The Ketosynthase Domain Constrains the Design of Polyketide Synthases

*Maja Klaus[a+], Lynn Buyachuihan[a+] and Martin Grininger[a]**

[a] *Institute of Organic Chemistry and Chemical Biology, Buchmann Institute for Molecular Life Sciences, Goethe University Frankfurt*
[+] *These authors contributed equally*
* *Corresponding author: grininger@chemie.uni-frankfurt.de*

Abstract
Modular polyketide synthases (PKSs) produce complex, bioactive secondary metabolites in assembly line-like multistep reactions. The sequence of reaction steps determines the identity of the compound and is defined by the order of modules within the assembly line. This type of chemical programming makes modular PKSs ideal templates for engineering multistep C-C bond forming reactions. However, engineering of modular PKSs by recombining intact modules to produce novel, biologically active compounds has often resulted in chimeras exhibiting poor turnover rates and decreased product yields. Recent findings demonstrate that the low efficiencies of modular chimeric PKSs also result from rate limitations in the translocation step; i.e., the transfer of the growing polyketide from one module to the other and further processing of the polyketide substrate by the ketosynthase (KS) domain.
In this study, we systematically vary protein:protein interfaces, substrate identity and substrate specificity of the KS involved reactions of bimodular chimeric PKSs. First, we broadened the substrate specificity of the C-C bond forming KS domains with the aim to increase the turnover rates of kinetically impaired chimeric PKSs. Diverse multipoint mutants were calculated with a Rosetta program suite to result in stable active site environments. Some mutations efficiently released the kinetic penalties imposed by the non-native substrate and especially one mutant exhibited unexpected higher turnover numbers than the reference system. Second, we changed the identity of the donor module in the chimeric constructs to also vary protein-protein interactions and substrate identity. We observed that the KS domain, specifically the translocation reaction, is sensitive towards both substrate identity and protein-protein interactions in a complex and often non-intuitive interplay of these constraints. Overall, our systematic study can explain why early oversimplified engineering strategies based on the plain shuffling of modules mostly failed and why more recent approaches show improved success rates.

INTRODUCTION

Multimodular polyketide synthases (PKSs) are large megasynthases, which are responsible for the production of a variety of pharmaceutically important compounds such as antibiotics, anti-fungal and anti-cholesterol agents (Figure 1A).[1] The modularity of PKSs has put them at the forefront of engineering efforts to achieve programmable access to novel compounds with new bioactivities. However, to date most of the engineered systems do not reach biotechnological relevant stages, because the turnover numbers and thus product yields are often too low.[2–4] Despite considerable efforts in the last 30 years, no generalizable rules for the generation of efficient chimeric PKSs exist today. For example, although the importance of protein-protein interactions e.g. between docking domains[5–7] or between the acyl carrier protein (ACP) and the ketosynthase (KS) at different stages in the catalytic cycle has been recognized (Figure 1B),[4,8–10] most strategies to modulate and adapt those interactions in chimeric modular PKSs have failed so far.[11–14]

The 6-deoxyerythronolide B synthase (DEBS) is the prototypical example of this enzyme class and has served as a platform to study the structure, mechanism, and engineering potential of these enzymes (Figure 1A).[15] In the past, numerous studies on DEBS have shed light on substrate tolerance,[6,16] protein-protein interaction specificities,[5,9,17,18] and structural characteristics[19–21] with broad relevance for the family of modular PKSs. Recently, we analyzed a library of bimodular chimeric PKSs, with modules mainly deriving from DEBS, to compare different chimeric interfaces (Figure 1C).[22] From our analysis, it became apparent that some systems with native domain interfaces across module boundaries (from the ACP of the upstream module to the KS of the downstream module) suffer a decrease in turnover rates, likely as a result of a too narrow substrate tolerance of the downstream module (in this case observed on the example of M6-TE).[22] In order for a PKS module to successfully process non-natural polyketide substrates, mutagenesis approaches are required that broaden the substrate tolerance of a given module.

There are multiple options for a downstream module to be rate-limiting when processing a non-cognate substrate. In case of our model PKS (Figure 1C, bottom), three reactions can in principle account for substrate specificity of M6-TE and cause the above-mentioned phenomenon: (i) either the KS6-catalyzed condensation reaction, (ii) the following NADPH-dependent reduction of the β-ketoester by KR6, or (iii) the TE-catalyzed lactonization to the triketide product. While no information is available on the steady-state kinetics of the individual reaction steps within M6-TE, previous analysis of the hydrolysis rate of two enantiomeric diketides by the TE domain,[23] comparison of the reduction rates of two KR domains[24] and kinetic analysis of the KS domain of DEBS M2[16,25] have led to the conclusion that the KS-catalyzed decarboxylative condensation likely limits the rate of a native module.[15,25] Identification of the KS domain as a bottleneck in PKS turnover is further supported by a recent mutagenesis study on DEBS KS3 in which single point active site mutagenesis resulted in significantly broadened substrate specificities and higher turnover rates.[26] Based on the observation that the KS domains of the mycolactone synthase (MYCL) share a >97% sequence identity, despite accepting substrates of different length and chemical modification, DEBS KS3 residues were replaced with their counterparts from MYCL, resulting in one particular point mutant ($A154W_{KS3}$) with significantly increased promiscuity compared to wild-type KS3.[26]

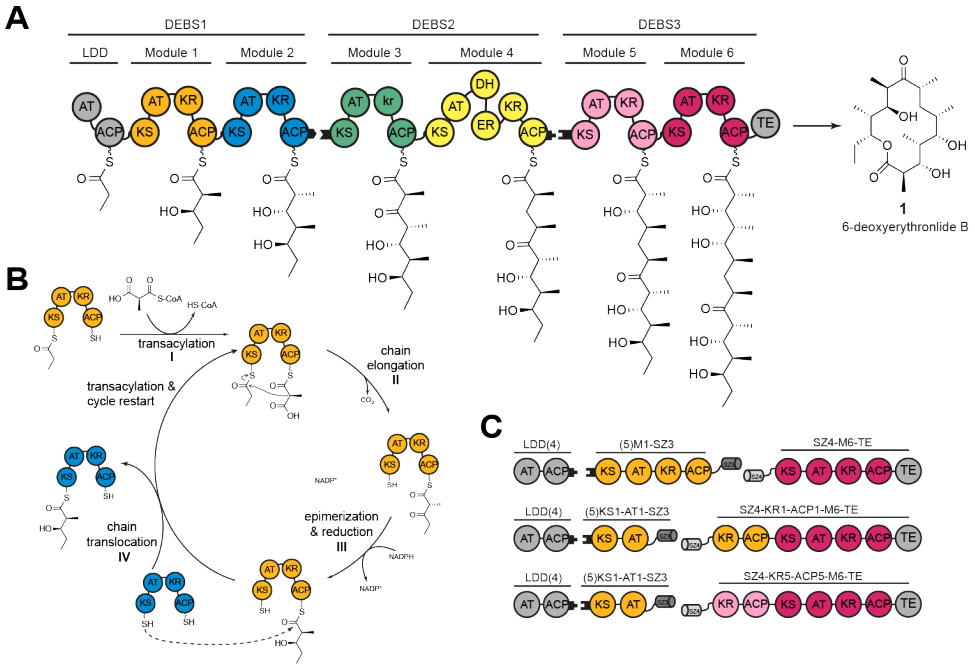


Figure 1. Schematic architecture of DEBS, its catalytic cycle and accessible interfaces for engineering approaches. (A) Architecture of DEBS. The three polypeptides (DEBS1-3), the encoded modules (M1-M6), as well as the loading didomain (LDD), the thioesterase domain (TE), and the final product (**1**) are depicted. Polyketide intermediates are shown as attached to the respective acyl carrier protein (ACP). Black tabs depict docking domains. Domain annotations are as follows: LDD- loading didomain, AT- acyltransferase, KS- ketosynthase, ACP- acyl carrier protein, KR- ketoreductase, DH- dehydratase, ER- enoylreductase, TE- thioesterase. Module 3 has a ketoreductase-like domain (denoted in lowercase) that lacks NADPH-dependent oxidoreductase activity but harbors C2 epimerase activity.[27,28] (B) Catalytic cycle on the example of M1 and M2. (I) AT-catalyzed transacylation using (2*S*)-methylmalonyl-CoA, followed by KS-mediated decarboxylative Claisen condensation (chain elongation, II) to give the (2*R*)-diketide.[29] In M1 the KR catalyzes the epimerization of the C2 methyl group, followed by diastereospecific reduction (III) to give the (2*S*,3*R*)-2-methyl-3-hydroxydiketide condensation product. The diketide is translocated to the KS domain of the downstream module (chain translocation, IV) from which the next round of reactions can start. In DEBS only KR1 and KR3 catalyze epimerization of the C2 methyl group.[30] (C) Subset of bimodular chimeric PKSs used previously.[22] SYNZIP domains (SZ3 and SZ4) assisted in construct design. Black tabs represent DEBS-derived docking domains (module numbers in brackets) installed at the protein termini to mediate weak interactions.

Having identified the KS as candidate enzymatic domain for additional engineering approaches, we sought to broaden its substrate specificity by multipoint mutations. Directed evolution experiments revealed that on average 10 mutations are required to improve the activity of an enzyme by factor 1000,[31] indicating that several mutations will likely be necessary to increase turnover of non-native substrates by PKSs. Due to the lack of practical selection methods, the application of directed evolution approaches is hampered in our experimental set-up. We therefore aimed at a more targeted approach to guide multipoint mutagenesis of KS domains, using the FuncLib server, a program that generates a variety of diverse multipoint mutants, by calculating stable networks of interacting active site residues.[32] FuncLib uses phylogenetic information to reveal possible, coupled mutations within a protein, by working on two levels: First, it uses phylogenetic information to suggest amino acid exchanges at selected positions, and second, it ranks the mutated proteins by calculating their stability via the Rosetta program suite. In this process, mutations that are predicted to destabilize the protein fold are discarded. The FuncLib approach does not target individual

substrate specificities, but delivers a set of stable variants that can then be screened for the specificities and activities of interest.

Here we identify KS domains as a bottleneck in assembly line PKSs and show that multipoint mutagenesis within the KS active site can increase the turnover rate of chimeric PKSs. Our data demonstrates that the translocation reaction in modular PKSs is constrained by protein-protein interactions and substrate specificity of the KS domain, and can explain why engineering approaches based on the simple shuffling of modules have often failed in the past.

RESULTS AND DISCUSSION

Design of KS Multipoint Mutants- To test the feasibility of KS engineering to broaden the substrate specificity of PKS modules, we decided to apply the FuncLib approach on KS domains of the downstream modules (also termed acceptor modules in the following) of two bimodular chimeric PKSs (Figure 2A). Specifically, we introduced the KS mutations in the context of the full-length modules M3-TE and M6-TE (DEBS TE in both systems) and compared the turnover rates of wild-type (non KS-mutated) vs. mutant modules (Figure 2B). In these bimodular systems, e.g. LDD(4)+(5)M1-SZ3+SZ4-M3-TE, docking domains (indicated as (4)+(5))[33] and SYNZIP domains (indicated as SZ3+SZ4)[34,35] were employed at the protein interfaces to mediate non-covalent interactions between separately purified modules (Figure 2B).

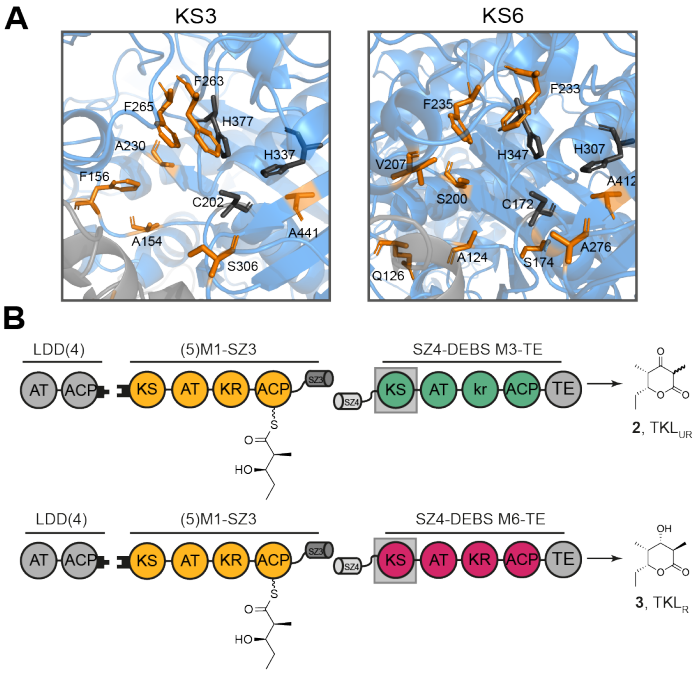


Figure 2. KS mutagenesis - Structural organization of KS active sites and setup of bimodular chimeric PKSs. (A) Structural organization of the active site of KS3 and KS6. The catalytic triad is depicted as gray sticks. Residues chosen for mutagenesis within a 12 Å-range of the active cysteine $C202_{KS3}$ or $C172_{KS6}$ are shown as orange sticks. (B) Bimodular chimeric PKSs using LDD(4), (5)M1-SZ3, and SZ4-M3-TE or SZ4-M6-TE as the final module. Matching DEBS-derived docking domains (black tabs; (4)+(5)) or matching SYNZIP domains (SZ3+SZ4) are installed at the protein termini to mediate non-covalent interactions. The chimeric ACP1 harbors the (2*S*,3*R*)-2-methyl-3-hydroxydiketide (natural diketide, NDK) as the incoming substrate for wild-type and mutant KS domains. Depending on the acceptor module either unreduced triketide lactone (TKL_{UR}) **2** or reduced TKL (TKL_{R}) **3** are the predicted products.

The choice of M3 and M6 was guided by their differing turnover rates with M1[22] and the previously observed substrate restrictions.[16] X-ray structural data exists for KS3-AT3,[36,37] which gave us the opportunity to run the FuncLib method for the calculation of stable active sites on authentic data. For KS6-AT6, a homology model was generated on the basis of a sequence similarity of ~55% to the structurally solved KS-AT didomains KS3-AT3[36] and KS5-AT5[19] and the model was also submitted to FuncLib. To determine suitable residues for KS active site mutagenesis, we considered residues within a first shell around the active cysteines that directly decorate the binding site for substrates and mapped them onto a multiple sequence alignment (Figure S1&S2). From the subset of non-conserved residues, all those pointing towards the active cysteines were chosen for mutagenesis, as they might be involved in substrate binding during the enzymatic reaction, leading to a total of 7 residues for KS3 and 9 residues for KS6 (Figure 2A).

Submission of the KS3-AT3 and KS6-AT6 didomains and selection of the target residues within FuncLib (including the possibility of a tryptophan at $A154_{KS3}$ and $A124_{KS6}$)[26] resulted in a sequence space of 152,826 designs for KS3-AT3 and 285,822 designs for KS6-AT6 with a large variability in mutated residues (Table S1). In a next step, Rosetta calculation revealed that from the initial designs 1,011 designs for KS3-AT3 and 3,683 designs for KS6-AT6 result in more stable proteins than the respective wild-type proteins. From the 50 highest ranked designs, we chose 12 for KS3-AT3 and 10 for KS6-AT6 for subsequent analysis, based on their large variability in mutated residues (Tables S2&S3). All designs harbored a minimum of three and a maximum of five mutations. In addition, the previously reported single point mutants $A154W_{KS3}$ and $A124W_{KS6}$ were also generated as a reference.[26] The mutations are predicted to change the geometry of the respective active sites, potentially leading to different properties in substrate binding or catalysis (Figures S3&S4).

Generation of Multipoint KS Mutants and Turnover Analysis-

To evaluate the performance of the different KS designs under PKS turnover conditions, all KS mutations were introduced in the background of the full-length modules SZ4-M3-TE and SZ4-M6-TE. All mutant proteins exhibited similar purification behavior and purity (Figure S5A&B), with yields comparable or higher than the wild-type modules (between 5 - 17 mg per liter *E. coli* culture, Table S4). Protein oligomerization, as measured by SEC, showed that besides Mut07 and Mut10 all mutant proteins eluted in a single peak indicative of a homodimeric protein fold (Figure S6&S7). In good agreement with these results, thermal shift assays of each mutant indicated no significant difference in thermal stability between wild-type and mutant modules (Table S5). By analyzing the protein quality of all generated mutants, we conclude that in our case FuncLib serves as a useful tool to calculate stable mutants with a large variety in active site residues without decreasing protein yield as well as oligomeric and thermal stability.

Turnover rates of bimodular chimeric PKSs harboring KS3 mutations

The turnover rates of all M3 mutants was first assessed in the bimodular chimeric PKS consisting of LDD(4), (5)M1-SZ3 and SZ4-M3-TE; a system harboring a non-cognate ACP1:KS3 interface and challenging KS3 with the DEBS-derived natural diketide (NDK, (2*S*,3*R*)-2-methyl-3-hydroxydiketide, Figure 3A). As measured before, the chimeric system with wild-type SZ4-M3-TE exhibited a >30-fold lower activity than the reference system with SZ4-M2-TE (Figure 3A, gray bar). Compared to wild-type M3, the single point mutant Mut01 ($A154W_{KS3}$), following the design of Murphy *et al.*,[26] had a roughly 2-fold increased activity. Among the multipoint mutants constructed by the FuncLib method, Mut06, Mut07, Mut10, Mut14, and Mut15 showed no improvement or even loss of activity (Mut07, Figure 3A). Similar to Mut01, Mut03 and Mut05 showed a 2-fold increase (Figure 3A). The multipoint mutants Mut04, Mut08, Mut09, Mut11 and Mut13 exhibited the largest increase in turnover rate (2.5-fold), narrowing the gap towards the reference system to a 15-fold difference. In good agreement

with the turnover rates, product analysis after overnight incubation confirmed the presence of the expected unreduced triketide lactone **2** (TKL_{UR}) in all systems with measurable turnover (Figure S8).

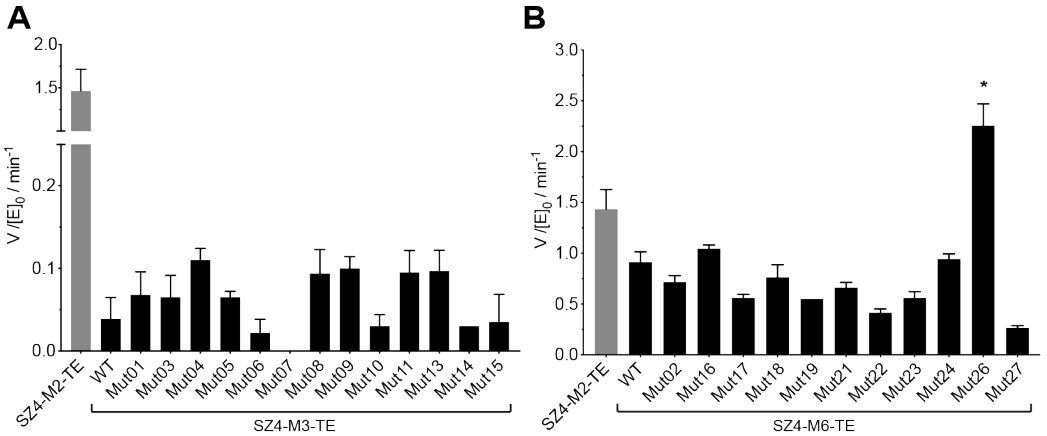


Figure 3. Turnover rates of bimodular chimeric PKSs using wild-type or mutant M3-TE (A) and M6-TE (B), respectively. Bimodular PKSs consisted of LDD(4), (5)M1-SZ3, and SZ4-M3-TE/SZ4-M6-TE (wild-type (WT) or one of their mutants). For both bimodular PKSs, the reference system consisted of LDD(4), (5)M1-SZ3, and SZ4-M2-TE (gray bar). All initial rate data was obtained at 4 μM enzyme concentration and non-limiting concentrations of propionyl-CoA, methylmalonyl-CoA, and NADPH. Measurements were performed in technical triplicate. Note that for further confidence, the turnover rate of Mut26 is displayed as an average of three independently purified protein samples (biological triplicate) measured in triplicate.

Turnover rates of bimodular chimeric PKSs harboring KS6 mutations

As a next step, we evaluated the turnover rates of bimodular chimeric PKSs using M6-TE or one of its mutants (Figure 3B). These bimodular systems harbor a non-cognate ACP1:KS6 interface. Analysis of wild-type M6-TE in the presence of (5)M1-SZ3 showed similar results as reported earlier.[22] In contrast to KS3 mutants, the single point mutant Mut02 ($A124W_{KS6}$)[26] and 8 out of 9 multipoint mutants revealed diminished turnover rates of the bimodular system compared to wild-type M6 (Figure 3B). To our surprise, one mutant, Mut26, did not only show significantly increased turnover compared to wild-type M6 (~2.5-fold improvement), but performed even better than the reference system M2-TE (1.5-fold improvement, Figure 3B). Mut26 contains only three KS mutations of which S174A and V207L are predicted to cause only minimal steric changes, while F235M is anticipated to enlarge the binding site of KS6 (Figure S4) and to increase the similarity between KS2 and KS6 (Figure S9).[22]

Note that for all acceptor modules shown in Figure 3B, a mixture of the expected reduced TKL_R **3** and unreduced TKL_{UR} **2** was obtained from the reaction mixture after a reaction time of only 10 min (Table S6, Figure S10). Similar data has been reported before for DEBS M6-TE.[38] As the turnover rates were measured by correlating the reduction rates of the KR domains (NADPH consumption) with product formation and assuming a 100% conversion to reduced TKL_R,[33] the absolute turnover numbers presented in Figure 3B would all be higher if corrected by the amount of produced unreduced TKL_{UR}. Yet, as the ratios of reduced to unreduced TKL are similar in all M6 samples (Table S6), the relative ratios of turnover rates are still captured in Figure 3B.

Using different donor ACP-substrate combinations to analyze the turnover rates of KS mutants

As the same donor module (5)M1-SZ3 was used to evaluate the performance of all generated bimodular PKSs, the analysis so far only gave insight into the effect of each mutant when presented with the chimeric ACP1 bearing NDK (Figure 2B). With the aim of additionally analyzing the contribution of ACP:KS recognition and substrate:KS interaction on the turnover of KS mutants, we

generated alternative M1 donor modules that would allow us to probe the remaining ACP-substrate combinations (Figure 4&5).

Based on the observations that KR domains can efficiently turn over with a variety of chimeric ACPs,[39] and that chain elongation at a chimeric ACP:KS interface in a fusion protein is not rate-limiting,[22] we generated 6 alternative M1 donors by permutating KR1 and ACP1. The exchange of ACP1 with ACP2 and ACP5, the native upstream ACPs of M3/M6, generated donor modules carrying NDK on the respective ACP (Figure 4A, donor 2 & Figure 5A, donor 5). Since KR domains are responsible for installing the stereogenic centers at the α- and β-position of the polyketide, an exchange of KR1 by KR2 allowed us to generate the DEBS-derived enantiomeric diketide (EDK, (2*R*,3*S*)-2-methyl-3-hydroxydiketide) on ACP1 (Figure 4A&5A, donor 3).[4] As a last engineering approach, both KR1 and ACP1 were exchanged by the native upstream domains of M3/M6, resulting in the formation of EDK on either ACP2 (Figure 4A, donor 4) or ACP5 (Figure 5A, donor 6). All donor proteins could be purified in similar quantities and were pure as judged by SEC and SDS-PAGE (Table S4, Figure S11&S5C).

With these donor modules in hand, we examined the activity of bimodular PKSs in the presence of either wild-type M3 or a subset of mutants that showed high turnover rates with donor 1 (Mut01, Mut08, Mut11, Mut13) (Figure 4A). Among all acceptors, donor 2, that reconstitutes the native ACP2:KS3 interaction and presents KS3 with the NDK substrate, emerged as the best donor module. In many cases, the second-best donor module appeared to be donor 4, also employing ACP2, but displaying the enantiomeric substrate. Overall, the comparison of different donors for wild-type M3 and Mut01 highlights the importance of protein-protein interactions at the ACP:KS interface in this chimeric PKS.[4] In addition, whereas wild-type M3 and the single point mutant Mut01 showed a strong substrate preference for NDK over EDK, some of the multipoint mutants efficiently broadened the substrate tolerance of KS3 and exhibited a released ACP preference.

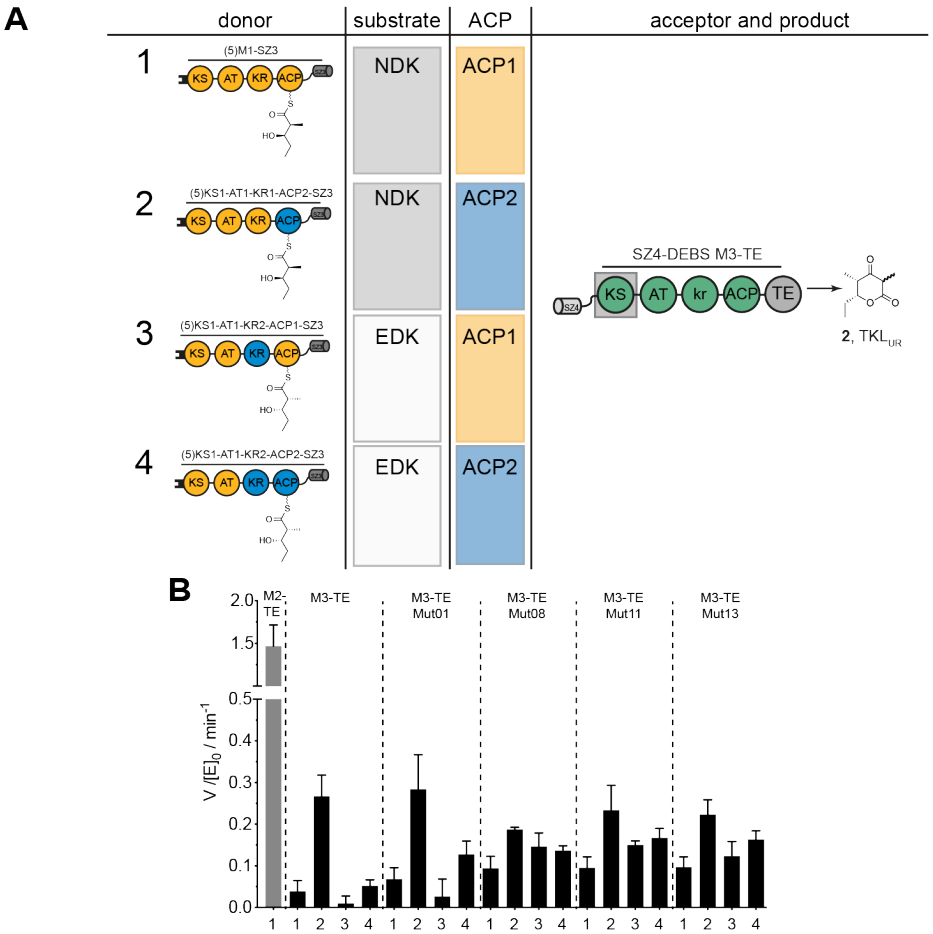
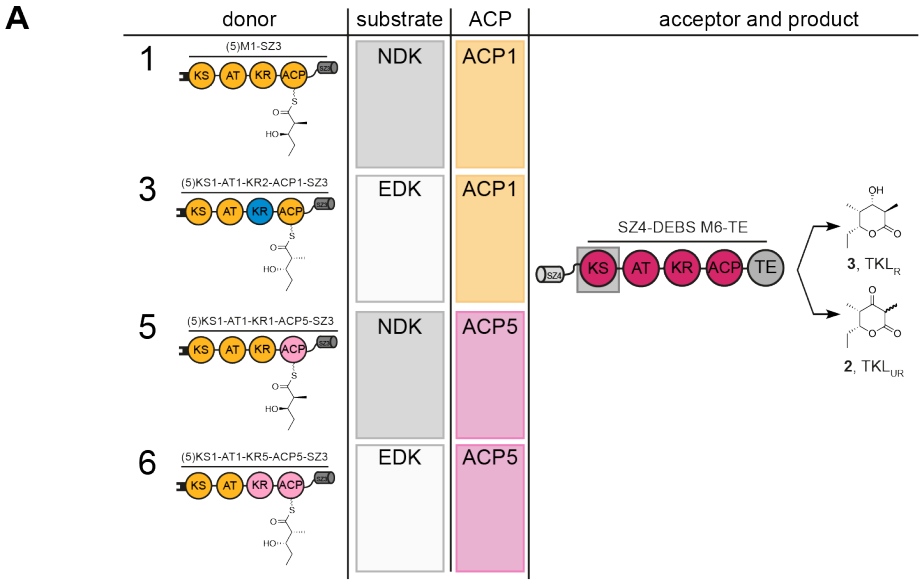


Figure 4. Turnover rates of bimodular chimeric PKSs using M3-TE and different donor modules. (A) The donor modules used to generate bimodular PKSs. Substrates and the employed ACPs are indicated. (B) Turnover rates of bimodular chimeric PKS using M3-TE as acceptor module. All bimodular PKSs consisted of LDD(4), followed by a donor module as indicated in (A), and SZ4-M3-TE or one of its mutants. The reference system consisted of LDD(4), (5)M1-SZ3, and SZ4-M2-TE (gray bar). All initial rate data was obtained at 4 µM enzyme concentration and non-limiting concentrations of propionyl-CoA, methylmalonyl-CoA, and NADPH. Measurements were performed in triplicate. The expected unreduced TKL_{UR} was detected in all samples after overnight incubation (Figure S12).

The alternative donor modules were further analyzed in bimodular chimeric PKSs with wild-type M6 and Mut26 as the acceptor modules (Figure 5A). Again, we note that all bimodular PKSs using wild-type M6 or Mut26 produced a mixture of TKL_R and TKL_{UR} (Figure 5B), which hampers kinetic analysis of the bimodular PKSs and, specifically, underestimates the actual number of turnovers (Figure 5C). Intriguingly, while similar ratios of TKL_R:TKL_{UR} were observed for both acceptors (Figure 5C, Table S7), the relative ratios differed between the used donors. This highlights that both wild-type M6 and Mut26 process the polyketide substrate in a similar manner. However, as the lactonization of TKL_{UR} can only occur after a successful reduction reaction in the donor module, the differences in product ratios indicate that the identity of the donor module affects the reduction reaction of KR6 within the acceptor module. For wild-type M6 and Mut26, the expected TKL_R is more efficiently generated with either ACP1 or ACP5 carrying EDK as the substrate (Figure 5B, donor 3&6), whereas a higher amount of TKL_{UR} accumulates in the presence of NDK bound to either ACP1 or ACP5 (Figure 5B, donor 1&5). Analysis of the product ratios highlights that the identity of the substrate impacts the reduction reaction by KR6. A pronounced substrate specificity of KR6 was observed before.[40] Overall, no

unambiguous correlation of turnover numbers with either the involved protein-protein interactions nor substrate:KS interactions can be drawn based on the turnover data for bimodular chimeric PKSs using different donor modules for loading of M6. While it seems that for wild-type M6, usage of donor 1 and 3 correlates with the highest activities, donor 1 appears to be the preferred for Mut26 (Figure 5C). In other words, while wild-type M6 only shows high turnover rates in the presence of ACP1 carrying either NDK or EDK, Mut26 exhibits high turnover numbers with (almost) all donors, but especially in the presence of ACP1-NDK as the donor.



B

acceptor	donor	substrate	ACP	TKL$_R$: TKL$_{UR}$
M6-TE	1, M1	NDK	ACP1	1 : 28.8
	5, KS1-AT1-KR1-ACP5	NDK	ACP5	N. A.
	3, KS1-AT1-KR2-ACP1	EDK	ACP1	1.7 : 1
	6, KS1-AT1-KR5-ACP5	EDK	ACP5	1.3 : 1
M6-TE_Mut26	1, M1	NDK	ACP1	1 : 39.2
	5, KS1-AT1-KR1-ACP5	NDK	ACP5	1 : 30.0
	3, KS1-AT1-KR2-ACP1	EDK	ACP1	7.3 : 1
	6, KS1-AT1-KR5-ACP5	EDK	ACP5	8.5 : 1

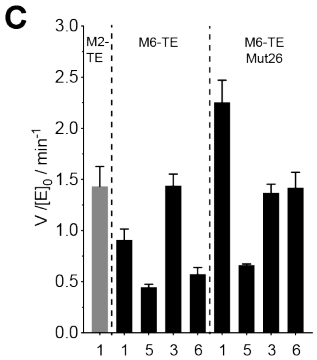


Figure 5. Turnover rates of bimodular chimeric PKSs using M6-TE or Mut26 and different donor modules. (A) The donor modules used to generate bimodular PKSs. Substrates and the employed ACPs are indicated. (B) Products of bimodular PKSs using M6-TE or Mut26. Different ratios of TKL_R-to TKL_{UR} were produced with different donors, yet wild-type M6 and Mut26 show the same trends. Calculation based on peak areas shown in Table S7. N. A., not applicable. (C) Turnover rates of bimodular chimeric PKS using M6-TE as the acceptor module. All bimodular PKSs consisted of LDD(4), followed by a donor module as indicated in (A), and SZ4-M6-TE or Mut26. The reference system consisted of LDD(4), (5)M1-SZ3, and SZ4-M2-TE (gray bar). All initial rate data was obtained at 4 µM enzyme concentration and non-limiting concentrations of propionyl-CoA, methylmalonyl-CoA, and NADPH. Measurements were performed in triplicate.

Conclusion

In this work, 12 multipoint mutants of KS3 and 10 mutants of KS6 were generated and tested for their ability to productively process two different substrates presented by either the natural upstream ACP or a chimeric ACP. Despite strong difference in active site residues, all mutants could

be purified as stable proteins and with high yields, highlighting the successful use of FuncLib to calculate stable active site mutants on the basis of a protein structure or a homology model. This is especially remarkable as purification of PKS modules can often be difficult. Analysis of the turnover rates of bimodular PKSs using the M3 and M6 acceptors in the presence of M1 revealed rate improvement or impairment for different mutants (Figure 3). FuncLib does not design mutations for turnover with specific substrates, but rather yields a large variety of mutations that can be screened for specificities and activities of interest.[32] In that light, the increase as well as the decrease in turnover rates was expected, provided that it is the substrate specificity of the acceptor KS that limits the rate of a chimeric PKS.[26,41]

In addition, we systematically varied the protein-protein interactions within the chimeric KS-mutated PKSs; i.e., the protein:protein interface emerging from the upstream ACP docking to the downstream KS for translocation and the structure of the substrate presented to the downstream KS for uptake and processing. Overall, our approach yielded a large body of data giving insight into mechanistic details of PKS-mediated product synthesis outlined in the following on three examples:

(i) The initial observation that wild-type M3 prefers NDK over EDK[16] has been confirmed in this study. While the low turnover rates with ACP1-NDK as the donor have been previously observed to stem from a rate-limiting effect based on the non-native ACP1:KS3 interface,[4] the usage of ACP2-NDK overcomes these rate limitations and results in the highest measured turnover rates of a bimodular PKS using wild-type M3-TE (Figure 4B). In this regard, the reconstitution of the native interface improved the efficiency of the system.

(ii) By taking a closer look at key residues mutated in the different M3 mutants, it is striking that from the 12 chosen designs, Mut08, Mut11, and Mut13 all include the S306P mutation (Table S2, Figure S3). Overall, a picture emerges in which the chimeric bimodular PKSs using M3-TE often benefits from retaining the native ACP2:KS3 interface (see (i)) and can be engineered towards higher turnover rates as well as broadened substrate specificities via KS multipoint mutagenesis.

(iii) Data on mutagenesis of KS6 provides further indication as to why DEBS M6-TE has been recognized as one of the best suited modules for engineering approaches.[2,4,22,38,42] Previously, the high sequence similarity between KS2 and KS6 has served as an explanation for wild-type M6 being almost as good as an acceptor as M2 with M1 as the donor module.[22] As such, it was not predicted that multipoint mutagenesis of KS6 would result in mutants with improved activity with M1, but rather lead to impaired mutants. Indeed, most of the screened mutants exhibited lower turnover numbers than wild-type M6, except for a single mutant (Mut26, Figure 3B). A closer look at the active site of this mutant revealed that the F235M mutation likely has the strongest impact on binding site geometry by enlarging the active site while also increasing the similarity between KS2 and KS6 even further (Figure S9). This data supports the relevance of the specificity of the downstream KS as a bottleneck in chimeric PKSs and the capability of KS mutagenesis to improve efficiencies (see (ii)).

Beyond the boundaries of the DEBS PKSs that served as a model system in this study, our data suggests limitations in engineering modular chimeric PKSs in general. As a yet unmatched systematic *in vitro* analysis of parameters involved in the KS-mediated uptake and processing of substrates (as part of the translocation and elongation reactions), it can highlight why conventional engineering strategies have failed in the past.[2–4] In most cases, turnover efficiency can neither be deduced from the identity of the ACP donor, nor from the incoming substrate, as the highest turnover rates cannot be correlated with either of these parameters. Even worse from a perspective of simple mix-and-match eingeering approaches, it appears that a cumulated effect and a non-intuitive interplay of the

identity of both ACP and substrate influence the turnover of many downstream modules. Future engineering strategies are likely to be more successful by either retaining substrate specificities,[43] or by preserving the natural domain-domain interfaces,[44–47] a task that might be achieved with the support from evolutionary analysis.[48]

METHODS

Reagents: CloneAmp HiFi PCR Premix was from Takara. Restriction enzymes were from New England Biolabs. All primers were synthesized by Sigma Aldrich. For DNA purification, the GeneJET Plasmid Miniprep Kit and the GeneJET Gel Extraction Kit from Thermo Scientific were used. Stellar Competent Cells were from Takara and One Shot® BL21 (DE3) Cells were from Invitrogen. All chemicals for buffer preparations were from SigmaAldrich. Isopropyl-β-D-1-thiogalactopyranoside (IPTG), kanamycin sulfate, and carbenicillin were from Sigma Aldrich/Carl Roth. LB broth (Lennox) and 2xYT media for cell cultures were from Carl Roth, Ni-NTA affinity resin was from Clontech. For anion exchange chromatography the HiTrapQ column was from GE Healthcare. For protein concentration Amicon Ultra centrifugal filters were from Millipore. Coenzyme A (CoA), reduced β-nicotinamide adenine dinucleotide 2'-phosphate (NADPH), sodium propionate, methylmalonic acid, and magenesium chloride hexahydrate were from Carl Roth. Adenosine-5'-triphosphate (ATP) was from SigmaAldrich. Reducing agent tris(2-carboxyethyl)-phosphine (TCEP) was from Thermo Scientific. UV-Star half area microtiter plates were from Greiner.

Plasmids: Plasmids harboring genes encoding individual PKS modules were either generated in this study via In-Fusion Cloning (Takara), site directed mutagenesis, or reported previously. Tables S8-S11 specify the cloning strategy, primer sequences, and the resulting plasmids. Plasmids sequences were verified by DNA sequencing (SeqLab). Protein sequences of donor modules newly generated in this study are presented in Table S12.

Bacterial Cell Culture and Protein Purification: All PKS proteins were expressed and purified using similar protocols. For *holo*-proteins (where the ACP domain is post-translationally modified with a phosphopantetheine arm) *E. coli* BL21 cells were co-transformed with a plasmid encoding for the phosphopantetheine transferase Sfp from *B. subtilis* (pAR357[49]). All proteins contained a C-terminal His_6-tag for purification. Cultures were grown on a 1-2 L scale in 2xYT media at 37 °C to an OD_{600} of 0.3, whereupon the temperature was adjusted to 18 °C. At OD_{600} of 0.6, protein production was induced with 0.1 mM IPTG, and the cells were grown for another 18 h. Cells were harvested by centrifugation at 5000 g for 15 min and lysed by French Press (lysis buffer: 50 mM sodium phosphate, 10 mM imidazole, 450 mM NaCl, 10% glycerol, pH 7.6). The cell debris was removed by centrifugation at 50,000 g for 50 min. The supernatant was further purified using affinity chromatography (Ni-NTA agarose resin, 5 mL column volumes). After applying the supernatant to the column, a first wash step was performed with the above lysis buffer (10 column volumes), followed by a second wash step using 10 column volumes of wash buffer (50 mM phosphate, 25 mM imidazole, 300 mM NaCl, 10 % glycerol, pH 7.6). Proteins were eluted with 6 column volumes elution buffer (50 mM phosphate, 500 mM, 10% glycerol, pH 7.6). The eluate was further purified by anion exchange chromatography using a HitrapQ column on an ÄKTA FPLC system. Buffer A consisted of 50 mM phosphate, 10% glycerol, pH 7.6, whereas buffer B contained 50 mM phosphate, 500 mM NaCl, 10% glycerol, pH 7.6. The yields of all protein per liter of culture are noted in Table S4. Enzymes MatB, SCME, and PrpE were purified as decribed.[33,50] Protein concentrations were determined with the BCA Protein Assay Kit (Thermo Scientific). Samples were stored as aliquots at −80 °C until further use.

Size Exclusion Chromatography: To determine the purity of the proteins after ion exchange chromatography, samples of each protein were analyzed by size exclusion chromatography on an

ÄKTA FPLC system using a Superose 6 Increase 10/300 GL column (buffer: 50 mM phosphate, 500 mM NaCl, 10% glycerol, pH 7.6). For protein purity by SDS-PAGE and SEC profiles refer to Figures S6, S7&S11.

Thermal Shift Assay: The melting temperatures of all proteins in a buffer consisting of 50 mM phosphate, 500 mM NaCl, 10% glycerol, pH 7.6 were measured according to published procedures using a protein concentration of 1.2 mg/mL.[51]

PKS Enzymatic Assays: PKS enzymatic assays were performed according to published procedures.[4] PKS enzymes were used at 4 µM final concentration, for *in situ* substrate generation enzymes MatB, PrpE, and SCME were used at final concentrations of 4 µM, 2 µM, and 8 µM, respectively.

Liquid Chromatography-Mass Spectrometry Analysis of Polyketides: Dried samples were reconstituted in 100 µL methanol, separated on a Acquity UPLC BEH C18 1.7 µm RP 2.1 x 50 mm column (Waters) with an Acquity UPLC BEH C18 1.7 µm RP 2.1 x 5 mm pre-column (Waters), connected to an Ultimate 3000 RSLC or an Ultimate 3000 LC (Dionex) HPLC over a 16 min linear gradient of acetonitrile from 5% to 95% in water, and subsequently injected into an Impact II qTof (Bruker) or an AmaZonX (Bruker) equipped with an ESI Source set to positive ionization mode. Reduced and unreduced triketide products were located by searching for the theoretical *m/z* for the $[M+H]^+$-ion. Unreduced triketides: $[M+H]^+$ = 171.098, reduced triketides $[M+H]^+$ =173.118.

Bioinformatical Analysis: Sequence alignments were generated with ClustalWS as implemented in *Jalview* 2.10.5.[52] Homology models were generated with SWISS-MODEL.[19] The FuncLib webserver was used to calculate diverse multipoint mutations within the KS active site.[32] The calculation for KS3 was based on the X-ray structure of the KS3-AT3 didomain.[36] The calculation for KS6 was based on the homology model generated with SWISS-MODEL based on the X-ray structure of KS5-AT5.[19] In the first step, for the calculation of the sequence space, chains A+B were included for KS3. To decrease the calculation effort only chain B was included for KS6. Seven residues were chosen for mutagenesis of KS3 and nine positions for KS6. In addition, residues of the catalytic triad (KS3: Cys202, H337 and H377; KS6: Cys172, H307 and H347) were selected not to be altered during simulations. The nomenclature of residues is derived from structural models. $C202_{KS3}$ corresponds to UniProt EryA2 (Q03132) position 202 and $C172_{KS6}$ to UniProt EryA3 (Q03133) position 1661. For generation of the multiple sequence alignment the default parameters were used during all calculations (Min ID 0.3, Max targets 3000, Coverage 0.6, and E value 0.0001). In the second calculation step, the parameters were chosen to yield between 500-500,000 designs. For KS3 152,826 and for KS6 285,822 designs harboring 3-5 mutations were selected for Rosetta calculation.

Acknowledgments:
We thank Helge Bode for use of their LC-MS instruments. We further want to thank Aleksandra Nivina and Chaitan Khosla for helpful discussions. In using the FuncLib method, we received fantastic support from Sarel Fleishman, Rosalie Lipsh and Jaime Prilusky.

Funding Sources:
This work was supported by a Lichtenberg grant of the Volkswagen Foundation to M.G. (grant number 85701). Further support was received from the LOEWE program (Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz) of the state of Hesse conducted within the framework of the MegaSyn Research Cluster, as well as the DFG grant GR3854/6-1.

Authors contribution:

The manuscript was written through contributions of all authors. M.K. conceived and supervised the project. M.K. cloned and purified alternative donor modules and analyzed the activity. L.B. designed the mutants, cloned and purified all mutant constructs and analyzed the activity, thermal stability, and oligomerization. M.G. designed the research and analyzed the data.

Supporting Information Available: Sequence alignments, structural models, SEC data, SDS-PAGE data, LC-MS data, FuncLib parameters, list of plasmids, construct design, protein sequences, and expression yields. This material is available free of charge via the internet at http://pubs.acs.org.

References

(1) Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep. 18*, 380–416.
(2) Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol. 23*, 1171–1176.
(3) Menzella, H. G., Carney, J. R., and Santi, D. V. (2007) Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol. 14*, 143–151.
(4) Klaus, M., Ostrowski, M. P., Austerjost, J., Robbins, T., Lowry, B., Cane, D. E., and Khosla, C. (2016) Protein Protein Interactions, not Substrate Recognition, Dominates the Turnover of Chimeric Assembly Line Polyketide Synthases. *J. Biol. Chem. 291*, 16404–16415.
(5) Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) Selective Protein-Protein Interactions Direct Channeling of Intermediates between Polyketide Synthase Modules. *Biochemistry 40*, 2326–2331.
(6) Wu, N., Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc. 123*, 6465–6474.
(7) Wu, N., Cane, D. E., and Khosla, C. (2002) Quantitative Analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry 41*, 5056–5066.
(8) Chen, A. Y., Schnarr, N. a., Kim, C. Y., Cane, D. E., and Khosla, C. (2006) Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc. 128*, 3067–3074.
(9) Kapur, S., Lowry, B., Yuzawa, S., Kenthirapalan, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2012) Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. 109*, 4110–4115.
(10) Mindrebo, J. T., Patel, A., Misson, L. E., Kim, W. E., Davis, T. D., Ni, Q. Z., La Clair, J. J., and Burkart, M. D. (2019) Structural Basis of Acyl-Carrier Protein Interactions in Fatty Acid and Polyketide Biosynthesis, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*. Elsevier.
(11) Bayly, C., and Yadav, V. (2017) Towards Precision Engineering of Canonical Polyketide Synthase Domains: Recent Advances and Future Prospects. *Molecules 22*, 235.
(12) Klaus, M., and Grininger, M. (2018) Engineering strategies for rational polyketide synthase design. *Nat. Prod. Rep. 35*, 1070–1081.
(13) Bozhüyük, K. A., Micklefield, J., and Wilkinson, B. (2019) Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol. 51*, 88–96.
(14) Yuzawa, S., Backman, T. W. H., Keasling, J. D., and Katz, L. (2018) Synthetic biology of polyketide synthases. *J. Ind. Microbiol. Biotechnol. 45*, 621–633.
(15) Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007) Structure and Mechanism of the 6-deoxyerythronolide B Synthase. *Annu. Rev. Biochem. 76*, 195–221.
(16) Wu, N., Kudo, F., Cane, D. E., and Khosla, C. (2000) Analysis of the molecular recognition features of individual modules derived from the erythromycin polyketide synthase. *J. Am. Chem. Soc. 122*, 4847–4852.
(17) Buchholz, T. J., Geders, T. W., Bartley III, F. E., Reynolds, K. a., Smith, J. L., and Sherman, D. H. (2009) Structural Basis for Binding Specificitybetween Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol. 4*, 41–52.
(18) Kapur, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. 107*, 22066–22071.

(19) Tang, Y., Kim, C., Mathews, I. I., Cane, D. E., and Khosla, C. (2006) The 2 .7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. 103*, 11124–11129.
(20) Chen, A. Y., Cane, D. E., and Khosla, C. (2007) Structure-Based Dissociation of a Type I Polyketide Synthase Module. *Chem. Biol. 14*, 784–792.
(21) Edwards, A. L., Matsui, T., Weiss, T. M., and Khosla, C. (2014) Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J. Mol. Biol. 426*, 2229–2245.
(22) Klaus, M., D'Souza, A. D., Nivina, A., Khosla, C., and Grininger, M. (2019) Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chem. Biol. 14*, 426–433.
(23) Gokhale, R. S., Hunziker, D., Cane, D. E., and Khosla, C. (1999) Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol. 6*, 117–125.
(24) Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J., and Leadlay, P. F. (2005) Molecular Basis of Celmer's Rules: Stereochemistry of Catalysis by Isolated Ketoreductase Domains from Modular Polyketide Synthases. *Chem. Biol. 12*, 1145–1153.
(25) Wu, J., Kinoshita, K., Khosla, C., and Cane, D. E. (2004) Biochemical analysis of the substrate specificity of the beta-ketoacyl-acyl carrier protein synthase domain of module 2 of the erythromycin polyketide synthase. *Biochemistry 43*, 16301–16310.
(26) Murphy, A. C., Hong, H., Vance, S., Broadhurst, R. W., and Leadlay, P. F. (2016) Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem. Commun. 52*, 8373–8376.
(27) Castonguay, R., He, W., Chen, A. Y., Khosla, C., and Cane, D. E. (2007) Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc. 129*, 13758–13769.
(28) Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C., and Cane, D. E. (2009) The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc. 131*, 18501–18511.
(29) Cane, D. E. (2010) Programming of Erythromycin Biosynthesis by a Modular Polyketide Synthase. *J. Biol. Chem. 285*, 27517–27523.
(30) Zheng, J., and Keatinge-Clay, A. T. (2013) The status of type I polyketide synthase ketoreductases. *Med. Chem. Commun. 4*, 34–40.
(31) Goldsmith, M., and Tawfik, D. S. (2017) Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol. 47*, 140–150.
(32) Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D. L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D. S., and Fleishman, S. J. (2018) Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell 72*, 178–186.
(33) Lowry, B., Robbins, T., Weng, C., Brien, R. V. O., Cane, D. E., and Khosla, C. (2013) In Vitro Reconstitution and Analysis of the 6-Deoxyerythronolide B Synthase. *J. Am. Chem. Soc. 135*, 16809–16812.
(34) Reinke, A. W., Grant, R. A., and Keating, A. E. (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc. 132*, 6025–6031.
(35) Thompson, K. E., Bashor, C. J., Lim, W. A., and Keating, A. E. (2012) SYNZIP Protein Interaction Toolbox: *in Vitro* and *in Vivo* Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synth. Biol. 1*, 118–129.
(36) Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E., and Khosla, C. (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol. 14*, 931–943.
(37) Li, X., Sevillano, N., La Greca, F., Deis, L., Liu, Y. C., Deller, M. C., Mathews, I. I., Matsui, T., Cane, D. E., Craik, C. S., and Khosla, C. (2018) Structure-Function Analysis of the Extended Conformation of a Polyketide Synthase Module. *J. Am. Chem. Soc. 140*, 6518–6521.
(38) Yuzawa, S., Deng, K., Wang, G., Baidoo, E. E. K., Northen, T. R., Adams, P. D., Katz, L., and Keasling, J. D. (2017) Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol. 6*, 139–147.
(39) Ostrowski, M. P., Cane, D. E., and Khosla, C. (2016) Recognition of acyl carrier proteins by ketoreductases in assembly line polyketide synthases. *J. Antibiot. (Tokyo). 69*, 507–510.
(40) Walker, M. C., Thuronyi, B. W., Charkoudian, L. K., Lowry, B., Khosla, C., and Chang, M. C. Y. (2013) Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science 341*, 1089–1094.
(41) Grote, M., Kushnir, S., Pryk, N., Möller, D., Erver, J., Ismail-Ali, A., and Schulz, F. (2019) Identification of crucial bottlenecks in engineered polyketide biosynthesis. *Org. Biomol. Chem. 17*, 6374–6385.
(42) Kalkreuter, E., Crowetipton, J. M., Lowell, A. N., Sherman, D. H., and Williams, G. J. (2019) Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender

Units. *J. Am. Chem. Soc. 141*, 1961–1969.
(43) Eng, C. H., Backman, T. W. H., Bailey, C. B., Magnan, C., García Martín, H., Katz, L., Baldi, P., and Keasling, J. D. (2018) ClusterCAD: A computational platform for type i modular polyketide synthase design. *Nucleic Acids Res. 46*, D509–D515.
(44) Sugimoto, Y., Ishida, K., Traitcheva, N., Busch, B., Dahse, H., and Hertweck, C. (2015) Freedom and Constraint in Engineered Noncolinear Article Freedom and Constraint in Engineered Noncolinear Polyketide Assembly Lines. *Chem. Biol.* 1–12.
(45) Peng, H., Ishida, K., Sugimoto, Y., Jenke-Kodama, H., and Hertweck, C. (2019) Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases. *Nat. Commun. 10*, 3918.
(46) Miyazawa, T., Hirsch, M., Zhang, Z., and Keatinge-Clay, A. T. (2020) An in vitro platform for engineering and harnessing modular polyketide synthases. *Nat. Commun. 11*, 80.
(47) Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-ya, K., and Abe, I. (2017) Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chemie - Int. Ed. 56*, 1740–1745.
(48) Nivina, A., Yuet, K. P., Hsu, J., and Khosla, C. (2019) Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* acs.chemrev.9b00525.
(49) Rittner, A., Paithankar, K. S., Drexler, D. J., Himmler, A., and Grininger, M. (2019) Probing the modularity of megasynthases by rational engineering of a fatty acid synthase Type I. *Protein Sci. 28*, 414–428.
(50) Lowry, B., Li, X., Robbins, T., Cane, D. E., and Khosla, C. (2016) A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS Cent. Sci. 2*, 14–20.
(51) Boivin, S., Kozak, S., and Meijers, R. (2013) Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif. 91*, 192–206.
(52) Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189–1191.

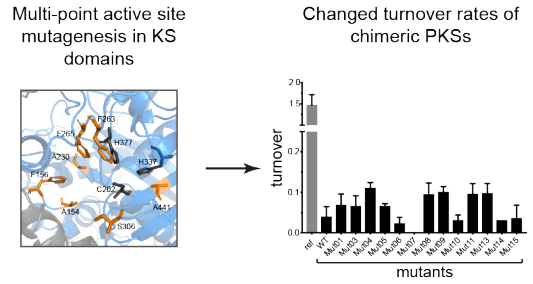



Table of contents graphic