ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Modular Construction of a Functional Artificial Epothilone Polyketide Pathway

Corina Oßwald,[†] Gregor Zipf,[‡] Gisela Schmidt,[‡] Josef Maier,[¶] Hubert S. Bernauer,[‡] Rolf Müller,[†] and Silke C. Wenzel*[,†]

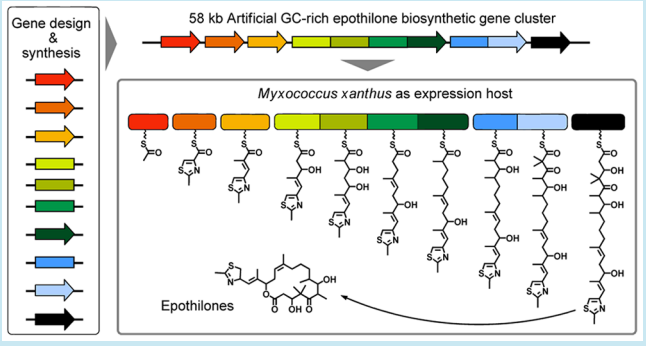
[†]Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research and Pharmaceutical Biotechnology, Saarland University, Saarbrücken, Germany

[‡]ATG Biosynthetics GmbH, Merzhausen, Germany

[¶]IStLS, Information Services to Life Science, Oberndorf a.N., Germany

Supporting Information

ABSTRACT: Natural products of microbial origin continue to be an important source of pharmaceuticals and agrochemicals exhibiting potent activities and often novel modes of action. Due to their inherent structural complexity chemical synthesis is often hardly possible, leaving fermentation as the only viable production route. In addition, the pharmaceutical properties of natural products often need to be optimized for application by sophisticated medicinal chemistry and/or biosynthetic engineering. The latter requires a detailed understanding of the biosynthetic process and genetic tools to modify the producing organism that are often unavailable. Consequently, heterologous expression of complex natural product pathways has been in the focus of development over recent years. However, piecing together existing DNA cloned from natural sources and achieving efficient expression in heterologous circuits represent several limitations that can be addressed by synthetic biology. In this work we have redesigned and reassembled the 56 kb epothilone biosynthetic gene cluster from *Sorangium cellulosum* for expression in the high GC host *Myxococcus xanthus*. The codon composition was adapted to a modified codon table for *M. xanthus*, and unique restriction sites were simultaneously introduced and others eliminated from the sequence in order to permit pathway assembly and future interchangeability of modular building blocks from the epothilone megasynthetase. The functionality of the artificial pathway was demonstrated by successful heterologous epothilone production in *M. xanthus* at significant yields that have to be improved in upcoming work. Our study sets the stage for future engineering of epothilone biosynthesis and production optimization using a highly flexible assembly strategy.



KEYWORDS: *epothilone, polyketide biosynthesis, heterologous expression, pathway engineering, synthetic biology, artificial gene cluster*

Today, the most successful pharmaceuticals are proteins (largely antibodies) and so-called small molecule drugs. The latter are usually natural products, natural product derived, or designed by medicinal chemistry from scratch with various intermediate classifications possible.[1] Natural products are thought to occupy a more sophisticated "structural space" due to their chemical complexity[2] and are also thought to be evolutionary optimized binders for biological targets, which makes them very promising compounds for drug development. However, these advantages also imply challenges as total synthesis of complex compounds for supply poses a demanding task for chemists and medicinal chemists aiming at structure−activity relationship studies and subsequent development of a candidate structure. Consequently, alternative production strategies are thought after. Here, in theory microbial natural products offer tremendous biotechnological opportunities because the producing organism can be grown in large bioreactors and genetic engineering of biosynthetic pathways can be achieved to deliberately alter the structure of the natural product once the biochemistry and the genetics of the underlying biosynthetic pathway are understood. This holds especially true for the multimodular assembly of polyketides and nonribosomal peptides, which are made in a stepwise process starting from simple building blocks. The underlying biochemistry of the involved polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) has been studied in great detail,[3,4] but significant challenges remain for engineering these complex megasynthetases as our understanding of assembly line structures and module/catalytic domain interactions is still very limited.[5] PKS and NRPS are widely

Special Issue: Natural Products

Received: August 31, 2012
Published: October 25, 2012

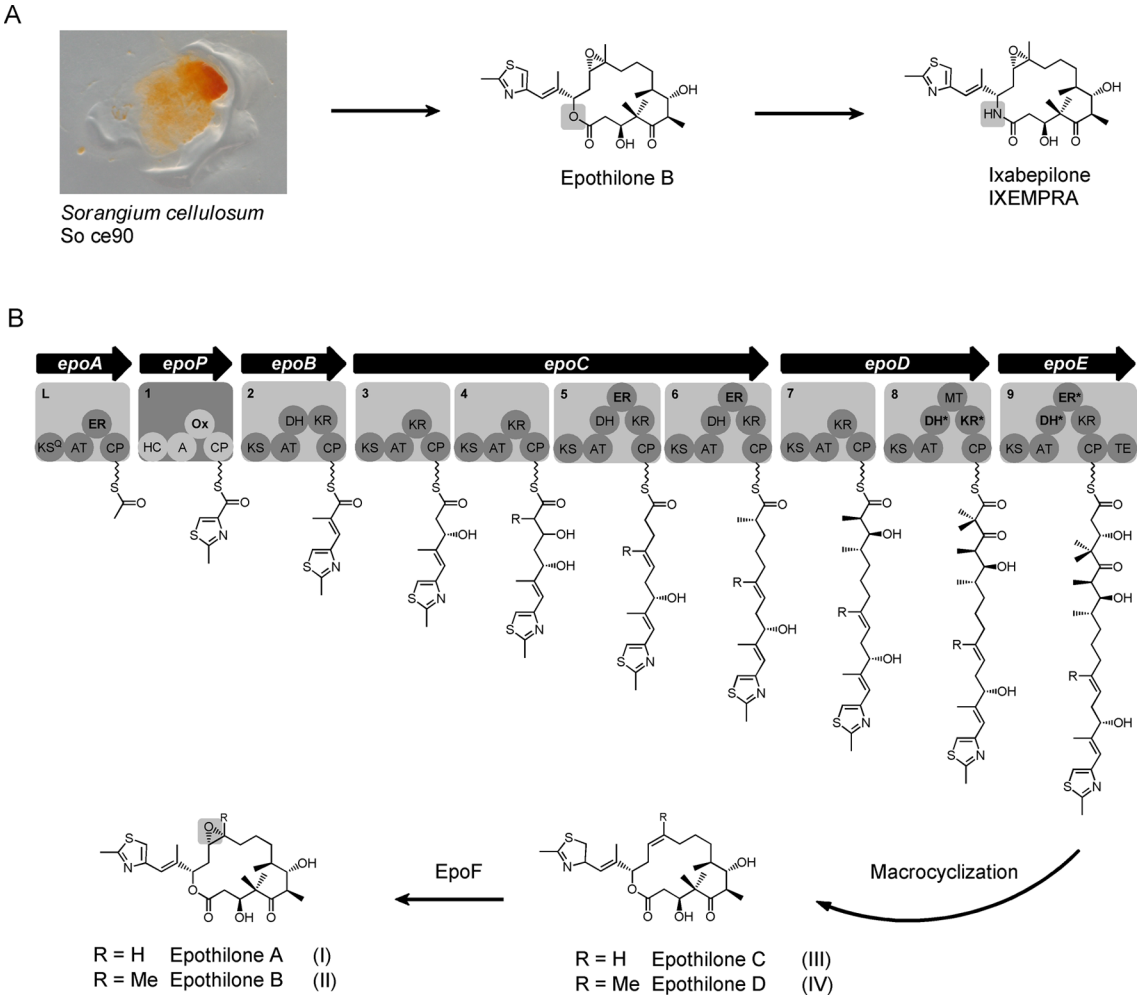


Figure 1. Biotechnological production of epothilones. (A) Generation of the anticancer drug IXEMPRA (Ixabepilone) based on large scale cultivations of the myxobacterial producer *S. cellulosum* So ce90. The picture shows a colony of the producer strain growing on VY2 medium and degrading the agar. After isolation of epothilone B from culture extracts the lactone is converted into the corresponding lactame via few medicinal chemistry steps to generate ixabepilone. (B) Epothilone biosynthetic gene cluster and the encoded biosynthetic machinery representing a multimodular polyketide/nonribosomal peptide megasynthetase. Genes are shown as black arrows, and modules of the biosynthetic machinery as gray boxes, each containing a set of catalytic domains illustrated as gray circles (KS, ketoacyl synthase; AT, acyl transferase; CP, acyl or peptidyl carrier protein; KR, β-ketoacyl reductase; DH, β-hydroxyacyl dehydratase; ER, enoyl reductase; MT, methyl transferase; TE, thioesterase). The main products, epothilones A–D (**I–IV**),[44] differ in one aliphatic residue (R = H or Me) as a result of the alternative incorporation of malonyl-CoA or methylmalonyl-CoA by module 4 and the presence or absence of the epoxide functionality installed by EpoF.

distributed and actually responsible for the formation of the majority of bioactive compounds isolated from microbes. Genome sequencing shows that an enormous potential for the production of further metabolites is encoded in microbial genomes awaiting exploitation by genome mining and synthetic biotechnology efforts.[6,7]

Therefore, to translate DNA sequences into real natural chemicals in the flask, efficient tools for structure and yield optimization are clearly required. However, this aim is notoriously hard to achieve due to the complexity of the involved megasynthetases and the difficulties encountered in the manipulation of biosynthetic pathways and producing organisms. Nevertheless engineering of PKS and NRPS assembly lines has already met with significant success,[8,9] and more positive results are expected once more and more efficient engineering tools are available.

In contrast to the actinomycetes, which have been studied for decades and for which numerous genetic tools are available, myxobacteria represent an underexploited source for natural products, although they are very well established producers of bioactive compounds.[10,11] Especially the genus *Sorangium* is a rich source of natural products with unique structures and mode of actions as approximately half of the known myxobacterial compounds have been isolated from *Sorangium* species including the epothilones, which are marketed as anticancer pharmaceuticals (ixabepilone - Ixempra).[12] Here, the production process involved large scale fermentation of *Sorangium cellulosum*, isolation of the natural product, and subsequent medicinal chemistry to yield the lactame ixabepilone from the lactone epothilone B (see Figure 1A). The natural producer, however, is notoriously difficult to manipulate genetically, is resistant to most antibiotics generally used for genetic engineering, and grows slowly. Thus, several attempts have been made to transfer the corresponding biosynthetic gene cluster into advantageous hosts to set the stage for engineering. This included expression in streptomycetes[13,14] and myxobacteria[15,16] but failed so far to generate an alternative industrial production process as yields did not

dx.doi.org/10.1021/sb300080t | *ACS Synth. Biol.* 2014, 3, 759–772

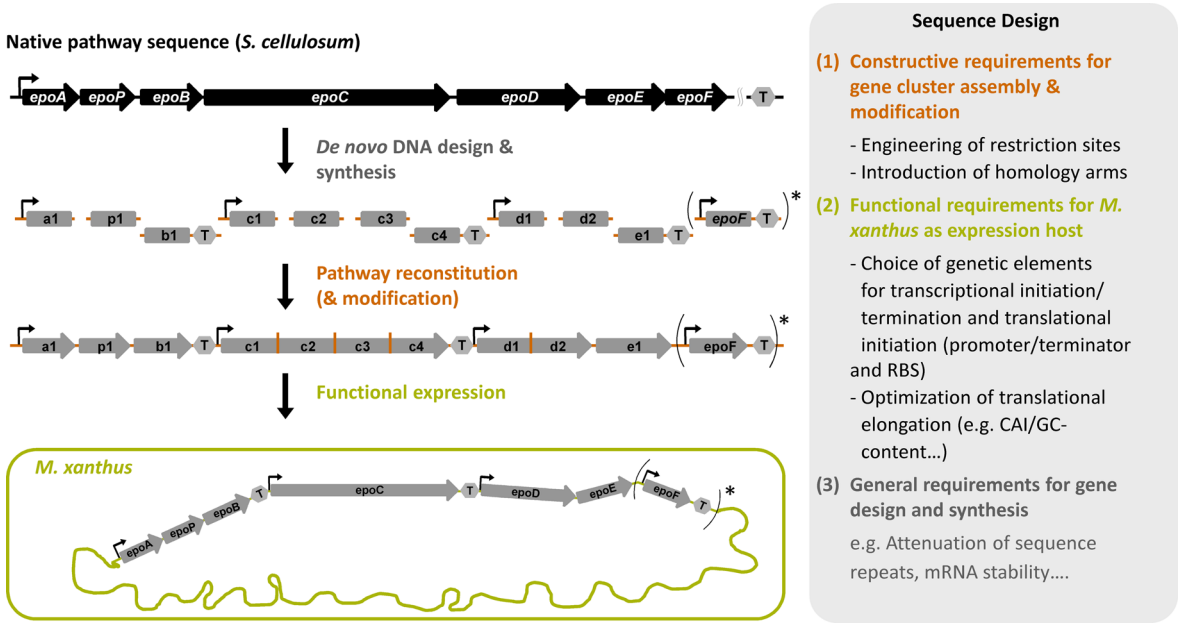


Figure 2. Overview of the sequence design process toward the generation and functional expression of an artificial epothilone biosynthetic gene cluster. Different types of sequence requirements had to be considered to redesign and construct a functional, artificial version of the epothilone pathway from *S. cellulosum* So ce90 for expression in *M. xanthus* DK1622. For gene synthesis, the pathway had to be subdivided into 10 building blocks (*11 after inclusion of *epoF*, which was optional).

compare to those in a traditionally optimized natural host. Best yields were achieved with *Myxococcus xanthus*, which is in agreement with other publications showing that this is a promising heterologous host that can efficiently produce non-native secondary metabolites from myxobacteria.[16−18] Modification of the bioprocess using *M. xanthus* provided up to 20 mg/L,[19] but this process includes the use of oil in the medium, which causes difficulties in the purification process. In addition to the still suboptimal yields, manipulation of the generated large native DNA sets for transfer into the heterologous host did not enable quantum leaps in the ability to manipulate the DNA for pathway engineering.

In theory, total synthesis of the complete gene cluster offers such opportunities as the pathway could be holistically optimized for the chosen production host including codon bias adaptation. Currently, there are only two examples for such work on complex multimodular assembly lines in the literature, both aiming at production in *Escherichia coli*: the 30 kb erythromycin PKS pathway as well as the epothilone PKS/NRPS pathway, which is about twice as large, have been redesigned, synthesized, and functionally expressed in *E. coli*.[20−22] After engineering of the host strain both experiments were successful, but the achieved production titers in *E. coli* were found very low.[23]

In light of the results achieved with native gene clusters and GC-rich hosts as mentioned above, we reasoned that synthetic design of an artificial epothilone pathway for expression in *M. xanthus* may provide a way forward to enable efficient genetic manipulation and at the same time allow production at reasonable yields. We therefore set out to start a pilot project in which to our knowledge the first high GC polyketide biosynthetic gene cluster was designed and subsequently assembled. Transfer of the resulting 58 kb pathway into *M. xanthus* enabled heterologous production of epothilone and now sets the stage for further optimization.

RESULTS AND DISCUSSION

Construction of a Functional Epothilone Polyketide Pathway by Synthetic Genes. An overview of our experimental strategy to recreate the epothilone biosynthesis pathway for the expression in *M. xanthus* by total gene synthesis is given in Figure 2. As two native epothilone biosynthetic gene clusters have been reported from *S. cellulosum* strains,[24,25] we initially performed a detailed sequence comparison between both pathways. In total, 693 differences in the DNA sequence were found resulting in changes in 285 amino acid positions of the encoded biosynthetic machineries (see Supplementary Figure S1). To exclude sequencing errors within the GenBank entry of the *S. cellulosum* So ce90 epothilone pathway, the respective positions were all resequenced and showed no deviation from the published gene cluster sequence. After these sequence verification efforts strategies for recreation of the *S. cellulosum* So ce90 epothilone pathway sequence and for gene cluster assembly based on synthetic building blocks were designed. Two different but partly interfering initial requirement specifications were defined and considered in the sequence design process: (1) constructive sequence requirements, resulting from specific sequence requirements reflecting the initial choice of (also) applying classical cloning approaches in order to assemble the sequences; (2) functional sequence requirements, resulting from specific demands relating to the selected host strain *M. xanthus*; and (3) in addition, general sequence requirements for gene design and gene synthesis (see Figure 2).

Functional Sequence Requirements for *M. xanthus* as Heterologous Expression Host. From the beginning of the project the minimum aim for the pathway construction was producing detectable amounts of epothilone in *M. xanthus* in order to enable continuous sequence improvement experiments from that point on. However, only little information was available on the questions of which sequence requirements implicitly are functional for an organism like *M. xanthus* and

dx.doi.org/10.1021/sb300080t | ACS Synth. Biol. 2014, 3, 759−772

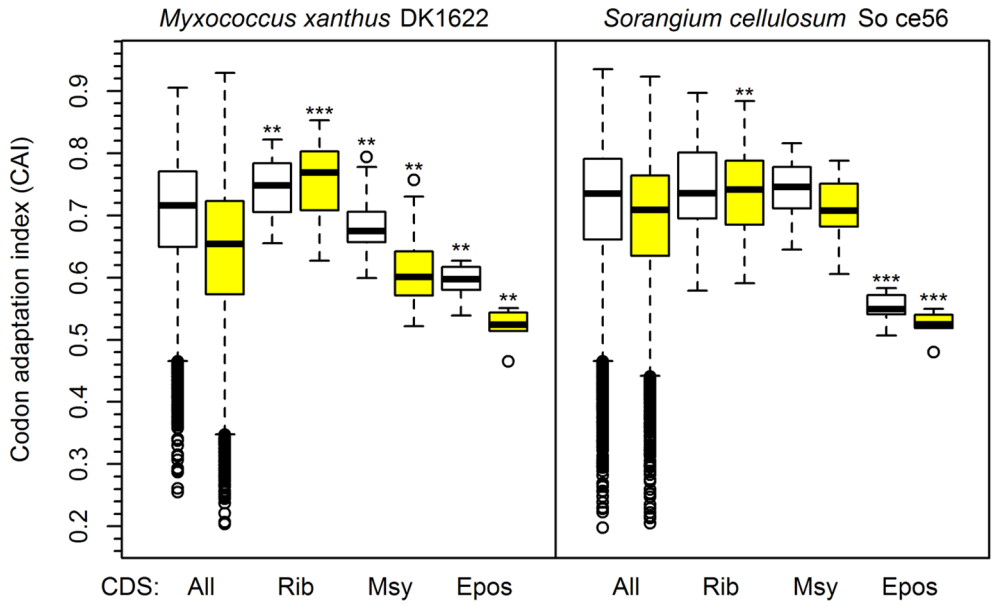


Figure 3. Boxplots showing the distribution of codon adaptation indices (CAI) of classes of coding sequence (CDS) based on codon usage tables constructed either from all CDS (in white) or from a subset encoding the presumably highly expressed ribosomal proteins (in yellow). The shown CDS classes are All, all CDS; Rib, ribosomal proteins; Msy, megasynthases; Epos, epothilone biosynthesis genes. Significant differences between CAI distributions of the classes and the respective distribution for all CDS were calculated using the Mann−Whitney test: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$. For further details see Methods.

what kind of modification would be needed for successful expression of the epothilone pathway genes derived from *S. cellulosum*.[26,27] Especially, for the multidomain biocatalysts like EpoC exhibiting large molecular weight (764 kDa) the factors involved in co-translational folding properties like slow codon clusters and RNA secondary structures were not clear. Neither was anything known about the impact of post-translational helper factors like chaperones on the functional expression of complex megasynthetases. Therefore a moderate conservative strategy was chosen, which first involved the screening for evolutionary conserved slow codon clusters.[28] However, an initial search for conserved slow codon clusters at homologous domain positions within the *epoC* gene revealed no suspect hits (data not shown). Consequently, the working hypothesis was that the epothilone pathway sequence does not contain functionally relevant rare codons with a potential impact on protein expression.

Basically, sequence modulating operations on the single base level for changing formal gene features with functional impact can be performed on the level of transcriptional initiation, translational initiation, and translational elongation, each having specific sequence parameter sets affecting the final outcome of expression levels and the specific activity of the gene products.[29] Furthermore, proper transcriptional and translational termination has to be ensured.[30−32] For efficient transcriptional and translational initiation of the epothilone biosynthetic genes in the heterologous host, an established standard promoter-5′-untranslated region (UTR) sequence (PTn5-5′-UTR), including a suitable Shine-Dalgarno (SD) sequence, was chosen to control the expression of the artificial transcription units. As the same setup already led to significant heterologous overproduction of myxochromides S at yields of >500 mg/L in *M. xanthus*,[16] neither further sequence modifications on promoter strength nor optimization of translational initiation parameters was envisaged in this initial experimental setup. To ensure proper termination in *M. xanthus*, the TD1 terminator[33] was integrated downstream of each artificial *epo* transcription unit. However, the impact of the applied default configurations (PTn5-5′-UTR, TD1) on the expression of the affected genes needs to be elucidated by further experimental sequence designs.

To minimize the number of parameters during sequence optimization, we decided to control only the most plausible parameters affecting translational elongation, in addition to parameters deriving from other distinct sequence requirements (see Figure 2). During the *in silico* sequence modulation process (see section "Design of an Artificial Epothilone Pathway Sequence"), the sensitivity of the system toward exchange of codons in a still functional artificial pathway is evaluated. The design of a codon usage table was one of the most important initial tasks for performing the multiparametric calculation. As basis for this, a comparative codon usage analysis between the expression host *M. xanthus* DK1622 and a close relative of the epothilone producer, *S. cellulosum* So ce56 (as no genome sequence of *S. cellulosum* So ce90 was available) was carried out (see Figure 3). Genes of *M. xanthus* DK1622 showed significant bias for their codon adaptation indices (CAI)[34] within different chromosomal gene functional classes, whereas *S. cellulosum* Soce 56 genes did not. Similar adaptations are generally found in fast-growing bacteria.[35] Genes from *M. xanthus* DK1622 encoding presumably highly expressed ribosomal proteins showed a higher CAI according to the codon usage table constructed from all CDS, whereas genes encoding megasynthetases showed a lower adaptation. The epothilone biosynthesis genes of *S. cellulosum* So ce90 showed a low adaptation for both organisms. Calculating the CAI for a codon usage table constructed from the ribosomal proteins alone (yellow boxplots in Figure 3) enhanced the respective CAI levels for ribosomal proteins and lowered that for megasynthetases, which was more pronounced in *M. xanthus* than in *S. cellulosum*. We concluded that the genes of *M. xanthus* are more biased and adapted to the respective codon

dx.doi.org/10.1021/sb300080t | *ACS Synth. Biol.* 2014, 3, 759−772

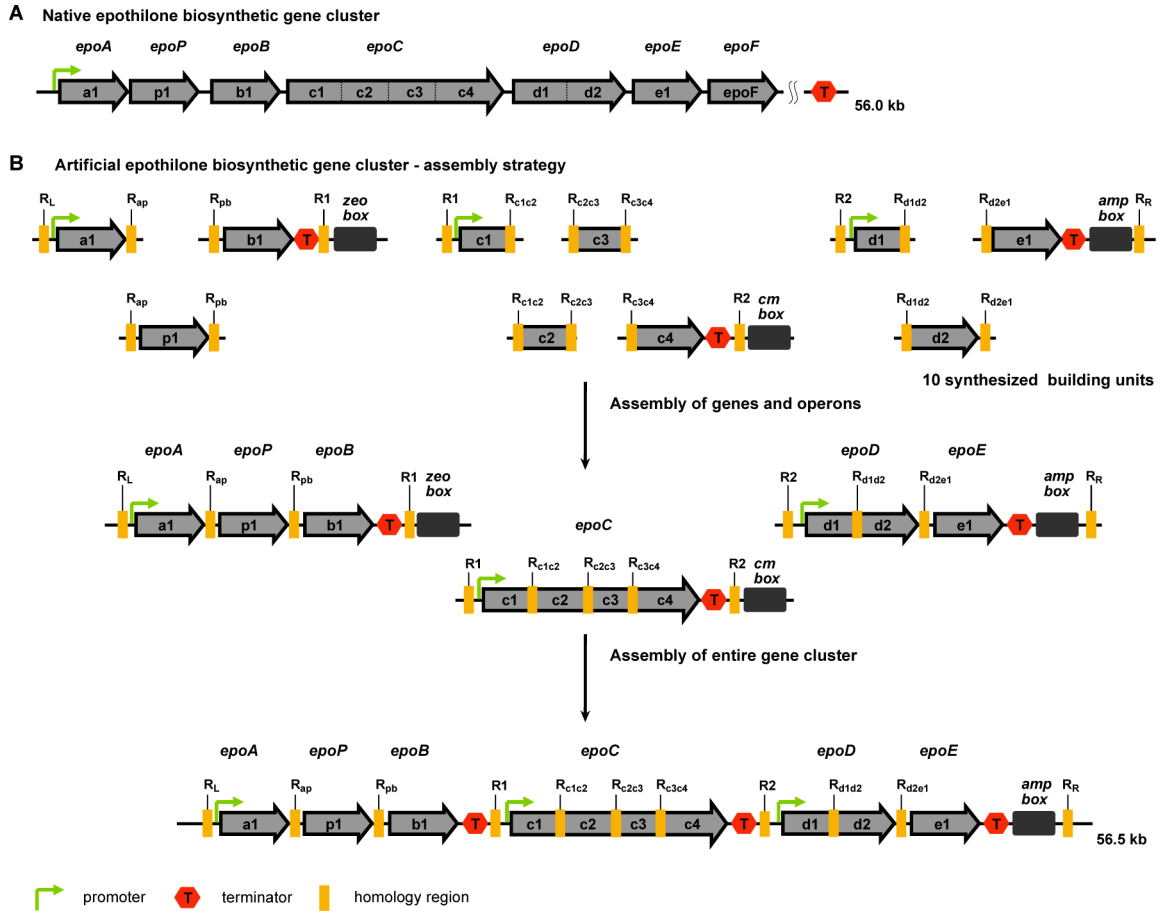


Figure 4. Assembly strategy to generate an artificial version of the epothilone biosynthetic gene cluster and comparison to the native pathway. (A) Gene arrangement of the native epothilone biosynthetic gene cluster from *S. cellulosum* So ce90. (B) Overview of the two-step assembly process to reconstruct the core gene cluster (without *epoF*) starting from 10 artificial building blocks generated by gene synthesis. Building blocks were defined according to the modular structure of the epothilone assembly line (see Figure 1) and were equipped with restriction sites (R_L, R_{ap}, R_{pb}, R_1, R_{c1c2}, R_{c2c3}, R_{c3c4}, R_2, R_{d1d2}, R_{d2e1}, and R_R) as well as homology arms to allow the application of different cloning technologies. The first assembly round results in the generation of the three artificial transcription units, which give rise to the entire gene cluster in the second assembly round facilitated by the presence of unique selection marker cassettes downstream of each transcription unit (*zeo box* = zeocin resistance gene cassette; *cm box* = chloramphenicol resistance gene cassette; *amp box* = ampicillin resistance gene cassette).

usage tables, presumably caused by the faster growth and concomitantly higher metabolic pressure on charged tRNA availability. On the basis of this analysis, we decided that the CAI of the epothilone biosynthesis genes should be at least optimized for reaching similar adaptation levels as observed for other chromosomal megasynthetases from *M. xanthus*. Adaptation to CAI levels as found for the ribosomal proteins for their respective codon usage table may lead to an unnecessary overoptimization restricting other constructive design parameters.

For gene optimization we constructed therefore a codon usage table from the 50 native megasynthetase genes of *M. xanthus* DK1622 (genes are listed in Supplementary Table S1). We additionally included three other megasynthetase genes from *Stigmatella aurantiaca* DW4/3-1 (*mchABC*), which were already functionally expressed in *M. xanthus* DK1622,[16] and which showed a similar CAI (0.618, 0.632, 0.643) as the native megasynthetases (0.681 ± 0.046). The resulting codon usage table ("myxo-PKS/NRPS", see Supplementary Figure S2A) shows five times more rare codons compared to, e.g., *E. coli*, and as expected, higher codon frequencies are predominantly allocated to the GC-rich codons. On the basis of the assumption that codons used with high frequency in *M. xanthus* are presumably correlated to high aminoacyl-tRNA pools, a codon exclusion threshold of 5% was applied during sequence optimization to avoid possible bottlenecks in aminoacyl-tRNA availabilities.

Constructive Sequence Requirements for Pathway Assembly and Modification. To construct a 58 kb artificial epothilone biosynthetic gene cluster sequence, the pathway had to be dissected into smaller segments of DNA, which could be supplied by gene synthesis companies despite their high GC content. At that time, commercial DNA fabrication was restricted to genetic cassettes not larger than 10 kb in size, while standard orders did not exceed the 1 kb benchmark. For the reconstitution of an artificial epothilone pathway based on a number of synthetic DNA fragments (<10 kb) we intended to design a strategy with two objectives: (1) allow various assembly technologies to combine the synthetic pathway segments and (2) permit facile interchangeability for subsequent combinatorial biosynthesis approaches as well as for the engineering of genetic elements (e.g., exchange of promoter structures). As illustrated in Figures 1B, the epothilone biosynthetic pathway consists of seven genes, encoding six subunits of the epothilone assembly line plus a modifying enzyme EpoF, which installs the epoxide function-

dx.doi.org/10.1021/sb300080t | *ACS Synth. Biol.* 2014, 3, 759−772

ality in epothilones A and B. The assembly line consists of 10 modules, nine PKS and one NRPS module, each of which containing a distinct set of catalytic domains required for the incorporation and modification of one specific precursor molecule. In light of our plans for future combinatorial biosynthesis approaches, we choose to dissect the pathway sequence according to its modular architecture. This led to the definition of 10 DNA segments (plus *epoF*), the largest one 8.6 kb in size (see Figure 4). Four out of these 10 fragments represent entire genes encoding monomodular subunits of the assembly line (EpoA, EpoP, EpoB, and EpoE). As the two other subunits (EpoC and EpoD) harbor multiple modules, the respective genes were subdivided into several fragments: c1/c2/c3/c4 and d1/d2. This was achieved by means of a detailed annotation of the domain architecture to locate intermodule linker regions as suitable fusion sites. Overall, the epothilone pathway sequence was redesigned in a way that the PKS/NRPS megasynthetase encoding genes form three individual transcription units (*epoAPB*, *epoC*, and *epoDE*) placed under control of the constitutive Tn5 promoter (PTn5) and the TD1 terminator (see Figure 4). Compared to the situation in the native epothilone pathway, this design offers various opportunities for future combinatorial biosynthesis approaches and also allows the co-expression of the three subunits on different physical entities ("multiplasmid approach"), e.g., if cloning of the entire pathway onto one construct fails. However, we initially aimed to establish a two-stage assembly process starting with the cloning of the three transcription units, each consisting of three to four synthetic fragments, followed by gene cluster assembly in the second stage. The latter step should be facilitated by the presence of unique selection markers inserted downstream of each transcription unit fragment, which allows for double selection when stitching the artificial cluster fragments.

To realize this strategy, we implemented a flexible and straightforward concept allowing for the fragment assembly in various ways: recombination based cloning (*in vitro* or *in vivo*) as well as restriction enzyme based cloning. For pathway reconstitution via conventional restriction/ligation based methods, each of the 10 module fragments was flanked with unique restriction sites. To facilitate future engineering approaches, restriction sites were also introduced between important genetic elements (e.g., promoter and terminator structures), and furthermore, each gene was flanked with a *Nde*I site (fused with the start codon) as well as a *Mly*I site (which would cut off the stop codon) to allow the subcloning of the individual genes into different expression vectors for future *in vitro* studies. As described below, bioinformatic tools were applied for restriction site engineering to eliminate perturbing sites and to introduce sites within the four intermodule linker regions of *epoC* and *epoD* by silent mutations. Based on the defined set of 24 restriction enzymes selected for pathway cloning and engineering, a total of 684 restriction sites had to be eliminated from the native epothilone gene cluster sequence. As this already represented a considerable constraint for the optimization of other sequence parameters (described above introducing the section "Functional Sequence Requirements for *M. xanthus* as Heterologous Expression Host" and below in "Design of an Artificial Epothilone Pathway Sequence"), we abandoned the aim to also flank each of the 50 catalytic domains from the epothilone megasynthetase with specific restriction sites. Next to the conventional cloning method, alternative assembly strategies based on *in vitro* or *in vivo* recombination (e.g., LIC, SLIC, Red/ET, and TAR cloning)[36,37] were considered by designing the 10 module fragments in a way that they contain overlaps to their neighboring DNA segments, each approximately 80 bp in size. In addition, both ends of each fragment were equipped with type IIS restriction sites to deliver suitable fragments for the *in vitro* stitching technologies. Overall, the assembly concept was thus set up in a highly flexible way allowing switches between different cloning technologies at almost all stages of the assembly process.

Design of an Artificial Epothilone Pathway Sequence. To create an artificial epothilone pathway sequence considering the functional and constructional requirements described above as well as standard requirements for gene synthesis (e.g., avoidance of sequence repeats, mRNA stabilities, just to name a few), all *epo* genes were subjected to the same algorithms of the evoMAG^{is} software package (ATG:biosynthetics GmbH). evoMAG^{is} applies concepts of evolutionary algorithms[38−40] to generate sequence variants whose formal sequence features satisfy given constraints without altering the amino acid sequence (see Figure 5), which is possible due to the degeneracy of the genetic code. Silent mutations were set for the random substitution of synonymous codons in order to

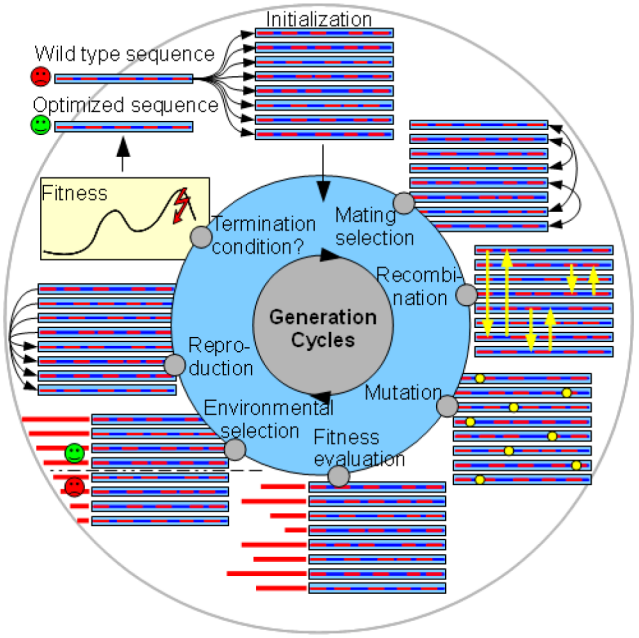


Figure 5. Possible scheme of a genetic algorithm for sequence optimization as used by the evoMAG core software module. A population of sequence variants is initialized from the wild type sequence by randomly assigning the synonymous codons to a number of random codon positions. During subsequent generation cycles of evolutionary improvement the sequence variants, all coding for the same protein, undergo recombination (with cut-points at the codon boundaries) followed by silent mutation and environmental selection in a "fitness landscape". At the end of each cycle the best variants are replicated into the next generation, by replacing the discarded ones with their offspring ("survival of the fittest"). The alternative selection scheme depicted in the illustration is called "truncation selection", similarly used by human breeders.[75] In this paper tournament selection was used (see text). Another optional feature is "mating selection" (not used here) where the selection of partners for recombination could depend for example on "phenotypic distance" (in sequence feature space) or "genotypic distance" (hamming distance).[76,77]

dx.doi.org/10.1021/sb300080t | *ACS Synth. Biol.* 2014, 3, 759−772

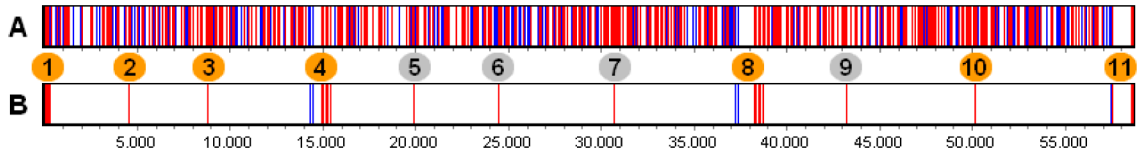


Figure 6. Representation of the distribution of the constructional restriction sites along the 58 kb native gene cluster sequence (A) and the result of calculating it out of sequence (B). Forward: blue bar. Reverse: red bar. (A) The 24 restriction sites, *Afl*II, *Age*I, *Bam*HI, *Bce*AI, *Bci*VI, *Bmg*BI, *Eco*RI, *Eco*RV, *Hind*III, *Kpn*I, *Mlu*I, *Mly*I, *Nde*I, *Nhe*I, *Not*I, *Pac*I, *Pst*I, *Sac*I, *Sal*I, *Sca*I, *Spe*I, *Stu*I, *Xba*I, and *Xho*I, which were selected for genetic engineering, were identified in altogether 684 positions of the native gene cluster sequence. (B) In the artificial sequence, the respective 684 restriction sites were removed, and constructive restriction sites for gene cluster assembly and modification (interchangeability of functional elements) were introduced in regions 1−11 (see Supplementary Table S2) including four sites in coding sequences (encircled gray). Sequence modulations by applying the codon sampling algorithm (based on the 'myxo-PKS/NRPS' codon usage table, see Supplementary Figure S2A) was used for converting possible cryptic R-sites to active ones by setting silent mutations in coding sequences.

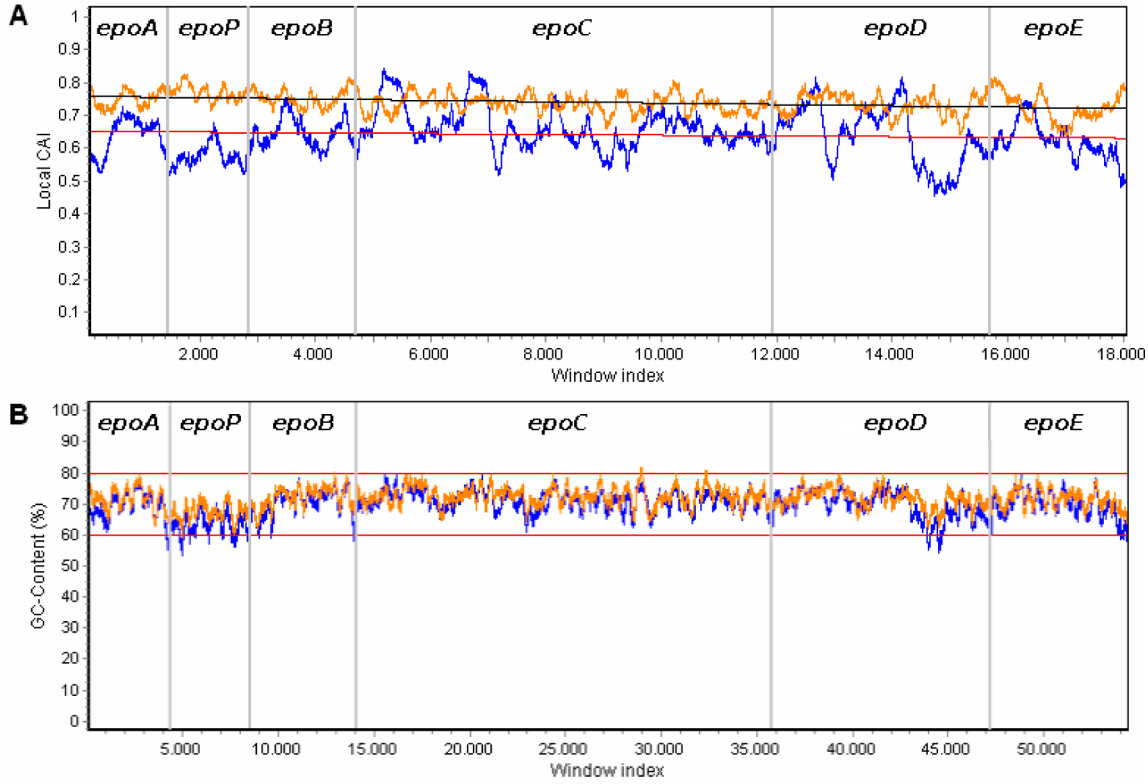


Figure 7. Comparison of the 58 kb native and artificial epothilone biosynthetic gene cluster sequences. The native sequence is shown in blue, and the artificial sequence is shown in orange. (A) Course of the local codon adaptation index (CAI). Window width is 200 codons. (B) Course of the local GC content. Window width is 200 bp. Untranslated regions between the protein coding regions were excluded from the graph.

generate the target sequence based on the natural template of the source organism *S. cellulosum* So ce90. Favorable mutation rates per codon per generation (∼cycle) were in the range of 0.1−1%. Two-point crossover was used as recombination operator with a crossover probability of 0.6 (60%). Based on a constant population of 60 sequence variants tournament selection[41] was performed as the selection scheme. A tournament size of four was applied to run competitions among four sequences chosen at random out of the population, whereas the winner (highest scoring sequence) of the competition replicates by a factor of 4 into the next generation, thereby replacing the three alternative solutions. The decision for winner sequences is based on better scoring values relative to predefined multivariate sequence parameter values. During the sequence modulation process, the *epo* genes were subjected to multiple objectives as follows: (1) incorporation of four constructional R-sites within coding sequences (linker regions 5, 6, 7, and 9, see Figure 6), (2) removal of 684 constructional R-sites from coding sequences (see Figure 6), except the four R-sites mentioned before, (3) exchange of synonymous codons by "codon sampling" to balance the synonymous codon fraction distribution according to the reference codon usage table thereby completely disregarding the synonymous codon frequency distribution of the native template, and (4) removal of extended secondary structures on the DNA level by use of silent mutations. In general, the smaller the set of used synonymous codons, the harder these aims are to achieve in parallel.

As discussed earlier in the section "Functional Sequence Requirements for *M. xanthus* as Heterologous Expression Host", an artificial codon usage table with a 5% exclusion threshold (see Supplementary Figure S2A) was applied. As this table only contains 46 sense codons (instead of 61) normalization of the synonymous codon fractions had to be

dx.doi.org/10.1021/sb300080t | *ACS Synth. Biol.* 2014, 3, 759−772

performed. Renormalization of the artificial table is necessary to ensure that the sum of each (sub)set of synonymous codon fractions (for each amino acid) equals 1. As shown in Supplementary Figure S2 almost 100% adaptation to the table (adapted "myxo-PKS/NRPS") could be achieved. Due to the high GC content of *M. xanthus* higher frequency codons are, in general, GC-rich. Consequently, the applied 5% cutoff for synonymous codon fraction distribution to exclude low frequency codons during the sequence modulation process led to a slight increase in the GC content of the optimized sequence compared to GC content of the native pathway sequence as shown in Figure 7. When comparing the course of the local CAI before and after sequence optimization, an overall slight increase by 0.122 (12.2%) could be observed (see Figure 7). It is also evident that the sequence optimization leads to a clearly smoothed shape of the local CAI showing no distinct peaks as found for the native sequence. This finding indicates that at least some of the sequence extremes with respect to the local CAI values might be not required for proper expression. The impact of the applied codon sampling procedure on the GC content as well as the CAI also became obvious when comparing these parameters for the individual CDS, showing a clear increase of the CAI values and a slight increase of the GC content values for the artificial (optimized) Epo CDS versions (see Supplementary Table S3). Although the optimization of translational initiation was not included in the first experimental setup (as discussed in section "Functional Sequence Requirements for *M. xanthus* as Heterologous Expression Host"), the $-\Delta G$ values of the +40 bps of the CDS effective on the RNA level in combination with the −20 bps of the nontranslated leader sequence around the Shine-Dalgarno sequence[42] have been compared as well (see Supplementary Table S3). In case of *epoP* and *epoD* the $-\Delta G$ values were changed significantly toward presumably reduced initiation frequencies, whereas for the other *epo* PKS genes no conspicuous differences could be observed. These data already show some potential for the optimization of the initiation frequency of all of the *epo* genes, which can be addressed in continuative sequence optimization approaches.

Another objective during sequence modulation was the removal of secondary structures on the DNA level. In the case of amino acid repeats on the protein level there is often no possibility for removal by silent mutations, because in many cases only one nucleotide exchange per codon is allowed. Furthermore, the detection of a sequence repeat depends on the choice of the homology threshold. Nevertheless, the reduction of extended secondary structures on the DNA level eased the gene synthesis process of contractors and reduced the price of synthesis. The modulation of codons according to the procedure of choice reduced the pattern of repeats as shown in Supplementary Figure S3.

In general, there is an enormous potential for optimizing gene expression by sequence modulations. However, the optimal calibration of all relevant and sometimes conflicting parameters to achieve best productivity requires considerable experimental support and several rounds of sequence optimization/experimental validation. Our initial sequence design was thus focused on the constructive part, whereas the functional optimization was mainly limited to efforts aiming at enhancing translational elongation. During the course of our work, additional findings of importance for the sequence design were published but could not be incorporated into our strategy at that time, which is also discussed below in the Outlook section.

Construction of an Artificial Epothilone Pathway. The *de novo* fabrication of the designed DNA was outsourced to different gene synthesis companies. However, the accurate, efficient, and rapid synthesis of the high GC and quite long DNA fragments turned out to be very challenging and became one time-critical step within this project as delivery times often significantly exceeded forecasts. After receiving the synthetic DNA segments, we aimed to build up complete PKS genes (*epoC* and *epoD*) and three individual transcription units in the first assembly stage and, in the next step, to reconstitute the entire pathway onto one construct. For both assembly stages shown in Figure 4, conventional cloning methods were applied (for details see Supporting Information). Two assembly vector backbones have been constructed, pUC18-kan for transcription unit assembly and a modified version, pUC18-spec, to enable rapid cloning by double selection without the need of fragment purification for gene cluster assembly (as each transcription unit was linked with a unique selection marker). In stepwise assembly processes, constructs pEpo28 (*PTn5-epoAPB-TD1-zeoR*), pEpo32 (*PTn5-epoC-TD1-cmR*), and pEpo36 (*PTn5-epoDE-TD1-ampR*) were generated. For gene cluster assembly, transcription unit *PTn5-epoAPB-TD1-zeoR* was subcloned onto the pUC18-spec backbone by spectinomycin/zeocin double selection to yield pEpo37, and the second transcription unit, *PTn5-epoC-TD1-cmR*, was added using spectinomycin/chloramphenicol as selection markers to yield pEpo38. In the next step, the third transcription unit, *PTn5-epoDE-TD1-ampR*, was added to the pEpo38 construct making use of the spectinomycin/ampicillin selection markers to generate a ∼60 kb construct, pEpo40, harboring the entire epothilone assembly line pathway (PTn5-*epoAPB-TD1-PTn5-epoC-TD1-PTn5-epoDE-TD1-ampR*). The latter construct turned out to be unstable already in *E. coli*, which may be due to various reasons (e.g., repetitive genetic elements like the 140 bp PTn5 sequence or the high copy origin of replication). After subcloning of the three transcription units into a low copy vector backbone to generate pEpo47 (p15A-derivative), similar stability problems could be observed. To circumvent these problems, we decided to clone the pathway on two separate expression constructs, one containing the PTn5-*epoAPB-TD1-PTn5-epoC-TD1-cmR* and the second containing the *PTn5-epoDE-TD1-ampR* pathway fragment. A prerequisite for the subsequent "multiplasmid approach" is the compatibility of both expression constructs regarding selection markers and genetic elements for propagation or integration into the host chromosome. As no origin of replication for the stable propagation of large plasmids was available for the host strain *M. xanthus*, both constructs had to be integrated into the chromosome. For this, each of the constructs was equipped with different integrase encoding genes (*mx8* or *mx9*)[33,43] including a phage attachment site for *Myxococcus* as well as suitable resistance genes (zeo^R or kan^R) for selection. The generated expression constructs (pEpo42 and pEpo44) harbor the entire assembly line gene set, which should already enable the production of the epothilone macrocycle. However, to complete the pathway, an artificial version of the modifying gene *epoF* was generated and introduced as independent transcription unit onto pEpo44 to finally give pEpo46.

Heterologous Expression in *M. xanthus*. The generated expression constructs, pEpo42 and pEpo44/pEpo46 (with *epoF*), were compatible and could thus subsequently be

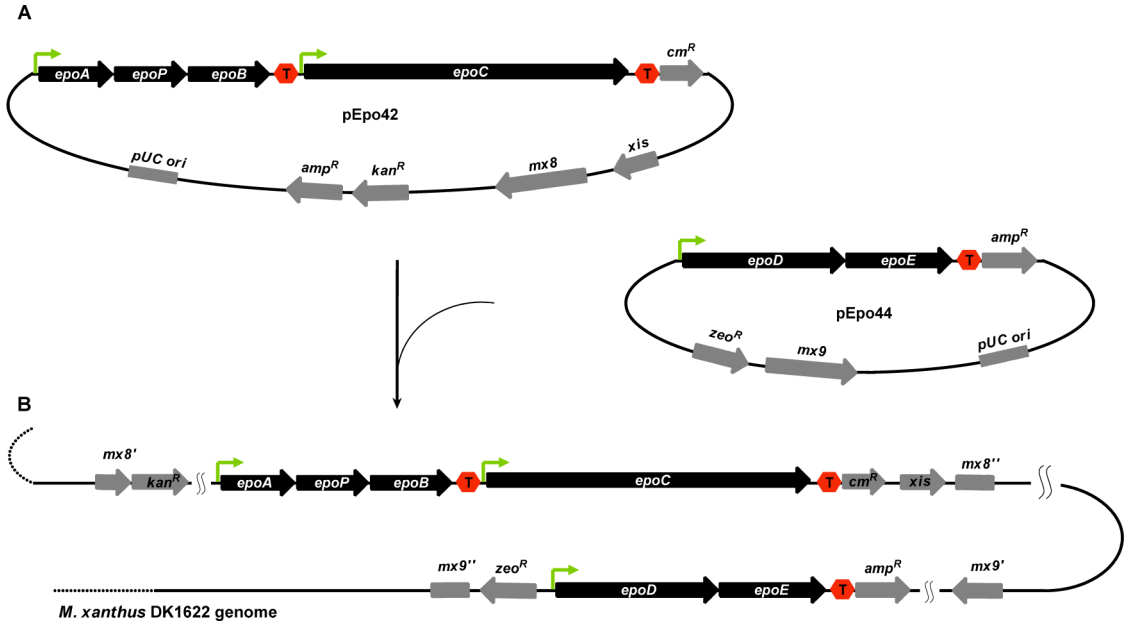


Figure 8. "Multiplasmid Approach" for the heterologous expression of the artificial epothilone pathway in *M. xanthus*. (A) Maps of the two compatible expression plasmids harboring parts of the biosynthetic gene cluster. pEpo42 contains the first two transcription units, whereas the third transcription unit is part of pEpo44. Both constructs encode different integrase genes (*mx8* and *mx9*) for site-specific integration into the *M. xanthus* chromosome and were equipped with different selection markers (zeo^R = zeocin resistance gene cassette; kan^R = kanamycin resistance gene cassette). (B) Genotype of the resulting *M. xanthus* mutant after integration of the artificial epothilone pathway into two different positions of the chromosome (Mx8 and Mx9 attachment site).

transformed into the host strain for heterologous expression. The correct integration into the *mx8* and *mx9* phage attachment sites, respectively, as shown in Figure 8, was analyzed by PCR (for details see Supporting Information). In addition, specific gene cluster regions located widespread along the expression constructs were amplified using different primer sets to confirm the integrity of the gene cluster. After genotypic verification, the *M. xanthus* mutants harboring artificial versions of the epothilone biosynthetic gene cluster were cultivated in parallel to the *M. xanthus* wild type strain. For epothilone production profiling, the culture extracts were analyzed by liquid chromatography (LC) coupled to high resolution mass spectrometry including the fragmentation of target ions (hrMS/MS). On the basis of this analysis and the comparison with fragmentation spectra of authentic reference substances, epothilone derivatives could be detected in extracts of the expression strains in contrast to *M. xanthus* wild type (see Figure 9). The genotype of the engineered host strain seemed to be stable, as robust epothilone production was observed after serial strain passaging in several cultivation experiments. The successful heterologous production unambiguously demonstrated the functionality of the redesigned and totally synthesized artificial epothilone biosynthetic gene cluster. As expected, epothilones C and D could be detected in mutants expressing the six subunits of the epothilone megasynthetase (EpoAPBCDE). The product spectra could be expanded by co-expression of the modifying enzyme EpoF, which was achieved by inserting the artificial transcription unit *PTn5-epoF-TD1* onto the expression construct pEpo44 resulting in pEpo46. This led to the production of the epoxidized derivatives epothilones A and B in addition to the nonoxidized forms, epothilones C and D, which could still be detected in trace amounts (see Supplementary Figure S8). As anticipated, the four epothilone family members (epothilones A–D), which

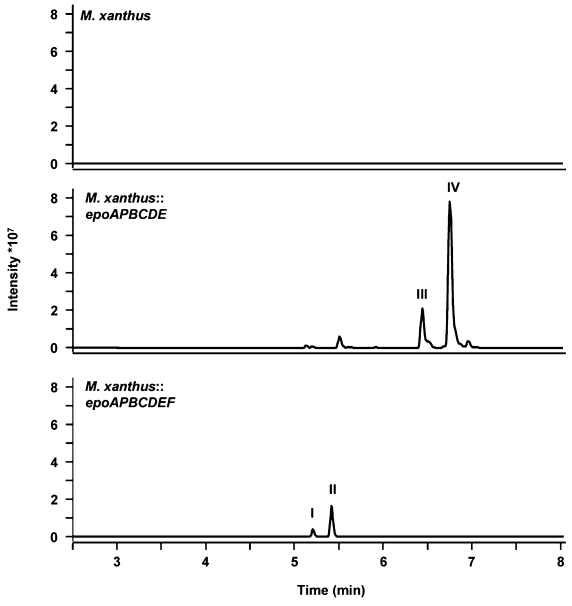


Figure 9. LC–hrMS analyses of heterologous epothilone production. Shown are base peak chromatograms at m/z = 494.2571 ($[M + H]^+$ of epothilone A), 508.2727 ($[M + H]^+$ of epothilone B), 478.2622 ($[M + H]^+$ of epothilone B), and 492.2778 ($[M + H]^+$ of epothilon D). Methanolic extracts from *M. xanthus* wild type and *M. xanthus* mutants harboring artificial epothilone pathway genes (*epoAPBCDE* or *epoAPBCDEF*) have been analyzed. As expected, epothilone C (**III**) and epothilone D (**IV**) could be detected when expressing *epoAPBCDE*. The additional expression of *epoF* resulted in the production of epothilone A (**I**) and B (**II**). In the latter case, trace amounts of epothilone C and D could also be detected by targeted fragmentations and analyses of MS^2 patterns (see Supplementary Figure S8).

dx.doi.org/10.1021/sb300080t | *ACS Synth. Biol.* 2014, 3, 759–772

were produced based on the artificial expression system, also represent major derivatives in the natural producer strain *S. cellulosum* So ce90.[44] The hitherto achieved production yields with the artificial epothilone pathway were moderate to low (approximately 100 μg/L), which might explain why epothilone variants described as minor compounds in the literature (>30 minor derivatives from So ce90)[44] could not be detected so far. In addition to epothilone, other secondary metabolites such as myxovirescin, myxalamid, DKxanthene, and myxochromide, which also (partially) originate from malonyl-CoA or methylmalonyl-CoA, can be detected in the host strain. Recent studies have shown that secondary metabolite production in *M. xanthus* can be increased by engineering pathway-specific regulators[45] or by engineering enzymatic activities responsible for posttranslational activation of natural product assembly lines.[46] Many of these experiments led to a shift in secondary metabolite profiles, and in most cases the total amount of the compounds mentioned above increased. We therefore assume that competition for precursor monomers is currently not the main limitation of epothilone production in our study. We cannot exclude, however, that any engineering or feeding of monomers may eventually increase production yields.

When comparing the product titers of synthetic gene clusters in *E. coli* (<0.001 mg/L)[22] and *M. xanthus* in our study, it seems justified to regard the latter species as showing the most promise for further development. Although productivity has to be increased by approximately 3 orders of magnitude to achieve levels of typical production processes for myxobacterial compounds, we believe that the work described here sets the stage for detailed studies to decipher bottlenecks of production and also generate derivatives of epothilones. This is underpinned by the fact that heterologous production of epothilone using native genes gave the best titers in *M. xanthus* as well[15,16,19] but did initially not exceed the productivity of the synthetic cluster.[19] A subsequent detailed bioprocess optimization for *M. xanthus* enabled production of 20 mg/L. However, this work included the addition of methyl oleate to the growth medium, obviously causing downstream problems with compound isolation. Nevertheless, it shows that significant production increases are feasible with the system.

Outlook. To achieve high product yields of complex secondary metabolites using microbial cell factories requires a concerted action of multiple genetic functions. In general, basic cellular features like available pools of aminoacyl-tRNAs and their regulation are highly important factors for production,[47] as recently demonstrated by an analysis of ribosomal occupancy of aminoacyl-tRNAs in *E. coli*.[48] Of similar importance is the analysis of physiological/biochemical bottlenecks, which in case of *M. xanthus* might be polyketide precursor (e.g., methylmalonyl-CoA) supply, or the cellular repertoire of chaperones and degrading proteases. A recent study on heterologous polyketide production in engineered *E. coli* strains revealed that the protease levels can vary from strain to strain and thus have significant impact on the overall production yields.[49] The importance of chaperones for the functional expression of high molecular weight PKS/NRPS proteins has already been demonstrated in expression studies with the epothilone megasynthetase in *E. coli*.[22] Epothilone production could only be achieved after co-expression of translational helper factors (GroEL/ES/Trigger-Factor) at low temperature conditions and under control of a tight, inducible promoter system. In addition, the largest subunit of the assembly line (~760 kDa) had to be split and expressed in two separate parts.

Complex sequence features are promoting the general fitness of the translational process as well as the cotranslational folding of proteins with direct impact on the specific activity and solubility of proteins.[50−54] The structural integrity of polysomes facilitates the transport of reactants of the translational process tremendously. Differential ribosomal density along the message controls and stabilizes the half-life of the mRNA providing structural prerequisites for a maximum efficiency of the translational machinery.[55,56] Different speeds of ribosomes along the message are responsible for correct cotranslational folding and therefore for the specific activity of gene products. This speed is discussed to be influenced by a variety of parameters: (1) slow codon clusters,[51] (2) internal SD sequences,[48] (3) RNA-secondary structures,[56] (4) amino acid composition of the nascent protein chain passing the ribosomal exit channel,[57−59] and (5) synonymous codon ordering, e.g., due to correlation of identical codons in direct proximity.[60] Another important factor to consider is the suppression of induced frame shifts[61,62] and premature translational termination, both of which are due to unfavorable sequence patterns rendering the translational process ineffective.[63] Altogether, the translational efficiency depends on various factors and represents a crucial key for achieving high production yields.

For future systematic optimization in molecular engineering, the multivariate property space should best be partitioned into separate hierarchically organized modules, which are nearly independent from each other and combine into emergent features (e.g., polyketide synthesis rate).[64] The ultimate goal will be a whole-cell multimodular simulation framework allowing *in silico* optimization for the production system built of hierarchical organized modules and containing all relevant cellular components. Such simulation was impressively realized recently for *Mycoplasma genitalium*,[65] whereas for hosts like *E. coli* or *M. xanthus* the required data and computing scheme have to be established.

In the future we aim to employ high-throughput strategies such as RNA-seq, proteomics, and metabolomics to increase our currently very limited understanding of the various contributing processes eventually leading to target compound formation. Integrated analysis of the gained data and appropriate modular modeling will provide us with engineering guidelines for further improvements of the contributing transcriptional, translational, and metabolic modules. The modular design as demonstrated here will allow for testing the optimization strategy described above in the future. In conjunction with optimized fermentation for promising hosts such as *M. xanthus*, this will lead to general procedural methods to employ synthetic biology including total synthesis of complex biosynthetic pathways for natural product optimization.

METHODS

Sequence Analysis and Design. A detailed sequence comparison of two native epothilone biosynthetic pathways described from *S. cellulosum* SMP44 (Acc. No. AF217189)[25] and *S. cellulosum* So ce90 (Acc. No. AF210843)[24] was performed. Differences in DNA sequence causing amino acid exchanges were analyzed by resequencing the respective regions in the *S. cellulosum* So ce90 pathway (for details see Supporting Information). After sequence verification, the So ce90 epothilone biosynthetic gene cluster was reannotated in detail (i.a. to define intermodule linker regions). All alignments and routine *in silico* DNA work was performed using Geneious v5.4

software package[66] (Biomatters Ltd.), and for reannotation the PKS/NRPS analyses web tool[67] and the basic local alignment search tool (BLAST)[68] were used.

Source of Coding Sequences. For *S. cellulosum* So ce56, 9384 coding sequences with 11321283 bp were retrieved from NCBI Genome RefSeq NC_010162.[69] For *M. xanthus* DK1622, 7331 coding sequences with 8362023 bp were retrieved from NCBI Genome RefSeq NC_008095.[26] A CDS subset of 33 ribosomal proteins (16719 bp) for *M. xanthus* and 42 for *S. cellulosum* (18186 bp), respectively, and of 50 chromosomal megasynthetases for *M. xanthus* (455181 bp) and for 22 of *S. cellulosum* (260940 bp), respectively, were extracted by their annotations and additionally curated manually (see Supplementary Tables S8−S11). The CDS of the six epothilone biosynthesis genes were extracted from the NCBI GenBank entry AF210843 (54489 bp)[24] (see Supplementary Table S12). Three additional megasynthetases of the myxochromide S biosynthesis gene cluster (*mchABC*) of *Stigmatella aurantiaca* DW4/3−1 were extracted from NCBI Genbank entry AJ698723[70] (see Supplementary Table S13). The subsets are listed in supplementary Tables S8−S13. All sequences were retrieved from the databases on March, 11, 2009.

Analysis of Codon Adaptation. Codon usage tables for the CDS subsets were calculated by the EMBOSS program cusp,[71] after removing FASTA format headers and line breaks within sequences and after addition of missing nucleotides as "N" from the FASTA-formatted input sequences by a Perl program. The EMBOSS program cai was used for calculating codon adaptation indices for all genes,[34] and the EMBOSS program codcmp was used for comparison of codon usage tables. Another Perl program was used for translating codon usage tables from EMBOSS to GCG format, which were imported by the EVOMAG program of ATG for gene design. The boxplot function of R[72] was used for calculating statistic parameters of CAI distributions (Figure 3), where the boxes around the median line extend to the first and third quartile of the data, and the whiskers extend to the last value within a range covering 1.5 times the distance from the median to the quartile edges. Outliers from these ranges are indicated by additional circles. Significant differences in the location parameters of the CAI distributions were evaluated with the Mann−Whitney test (function wilcox.test for two samples in R).

Analysis of Slow Codon Clusters. Slow codon clusters can assist correct cotranslational folding of multidomain proteins by slowing elongation at domain borders and allowing the correct folding of critical intermediates.[51] Investigation of the evolutionary conservation of clusters of rare codons at homologous positions in organisms having different codon usage characteristics can reveal functional clusters.[28] However, for the epothilone biosynthesis genes, no closely related megasynthetases were found by us, so that only clusters within the repeated domain structures of the same protein could be compared with each other. Moreover, the codon usage of the *epo* genes is so significantly different from other chromosomal megasynthetases of the native host *S. cellulosum* (Figure 3), that a comparison with the codon usage table of these megasynthetases or the general codon usage table of this organism is not useful. Any conserved and functionally important rare codon cluster may be reprogrammed for being correctly interpreted by the host translation machinery and may be therefore better for the expression of functional megasynthetases. Therefore, based on the codon usage table of the six epothilone biosynthesis genes themselves, rare codon clusters were listed by a Perl program and aligned with known domain borders for the domain-richest gene *epoC*, which has four modules containing 20 domains (see Figure 1). The keto synthase, acyl transferase, keto reductase and acyl carrier protein domains each occur four times, respectively, at sequentially homologous positions. However, not one of the rare codon clusters observed was conserved at all four homologous positions for any of the repeated domains or their interdomain linkers (data not shown).

Strains and Culture Conditions. *Myxococcus xanthus* DK1622[26] was used as heterologous host for epothilone production. Wild type cultures and their mutants (harboring the artificial epothilone pathway) were cultivated at 30 °C in CTT medium[73] for routine growth and production. *E. coli* strains used in this study were cultivated in LB medium containing 5 g/L NaCl,[74] at 30 or 37 °C. For general molecular biology experiments *E. coli* HS996 (Invitrogen) was employed. Red/ET recombineering experiments were performed in *E. coli* HS996/pSC101BADgbaA-tet (Gene Bridges GmbH). Antibiotics were added to the cultures if necessary in the following final concentrations: ampicillin 50 μg/mL, kanamycin 50 μg/mL, tetracycline 6,5 μg/mL, spectinomycin 100 μg/mL, chloramphenicol 34 μg/mL and zeocin 25 μg/mL for *M. xanthus*, 50 μg/mL for *E. coli*, respectively.

DNA Manipulation, Analysis, and PCR. All enzymes were obtained from Fermentas, except for Triple Master enzyme mix that was purchased from Eppendorf. Oligonucleotides were obtained from Sigma Genosys. Polymerase chain reaction (PCR) products were purified by agarose gel electrophoresis and subsequent gel extraction with NucleoSpin Extract II Kit (Macherey-Nagel). An alkaline lyses method[74] was used to isolate and purify plasmid DNA from *E. coli*. For stepwise cloning of the synthetic modules into the whole artificial biosynthetic epothilone gene cluster donor and acceptor vectors were hydrolyzed at the respective restriction sites. For acceptor vectors this was followed by a dephosphorylation step using shrimp alkaline phosphatase. Cleaved DNA was separated by agarose gel electrophoresis. For purification PeqGOLD gel extraction kit (PeqLab) was used. Ligation was carried out by T4 DNA ligase and the reaction was extracted with chloroform or heat inactivated prior to electroporation into *E. coli*. All methods were performed as recommended by the manufacturers and according to standard protocols.[74] For details regarding the cloning vectors, constructed plasmids and expression constructs see Supporting Information.

Heterologous Expression of the Artificial Epothilone Biosynthetic Pathway. The expression constructs were transformed into *M. xanthus* DK1622 by electroporation according to previously established methods.[16] Genomic DNA of obtained mutants resistant to kanamycin and zeocin was isolated (Gentra Puregene Yeast/Bact. Kit from Qiagen) and used as template for verification in genotypic analyses by PCR. Applied reaction conditions and used primers are described in the Supporting Information. Production cultures of verified *M. xanthus* DK1622 mutants were inoculated 1:50 with a one day old preculture and grown in 50 mL CTT medium supplemented with the appropriate antibiotic in 300 mL shake flasks for 2−3 d at 30 °C. At the end of cultivation 2% AMBERLITE XAD16 adsorber resin (Rohm & Haas) was added and incubated on a rotary shaker at 160 rpm for 30 min. The cell mass and the adsorber resin were harvested by

centrifugation (20 min, 4000 rpm). For extraction cells and XAD16 were stirred for 2 h in acetone/methanol (1:1), solvents were removed *in vacuo*, and the extracts were redissolved in 250−500 μL methanol for analysis of epothilone production. For comparison the wild type was grown and processed applying the same conditions.

Epothilone Production Analysis. High-performance liquid chromatography coupled to high resolution mass spectrometry (HPLC−hrMS) was used to analyze mutant and wild-type extracts. All measurements were performed on an Ultimate 3000 RSLC system (Dionex). Chromatographic separation was carried out on an ACQUITY UPLC BEH C18 column (50 by 2.1 mm, 1.7 μm particle size, Waters). A linear mobile phase gradient was used with solvent A (H_2O + 0.1% formic acid) and solvent B (acetonitrile + 0.1% formic acid) at a flow rate of 600 μL/min and 45 °C. The gradient was initiated by a 0.33 min isocratic step at 5% of solvent B, followed by an increase to 95% solvent B in 9 min to end up with a 1 min flush step at 95% solvent B before re-equilibration with initial conditions. HPLC−MS coupling was supported by a TriVersa Nanomate nano-ESI system (Advion) attached to a Thermo Fisher Orbitrap. Mass spectra were acquired in centroid mode ranging from 200 to 2000 m/z at a resolution of R = 30000. For the identification of epothilones, retention times as well as MS^2 fragmentation pattern were compared to authentic reference substances. The collision induced dissociation spectra were acquired using the linear ion trap. Parent ions were isolated within a 2 m/z window and fragmented at normalized collision energy of 35. Quantification of a 1:100 dilution of the extracts was performed by peak integration of the respective masses using the Thermo Xcalibur software package (Thermo Fischer) based on a calibration curve calculated from serial dilution measurements with authentic reference compounds (in a range of 1−0.03 μg/mL).

■ ASSOCIATED CONTENT

Supporting Information

Tables of CDS data for ribosomal proteins of *M. xanthus* DK1622, megasynthetases of *M. xanthus* DK1622, ribosomal proteins of *S. cellulosum* So ce56, megasynthetases of *S. cellulosum* So ce56, epothilone biosynthesis of *S. cellulosum* So ce90, and myxochromide S biosynthesis of *Stigmatella aurantiaca*; FASTA-file of DNA sequences of designed artificial building blocks. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: s.wenzel@mx.uni-saarland.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Thomas Hoffmann (Helmholtz Institute for Pharmaceutical Research Saarland & Saarland University) for skillful help with various analytical technologies. This work was generously supported by a grant of the Bundesministerium für Bildung und Forschung (FKZ 0315488E) to Saarland University and ATG:biosynthetics GmbH (FKZ0351488A).

■ REFERENCES

(1) Newman, D. J., and Cragg, G. M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod. 75*, 311−335.
(2) Bon, R. S., and Waldmann, H. (2010) Bioactivity-guided navigation of chemical space. *Acc. Chem. Res. 43*, 1103−1114.
(3) Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angew. Chem., Int. Ed. 48*, 4688−4716.
(4) Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) Nonribosomal peptides: From genes to products. *Nat. Prod. Rep. 20*, 275−287.
(5) Weissman, K. J., and Müller, R. (2008) Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem 9*, 826−848.
(6) Bode, H. B., and Müller, R. (2005) The impact of bacterial genomics on natural product research. *Angew. Chem., Int. Ed. 44*, 6828−6846.
(7) Fischbach, M., and Voigt, C. A. (2010) Prokaryotic gene clusters: a rich toolbox for synthetic biology. *Biotechnol. J. 5*, 1277−1296.
(8) Weissman, K. J., and Leadlay, P. F. (2005) Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol. 3*, 925−936.
(9) Baltz, R. H. (2006) Molecular engineering approaches to peptide, polyketide and other antibiotics. *Nat. Biotechnol. 24*, 1533−1540.
(10) Wenzel, S. C., and Müller, R. (2009) Myxobacteria–'microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst. 5*, 567−574.
(11) Weissman, K. J., and Müller, R. (2010) Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep. 27*, 1276−1295.
(12) Altmann, K. H., Gaugaz, F. Z., and Schiess, R. (2011) Diversity through semisynthesis: the chemistry and biological activity of semisynthetic epothilone derivatives. *Mol. Diversity 15*, 383−399.
(13) Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science 287*, 640−642.
(14) Park, S. R., Park, J. W., Jung, W. S., Han, A. R., Ban, Y. H., Kim, E. J., Sohng, J. K., Sim, S. J., and Yoon, Y. J. (2008) Heterologous production of epothilones B and D in *Streptomyces venezuelae*. *Appl. Microbiol. Biotechnol. 81*, 109−117.
(15) Julien, B., and Shah, S. (2002) Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*. *Antimicrob. Agents Chemother. 46*, 2772−2778.
(16) Fu, J., Wenzel, S. C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A. F., Müller, R., and Zhang, Y. (2008) Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res. 36*, 1−14.
(17) Perlova, O., Fu, J., Kuhlmann, S., Krug, D., Stewart, F., Zhang, Y., and Müller, R. (2006) Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Appl. Environ. Microbiol. 72*, 7485−7494.
(18) Chai, Y., Shan, S., Weissman, K. J., Hu, S., Zhang, Y., and Müller, R. (2012) Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. *Chem. Biol. 19*, 361−371.
(19) Lau, J., Frykman, S., Regentin, R., Ou, S., Tsuruta, H., and Licari, P. (2002) Optimizing the heterologous production of epothilone D in Myxococcus xanthus. *Biotechnol. Bioeng. 78*, 280−288.
(20) Kodumal, S. J., Patel, K. G., Reid, R., Menzella, H. G., Welch, M., and Santi, D. V. (2004) Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U.S.A. 101*, 15573−15578.
(21) Menzella, H. G., Reisinger, S. J., Welch, M., Kealey, J. T., Kennedy, J., Reid, R., Tran, C. Q., and Santi, D. V. (2006) Redesign, synthesis and functional expression of the 6-deoxyerythronolide B polyketide synthase gene cluster. *J. Ind. Microbiol. Biotechnol. 33*, 22−28.
(22) Mutka, S. C., Carney, J. R., Liu, Y., and Kennedy, J. (2006) Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry 45*, 1321−1330.

(23) Gao, X., Wang, P., and Tang, Y. (2010) Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Appl. Microbiol. Biotechnol. 88*, 1233−1242.
(24) Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D., Gorlach, J., Mayo, J. M., Hu, A., Goff, S., Schmid, J., and Ligon, J. M. (2000) The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol. 7*, 97−109.
(25) Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. *Gene 249*, 153−160.
(26) Goldman, B. S., Nierman, W. C., Kaiser, D., Slater, S. C., Durkin, A. S., Eisen, J., Ronning, C. M., Barbazuk, W. B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H. S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S. A., Vaudin, M., Wiegand, R., and Kaplan, H. B. (2006) Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Natl. Acad. Sci. U.S.A. 103*, 15200−15205.
(27) Karlin, S., Brocchieri, L., Mrazek, J., and Kaiser, D. (2006) Distinguishing features of {delta}-proteobacterial genomes. *Proc. Natl. Acad. Sci. U.S.A. 103*, 11352−11357.
(28) Widmann, M., Clairo, M., Dippon, J., and Pleiss, J. (2008) Analysis of the distribution of functionally relevant rare codons. *BMC Genomics 9*, 1−8.
(29) Plotkin, J. B., and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet. 12*, 32−42.
(30) Allert, M., Cox, J. C., and Hellinga, H. W. (2010) Multifactorial determinants of protein expression in prokaryotic open reading frames. *J. Mol. Biol. 402*, 905−918.
(31) Henkin, T. M. (2000) Transcription termination control in bacteria. *Curr. Opin. Microbiol. 3*, 149−153.
(32) Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem. 55*, 339−372.
(33) Magrini, V., Creighton, C., and Youderian, P. (1999) Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: Genetic elements required for integration. *J. Bacteriol. 181*, 4050−4061.
(34) Sharp, P. M., and Li, W. H. (1987) The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res. 15*, 1281−1295.
(35) Karlin, S., and Mrazek, J. (2000) Predicted highly expressed genes of diverse prokaryotic genomes. *J. Bacteriol. 182*, 5238−5250.
(36) Czar, M. J., Anderson, J. C., Bader, J. S., and Peccoud, J. (2009) Gene synthesis demystified. *Trends Biotechnol. 27*, 63−72.
(37) 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.
(38) Holland, J. H. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
(39) Bäck, T., and Schwefel, H. P. (1993) An overview of evolutionary algorithms for parameter optimization. *Evol. Comput. 1*, 1−23.
(40) Goldberg, D. E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Boston.
(41) Miller, B. L., and Goldberg, D. E. (1995) Genetic algorithms, tournament selection, and the effects of noise. *Complex Syst. 9*, 193−212.
(42) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946−950.
(43) Julien, B. (2003) Characterization of the integrase gene and attachment site for the Myxococcus xanthus bacteriophage Mx9. *J. Bacteriol. 185*, 6325−6330.
(44) Hardt, I. H., Steinmetz, H., Gerth, K., Sasse, F., Reichenbach, H., and Höfle, G. (2001) New natural epothilones from *Sorangium cellulosum*, strains So ce90/B2 and So ce90/D13: isolation, structure elucidation, and SAR studies. *J. Nat. Prod. 64*, 847−856.
(45) Volz, C., Kegler, C., and Müller, R. (2012) Enhancer binding proteins act as hetero-oligomers and link secondary metabolite production to myxococcal development, motility and predation. *Chem. Biol.* http://dx.doi.org/10.1016/j.chembiol.2012.09.010.
(46) Meiser, P., and Müller, R. (2008) Two functionally redundant Sfp-type 4′-phosphopantetheinyl transferases differentially activate biosynthetic pathways in *Myxococcus xanthus*. *ChemBioChem 9*, 1549−1553.
(47) Sorensen, H. P., and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J. Biotechnol. 115*, 113−128.
(48) Li, G. W., Oh, E., and Weissman, J. S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature 484*, 538−541.
(49) Wu, J., Boghigian, B. A., Myint, M., Zhang, H., Zhang, S., and Pfeifer, B. A. (2010) Construction and performance of heterologous polyketide-producing K-12- and B-derived Escherichia coli. *Lett. Appl. Microbiol. 51*, 196−204.
(50) Spencer, P. S., Siller, E., Anderson, J. F., and Barral, J. M. (2012) Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol. 422*, 328−335.
(51) Zhang, G., Hubalewska, M., and Ignatova, Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol. 16*, 274−280.
(52) Fredrick, K., and Ibba, M. (2010) How the sequence of a gene can tune its translation. *Cell 141*, 227−229.
(53) Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell 141*, 344−354.
(54) Cannarozzi, G., Schraudolph, N. N., Faty, M., von, R. P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G., and Barral, Y. (2010) A role for codon order in translation dynamics. *Cell 141*, 355−367.
(55) Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F. U., and Baumeister, W. (2009) The native 3D organization of bacterial polysomes. *Cell 136*, 261−271.
(56) Katz, L., and Burge, C. B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res. 13*, 2042−2051.
(57) Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol. 16*, 589−597.
(58) Jha, S., and Komar, A. A. (2011) Birth, life and death of nascent polypeptide chains. *Biotechnol. J. 6*, 623−640.
(59) Lu, J., and Deutsch, C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol. 384*, 73−86.
(60) Shao, Z. Q., Zhang, Y. M., Feng, X. Y., Wang, B., and Chen, J. Q. (2012) Synonymous codon ordering: a subtle but prevalent strategy of bacteria to improve translational efficiency. *PLoS ONE 7*, e33547.
(61) Huang, Y., Koonin, E. V., Lipman, D. J., and Przytycka, T. M. (2009) Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage. *Nucleic Acids Res. 37*, 6799−6810.
(62) Tholstrup, J., Oddershede, L. B., and Sorensen, M. A. (2012) mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res. 40*, 303−313.
(63) Warnecke, T., Huang, Y., Przytycka, T. M., and Hurst, L. D. (2010) Unique cost dynamics elucidate the role of frameshifting errors in promoting translational robustness. *Genome Biol. Evol. 2*, 636−645.
(64) Yadav, V. G., De, M. M., Lim, C. G., Ajikumar, P. K., and Stephanopoulos, G. (2012) The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab. Eng. 14*, 233−241.
(65) Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Jr., Assad-Garcia, N., Glass, J. I., and Covert,

M. W. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell 150*, 389−401.
(66) Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., and Wilson, A. (2011) *Geneious v5*, 6.
(67) Bachmann, B. O., and Ravel, J. (2009) Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enymol. 458*, 181−217.
(68) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389−3402.
(69) Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M. O., Bartels, D., Bekel, T., Beyer, S., Bode, E., Bode, H. B., Bolten, C. J., Choudhuri, J. V., Doss, S., Elnakady, Y. A., Frank, B., Gaigalat, L., Goesmann, A., Groeger, C., Gross, F., Jelsbak, L., Jelsbak, L., Kalinowski, J., Kegler, C., Knauber, T., Konietzny, S., Kopp, M., Krause, L., Krug, D., Linke, B., Mahmud, T., Martinez-Arias, R., McHardy, A. C., Merai, M., Meyer, F., Mormann, S., Munoz-Dorado, J., Perez, J., Pradella, S., Rachid, S., Raddatz, G., Rosenau, F., Rückert, C., Sasse, F., Scharfe, M., Schuster, S. C., Suen, G., Treuner-Lange, A., Velicer, G. J., Vorhölter, F. J., Weissman, K. J., Welch, R. D., Wenzel, S. C., Whitworth, D. E., Wilhelm, S., Wittmann, C., Blöcker, H., Pühler, A., and Müller, R. (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol. 25*, 1281−1289.
(70) Wenzel, S. C., Kunze, B., Höfle, G., Silakowski, B., Scharfe, M., Blöcker, H., and Müller, R. (2005) 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.
(71) Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: The European molecular biology open software suite. *Trends Genet. 16*, 276−277.
(72) R Development Core Team (2011) *R: a language and environment for statistical computing*, http://www.R-project.org/.
(73) Kroos, L., Kuspa, A., and Kaiser, D. (1986) A global analysis of developmentally regulated genes in Myxococcus xanthus. *Dev. Biol. 117*, 252−266.
(74) Sambrook, J. Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
(75) Mühlenbein, H., and Schlierkamp-Voosen, D. (1993) Predictive models for the breeder genetic algorithm I. Continuous parameter optimization. *Evol. Comput. 1*, 25−49.
(76) De, S., Pal, S. K., and Gosh, A. (1998) Genotypic and phenotypic assortative mating in genetic algorithm. *Inf. Sci. 105*, 209−226.
(77) Booker, L. (1985) Improving the performance of genetic algorithms in classifier systems, in *Proceedings of the First International Conference on Genetic Algorithms*, pp 80−92, Lawrence Erlbaum Associates, Mahwah, NJ.