* **17**, 1232-1240 (2010).
123. Pfeifer,B.A., Wang,C.C.C., Walsh,C.T., & Khosla,C. Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host. *Appl. Environ. Microbiol.* **69**, 6698-6702 (2003).
124. Gross,F., Gottschalk,D., & Müller,R. Posttranslational modification of myxobacterial carrier protein domains in *Pseudomonas* sp. by an intrinsic phosphopantetheinyl transferase. *Appl. Microbiol. Biotechnol.* **68**, 66-74 (2005).
125. Finking,R. *et al.* Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. *J. Biol. Chem.* **277**, 50293-50302 (2002).
126. Gross,F. *et al.* Metabolic engineering of *Pseudomonas putida* for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation. *Chem. Biol.* **13**, 1253-1264 (2006).
127. Gross,F. *et al.* Bacterial type III polyketide synthases: Phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expession in pseudomonads. *Arch. Microbiol.* **185**, 28-38 (2006).
128. Velasco,A. *et al.* Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds. *Mol. Microbiol.* **56**, 144-154 (2005).
129. Julien,B. & Shah,S. Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*. *Antimicrob. Agents Chemother.* **46**, 2772-2778 (2002).
130. Stevens,D.C., Henry,M.R., Murphy,K.A., & Boddy,C.N. Heterologous Expression of the Oxytetracycline Biosynthetic Pathway in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **76**, 2681-2683 (2010).
131. Baltz,R.H. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.* **37**, 759-772 (2010).
132. Gottelt,M., Kol,S., Gomez-Escribano,J.P., Bibb,M., & Takano,E. Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2). *Microbiology-(UK)* **156**, 2343-2353 (2010).
133. Flinspach,K. *et al.* Heterologous expression of the biosynthetic gene clusters of coumermycin A(1), clorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. *Biopolymers* **93**, 823-832 (2010).
134. Gomez-Escribano,J.P. & Bibb,M.J. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **4**, 207-215 (2011).
135. Komatsu,M., Uchiyama,T., Omura,S., Cane,D.E., & Ikeda,H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA* **107**, 2646-2651 (2010).
136. Wang,Y. & Pfeifer,B.A. 6-Deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in E. coli. *Metab. Eng.g* **10**, 33-38 (2008).
137. Baltz,R.H. Streptomyces temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). *J. Ind. Microbiol. Biotechnol.* **39**, 661-672 (2012).
138. Sattely,E.S., Fischbach,M.A., & Walsh,C.T. Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Nat. Prod. Rep.* **25**, 757-793 (2008).
139. Goss,R.J., Shankar,S., & Fayad,A.A. The generation of "unNatural" products: Synthetic biology meets synthetic chemistry. *Nat. Prod. Rep.* **29**, 870-889 (2012).
140. Kennedy,J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat. Prod. Rep.* **25**, 25-34 (2008).

141. Weissman,K.J. Mutasynthesis - uniting chemistry and genetics for drug discovery. *Trends Biotechnol.* **25**, 139-142 (2007).
142. Patel,K. *et al.* Engineered biosynthesis of geldanamycin analogs for hsp90 inhibition. *Chem. Biol.* **11**, 1625-1633 (2004).
143. Petkovic,H. *et al.* A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase. *J. Antibiot.* **56**, 543-551 (2003).
144. Stachelhaus,T., Schneider,A., & Marahiel,M.A. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**, 69-72 (1995).
145. Nguyen,K.T. *et al.* Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. USA* **103**, 17462-17467 (2006).
146. Nguyen,K.T. *et al.* Genetically Engineered Lipopeptide Antibiotics Related to A54145 and Daptomycin with Improved Properties. *Antimicrob. Agents Chemother.*(2010).
147. Miao,V. *et al.* Genetic engineering in Streptomyces roseosporus to produce hybrid lipopeptide antibiotics. *Chem. Biol.* **13**, 269-276 (2006).
148. Schneider,A., Stachelhaus,T., & Marahiel,M.A. Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Mol. Gen. Genet.* **257**, 308-318 (1998).
149. Rowe,C. *et al.* Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem. Biol.* **8**, 475-485 (2001).
150. Mootz,H.D., Schwarzer,D., & Marahiel,M.A. Construction of hybrid peptide synthetases by module and domain fusions. *Proc. Natl. Acad. Sci. USA* **97**, 5848-5853 (2000).
151. Mootz,H.D. *et al.* Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. *J. Am. Chem. Soc.* **124**, 10980-10981 (2002).
152. Martin,C.J. *et al.* Heterologous expression in *Saccharopolyspora erythraea* of a pentaketide synthase derived from the spinosyn polyketide synthase. *Org. Biomol. Chem.* **1**, 4144-4147 (2003).
153. Gaisser,S. *et al.* Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of *Streptomyces avermitilis* ATCC31272. *Org Biomol Chem* **1**, 2840-2847 (2003).
154. Arslanian,R.L. *et al.* A new cytotoxic epothilone from modified polyketide synthases heterologously expressed in Myxococcus xanthus. *J. Nat. Prod.* **65**, 1061-1064 (2002).
155. Tang,L. *et al.* Elucidating the mechanism of cis double bond formation in epothilone biosynthesis. *J. Am. Chem. Soc.* **126**, 46-47 (2004).
156. Tang,L. *et al.* Generation of new epothilones by genetic engineering of a polyketide synthase in *Myxococcus xanthus*. *J. Antibiot.* **58**, 178-184 (2005).
157. Watanabe,K. *et al.* Rationally Engineered Total Biosynthesis of a Synthetic Analogue of a Natural Quinomycin Depsipeptide in *Escherichia coli*. *ChemBioChem* **10**, 1965-1968 (2009).
158. Keller,U. & Schauwecker,F. Combinatorial biosynthesis of non-ribosomal peptides. *Comb. Chem. High Throughput Screen* **6**, 527-540 (2003).
159. Sieber,S.A. & Marahiel,M.A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715-738 (2005).
160. Eustaquio,A.S. & Moore,B.S. Mutasynthesis of fluorosalinosporamide, a potent and reversible inhibitor of the proteasome. *Angew. Chem. Int. Ed.* **47**, 3936-3938 (2008).
161. Kusebauch,B., Brendel,N., Kirchner,H., Dahse,H.M., & Hertweck,C. Assessing Oxazole Bioisosteres as Mutasynthons on the Rhizoxin Assembly Line. *ChemBioChem* **12**, 2284-2288 (2011).
162. Weist,S. *et al.* Fluorobalhimycin - A new chapter in glycopeptide antibiotic research. *Angew. Chem. Int. Ed.* **41**, 3383-3385 (2002).
163. Weist,S. *et al.* Mutasynthesis of glycopeptide antibiotics: Variations of vancomycin's AB-ring amino acid 3,5-dihydroxyphenylglycine. *J. Am. Chem. Soc.* **126**, 5942-5943 (2004).
164. Gregory,M.A. *et al.* Mutasynthesis of rapamycin analogues through the manipulation of a gene governing starter unit biosynthesis. *Angew. Chem. Int. Ed.* **44**, 4757-4760 (2005).
165. Ragab,A.E., Gruschow,S., Rackham,E.J., & Goss,R.J.M. New pacidamycins biosynthetically: probing N- and C-terminal substrate specificity. *Org. Biomol. Chem.* **8**, 3128-3129 (2010).

166. Amir-Heidari,B., Thirlway,J., & Micklefield,J. Auxotrophic-precursor directed biosynthesis of nonribosomal lipopeptides with modified tryptophan residues. *Org. Biomol. Chem.* **6**, 975-978 (2008).
167. Powell,A., Al Nakeeb,M., Wilkinson,B., & Micklefield,J. Precursor-directed biosynthesis of nonribosomal lipopeptides with modified glutamate residues. *Chem. Commun.*2683-2685 (2007).
168. Bode,H.B. *et al.* Mutasynthesis-derived myxalamids and origin of the isobutyryl-CoA starter unit of myxalamid B. *ChemBioChem* **8**, 2139-2144 (2007).
169. Eichner,S., Floss,H.G., Sasse,F., & Kirschning,A. New, Highly Active Nonbenzoquinone Geldanamycin Derivatives by Using Mutasynthesis. *ChemBioChem* **10**, 1801-1805 (2009).
170. Kim,W. *et al.* Rational Biosynthetic Engineering for Optimization of Geldanamycin Analogues. *ChemBioChem* **10**, 1243-1251 (2009).
171. Menzella,H.G. *et al.* Potent Non-Benzoquinone Ansamycin Heat Shock Protein 90 Inhibitors from Genetic Engineering of *Streptomyces hygroscopicus*. *J. Med. Chem.* **52**, 1518-1521 (2009).
172. Eichner,S. *et al.* Broad Substrate Specificity of the Amide Synthase in *S. hygroscopicus*-New 20-Membered Macrolactones Derived from Geldanamycin. *J. Am. Chem. Soc.* **134**, 1673-1679 (2012).
173. Harvey,C.J.B., Puglisi,J.D., Pande,V.S., Cane,D.E., & Khosla,C. Precursor Directed Biosynthesis of an Orthogonally Functional Erythromycin Analogue: Selectivity in the Ribosome Macrolide Binding Pocket. *J. Am. Chem. Soc.* **134**, 12259-12265 (2012).
174. Taft,F. *et al.* Highly active ansamitocin derivatives: mutasynthesis using an AHBA-blocked mutant. *ChemBioChem* **9**, 1057-1060 (2008).
175. Taft,F. *et al.* Ansamitocin libraries by combining mutasynthesis with chemical synthesis; a new version of total synthesis. *Synlett.* 23, 1416-1426 (2012).
176. Eichner,S. *et al.* The Interplay between Mutasynthesis and Semisynthesis: Generation and Evaluation of an Ansamitocin Library. *Angew. Chem. Int. Ed.* **51**, 752-757 (2012).
177. Taft,F. *et al.* Combined Muta- and Semisynthesis: A Powerful Synthetic Hybrid Approach to Access Target Specific Antitumor Agents Based on Ansamitocin P3. *Chem. Eur. J.* **18**, 880-886 (2012).
178. Knobloch,T. *et al.* Mutational biosynthesis of ansamitocin antibiotics: a diversity-oriented approach to exploit biosynthetic flexibility. *ChemBioChem* **12**, 540-547 (2011).
179. Beck,Z.Q., Aldrich,C.C., Magarvey,N.A., Georg,G.I., & Sherman,D.H. Chemoenzymatic synthesis of cryptophycin/arenastatin natural products. *Biochemistry-US* **44**, 13457-13466 (2005).
180. Ding,Y.S., Rath,C.M., Bolduc,K.L., Hakansson,K., & Sherman,D.H. Chemoenzymatic Synthesis of Cryptophycin Anticancer Agents by an Ester Bond-Forming Non-ribosomal Peptide Synthetase Module. *J. Am. Chem. Soc.* **133**, 14492-14495 (2011).
181. Ziehl,M., He,J., Dahse,H.M., & Hertweck,C. Mutasynthesis of aureonitrile: An aureothin derivative with significantly improved cytostatic effect. *Angew. Chem. Int. Ed.* **44**, 1202-1205 (2005).
182. Rachid,S. *et al.* Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *ChemBioChem* **12**, 922-931 (2011).
183. Steinmetz,H. *et al.* Precursor-directed syntheses and biological evaluation of new elansolid derivatives. *ChemBioChem* **13**, 1813-1817 (2012).
184. Baltz,R.H. Biosynthesis and Genetic Engineering of Lipopeptides in *Streptomyces Roseosporus*. *Methods Enzymol.* **458**, 511-531 (2009).
185. Alexander,D.C. *et al.* Development of a Genetic System for Combinatorial Biosynthesis of Lipopeptides in *Streptomyces fradiae* and Heterologous Expression of the A54145 Biosynthesis Gene Cluster. *Appl. Environ. Microbiol.* **76**, 6877-6887 (2010).
186. Baltz,R.H. Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.* doi: 10.1021/sb3000673, (2012).
187. deFerra,F., Rodriguez,F., Tortora,O., Tosi,C., & Grandi,G. Engineering of peptide synthetases - Key role of the thioesterase-like domain for efficient production of recombinant peptides. *J. Biol. Chem.* **272**, 25304-25309 (1997).
188. Zhang,M.Q. *et al.* Optimizing natural products by biosynthetic engineering: Discovery of nonquinone Hsp90 inhibitors. *J. Med. Chem.* **51**, 5494-5497 (2008).

189. Watanabe,K. *et al. Escherichia coli* Allows Efficient Modular Incorporation of Newly Isolated Quinomycin Biosynthetic Enzyme into Echinomycin Biosynthetic Pathway for Rational Design and Synthesis of Potent Antibiotic Unnatural Natural Product. *J. Am. Chem Soc.*131, 9347-9353 (2009).
190. Eustaquio,A.S. *et al.* Production of 8'-halogenated and 8'-unsubstituted novobiocin derivatives in genetically engineered *Streptomyces coelicolor* strains. *Chem. Biol.* **11**, 1561-1572 (2004).
191. Boddy,C.N., Hotta,K., Tse,M.L., Watts,R.E., & Khosla,C. Precursor-directed biosynthesis of epothilone in *Escherichia coli*. *J. Am. Chem. Soc.* **126**, 7436-7437 (2004).
192. Wang,J. *et al.* An improved recombineering approach by adding RecA to lambda Red recombination. *Mol. Biotechnol* **32**, 43-53 (2006).
193. Copeland,N.G., Jenkins,N.A., & Court,D.L. Recombineering: A powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769-779 (2001).
194. Sharan,S.K., Thomason,L.C., Kuznetsov,S.G., & Court,D.L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protocols* **4**, 206-223 (2009).
195. Zhang,Y., Muyrers,J.P., Rientjes,J., & Stewart,A.F. Phage annealing proteins promote oligonucleotide-directed mutagenesis in *Escherichia coli* and mouse ES cells. *BMC Mol. Biol.* **4**, 1 (2003).
196. Muyrers,J.P., Zhang,Y., Buchholz,F., & Stewart,A.F. RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.* **14**, 1971-1982 (2000).
197. Kovall,R. & Matthews,B.W. Toroidal structure of lambda-exonuclease. *Science* **277**, 1824-1827 (1997).
198. Erler,A. *et al.* Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J. Mol. Biol.* **391**, 586-598 (2009).
199. Iyer,L.M., Koonin,E.V., & Aravind,L. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**, 8 (2002).
200. Kuzminov,A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol.Mol. Biol. Rev.* **63**, 751-813 (1999).
201. Fu,J., Teucher,M., Anastassiadis,K., Skarnes,W., & Stewart,A.F. A recombineering pipeline to make conditional targeting constructs. *Methods Enzymol.* **477**, 125-144 (2010).
202. Sarov,M. *et al.* A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nat. Methods* **3**, 839-844 (2006).
203. Testa,G. *et al.* Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat. Biotechnol.* **21**, 443-447 (2003).
204. Muyrers,J.P.P., Zhang,Y.M., & Stewart,A.F. Techniques: Recombinogenic engineering - new options for cloning and manipulating DNA. *Trends Biochem. Sci.* **26**, 325-331 (2001).
205. Ellis,H.M., Yu,D.G., DiTizio,T., & Court,D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* **98**, 6742-6746 (2001).
206. Bird,A.W. *et al.* High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. *Nat. Methods* **9**, 103-109 (2011).
207. Heermann,R., Zeppenfeld,T., & Jung,K. Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red /ET recombination. *Microb. Cell Fact.* **7**, 14 (2008).
208. Murphy,K.C. Phage recombinases and their applications. *Adv. Virus Res.* **83**, 367-414 (2012).
209. Skarnes,W.C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-U361 (2011).
210. Eustaquio,A.S. *et al.* Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. *Appl. Environ. Microbiol.* **71**, 2452-2459 (2005).
211. Fu,J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440-446 (2012).
212. Shan,S. Cloning, heterologous expression and their structural and functional study of bacterial active product gene clusters. 2012. Ref Type: Thesis/Dissertation, Changsha, China
213. Mootz,H.D. & Marahiel,M.A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843-6850 (1997).

214. Crawford,J.M., Portmann,C., Kontnik,R., Walsh,C.T., & Clardy,J. NRPS substrate promiscuity diversifies the xenematides. *Org. Lett.* **13**, 5144-5147 (2011).
215. Donadio,S., Monciardini,P., & Sosio,M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep.* **24**, 1073-1109 (2007).
216. Schellenberg,B., Bigler,L., & Dudler,R. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. *Environ. Microbiol.* **9**, 1640-1650 (2007).
217. Joyce,S.A. *et al.* Bacterial biosynthesis of a multipotent stilbene. *Angew. Chem. Int. Ed.* **47**, 1942-1945 (2008).
218. Bian,X., Plaza,A., Zhang,Y., & Müller,R. Luminmycins A-C, Cryptic Natural Products from *Photorhabdus luminescens* Identified by Heterologous Expression in *Escherichia coli*. *J. Nat. Prod.***75**, 1652-1655 (2012).
219. Imker,H.J., Krahn,D., Clerc,J., Kaiser,M., & Walsh,C.T. *N*-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem. Biol.* **17**, 1077-1083 (2010).
220. Weber,M.M., French,C.L., Barnes,M.B., Siegele,D.A., & Mclean,R.J.C. A previously uncharacterized gene, yjfO (bsmA), influences *Escherichia coli* biofilm formation and stress response. *Microbiology-(UK)* **156**, 139-147 (2010).
221. Smirnov,S.V. *et al.* Metabolic engineering of Escherichia coli to produce (2S, 3R, 4S)-4-hydroxyisoleucine. *Appl. Microbiol. Biotechnol.* **88**, 719-726 (2010).
222. Zhang,W., Heemstra,J.R., Walsh,C.T., & Imker,H.J. Activation of the Pacidamycin PacL Adenylation Domain by MbtH-Like Proteins. *Biochemistry-US*(2010).
223. Numata,K. *et al.* Enhanced Production of the Minor Components of Glidobactins in *Polyangium-Brachysporum*. *J. Antibiot.* **41**, 1358-1365 (1988).
224. Amrein,H. *et al.* Functional analysis of genes involved in the synthesis of syringolin A by *Pseudomonas syringae* pv. *syringae* B301D-R. *Mol. Plant-Microbe Interact.* **17**, 90-97 (2004).
225. Krahn,D., Ottmann,C., & Kaiser,M. The chemistry and biology of syringolins, glidobactins and cepafungins (syrbactins). *Nat. Prod. Rep.* **28**, 1854-1867 (2011).
226. Groll,M. *et al.* A plant pathogen virulence factor inhibits the eukaryotic proteasome by a novel mechanism. *Nature* **452**, 755-758 (2008).
227. Clerc,J. *et al.* The natural product hybrid of Syringolin A and Glidobactin A synergizes proteasome inhibition potency with subsite selectivity. *Chem. Commun.* **47**, 385-387 (2011).
228. Bian,X. *et al.* Direct Cloning, Genetic Engineering, and Heterologous Expression of the Syringolin Biosynthetic Gene Cluster in *E. coli* through Red/ET Recombineering. *Chembiochem.* **13**, 1946-1952 (2012).
229. Wuest,W.M., Krahn,D., Kaiser,M., & Walsh,C.T. Enzymatic timing and tailoring of macrolactamization in syringolin biosynthesis. *Org. Lett.* **13**, 4518-4521 (2011).
230. Buntin,K. *et al.* Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. *Chem. Biol.* **17**, 342-356 (2010).

E. Appendix

Author's efforts in publications presented in this work

Chapter 1: The author performed the construction of a heterologous host and plasmids for expression, annotation and bioinformatic analysis of cloned biosynthetic genes, heterologous expression of unknown biosynthetic gene cluster including fermentation, extraction, and comparative LC/MS analysis of extracts, isolation and structural elucidation of luminmides A and B by using HRMS, 2D-NMR and advanced Marfey's method, and structural prediction of luminmycin A by comparative MS/MS fragmentation experiments.

Chapter 2: The author generated the constructs for higher expression, isolated the pure luminmycins A-C and glidobactin A, determined the structures by using HRMS, 2D-NMR and advanced Marfey's method.

Chapter 3: The author designed and performed the single base pair substitution in *plu3263* in pBR322, fermentation, extraction, LC/MS analysis of mutants.

Chapter 4: The author designed and generated all the mutated plasmids, and analyzed the metabolic profiles of luminmides. Further the author purified luminmide C and identified its structure by NMR. The structures of luminmides D-G were also proposed by the author based on the feeding experiments and MS/MS fragmentation experiments.

Chapter 5: The author constructed the expression plasmids and heterologous host for syringolin biosynthetic gene cluster, performed the fermentation, extraction, and LC/MS analysis of mutants, and analyzed the MS/MS fragmentation data for structural elucidation of new syringolins.

Patent

ZHANY, Youming; FU, Jun; **BIAN, Xiaoying**; MÜLLER, Rolf; STEWART, Francis. (2010) Heterologous expression of PKS/NRPS gene cluster and targeted anticancer drug delivery. ***International patent application No: PCT/IB2010/055923***

Award

'HZI – Paper of the Month' in July, 2012.

Conference contributions

Bian, Xiaoying; Wenzel, Silke C.; Zhang, Youming; Müller, Rolf. Efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives *in vivo* (**Poster**). International VAAM-Workshop 2012: "Biology and Chemistry of Antibiotic-Producing Bacteria and Fungi". September, 2012, Braunschweig, Germany.

Bian, Xiaoying; Fu, Jun; Plaza, Alberto; Müller, Rolf; Zhang, Youming. Rapid Cloning and engineering of natural product biosynthetic pathway via Red/ET recombineering (**Poster**). ESF-EMBO Symposium: Synthetic Biology of Antibiotic Production. October, 2011, Sant Deliu de Guixols, Spain.

Bian, Xiaoying. Mining and engineering of natural product biosynthetic pathways by Red/ET recombineering (**Oral presentation**). International VAAM-Workshop 2011: "Biology of Bacteria Producing Natural Products". September, 2011, Bonn, Germany.

Bian Xiaoying; Klefisch, Thorsten; Huang, Fan; Wenzel, Silke C.; Müller, Rolf; Zhang, Youming. Heterologous expression of two large natural product biosynthetic gene clusters in *Burkholderia* K481-B101 **(Poster).** International VAAM-Workshop 2011: "Biology of Bacteria Producing Natural Products". September, 2011, Bonn, Germany.

Bian, Xiaoying; Fu, Jun, Stewart, A. Francis; Müller, Rolf; Zhang, Youming. Direct Cloning & Heterologous Expression: A Strategy to Study Unknown Biosynthetic Pathways (**Poster**). International VAAM-Workshop 2010: "Biology of Bacteria Producing Natural Products". September, 2010, Tübingen, Germany.

Zhang, Youming; Fu, Jun; Hu, Shengbiao; **Bian, Xiaoying**; Müller, Rolf; Stewart, A. Francis. Direct Cloning of Large Size Gene Clusters from Bacterial Genomic DNA Using Red/ET Recombineering (**Oral presentation**). 37th International Conference on the Biology of Myxobacteria. Augest, 2010, Nonnweiler-Otzenhausen, Saarland, Germany.

Bian, Xiaoying. A Novel Method for Direct Cloning Silent Biosynthetic Gene Clusters (**Oral presentation**). 1st Life Science PhD Student Day. Saarland University. August, 2009, Saarbrücken, Germany.