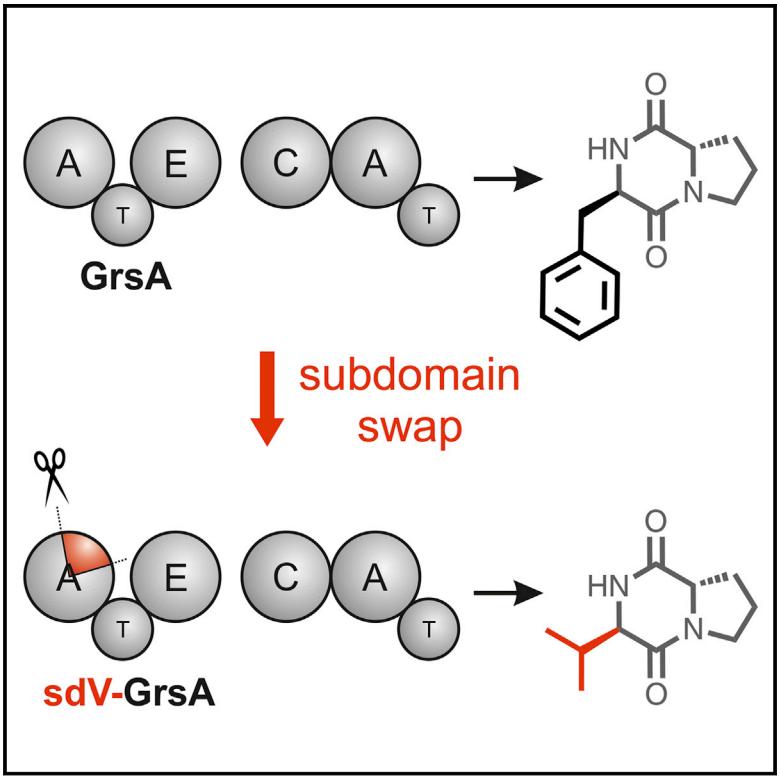
Article

Chemistry & Biology

A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides

Graphical Abstract



Authors

Hajo Kries, David L. Niquille, Donald Hilvert

Correspondence

hilvert@org.chem.ethz.ch

In Brief

Nonribosomal peptide synthetases (NRPSs) produce a plethora of bioactive natural products that can be further diversified by enzyme engineering. Kries et al. successfully employ swapping of short subdomains to transfer specificity from one NRPS module to another. This approach complements other engineering strategies and may facilitate combinatorial biosynthesis of novel peptides.

Highlights

- Specificity of nonribosomal peptide synthetases (NRPSs) is encoded on subdomains
- Subdomain swaps can be used to reprogram the specificity of a dipeptide synthetase
- An engineered construct successfully incorporates valine into a dipeptide
- Transplanting short subdomains may be advantageous for combinatorial screening

Kries et al., 2015, Chemistry & Biology *22*, 640–648
May 21, 2015
http://dx.doi.org/10.1016/j.chembiol.2015.04.015

CellPress

CellPress

Chemistry & Biology
Article

A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides

Hajo Kries,[1,2,3] David L. Niquille,[1,2] and Donald Hilvert[1,*]
[1]Laboratory of Organic Chemistry, ETH Zurich, 8093 Zürich, Switzerland
[2]Co-first author
[3]Present address: Department of Biological Chemistry, The John Innes Center, Norwich NR4 7UH, UK
*Correspondence: hilvert@org.chem.ethz.ch
http://dx.doi.org/10.1016/j.chembiol.2015.04.015

SUMMARY

Nonribosomal peptide synthetases (NRPSs) protect microorganisms from environmental threats by producing diverse siderophores, antibiotics, and other peptide natural products. Their modular molecular structure is also attractive from the standpoint of biosynthetic engineering. Here we evaluate a methodology for swapping module specificities of these mega-enzymes that takes advantage of flavodoxin-like subdomains involved in substrate recognition. Nine subdomains encoding diverse specificities were transplanted into the Phe-specific GrsA initiation module of gramicidin S synthetase. All chimeras could be purified as soluble protein. One construct based on a Val-specific subdomain showed sizable adenylation activity and functioned as a Val-Pro diketopiperazine synthetase upon addition of the proline-specific GrsB1 module. These results suggest that subdomain swapping could be a viable alternative to previous NRPS design approaches targeting binding pockets, domains, or entire modules. The short length of the swapped sequence stretch may facilitate straightforward exploitation of the wealth of existing NRPS modules for combinatorial biosynthesis.

INTRODUCTION

The evolutionary history of proteins provides a rich source of ideas worthy of imitation by protein engineers (Glasner et al., 2007). These ideas extend beyond single point mutations. Protein evolution "one amino acid at a time" (Bloom and Arnold, 2009; Tracewell and Arnold, 2009) alone cannot account for the vast diversity of proteins found in the biological world. In addition to incremental sequence optimization, genetic events that perturb protein folds more drastically present shortcuts to remote areas of the fitness landscape (Grishin, 2001; Lupas et al., 2001). For example, gene fusions, circular permutations, and illegitimate recombinations between unrelated genes (Lupas et al., 2001) mix and match folding units of proteins. Comparisons of bacterial genomes show that protein domains from one species are sometimes found as freestanding proteins in another species (Enright et al., 1999). In other cases, the freestanding progenitors may no longer exist but can be inferred from phylogenetic analyses and structural studies. The widespread $(\beta/\alpha)_8$ fold of TIM barrel enzymes, for instance, is believed to have arisen from fusion of two no longer extant half barrels (Farías-Rico et al., 2014; Lang et al., 2000). Insertion-deletions can also alter protein topology in a dramatic fashion (Grishin, 2001).

Protein engineering can be very successful when designers turn to nature for inspiration. For instance, introduction of point mutations, screening, selection, and recombination in directed evolution experiments is nothing more than a recapitulation of Darwinian evolution in the laboratory (Jäckel and Hilvert, 2010). In analogy to homologous recombination in nature, family shuffling of genes takes advantage of the natural diversity in a protein family to boost activity in laboratory evolution (Crameri et al., 1998; Minshull and Stemmer, 1999). Circular permutation, too, is a tool that has been used for protein engineering, for example to modulate substrate specificity of the ene-reductase old yellow enzyme (Daugherty et al., 2013). In another impressive design effort, a $(\beta/\alpha)_8$ barrel was illegitimately recombined with a $(\beta/\alpha)_4$ motif excised from a flavodoxin fold (Bharat et al., 2008), thereby swapping four strands of the barrel. Only a few mutations were needed to restore stability and ligand binding capability to the artificial fusion protein (Eisenbeis et al., 2012). Mixing and matching of protein fragments has also been successfully applied to multimodular natural product synthetases (Calcott and Ackerley, 2014; Cane et al., 1998; Williams, 2013).

Domains of modular nonribosomal peptide synthetases (NRPSs) (Tanovic et al., 2008), polyketide synthetases, and fatty acid synthetases (Smith and Tsai, 2007) lend themselves to engineering by fragment recombination because they are connected like beads on a string. NRPSs, for example, are multienzyme clusters in which the number and order of domains usually determines the sequence of the synthesized peptide. Adenylation (A), condensation (C), and thiolation (T) domains work together in modules that elongate nascent peptides by successive addition of amino acids. Permutation of NRPS domains and individual modules has shown great promise for the purposeful, combinatorial biosynthesis of novel peptides (Calcott and Ackerley, 2014; Cane et al., 1998; Kries and Hilvert, 2011; Williams, 2013).

In a seminal study, Stachelhaus et al. engineered surfactin A synthetase by transplanting A domains to alter amino acid specificities (Stachelhaus et al., 1995). Module deletions (Mootz et al., 2002), insertions (Butz et al., 2008), and fusions of unrelated

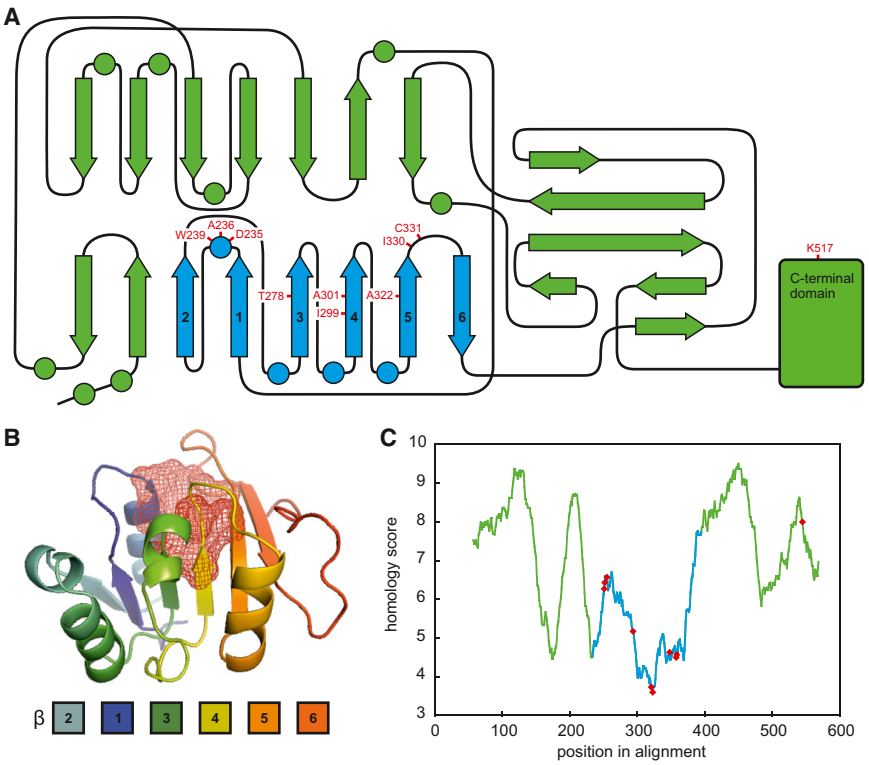


Figure 1. A Flavodoxin-like Subdomain Is Responsible for Substrate Binding in A Domains

(A) Topological map of the A domain fold, adapted from Conti et al. (1997). Circles and arrows symbolize helices and β-strands, respectively. Approximate positions of binding pocket residues are indicated in red. The subdomain of the protein containing all variable binding pocket residues is shown in blue.

(B) Cartoon structure of the subdomain in rainbow colors from the N terminus (blue) to the C terminus (red). The surface of the residues lining the L-Phe binding pocket is shown as a red mesh. The secondary structure elements are characteristic of a flavodoxin-type fold.

(C) Homology score of an alignment of $GrsA_A$ with all A domains from GrsB. The homology score was calculated with jalview (Waterhouse et al., 2009) taking into account the physicochemical similarity of residues (Livingstone and Barton, 1993) and averaged over a window of 30 sequence positions. Binding pocket residues are shown as red diamonds.

modules (Duerfahrt et al., 2003; Mootz et al., 2000) identified additional degrees of freedom for combinatorial biosynthesis. Mechanistic investigations of individual domains helped to circumvent unproductive module combinations (Belshaw et al., 1999). Later, the discovery of small, terminal domains that mediate noncovalent module communication enabled the functional connection of noncognate modules (Hahn and Stachelhaus, 2004).

Cutting and pasting individual domains or entire modules has become standard practice in NRPS engineering (Baltz, 2012; Calcott et al., 2014; Duerfahrt et al., 2003; Mootz et al., 2000, 2002; Schauwecker et al., 2000; Schneider et al., 1998; Stachelhaus et al., 1995). Nevertheless, this approach is not always robust. In the past, the focus has been on changing the order of the beads on the string—essentially the individual C, A, and T domains—because the flexible linkages between them are believed to be rather insensitive to changes (Udwary et al., 2002). However, the bead on a string metaphor is not necessarily the best guide for NRPS engineering. Spatial and temporal coordination of domain interactions is still poorly understood. As a consequence, domain and module swaps often result in low activities (Fischbach et al., 2007; Schneider et al., 1998; Stachelhaus et al., 1995). Alternatively, synthetases can be tailored by fine-tuning individual domains, usually the specificity-determining A domains (Eppelmann et al., 2002; Evans et al., 2011; Kries et al., 2014; Thirlway et al., 2012), but the specificity changes that can be achieved are typically conservative.

Here, we investigate an alternative strategy for modulating A domain specificity in NRPSs that takes advantage of modularity on a subdomain structural level. We identify a compactly folded subdomain of the A domain that encompasses the specificity-determining binding pocket region. Transplantation of such subdomains into a phenylalanine-encoding A domain affords a chimera that functions in peptide formation.

RESULTS

Engineering Strategy

We scanned the $GrsA_A$ adenylation domain from the first module of gramicidin S synthetase for a sequence stretch suitable for specificity engineering of NRPSs. Such a sequence stretch should meet several criteria. First, all residues relevant for substrate recognition in the A domain should be included. Second, it should be short. Third, the tertiary structure should be compact. And fourth, functionally relevant domain interfaces should be avoided.

Inspection of the $GrsA_A$ structure (PDB: 1AMU) (Conti et al., 1997) suggested a 132-residue-long segment encompassing the active site (T221 to I352) as a possible binding subdomain (Figure 1). This protein fragment has a flavodoxin-like topology (Figure 1): a parallel, five-stranded β-sheet sandwiched by α-helices with the first two strands of the sheet in inverted order (Eisenbeis et al., 2012). A sixth strand at the C terminus was included because it is part of the same β-sheet. The flavodoxin fold has been classified as one of nine ancestral protein folds to which all fold diversity observed today can be traced (Caetano-Anollés et al., 2007). NRPS adenylation domains belong to the ANL superfamily of enzymes (Gulick, 2009), which also includes acyl-CoA synthetases and firefly luciferase. It is tempting to speculate that the evolutionary history of the ANL superfamily might have begun with a freestanding flavodoxin-like subunit that later became embedded within a larger scaffold.

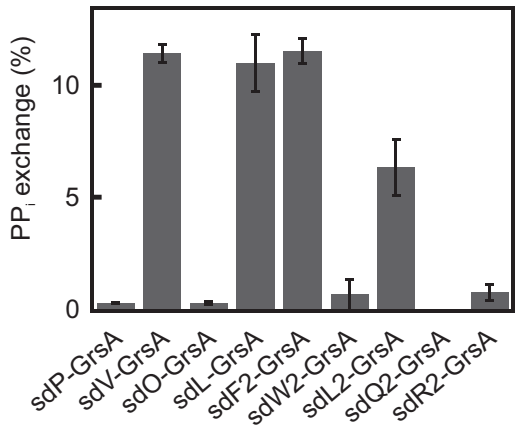


Figure 2. Pyrophosphate Exchange Assay with Subdomain-Swapped GrsA Variants

Exchange between 1 mM ^{32}P-labeled pyrophosphate (PP$_i$) and 5 mM ATP was measured for 2 hr catalyzed by 5 μM enzyme in the presence of 1 mM amino acid substrate. Background without amino acid is subtracted. Error bars indicate the SD of four replicates with one batch of protein. With sdV-GrsA additionally purified by anion exchange chromatography, PP$_i$ exchange increases to 18%.

The sequence of the binding subdomain comprises only 24% of the 558 A domain residues, but includes nine of ten binding pocket residues correlated with substrate specificity (Challis et al., 2000; Stachelhaus et al., 1999). In the absence of information about NRPS dynamics, it is unclear whether the binding subdomain is part of interaction networks relevant for substrate recognition and catalysis. In the only available crystal structure of an entire module (SrfA-C; PDB: 2VSQ) (Tanovic et al., 2008), inevitably a static picture, the binding subdomain does not engage in interdomain contacts. The structure of an A-T bidomain suggests, however, that certain subdomain residues are probably involved in T domain binding (Sundlov et al., 2012).

Subdomain swapping occurred at least once during the evolution of natural NRPSs. Genetic analysis of the NRPS for hormaomycin revealed A domains with nucleotide identities in the range of 90% but different specificities (Höfer et al., 2011). Within these almost identical A domain genes, only a stretch of ca. 400 base pairs (bp) encoding active site residues showed lower identities. This observation was interpreted as the likely result of a genetic recombination event that transferred the specificity-determining portion of one domain into the protein scaffold of another. Indeed, using the sequence boundaries inferred from the natural recombination event (K205 to A322 in GrsA$_A$ numbering), evolution-guided genetic recombination of hormaomycin domains in the laboratory succeeded in transplanting A domain specificity in three of five cases (Crüsemann et al., 2013).

Constructs

The feasibility of exploiting subdomain swaps for NRPS engineering with our structure-guided boundaries was tested with nine constructs based on the D-Phe-encoding initiation module GrsA. In addition to the canonical AT domain, GrsA contains an E domain for epimerization of the amino acid. Subdomains were excised from all four A domains of the second protein from gramicidin S synthetase, GrsB, yielding constructs sdX-GrsA, where X stands for the amino acid specificity of the subdomain (X = P, V, O, L). The recombination partners and GrsA$_A$ had pairwise sequence identities in the range of 44%–48% (Table S1) and showed relatively low sequence conservation in the region of the subdomain (Figure 1C; Table S1). An additional set of five subdomains was obtained from several other organisms, yielding constructs sdX2-GrsA (X = R, Q, L, F, and W). For the second set, recombination partners possessing similarly high levels of identity to GrsA$_A$ as the first set (38%–48%; Table S1) but different specificities were retrieved from an annotated collection of partial A domain sequences (K198 to T334) (Röttig et al., 2011) by a BLAST search (Altschul et al., 1990).

Table 1. Michaelis-Menten Parameters of NRPS Modules

Enzyme	Substrate	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (mM^{-1} min^{-1})
GrsB2	L-Val	100 ± 30	21 ± 3	4.5 ± 0.7
sdV-GrsA	L-Val	50 ± 20	160 ± 30	0.3 ± 0.1
sdV-GrsA	L-Phe	n.d.[a]	n.d.	0.07 ± 0.03

Michaelis-Menten parameters were measured for the adenylation partial reaction in a pyrophosphate exchange assay (Otten et al., 2007). Errors are given as the SD of at least four independent measurements with different batches of protein.

[a]n.d., not determined due to the absence of substrate saturation.

All nine constructs were expressed as phosphopantetheinylated holoenzymes in *Escherichia coli* strain HM0079 (Gruenewald et al., 2004) and purified by NiNTA affinity and anion exchange chromatography. Each construct was isolated as soluble protein in quantities ranging from 5 to 21 mg/l (GrsA, 26 mg/l; Figures S1A and S1B). Successful incorporation of the phosphopantetheine prosthetic group into sdV-GrsA was indicated by electrospray ionization mass spectrometry (ESI-MS) (Figure S2: calc., 128,848 Da; meas., 128,880 Da). Corroborating evidence was obtained by MALDI-tandem mass spectrometry (MS/MS) analysis of the trypsin digested protein, which yielded the mass for the phosphopantetheinylated peptide IGIKDNFYALGGDS(ppant)IK (calc., 2050.992 Da; meas., 2050.869 Da) with the cofactor attached to Ser578, as expected. A fragment with a mass corresponding to the unmodified peptide was not found. Together, these results clearly show that the variant created by subdomain swapping is efficiently phosphopantetheinylated by the genomically encoded Sfp. Circular dichroism (CD) spectra of six constructs, including sdV-GrsA, resemble that of GrsA (Figure S1C). Evidently, despite the heavy mutational load of more than 80 amino acid substitutions and two to four insertions or deletions, structural integrity was not substantially compromised.

Adenylation Activity

We measured adenylation activities of all nine constructs using a pyrophosphate exchange assay (Otten et al., 2007). In an initial screen, sdV-GrsA, sdL-GrsA, sdF2-GrsA, and sdL2-GrsA showed significant pyrophosphate exchange activity stimulated by the respective target amino acid at 1 mM (Figure 2). Kinetic analysis of sdV-GrsA revealed a slightly reduced k_{cat} (50 min^{-1}) for valine and an elevated K_M (160 mM) compared with GrsB2, the source of the subdomain (k_{cat} = 100 min^{-1}; K_M = 21 mM; Table 1). As a consequence, the apparent second-order rate constant k_{cat}/K_M for the chimera was 15-fold lower than for

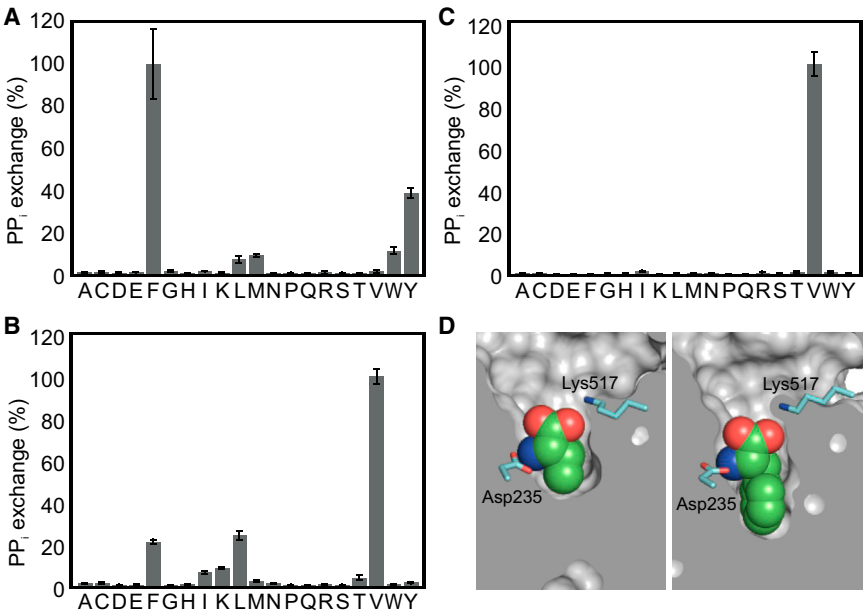


Figure 3. Specificity of sdV-GrsA Is Reminiscent of Wild-Type A Domains

Adenylation kinetics of (A) GrsA, (B) sdV-GrsA, and (C) GrsB2 were measured with all proteinogenic amino acids as substrates (abbreviated in one-letter code) at 1 mM substrate concentration. Reactions with the preferred amino acid substrate were quenched at 10%–15% conversion and all others were normalized to this value. Error bars depict the SD from four measurements with the same enzyme batch.

(D) A homology model of sdV-GrsA$_A$ (left) in comparison with the crystal structure of GrsA$_A$ (right) (Conti et al., 1997). The protein is cut away in the plane of the amino acid binding pocket. Amino acid ligands are shown as spheres (left, L-Val; right, L-Phe; green, carbon; red, oxygen; blue, nitrogen) and protein sidechains as cyan sticks.

GrsB2. Detailed kinetic characterization of other active constructs proved challenging probably due to low k_{cat} values.

In analogous experiments, Crüsemann et al. (2013) obtained active A domains upon swapping subdomains within the hormaomycin NRPS. However, constructs with subdomains from unrelated synthetases showed no activity. The observation of substantial adenylation activity in sdF2-GrsA and sdL2-GrsA (Figure 2) demonstrates that species borders need not be prohibitive for a subdomain swap approach.

Adenylation Specificity

Confronted with a plethora of cellular metabolites, including many amino acids, NRPSs have to be specific, a property generally attributed to A domains (Villiers and Hollfelder, 2009), which activate the appropriate amino acid and covalently tether it to the assembly line. We mapped the substrate specificity of GrsA, sdV-GrsA, and the subdomain donor GrsB2 with all 20 standard amino acids at 1 mM substrate concentration (Figures 3A–3C). The specificity profile shows that the sdV-GrsA chimera is as selective as the native GrsA domain under these conditions (Villiers and Hollfelder, 2009) but less discriminating than GrsB2. L-Val gave the highest activity, as expected by design, followed by L-Leu and L-Phe. However, because the kinetics were measured at a single substrate concentration, this specificity profile is only a snapshot of the complex situation in the cell. The preference of sdV-GrsA for L-Val over L-Phe, in terms of k_{cat}/K_M parameters (Table 1), is roughly four-fold.

A homology model of sdV-GrsA$_A$, constructed on the Swiss-Model server (Biasini et al., 2014) using the crystal structure of the GrsA$_A$ PheA domain (PDB: 1AMU) (Conti et al., 1997) as a template, rationalizes the substrate preferences of the chimera. After manually docking the L-Val substrate and AMP into the structure, the complex was optimized with the Rosetta Relax protocol (Leaver-Fay et al., 2011). In qualitative agreement with the measured substrate profile (Figure 3B), L-Val fits snuggly in the modeled binding pocket, which is contracted compared with GrsA$_A$ (Figure 3D), whereas L-Phe would experience severe steric clashes.

DKP Synthesis

The ability of the subdomain-swapped sdV-GrsA to communicate with other domains and modules was tested in an artificial diketopiperazine (DKP) synthetase consisting of sdV-GrsA and GrsB1, the second module from gramicidin S synthesis (Figure 4A). GrsB1 is a typical proline-specific elongation module with C, A, and T domains. DKP synthetases provide a convenient sensor for peptide formation activity because the required modules are reasonably small and can be expressed and assayed in vitro (Gruenewald et al., 2004; Kries et al., 2014; Stachelhaus et al., 1998). The cyclized DKP products are readily detected by liquid chromatography (LC)-MS.

When combined with GrsB1, sdV-GrsA stimulated significant DKP formation. The expected product, D-Val-L-Pro DKP, was detected by LC-MS along with 25% unepimerized L-Val-L-Pro DKP (Figure 4B). The engineered sdV-GrsA/GrsB1 synthetase incorporated L-Val with a k_{obs} of 0.003 min^{-1} (Figure 4C), 300 times slower than the wild-type GrsA/GrsB1 system, which produces D-Phe-L-Pro DKP with a k_{obs} of 0.9 min^{-1} (Kries et al., 2014). Despite the large structural transition from the native substrate L-Phe to L-Val, all domains remain at least partially functional.

DISCUSSION

Combinatorial biosynthesis of natural products is a powerful strategy employed by natural evolution. The notion that modular machines like NRPSs and PKSs are also a valuable resource for biosynthetic engineering in the laboratory has attracted broad attention over the last two decades (Calcott and Ackerley, 2014; Cane et al., 1998; Hur et al., 2012; Kries and Hilvert, 2011; Williams, 2013). While the chemical logic underlying efficient assembly of polyketides and nonribosomal peptides has been elucidated in detail, modifying these mega-enzymes for

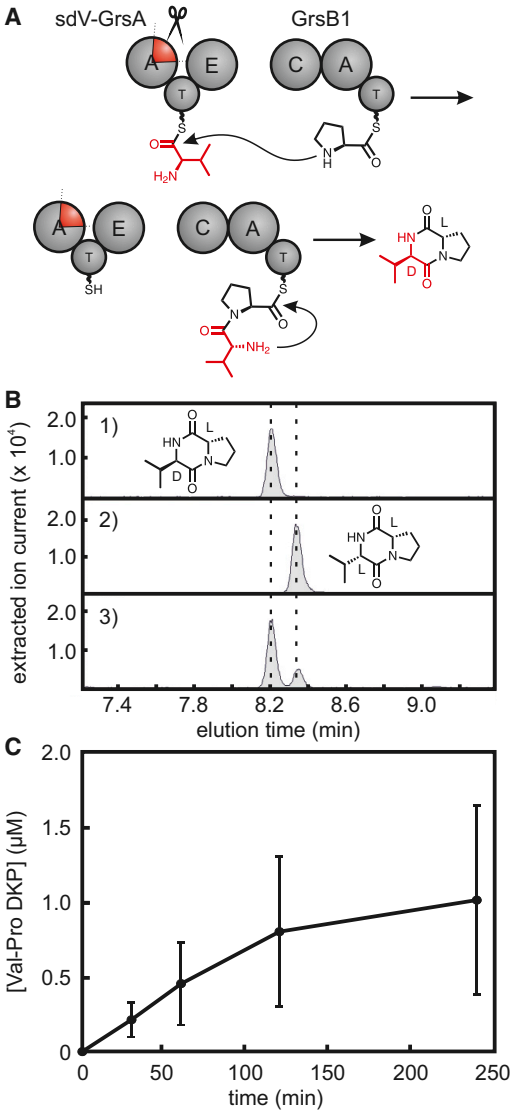


Figure 4. Peptide Synthesis by sdV-GrsA/GrsB1

(A) Mechanism of DKP formation. A, adenylation domain; E, epimerization domain; C, condensation domain; T, thiolation domain.

(B) Extracted ion chromatograms (*m/z* 197.12 ± 0.5) of a synthetic standard for D-Val-L-Pro DKP (1) and L-Val-L-Pro DKP (2), and an enzymatic reaction with sdV-GrsA and GrsB1 (3).

(C) Kinetics of Val-Pro DKP formation monitored by LC-MS. Error bars depict the SD of three independent measurements with three batches of protein. Errors were below 11% when measurements were repeated with the same set of proteins. The large batch-to-batch variation is probably due to impurities in GrsB1, which was obtained in low yield.

the production of novel antibiotics and other drugs has proved challenging (Wilkinson and Micklefield, 2007).

Efforts directed at biosynthetic engineering of natural products should be redoubled for several reasons. First, the number of building blocks for biosynthetic engineering is growing exponentially in the postgenomic era with every microbial genome deposited in a public database. Second, the field of enzyme design and redesign is maturing as increasingly robust methods for the design and optimization of biocatalysts are emerging (Kries et al., 2013). Third, technological innovations are helping to unravel the dynamic processes in natural product synthetases (Dutta et al., 2014; Whicher et al., 2014). Many problems associated with combinatorial biosynthesis today may thus be alleviated in the near future.

What would a reliable platform for the biosynthetic diversification of nonribosomal peptides look like? Juggling genetic fragments several kilobases long is clearly not an easy task. Our findings support the feasibility of an alternative approach based on genetic parts—binding subdomains—that are only ~400 bp long and, in contrast to genes of individual A domains or entire modules, amenable to gene synthesis. Modern sequencing technologies are providing metagenomic data from heterogeneous environmental samples, circumventing the need for monoclonal DNA constructs. As a consequence, sequence databases are continuously growing and include data from organisms that cannot be cultivated (Venter et al., 2004; Wilson and Piel, 2013). Subdomain swaps with synthetic genes provide a potentially powerful means to exploit these databases for combinatorial biosynthesis.

Subdomain exchange in NRPSs is likely to be an ancient evolutionary strategy. Crüsemann et al. (2013) have argued that swapping subdomains by evolutionary recombination events shaped the hormaomycin gene cluster. Based on the presence of a flavodoxin-like folding unit embedded within the A domain structure (Figure 1), we inferred slightly different subdomain boundaries than those deduced for hormaomycin. Both frames are similar but our subdomain stretch is shifted by approximately one secondary structural element toward the C terminus. Interestingly, Crüsemann et al.'s more N-terminal evolution-guided frame does not include two residues of the specificity code proposed by Stachelhaus et al. (1999) (residues 330 and 331 in $GrsA_A$). One of these residues (amino acid 331) is also missing in the specificity code proposed by Challis et al. (2000) because it was argued to point away from the substrate. Adenylation activities and specificity profiles of subdomain-swapped constructs created with the more N-terminal frame indicate that these residues are not necessary for specificity transplantation in every case. Nevertheless, the beta hairpin motif containing these residues directly contacts the substrate, whereas the first helix of the evolution-guided subdomain, which is not included in our structure-guided frame, contains only second-shell residues. The more C-terminal frame is altogether more proximal to the substrate binding pocket. In the future, experimentally comparing different frames from one subdomain source may help delineate the best frame for efficient swaps.

Our results extend those of Crüsemann et al. (2013) and provide a proof-of-principle demonstration that a subdomain swap strategy can yield active and selective A domain chimeras. In particular, our experiments show that such chimeric A domains can function in the context of a dipeptide synthetase. Although the yields are still low, the kinetic parameters determined for sdV-GrsA (Table 1) indicate that adenylation is not the kinetic bottleneck for DKP synthesis: sdV-GrsA has a k_{cat} of 50 min^{-1}, which exceeds DKP formation rates measured with TycA/TycB1 or GrsA/GrsB1 (Belshaw et al., 1999; Gruenewald et al., 2004; Kries et al., 2014) by more than an order of magnitude. Hence, DKP formation in sdV-GrsA/GrsB1 is

probably limited by one of the partial reactions following adenylation.

Even though catalysis in the chimeras we studied was either impaired to some extent or completely destroyed, repair of these constructs might be relatively straightforward. A subdomain swap is expected to perturb only a limited number of residues at the hydrophobic interface between the subdomain and the peripheral structure. Consequently, computational modeling, coupled with laboratory evolution, might be exploited to identify and remove inadvertently created voids and clashes (Evans et al., 2011; Fischbach et al., 2007; Villiers and Hollfelder, 2011).

More detailed guidelines describing the individual strengths and weaknesses of binding pocket, subdomain, domain, and module swap for NRPS engineering should be established in the future based on comprehensive and quantitative data. We speculate that binding pocket mutagenesis will be the preferred instrument to achieve small structural changes in substrate preference (Eppelmann et al., 2002; Evans et al., 2011; Kries et al., 2014; Thirlway et al., 2012) since this approach is least likely to wreak havoc on the structural integrity of the synthetase. However, second-shell and long-range interactions that may be crucial for larger changes in specificity are ignored by this approach. If longer fragments are swapped, more profound changes of the NRPS become feasible albeit at the expense of possible disruptions of the surrounding machinery. Subdomain swapping represents a compromise that allows transplantation of virtually all the selectivity determining residues and minimization of larger disturbances to the enzymatic machinery, but activity loss is still a significant risk. Selectivity filters in downstream domains, for instance the C domain, may cause additional problems.

Further information will be required to ascertain why some swaps yield active enzymes and others do not. The best predictor of success might be sequence identity of the donor and acceptor A domains (Table S1). Structural similarity of the donor and acceptor substrate likely plays a role as well (Figure 2). An interesting question is whether subdomain swaps also transfer binding elements for MbtH-like proteins, which often copurify with A domains and are, in certain cases, required for activity (Felnagle et al., 2010; Zhang et al., 2010). Structural analysis of an A domain bound to an MbtH-like protein shows that the subdomain is far away from the binding interface (Herbst et al., 2013). If binding subdomains do not participate in MbtH binding, their transplantation might be more successful than domain or module exchanges.

SIGNIFICANCE

Our results suggest that a design strategy inspired by the fold architecture of A domains can be a viable alternative to previous NRPS design approaches focusing on domains, modules, and binding pocket mutagenesis. Since binding subdomains are considerably shorter than domains or modules, subdomain swapping could pave the way for NRPS engineering based on bioinformatic searches and gene synthesis. A simplified procedure for combinatorial biosynthesis of nonribosomal peptides based on subdomain swaps may accelerate the development of new peptide drugs in the future.

EXPERIMENTAL PROCEDURES

General Cloning

Procedures for cloning and media preparation were adapted from standard protocols (Russell and Sambrook, 2001). General cloning was carried out in *Escherichia coli* strain XL1-Blue (Stratagene). Microsynth AG synthesized the custom oligonucleotides that were used as PCR primer (Table S2) and performed Sanger sequencing of all inserts amplified by PCR. Restriction enzymes and DNA polymerases were purchased from New England BioLabs Inc. Detailed protocols for plasmid isolation, gel electrophoresis, transformation, PCR, and ligation can be found in the Supplemental Information.

pSU18_grsB2_AT

For the cloning of expression plasmid pSU18_grsB2_AT encoding the AT bidomain of the internal module GrsB2 (Table S3), the gene was amplified by PCR from *Aneurinibacillus migulanus* genomic DNA in two fragments *a* and *b*. Fragment *a* was amplified with primer pair grsb2_a_f/grsb2_EcoRI_del_r. Fragment *b* was amplified with primer pair grsb2_EcoRI_del_f/grsb2_b_r. The primer grsb2_EcoRI_del_f introduces a silent mutation in order to delete an EcoRI site inside the gene. Fragments *a* and *b* were assembled by PCR using primer pair grsb2_a_f/grsb2_b_r. The full-length gene was ligated into the pSU18 vector via the EcoRI and BamHI restriction sites. The resulting plasmid encodes residues K1549 to G2083 of GrsB2 (UniProt: P0C063), an additional N-terminal methionine and a C-terminal –GSRSH$_6$ tag.

Subdomain-Swapped pSU18_mgrsA Constructs

Construction of plasmids pMG211_mgrsAA′, pSU18_mgrsA, and pTrc99a_grsB1 has been described previously (Kries et al., 2014). In plasmids pMG211_mgrsAA′ and pSU18_mgrsA, the subdomain stretch of *grsA*$_A$ is flanked by restriction sites that can be harnessed for exchanging the subdomain. The subdomain encoding sequences of *grsB* were amplified from *A. migulanus* genomic DNA by PCR using primers mXbeg_f and mXend_r, where X indicates the specificity of the subdomain in *grsB* in one-letter amino acid code (X = V, L, O, P). At the 5′-end of the subdomain, the mXbeg_f primer spans the EcoRI restriction site and the beginning of the subdomain. In order to append a stretch between the 3′-end of the subdomain and the SacI restriction site, a second fragment was amplified with primers mend_f and T7TR. The overlapping fragments were assembled by PCR and cloned into pMG211_mgrsAA′ via the EcoRI and SacI sites. From pMG211_mXgrsAA′ constructs, AflII/SacI fragments encoding the subdomain were cut out and cloned into pSU18_mgrsA.

A second generation of subdomain swap variants was constructed with synthetic, codon-optimized gene fragments ordered from ATG:biosynthetics GmbH that encode subdomains for the substrates Phe, Trp, Leu, Gln, and Arg (Table S4). Sequences of these subdomains were retrieved from databases (Table S3).

Protein Production, Purification and Mass Spectrometry

Phosphopantetheinylated proteins were expressed in *E. coli* HM0079 (Gruenewald et al., 2004) and purified by affinity chromatography on NiNTA similar to a previously described procedure (Kries et al., 2014). Bacterial cultures were grown in LB medium containing 20 μg/ml chloramphenicol for pSU18 constructs and 150 μg/ml ampicillin for pTrc99a_grsB1. Protein production was induced by adding 250 μM isopropyl β-D-1-thiogalactopyranoside to 500 ml of culture incubated in a rotary shaker at 37°C and 250 rpm. Cells were harvested by centrifugation after 16–20 hr at 18°C and lysed by sonication in 50 mM Tris-HCl (pH 7.4), 0.5 M NaCl supplemented with 1 mg/ml chicken egg white lysozyme, 1 mM tris(2-carboxyethyl)phosphine (TCEP), and protease inhibitor (Sigma, P8849). The cleared lysate was applied to NiNTA columns. The columns were washed with 50 mM Tris-HCl (pH 7.4), 0.5 M NaCl containing 20 mM imidazole, and 1 mM TCEP and eluted with the same buffer containing 300 mM imidazole. Protein was washed with assay buffer in Amicon Ultra-15 centrifugal filters (30 kDa cut-off, Millipore). Concentrations were measured spectrophotometrically using predicted extinction coefficients at 280 nm. For determining Michaelis-Menten parameters, protein was additionally purified by anion exchange chromatography on Mono Q HR 10/10 columns connected to a Biologic Duo Flow fast protein liquid chromatography system (Bio-Rad Laboratories Inc.). Protein was

eluted with a gradient of 20–400 mM NaCl in 20 mM Tris-HCl (pH 8). The molecular mass of sdV-GrsA was determined after extensive washing with 0.1% acetic acid by ESI-MS on a Bruker maXis ESI-Qq-TOF-MS. For MALDI-MS/MS, the sample was precipitated with 10% trichloroacetic acid, washed with cold acetone, dissolved in 10 mM Tris (pH 8.2) buffer supplemented with 2 mM $CaCl_2$, and digested with 500 ng of trypsin for 34 min at 60°C in a microwave. The resulting peptide fragments were measured on a Bruker Ultraflextreme.

Adenylation Kinetics

Adenylation kinetics were recorded with a 96-well pyrophosphate exchange assay as described previously (Kries et al., 2014; Otten et al., 2007). Reactions were conducted at room temperature (22–24°C) in B&W Isoplate-96 plates (Perkin Elmer) in a volume of 60 µl of buffer (50 mM Tris-HCl [pH 7.4], 10 mM $MgCl_2$, 1 mM TCEP, 0.1 mg/ml BSA) in the presence of 0.1 mM inorganic pyrophosphate (PP_i) labeled with ca. 0.02 µCi ^{32}P-PP_i (Perkin Elmer), 2 mM ATP (Sigma A7699), and varying enzyme and substrate concentrations. In the activity screen (Figure 2), higher concentrations of PP_i (1 mM) and ATP (5 mM) were employed. Reactions were quenched with 60 µl of a 1.6% charcoal suspension in an aqueous solution of 3.5% $HClO_4$ and 3.5% $Na_4P_2O_7$. The charcoal was pelleted by centrifugation and washed twice with 200 µl of water. After resuspending the charcoal in 150 µl of scintillation mix (Optiphase Supermix, Perkin Elmer), scintillation was measured in a Microplate Scintillation & Luminescence Counter TopCount NXT (Perkin Elmer). Steady-state parameters were determined by fitting initial velocities of ^{32}P-PP_i incorporation to the Michaelis-Menten equation in Kaleidagraph (Synergy Software).

CD Spectroscopy

CD of the protein samples was measured on a JASCO J-715 spectropolarimeter from 200 to 260 nm in 0.5-nm steps at a scanning speed of 20 nm/min and a response time of 2 s. Samples were kept at 25°C in a quartz cuvette with a pathlength of 0.2 cm. Protein was diluted in low-salt phosphate buffer (10 mM Na_2HPO_4 [pH 7.5], 10 mM NaCl). A buffer blank was subtracted and curves were smoothed by averaging ellipticities from four wavelengths. Mean residual ellipticities (MRE) were calculated according to the following equation: $MRE = \Theta_{obs}/(10 \times lcn)$, where Θ_{obs} is the ellipticity measured in degrees, l the optical pathlength in cm, c is the molar enzyme concentration, and n is the number of protein residues.

DKP Formation Assay

Module sdV-GrsA was tested for DKP formation in combination with GrsB1 at 5 µM concentration of each protein in 50 mM HEPES buffer supplemented with 10 mM $MgCl_2$, 100 mM NaCl, 5 mM ATP, 1 mM TCEP, and 1 mM L-Val and L-Pro. HEPES buffer, $MgCl_2$, and NaCl were added from a 20-fold stock solution adjusted to pH 8.0. Reactions were conducted at 37°C in a volume of 500 µl. Aliquots of 100 µl were removed at the indicated time points and quenched by heat denaturation for 3 min at 95°C. For LC-MS analysis, protein was removed by centrifugation for 5 min at 16,000 *g*. Samples were analyzed on an Agilent HPLC system (1200) equipped with a 3 × 30 mm Agilent RP-C18 column (3.5 µm particle size) connected to a Bruker maXis ESI-Qq-TOF mass spectrometer. Comparison of peak areas in extracted ion chromatograms to a linear regression curve obtained with an authentic standard allowed quantification of the DKP product (Figure 4C). Diastereomers of Val-Pro DKP were separated on an Xbridge C18 column with 3.5-µm particle size (Figure 4B).

DKP Standards

L-Val-L-Pro DKP was purchased from Bachem (G-4730) as a standard for LC-MS. D-Val-L-Pro DKP was synthesized by standard Boc solid-phase peptide synthesis on a Merrifield resin according to published protocols (Gisin and Merrifield, 1972). All synthesis steps were conducted at room temperature in a 15-ml reaction vessel equipped with a glass frit and a nitrogen inlet for agitation. Boc-protected L-Pro precoupled to Merrifield resin (Bachem, 0.9 µmol/g, 200–400 mesh) was deprotected with trifluoroacetic acid (TFA) in CH_2Cl_2 (1:1, v/v) (2 × 5 ml for 15 min), washed with CH_2Cl_2 (3 × 5 ml for 2 min), neutralized with diisopropylethyl amine (DIPEA) in CH_2Cl_2 (1:19, v/v; 2 × 5 ml for 3 min), and washed again with CH_2Cl_2 (3 × 5 ml for 2 min). Boc-protected D-Val (Bachem) was activated with *N*,*N*′-diisopropylcarbodiimide (1.1 eq.) and DIPEA (3 eq.) for 2 min and coupled to the deprotected L-Pro resin for 4 hr. After alternating washing with CH_2Cl_2 and dimethylformamide (each 3 × 5 ml for 2 min), the coupling step was repeated overnight. The resulting Boc-protected dipeptide was deprotected with TFA in CH_2Cl_2 (1:1, v/v) and washed with CH_2Cl_2 (3 × 5 ml for 2 min). The deprotected dipeptide was cleaved from the resin by cyclization with acetic acid in CH_2Cl_2 (0.1 M) for 1 hr and the resin was washed with CH_2Cl_2. Evaporation of solvent in vacuo yielded pure D-Val-L-Pro DKP as a white solid (88% yield) without further purification. The ^{1}H-NMR spectrum was in good agreement with a published reference (Hendea et al., 2006). ^{1}H-NMR (300 MHz, CD_3OD): δ 4.14 (dd, J = 9.9, 6.5 Hz, 1H), 3.50 (dd, J = 6.1, 0.8 Hz, 1H), 3.59–3.32 (m, 2H), 2.32–2.16 (m, 1H), 2.04 (dq, J = 13.6, 6.8 Hz, 1H), 1.96–1.68 (m, 3H), 0.91 (dd, J = 9.8, 6.8 Hz, 6H). LC-MS, 197.13 *m/z* ($M + H^+$, 197.12 *m/z*).

SUPPLEMENTAL INFORMATION

Supplemental Information includes supplemental protocols, two figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2015.04.015.

AUTHOR CONTRIBUTIONS

H.K., D.N., and D.H. designed the research; D.N. performed the experiments including cloning, protein production, kinetic assays, CD spectroscopy, and chemical synthesis with assistance from H.K.; H.K. performed sequence analyses and computational docking; H.K., D.N., and D.H. analyzed the data and wrote the paper.

ACKNOWLEDGMENTS

We gratefully acknowledge the gift of strain HM0079 from Professor Mohamed A. Marahiel (Philipps Universität Marburg). We thank Professor Hans-Martin Fischer (ETH Zürich) for assistance with the radiochemical experiments. We also thank Louis Bertschi and Dr. Xiangyang Zhang (Mass Spectrometry Service Facility, Laboratory of Organic Chemistry, ETH Zürich) and the Functional Genomics Center Zurich for performing the MS analyses. This work was financially supported by the Studienstiftung des deutschen Volkes (H.K.), the Stipendienfonds der Schweizer Chemischen Industrie (H.K.), the Schweizer Nationalfonds and ETH Zürich.

Received: January 27, 2015
Revised: March 31, 2015
Accepted: April 15, 2015
Published: May 21, 2015

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Baltz, R.H. (2012). Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. ACS Synth. Biol. *3*, 748–758.

Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science *284*, 486–489.

Bharat, T.A.M., Eisenbeis, S., Zeth, K., and Höcker, B. (2008). A βα-barrel built by the combination of fragments from different folds. Proc. Natl. Acad. Sci. USA *105*, 9942–9947.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. *42*, W252–W258.

Bloom, J.D., and Arnold, F.H. (2009). In the light of directed evolution: pathways of adaptive protein evolution. Proc. Natl. Acad. Sci. USA *106*, 9995–10000.

Butz, D., Schmiederer, T., Hadatsch, B., Wohlleben, W., Weber, T., and Süssmuth, R.D. (2008). Module extension of a non-ribosomal peptide synthetase of the glycopeptide antibiotic balhimycin produced by *Amycolatopsis balhimycina*. ChemBioChem *9*, 1195–1200.

Caetano-Anollés, G., Kim, H.S., and Mittenthal, J.E. (2007). The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. Proc. Natl. Acad. Sci. USA *104*, 9358–9363.

Calcott, M.J., and Ackerley, D.F. (2014). Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. Biotechnol. Lett. *36*, 2407–2416.

Calcott, M.J., Owen, J.G., Lamont, I.L., and Ackerley, D.F. (2014). Biosynthesis of novel pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. *80*, 5723–5731.

Cane, D.E., Walsh, C.T., and Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. Science *282*, 63–68.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. *7*, 211–224.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J. *16*, 4174–4183.

Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P.C. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature *391*, 288–291.

Crüsemann, M., Kohlhaas, C., and Piel, J. (2013). Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. Chem. Sci. *4*, 1041–1045.

Daugherty, A.B., Govindarajan, S., and Lutz, S. (2013). Improved biocatalysts from a synthetic circular permutation library of the flavin-dependent oxidoreductase old yellow enzyme. J. Am. Chem. Soc. *135*, 14425–14432.

Duerfahrt, T., Doekel, S., Sonke, T., Quaedflieg, P., and Marahiel, M.A. (2003). Construction of hybrid peptide synthetases for the production of α-L-aspartyl-L-phenylalanine, a precursor for the high-intensity sweetener aspartame. Eur. J. Biochem. *270*, 4555–4563.

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

Eisenbeis, S., Proffitt, W., Coles, M., Truffault, V., Shanmugaratnam, S., Meiler, J., and Höcker, B. (2012). Potential of fragment recombination for rational design of proteins. J. Am. Chem. Soc. *134*, 4019–4022.

Enright, A.J., Iliopoulos, I., Kyrpides, N.C., and Ouzounis, C.A. (1999). Protein interaction maps for complete genomes based on gene fusion events. Nature *402*, 86–90.

Eppelmann, K., Stachelhaus, T., and Marahiel, M.A. (2002). Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochemistry *41*, 9718–9726.

Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011). Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. Chem. Biol. *18*, 601–607.

Farías-Rico, J.A., Schmidt, S., and Höcker, B. (2014). Evolutionary relationship of two ancient protein superfolds. Nat. Chem. Biol. *10*, 710–715.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W., and Thomas, M.G. (2010). MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. Biochemistry *49*, 8815–8817.

Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T., and Liu, D.R. (2007). Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. Proc. Natl. Acad. Sci. USA *104*, 11951–11956.

Gisin, B.F., and Merrifield, R.B. (1972). Carboxyl-catalyzed intramolecular aminolysis. A side reaction in solid-phase peptide synthesis. J. Am. Chem. Soc. *94*, 3102–3106.

Glasner, M.E., Gerlt, J.A., and Babbitt, P.C. (2007). Mechanisms of protein evolution and their application to protein engineering. Adv. Enzymol. Relat. Areas Mol. Biol. *75*, 193–239.

Grishin, N.V. (2001). Fold change in evolution of protein structures. J. Struct. Biol. *134*, 167–185.

Gruenewald, S., Mootz, H.D., Stehmeier, P., and Stachelhaus, T. (2004). In vivo production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. Appl. Environ. Microbiol. *70*, 3282–3291.

Gulick, A.M. (2009). Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem. Biol. *4*, 811–827.

Hahn, M., and Stachelhaus, T. (2004). Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. Proc. Natl. Acad. Sci. USA *101*, 15585–15590.

Hendea, D., Laschat, S., Baro, A., and Frey, W. (2006). Diastereoselective alkylation of a proline-derived bicyclic lactim ether. Helv. Chim. Acta *89*, 1894–1909.

Herbst, D.A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013). Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. J. Biol. Chem. *288*, 1991–2003.

Höfer, I., Crüsemann, M., Radzom, M., Geers, B., Flachshaar, D., Cai, X., Zeeck, A., and Piel, J. (2011). Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. Chem. Biol. *18*, 381–391.

Hur, G.H., Vickery, C.R., and Burkart, M.D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. Nat. Prod. Rep. *29*, 1074–1098.

Jäckel, C., and Hilvert, D. (2010). Biocatalysts by evolution. Curr. Opin. Biotechnol. *21*, 753–759.

Kries, H., and Hilvert, D. (2011). Tailor-made peptide synthetases. Chem. Biol. *18*, 1206–1207.

Kries, H., Blomberg, R., and Hilvert, D. (2013). De novo enzymes by computational design. Curr. Opin. Chem. Biol. *17*, 221–228.

Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D., and Hilvert, D. (2014). Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. Angew. Chem. Int. Ed. Engl. *53*, 10105–10108.

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000). Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. Science *289*, 1546–1550.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., et al. (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. *487*, 545–574.

Livingstone, C.D., and Barton, G.J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput. Appl. Biosci. *9*, 745–756.

Lupas, A.N., Ponting, C.P., and Russell, R.B. (2001). On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? J. Struct. Biol. *134*, 191–203.

Minshull, J., and Stemmer, W.P.C. (1999). Protein evolution by molecular breeding. Curr. Opin. Chem. Biol. *3*, 284–290.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2000). Construction of hybrid peptide synthetases by module and domain fusions. Proc. Natl. Acad. Sci. USA *97*, 5848–5853.

Mootz, H.D., Kessler, N., Linne, U., Eppelmann, K., Schwarzer, D., and Marahiel, M.A. (2002). Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. J. Am. Chem. Soc. *124*, 10980–10981.

Otten, L.G., Schaffer, M.L., Villiers, B.R.M., Stachelhaus, T., and Hollfelder, F. (2007). An optimized ATP/PP_i-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis. Biotechnol. J. *2*, 232–240.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011). NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. *39*, W362–W367.

Russell, D.W., and Sambrook, J. (2001). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor).

Schauwecker, F., Pfennig, F., Grammel, N., and Keller, U. (2000). Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. Chem. Biol. *7*, 287–297.

Schneider, A., Stachelhaus, T., and Marahiel, M.A. (1998). Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. Mol. Gen. Genet. *257*, 308–318.

Smith, S., and Tsai, S.-C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat. Prod. Rep. *24*, 1041–1072.

Stachelhaus, T., Schneider, A., and Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science *269*, 69–72.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J. Biol. Chem. *273*, 22773–22781.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. *6*, 493–505.

Sundlov, J.A., Shi, C., Wilson, D.J., Aldrich, C.C., and Gulick, A.M. (2012). Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. Chem. Biol. *19*, 188–198.

Tanovic, A., Samel, S.A., Essen, L.-O., and Marahiel, M.A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. Science *321*, 659–663.

Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A.-W., Smith, C.P., and Micklefield, J. (2012). Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. Angew. Chem. Int. Ed. Engl. *51*, 7181–7184.

Tracewell, C.A., and Arnold, F.H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr. Opin. Chem. Biol. *13*, 3–9.

Udwary, D.W., Merski, M., and Townsend, C.A. (2002). A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. J. Mol. Biol. *323*, 585–598.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science *304*, 66–74.

Villiers, B.R.M., and Hollfelder, F. (2009). Mapping the limits of substrate specificity of the adenylation domain of TycA. ChemBioChem *10*, 671–682.

Villiers, B., and Hollfelder, F. (2011). Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. Chem. Biol. *18*, 1290–1299.

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.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., et al. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature *510*, 560–564.

Wilkinson, B., and Micklefield, J. (2007). Mining and engineering natural-product biosynthetic pathways. Nat. Chem. Biol. *3*, 379–386.

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

Wilson, M.C., and Piel, J. (2013). Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem. Biol. *20*, 636–647.

Zhang, W., Heemstra, J.R., Jr., Walsh, C.T., and Imker, H.J. (2010). Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. Biochemistry *49*, 9946–9947.

