Chemistry & Biology
Article

Directed Evolution of a Gatekeeper Domain in Nonribosomal Peptide Synthesis

Benoit Villiers[1] and Florian Hollfelder[1,*]
[1]Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK
*Correspondence: fh111@cam.ac.uk
DOI 10.1016/j.chembiol.2011.06.014

SUMMARY

Modular natural products are biosynthesized by series of enzymes that activate, assemble, and process a nascent chain of building blocks. Adenylation domains are gatekeepers in nonribosomal peptide biosynthesis, providing the entry point for assembly of typical peptide-based natural products. We report the directed evolution of an adenylation domain based on a strategy of using a weak, promiscuous activity as a springboard for reprogramming the biosynthetic assembly line. Randomization of residues invoked in a "specificity-conferring code" and selection for a non-native substrate lead to mutant G2.1, favoring smaller amino acids with a specificity change of 10^5: a 170-fold improvement for L-alanine corresponds to a 10^3-fold decrease for its original substrate (L-phenylalanine). These results establish directed evolution as a method to change gatekeeper domain specificity and suggest that adaptation of modules in combinatorial biosynthesis is achievable with few mutations during evolution.

INTRODUCTION

The modular structure of many natural products is reflected in the modular organization of their biosynthetic systems, such as polyketide synthases (PKSs) or nonribosomal peptide synthetases (NRPSs). These enzymatic assembly lines produce a large number of diverse bioactive molecules using a small set of homologous catalytic domains (Finking and Marahiel, 2004; Grünewald and Marahiel, 2006; Meier and Burkart, 2009; Sieber and Marahiel, 2005; Walsh, 2004). Figure 1 illustrates these principles for the biosynthesis of the cyclic decapeptide antibiotic tyrocidine A, in which amino acid building blocks are sequentially incorporated into a nascent peptide chain by a NRPS system. Combinatorial biosynthesis has the ambition to alter the biosynthetic pathway based on this modularity and produce nonnatural analogs of the biosynthesized compounds. Change or alteration of enzymatic modules would lead to a corresponding change in the structural features of an altered natural product.

Two main strategies have been pursued to allow the incorporation of alternative building blocks: (1) swapping of modules (or domains) from other biosynthetic assembly lines or, alternatively, (2) engineering the specificity of individual existing modules (or domains).

In nonribosomal peptide biosynthesis, initial attempts to follow the first strategy have involved adenylation (A)-thiolation (T) didomain exchange achieving novel products, albeit with dramatically reduced product yield (Stachelhaus et al., 1995). The identification of the linker regions between NRPS domains enabled more successful attempts in module fusion (Doekel et al., 2008; Mootz et al., 2000), deletion (Mootz et al., 2002), exchange (Nguyen et al., 2006), and extension (Butz et al., 2008) by genetic engineering. Efforts to redesign NRPS assembly lines also focused on the interactions between partner modules in *trans*, and in some cases matching pairs of short COM domains were shown to control this type of communication (Hahn and Stachelhaus, 2004). Despite progress in the understanding of NRPS interdomain contacts (Frueh et al., 2008; Samel et al., 2007; Tanovic et al., 2008), the engineering of interfaces between domains still represents a considerable challenge due to their unknown modes of association or potentially complex conformational changes. In cases where the rather drastic modification of the synthetase machinery by module swapping still appears difficult, as evidenced by significantly reduced product yield, the second strategy of module adaptation becomes attractive, avoiding invasive modification of protein structure. Potential targets for such an approach are the substructures of NRPSs that have been highlighted as selectivity filters, i.e., the adenylation (A), the condensation (C), and the thioesterase (TE) domains (Finking and Marahiel, 2004; Grünewald and Marahiel, 2006; Sieber and Marahiel, 2005). Specifically A-domains have been targets of protein engineering strategies because the specific activation of amino acid substrates allows them to enter the NRPS assembly line. The definition of a specificity-conferring code (Challis et al., 2000; Stachelhaus et al., 1999) based on A-domain sequence alignments and structural information (Conti et al., 1997) has provided a heuristic method for assignment of A-domain specificity. Unfortunately using this code for redesign is not straightforward and thus far only a few, distinctly conservative specificity changes have been reported, e.g., from L-Phe to L-Leu for GrsA (Stachelhaus et al., 1999), from L-Glu to L-Gln for SrfA-A1 (Eppelmann et al., 2002), and from L-Asp to L-Asn for SrfA-B5 (Eppelmann et al., 2002) and CDAPS2-1 (Uguru et al., 2004). More recently, complex computational redesign has been used successfully to alter specificity (e.g., from L-Phe to L-Leu (Chen et al., 2009) or to L-Tyr (Stevens et al., 2006) for GrsA) probing a larger conformational space.

It was proposed more than a decade ago that directed evolution should play a central role in combinatorial manipulation of

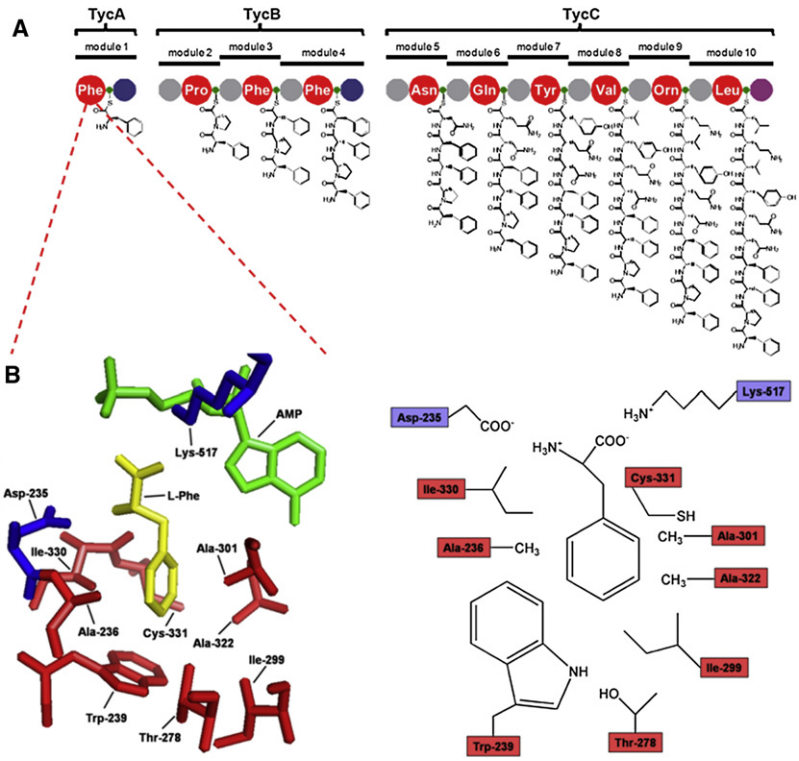

Figure 1. The Tyrocidine Biosynthetic System and the Putative Substrate Binding Interactions in the TycA A-Domain

(A) The enzymatic assembly line of the cyclic decapeptide antibiotic tyrocidine A, consists of three polypeptide chains (TycA, TycB, and TycC) that are composed of one, three, and six modules, respectively, responsible for the incorporation of one monomeric amino acid into the nascent peptide chain. In each module, individual catalytic domains can be assigned: adenylation (A, red), thiolation (or peptidyl-carrier protein, T, green), condensation (C, gray), epimerization (E, blue), and, cyclization/thioesterase (TE, violet) domains (Sieber and Marahiel, 2005). TycA, is composed of an A-, T-, and an E-domain.

(B) The structure of the GrsA A-domain (Conti et al., 1997) is used as a model of the TycA A-domain (based on 79% amino acid sequence homology or 64% amino acid identity, determined by BLAST (Tatusova and Madden, 1999)). The homology modeling of the TycA A-domain to the GrsA A-domain with SWISS-MODEL (Arnold et al., 2006) showed a perfect superposition of the binding pocket residues between both structures. The left panel displays the 3-D structure generated from 1AMU.pdb (Conti et al., 1997) using PyMOL (v0.99). The two conserved A-domain binding pocket residues (Stachelhaus et al., 1999) are shown in blue, the mutated ones in red, the natural amino acid substrate L-Phe in yellow, and the reaction product AMP in green. The right panel shows a 2-D representation of the binding pocket with L-Phe. The labels follow the numbering in GrsA, i.e., Asp-235 in TycA corresponds to Asp-223, Ala-236 to Ala-224, Trp-239 to Trp-227, Thr-278 to Thr-266, Ile-299 to Ile-287, Ala-301 to Ala-289, Ala-322 to Ala-310, Ile-330 to Ile-318, Cys-331 to Cys-319, and Lys-517 to Lys-506.

modular biosynthetic systems (Cane et al., 1998). However, while directed evolution has become a viable alternative to rational protein (re-)design in many other areas of protein science, this approach has not been successfully applied to change the specificity of gatekeeper domains in nonribosomal peptide biosynthesis. Here, we report directed evolution of substrate specificity of this entry point for assembly of a natural product in the A-domain of tyrocidine synthetase 1 (TycA; Figure 1A) (Mootz and Marahiel, 1997), the preferred substrate of which is L-phenylalanine. Randomization of residues previously suggested to be responsible for A-domain specificity (Challis et al., 2000; Stachelhaus et al., 1999) and screening of a mutant library for the ability to activate small amino acids for adenylation (Otten et al., 2007; Villiers and Hollfelder, 2009) lead to the identification of mutants that activate L-alanine efficiently. A very weak promiscuous activity for the new substrate was enhanced 170-fold and the specificity (L-Ala versus L-Phe) altered by 10^5, establishing directed evolution as a method for changing and improving the activity of a gatekeeper domain.

RESULTS

Screening for TycA Mutants that Prefer Smaller Substrates

Library Design and Screening

To limit library diversity, the eight nonconserved binding pocket residues described in a specificity-conferring code (Challis et al., 2000; Stachelhaus et al., 1999) (Figure 1B) were mutated by successive saturation mutagenesis (SM) using an NNS codon (Miyazaki and Arnold, 1999). The classic ATP/PP_i-exchange assay (e.g., employed for the assessment of aminoacyl-tRNA synthetase specificity) (Lee and Lipmann, 1975) was used for the directed evolution of the TycA A-domain substrate specificity. This sensitive assay has been adapted for 96-well format (Otten et al., 2007), allowing up to 400 assays to be carried out per day. In order to limit the screening redundancy, 45 clones were screened for each of the initial eight libraries, statistically ensuring an amino acid completeness of 84% for each position, i.e., 17 of 20 amino acids (Firth and Patrick, 2005) (http://guinevere.otago.ac.nz/cgi-bin/aef/glue-IT.pl). Including the controls, four 96-well plates were screened to cover the eight positions. We chose to screen initially against an amino acid substrate (L-Thr) that is much smaller than the native substrate (L-Phe) (Figure 2) and later also characterized activity against other small, but hydrophobic substrates for selected clones (Table 1). TycA exhibits a weak, yet detectable promiscuous activity toward L-Thr, and small improvements can be reliably detected by the ATP/PP_i-exchange assay. The promiscuous k_{cat}/K_M measured for the wild-type (WT) enzyme is 2×10^6-fold lower for L-Thr (Table 1; see Figure S1 available online), and 5×10^5 lower for L-Ala (the smallest amino acid tested) compared with L-Phe (Table 1).

After the adenylation assay, clones with enhanced activity were sequenced and the relevant mutations were recombined. The activity of the recombined mutant was verified, and used

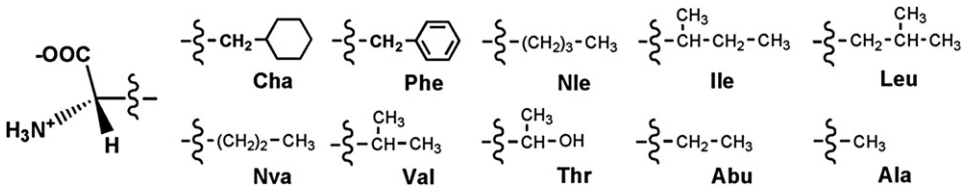

Figure 2. Amino Acid Substrates of the TycA A-Domain

The amino acid substrates are ranked by side-chain volume: β-cyclohexyl-L-alanine (Cha; 117.0 Å^3), L-Phe (Phe; 87.0 Å^3), L-norleucine (Nle; 75.7 Å^3), L-Ile (Ile; 75.7 Å^3), L-Leu (Leu; 75.6 Å^3), L-norvaline (Nva; 58.5 Å^3), L-Val (Val; 58.5 Å^3), L-Thr (Thr; 48.5 Å^3), L-2-aminobutyric acid (Abu; 41.2 Å^3), L-Ala (Ala; 24.3 Å^3). The side-chain volume was calculated with ChemBio3D Ultra 11.0 and corresponds to the Connolly solvent-excluded volume.

as a template for the following round of SM. The strategy for altering substrate specificity was to retain the maximum number of mutations, thus deliberately moving away from WT and potentially decreasing the native activity of TycA toward L-Phe. The directed evolution procedure is schematically shown in Figure 3.

Analysis of the Selected Mutants

The selected TycA mutants were expressed, purified in quantity and kinetically characterized. The Michaelis-Menten curves and the corresponding kinetic parameters of the TycA variants with L-Thr are shown in Figure S1. The amount of soluble protein did not change significantly during the rounds of SM. However, the average activity of the SM libraries significantly increased (Table S1), suggesting adaptability for mutations. All selected mutations appeared beneficial during the screening, except W239M (Figure 3).

After the first round of SM on TycA WT (G0), two mutations were selected: A301C (1.2 ± 0.3-fold increase of catalytic efficiency for G1.1 compared with WT) and C331I (11.5 ± 3.8-fold increase for G1.2 compared with WT). These mutations were recombined in G2 (9.6 ± 4.8-fold increase compared with WT). G2 was used as template for the second round of SM, following which mutants G2.1 containing I330V, with a 12 ± 1.9-fold increase compared with WT, and G2.2 containing W239M with 3.2 ± 1.5-fold increase were selected. Recombination of these mutations in G3 and further rounds of evolution starting from G3 as a template did not give further improvements (Figure S1). Due to the deleterious mutation W239M, the mutants G2.2 and G3 were less active than their respective template G2 and G2.1, but still more active toward smaller substrates than WT (Figures S1 and S2). Thus, the mutant G2.1, which exhibited the most drastic specificity change, was further characterized.

Narrowing of the Substrate Size Exclusion Limit in the TycA Mutant G2.1

To investigate whether the imperative to favor the acceptance of smaller substrates was achieved in the selected mutants, the kinetic parameters of all the TycA variants were determined for three substrates of different side-chain volumes but with similar chemical properties (i.e., hydrophobic) (Figure 2): L-Phe, L-norleucine and L-Ala. Figure 4 displays the evolution of the catalytic efficiency (k_{cat}/K_M) with the introduction of the different mutations in TycA and the corresponding kinetic parameters are shown in Table 1 (the effects of subsequent mutations in the mutants G2.2 and G3 are shown in Figures S2 and Table S2).

Throughout directed evolution cycles the native activity (toward L-Phe) decreased steadily with the accumulation of the selected mutations, eventually to be lowered by three orders of magnitude in mutant G2.1. The effects of mutations A301C and C331I were additive for the activity decrease with the largest substrate tested, L-Phe (i.e., 6- and 30-fold for each and 340-fold for the combination mutant). For the smaller L-norleucine (Figure 2), the activity decrease was less pronounced, i.e., 40-fold in G2.1. Finally, for L-Ala, the smallest substrate tested, the activity increased most strongly, amounting to a 170-fold increase for G2.1. This increase was mostly due to the beneficial effects of mutations A301C and C331I. As the activity toward L-Phe decreased, the combined increases toward L-Ala were nearly additive (i.e., 2- and 50-fold for each and 150-fold for the combination mutant).

To characterize the new exclusion limit for substrate size in mutant G2.1, the kinetic parameters for nine substrates were determined (Table 2). These substrates share hydrophobic character, yet differ in their side-chain volumes (Figure 2). Figure 5 compares the catalytic efficiencies (k_{cat}/K_M) of TycA WT and G2.1 toward these substrates.

Consistent with selection for a smaller substrate the catalytic efficiency (k_{cat}/K_M) in G2.1 decreases for larger substrates (230-fold for β-cyclohexyl-L-alanine; 1000-fold for L-Phe; 37-fold for L-norleucine; 240-fold for L-Ile; 42-fold for L-Leu) and increases for smaller substrates (1.4-fold for L-norvaline; 25-fold for L-Val; 57-fold for L-2-amino butyric acid; 170-fold for L-Ala). For L-Val, L-2-amino butyric acid and L-Ala the activity increases lead to a promiscuous A-domain (Figure 5A) with a catalytic efficiency for adenylation of 3 $min^{-1}mM^{-1}$. This value is comparable to native activities previously reported for other WT A-domains (Eppelmann et al., 2002).

For these three substrates the specificity was changed by 2.6 × 10^4-fold, 6.0 × 10^4-fold, and 1.8 × 10^5-fold compared with the native substrate L-Phe, respectively. These substantial specificity changes enabled G2.1 to display promiscuous activity levels that were comparable to its remaining activity toward L-Phe. The k_{cat} of G2.1 for L-Ala is 2.4-fold higher than its corresponding value for the original substrate L-Phe.

Reduced Solvent Exposure of the Hydrophobic Binding Pocket

Hydrophobic effects have been shown to be key in substrate binding for TycA (Villiers and Hollfelder, 2009). The exposure of the binding pocket to the solvent in the different TycA variants was measured by recording the emission spectra of a non-specific fluorescent hydrophobic probe (SYPRO Orange) (the corresponding experimental procedures are detailed in Supplemental Experimental Procedures). Such extrinsic fluorescent dyes have been used to assess surface hydrophobicity, protein

Table 1. Michaelis-Menten Parameters of the TycA Variants for L-Phe, L-Norleucine, L-Ala, and L-Thr

TycA Variant	Mutations	k_{cat} (min^{-1})				K_M (mM)				k_{cat}/K_M ($min^{-1}mM^{-1}$)			
		Phe	Nle	Ala	Thr	Phe	Nle	Ala	Thr	Phe	Nle	Ala	Thr
WT	–	119 ± 1	62 ± 2	nd	nd	12.0 ± 0.3 × 10^{-3}	9.0 ± 1.0	nd	nd	9887 ± 325	6.8 ± 0.9	18.9 ± 0.4 × 10^{-3}	4.6 ± 0.3 × 10^{-3}
G1.1	A301C	125 ± 4	36 ± 1	7.7 ± 0.9	3.5 ± 0.3	71 ± 10 × 10^{-3}	15.1 ± 1.2	220 ± 50	624 ± 71	1751 ± 293	2.4 ± 0.3	35 ± 12 × 10^{-3}	5.7 ± 1.1 × 10^{-3}
G1.2	C331I	144 ± 6	130 ± 3	303 ± 13	9.9 ± 0.9	0.43 ± 0.04	12.4 ± 1.0	328 ± 23	189 ± 31	337 ± 49	10.5 ± 1.1	0.93 ± 0.11	53 ± 14 × 10^{-3}
G2	A301C	179 ± 3	33 ± 3	317 ± 8	8.9 ± 1.3	6.1 ± 0.3	83 ± 10	109 ± 7	204 ± 53	29 ± 2	0.41 ± 0.08	2.9 ± 0.3	44 ± 19 × 10^{-3}
	C331I												
G2.1	A301C	119 ± 3	nd	283 ± 12	15.8 ± 0.5	12.7 ± 0.9	nd	89 ± 6	286 ± 16	9.4 ± 1.0	0.186 ± 0.006	3.2 ± 0.3	55 ± 5 × 10^{-3}
	C331I												
	I330V												

See also Table S2. nd, not detectable.

denaturation and other protein characteristics (Hawe et al., 2008). The fluorescence emission of this probe increases when it binds to hydrophobic regions of a protein and can thus be used to infer the accessibility for a hydrophobic substrate. In TycA variants, the fluorescence emission decreased steadily with the introduction of mutations (Figure S3) although the character of the selected amino acids led to a small increase in hydrophobicity (Table S3). These data suggest that selections favored binding pockets with reduced accessibility, but not necessarily of more hydrophobic character.

DISCUSSION

Evolution of a New Steric Cutoff Limit

To illustrate and quantitate the kinetic consequences of the new steric cutoff, the difference in catalytic efficiency between the best mutant G2.1 and WT was plotted against the side-chain volume of the substrates (Figure 5B). A sharp transition was observed between large and small substrates, allowing a rather precise determination of the volume where a substrate ceases to be tolerated by the evolved active site. The observed limit of around 60 $Å^3$ is close to the side-chain volume of the substrate used in evolution cycles (Figure 2). The observation of a cutoff in the region of the side-chain volume of the amino acid used for screening, suggests that the steric imperative in selection is directly reflected in the outcome of the evolution experiment. While the steric requirement could be fulfilled, the electronic character of L-Thr, the small substrate used for initial screening, could not be matched by cognate recognition features. The higher charge density in the side chain of L-Thr (compared with an apolar aliphatic amino acid side chain) could apparently not readily be accommodated in the hydrophobic pocket by the emergence of hydrogen bond donors/acceptors, leading to a smaller rate increase.

The selected mutations have contrasting effects on the character of the binding pocket in terms of hydrophobicity and size. The best mutation (C331I) occupies a larger volume, reducing binding pocket size and allowing a tighter fit of the small substrates (e.g., L-Ala). However, a small, polar substrate such as L-Thr benefits less from the tighter fit, suggesting that the increased hydrophobicity in C331I does not support recognition of L-Thr to the same extent as L-Ala.

Evolution appears first to take advantage of the intrinsically more promiscuous substrate binding contributions by entropy-driven hydrophobic effects, while introducing, e.g., hydrogen bonds with their more stringent directionality and distance requirements requires exploration of larger or differently constructed libraries. To alter the character of the binding pocket further by an "adaptive walk" (Orr, 2005) may require more variation than can be programmed by the residues directly forming the active site as identified by the "NRPS code," starting, e.g., with more remote second shell residues.

Even without negative selection the native activity with L-Phe decreased by three orders of magnitude (in G2.1) with the introduction of beneficial mutations for the smaller amino acids. The optimal shape complementarity between L-Phe and the binding pocket was estimated previously to account for two orders of magnitude in catalytic efficiency (Villiers and Hollfelder, 2009) and is lost after mutation leading to a total decrease of 10^3.

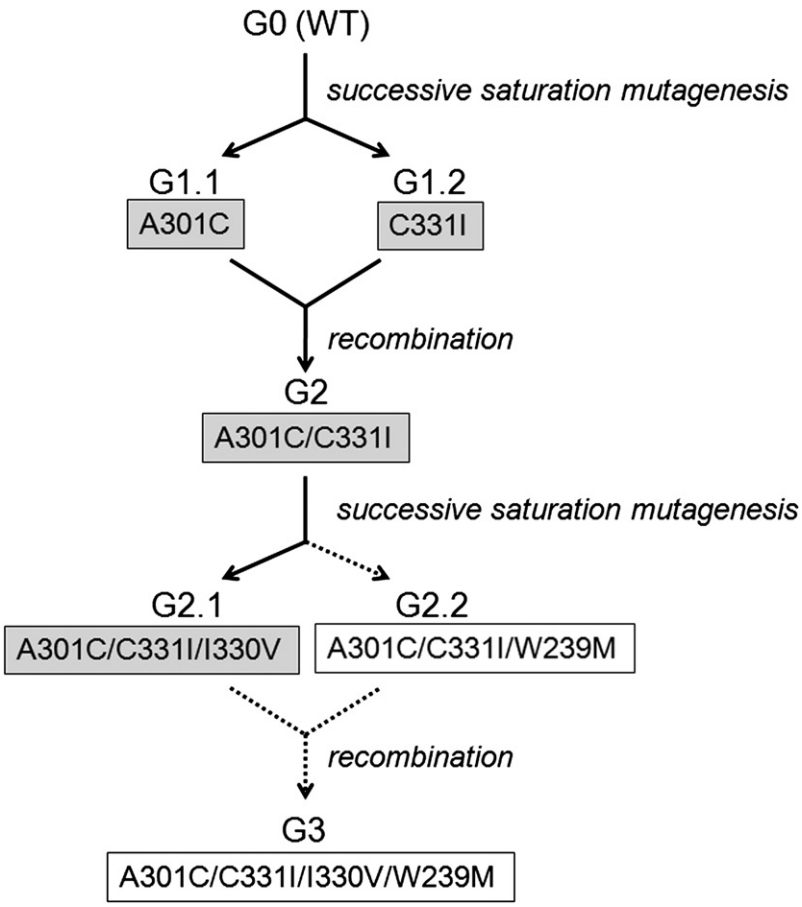

Figure 3. Directed Evolution Procedure

After the first round of successive saturation mutagenesis at the active site positions of TycA WT (G0), two beneficial mutations were selected (A301C and C331I) and recombined to provide the template for the second round (G2: A301C/C331I). After the second round, one beneficial mutation was selected (I330V) to give the best mutant G2.1. The mutation W239M appeared slightly deleterious in the screening of the successive saturation mutagenesis library of G2. It was introduced in the best mutant G2.1 (to obtain G3) in order to replace the key hydrophobic residue Trp-239 by a less hydrophobic and more flexible residue (Met), for future potential activity increases with the screening substrate L-Thr. However, successive saturation mutagenesis on G3 gave no further improvements. The arrows leading to G2.2 and G3 are shown as dotted lines because these variants are less active than their templates G2 and G2.1 (see also Figure S1).

Despite being larger, the initially suboptimal substrate β-cyclohexyl-L-alanine suffers a smaller activity decrease (approximately 200-fold).

A Promiscuous, Generalist A-Domain

While directed evolution has been used once previously to improve the overall activity of chimeric NRPSs 10-fold (Fischbach et al., 2007), this study is the first reported example, to our knowledge, of directed evolution of the specificity of a gatekeeper A-domain. The specificity changes reported here are also larger by one to two orders of magnitude than those achieved recently by a computational approach that appears to lead to strong ground state binding (consistent with a low K_M) (Chen et al., 2009), while our directed evolution is better able to achieve transition state binding. (In the work by Chen et al. a pyrophosphate release assay was used, instead of the pyrophosphate exchange employed here. While relative specificity changes are comparable, the individual kinetic parameters may not strictly refer to the same process.)

We obtained a promiscuous TycA variant (G2.1) showing comparable catalytic efficiency for substrates with very different steric demands, i.e., large (β-cyclohexyl-L-alanine, L-Phe) or small (L-norvaline, L-Val, L-2-aminobutyric acid, L-Ala) (Figures 2 and 5). For a group of substrates relative steric tolerance was observed: L-norvaline (58.5 Å^3), L-Val (58.5 Å^3), L-2-aminobutyric acid (41.2 Å^3), and L-Ala (24.3 Å^3) showed a comparable activity (Figure 5). However, substrates larger than this group (>60 Å^3), such as L-norleucine, L-Ile or L-Leu were subjected to steric discrimination in G2.1 that was absent in WT. This suggests that a new steric cutoff was created as a response to directed evolution for processing smaller amino acids. This cutoff does not work like a rigid sieve, completely disallowing entry of large substrates into the active site: G2.1 still has activity toward the original substrate L-Phe, albeit 10^3-fold reduced. The role of conformational diversity has been demonstrated in the specificity of α-lytic protease (Bone et al., 1989; Miller and Agard, 1999) and invoked more generally (Tokuriki and Tawfik, 2009). In the case of TycA, substrate-induced subtle conformational changes, leading to alteration of specificity by plasticity of the enzyme pocket, would also provide a basis for the observed promiscuity (Villiers and Hollfelder, 2009). Mutants elicited during directed evolution might then stabilize alternative conformations that recognize substrates in different ways from WT. Indeed, a conformational change upon substrate binding has been postulated in the related PheA domain of gramicidin synthetase (Stevens et al., 2006). Although absolute rejection of Phe was not achieved, G2.1 may be *en route* to such an outcome mimicking the promiscuous progenitors that have been hypothesized to exist as evolutionary way stations prior to further specialization and eventual divergence (Khersonsky and Tawfik, 2010; Khersonsky et al., 2006; O'Brien and Herschlag, 1999).

Low donor selectivity of a related condensation domain (e.g., observed between L-Ala and L-Phe; Belshaw et al., 1999) suggests that processing of alternative substrates is possible, but other domains (such as the GrsA epimerization domain) (Luo et al., 2001) may act as additional selectivity filters and limit progress toward a complete natural product. At a minimum a generalist A-domain may be useful to address the specificity of downstream modules that require substrate presentation from an upstream domain, without the need to synthesize intermediates.

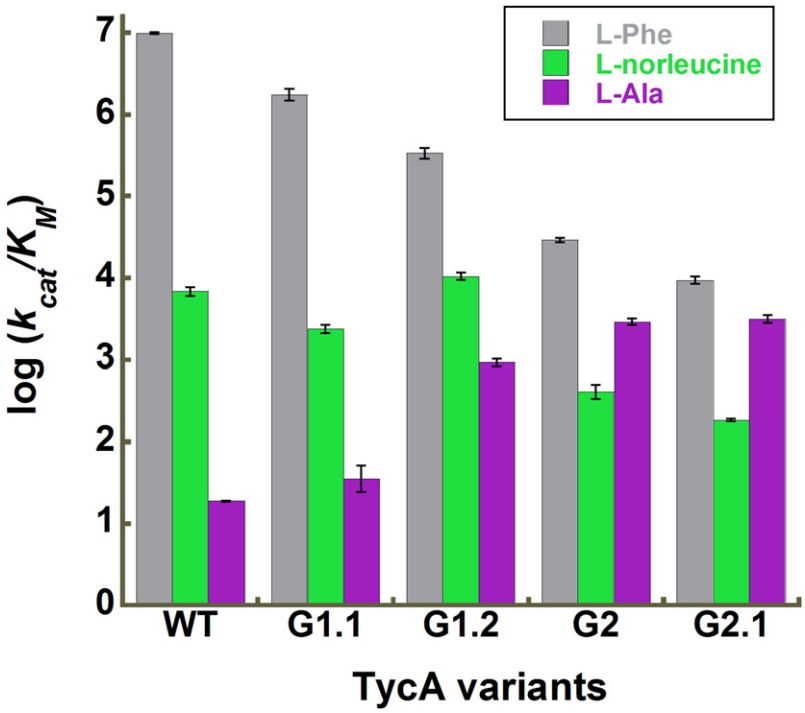

Figure 4. Development of Catalytic Efficiency (log k_{cat}/K_M) in Evolved TycA Variants

The evolution of catalytic efficiency with substrates of different steric properties (L-Phe, L-norleucine and L-Ala; Figure 2) is shown until the best mutant G2.1. The kinetic parameters were determined by ATP/PP_i-exchange assay (Villiers and Hollfelder, 2009) and are shown in Table 1. The catalytic efficiencies of all the TycA variants (including the mutants G2.2 and G3) with the same substrates are shown in Figure S2 and the corresponding kinetic parameters of G2.2 and G3 are shown in Table S2.

The NRPS Code: Prediction versus Evolution

By targeting only the eight nonconserved binding pocket residues of the TycA A-domain, its substrate specificity for small polar (L-Thr) or apolar (L-Val, L-2-amino butyric acid, L-Ala) amino acid side chains was changed by up to five orders of magnitude compared with the original substrate (L-Phe). However, the previous systematic analysis of the NRPS code did not predict the evolutionary outcome. The tolerance to mutations in terms of production of soluble protein or adenylation activity does not correlate with the amino acid residue variability observed in NRPS A-domains at the binding pocket positions. For example, positions 278 and 299, which were suggested to be the most adaptable and crucial in the specificity-conferring codes (Challis et al., 2000; Stachelhaus et al., 1999), were the least tolerant to mutations (measured by the amount of soluble protein produced after the first round of SM, Table S1). This observation suggests that the specificity-conferring residues are not necessarily independently adaptable positions but rather that their potential for functional switches is context dependent. As a consequence, the utility of the specificity-conferring code for altering A-domain specificity appears limited by effects that transcend the residues considered thus far. For example, conservation of second shell residues between a phylogenetically closely-related target and the model A-domain may facilitate a switch, but an extension of this approach may not be successful in phylogenetically more distantly related A-domains.

There are very few studies in which the NRPS code has been used to reprogram A-domain specificity and the quantitative analysis of the improvements suggests that directed evolution is more powerful in achieving this goal. For example, a 10-fold enhancement and 60-fold specificity change has been reported for a switch from L-Phe to L-Leu, a 2-fold improvement and 200-fold specificity change for an L-Asp to L-Asn conversion (Stachelhaus et al., 1999), and a 20-fold improvement and a 380-fold increase in specificity change from L-Glu to L-Gln (Eppelmann et al., 2002). The switch from L-Asp to L-Asn was only achieved by abolishing the activity for the original substrate (Eppelmann et al., 2002).

A comparison of the mutations selected in this study (A301C, C331I, I330V, W239M) with the binding-pocket residues that have been selected by natural evolution and classified in the specificity-conferring codes (Challis et al., 2000; Stachelhaus et al., 1999) is consistent with this view. Of the four selected mutations, only I330V corresponds to an alternative consensus residue described previously in the specificity-conferring codes (Challis et al., 2000; Stachelhaus et al., 1999). Indeed, in sequenced A-domains Val is the most frequently represented residue at position 330 (Stachelhaus et al., 1999). For a more comprehensive analysis, the specificities of the six selected mutants (G1.1, G1.2, G2, G2.1, G2.2, G3) were compared with the predictions from two methods (both available at http://www-ab.informatik.uni-tuebingen.de/software/NRPSpredictor): (1) a transductive support vector machine (TSVM)-based method developed by Rausch et al. (2005) (Table S4), and (2) a phylogenetic method based on the alignment of the ten amino acid code defined by Stachelhaus et al. (1999) to the database of known specificities (Table S5). The similar sequence-based prediction model developed by Challis et al. (2000) (http://nrps.igs.umaryland.edu/nrps/blast.html) could not be used directly since position 331 was not part of the model. The TSVM predictor (Rausch et al., 2005) considers all the binding-pocket residues in a diameter of approximately 8 Å around the substrate amino acid and clusters the extracted amino acid sequences with similar substrate specificities (in terms of physicochemical properties) into composite specificities. For large clusters a larger number of substrate specificities was grouped together whereas for small clusters fewer specificities were combined. The small cluster method has the advantage of allowing more precise predictions, but is less reliable and more phylogenetically dependent due to a limited amount of positive training data. In our case, the large cluster method successfully identified the significant specificity change resulting from the best mutation C331I, which lowered the steric exclusion limit of the binding pocket for better recognition of apolar, aliphatic side chains (Table S4). However, the small cluster method did not allow prediction of this change

Table 2. Michaelis-Menten Parameters of TycA WT and the Evolved Mutant G2.1 for Hydrophobic Substrates with Different Side-Chain Volumes

		Cha	Phe	Nle	Ile	Leu	Nva	Val	Abu	Ala
WT	k_{cat} (min^{-1})	18.8 ± 0.3	119 ± 1	62 ± 2	28 ± 1	13.8 ± 0.3	36 ± 1	13.5 ± 0.2	24 ± 1	nd
	K_M (mM)	83 ± 6 × 10^{-3}	12.0 ± 0.3 × 10^{-3}	9.0 ± 1.0	8.3 ± 0.7	11.0 ± 0.8	53 ± 3	103 ± 4	506 ± 28	nd
	k_{cat}/K_M ($min^{-1}mM^{-1}$)	227 ± 20	9887 ± 325	6.8 ± 0.9	3.4 ± 0.4	1.26 ± 0.12	0.68 ± 0.05	0.130 ± 0.008	47 ± 4 × 10^{-3}	18.9 ± 0.4 × 10^{-3}
G2.1	k_{cat} (min^{-1})	3.8 ± 0.1	119 ± 3	nd	1.04 ± 0.07	1.55 ± 0.12	191 ± 10	534 ± 37	404 ± 18	283 ± 12
	K_M (mM)	3.8 ± 0.2	12.7 ± 0.9	nd	72 ± 8	51 ± 7	200 ± 18	166 ± 20	152 ± 15	89 ± 6
	k_{cat}/K_M ($min^{-1}mM^{-1}$)	1.00 ± 0.06	9.4 ± 1.0	0.186 ± 0.006	14.4 ± 2.6 × 10^{-3}	30 ± 6 × 10^{-3}	0.96 ± 0.13	3.2 ± 0.6	2.7 ± 0.4	3.2 ± 0.3

nd, not detectable. β-cyclohexyl-L-alanine (Cha), L-Phe (Phe), L-norleucine (Nle), L-Ile (Ile), L-Leu (Leu), L-norvaline (Nva), L-Val (Val), L-2-aminobutyric acid (Abu), L-Ala (Ala).

(or its consequences on specificity), probably due to a lack of similar known sequences in nature (Table S5).

SIGNIFICANCE

The range of bioactivities exhibited by nonribosomal peptides is largely dependent on the functional diversity of their building blocks, which encompasses proteinogenic and nonproteinogenic amino acids with various steric and chemical properties. This diversity has been accessible through the natural evolution of the gatekeeper A-domain around a conserved fold (Stachelhaus et al., 1999). Our data from directed evolution experiments suggest that adaptation could have taken place in a minimalist way, requiring very few mutations (Toscano et al., 2007). Library synthesis based on the NRPS code positions allowed adaptation of the new function, despite the relatively small number of mutants screened. This work demonstrates that A-domains (relying on the intrinsically promiscuous hydrophobic effects that are largely driven by entropic contributions) are good starting points for functional diversification (Babtie et al., 2010; Villiers and Hollfelder, 2009). Indeed, a generalist A-domain such as G2.1 that activates multiple amino acids would have enabled activation of distinct building blocks, to potentially access diversity in the natural products produced, assuming tolerant downstream processing. Understanding the evolutionary mechanisms that led to specialization and despecialization in the A-domain substrate specificity repertoire and developing methods to mimic them is a necessary first step to reach the long-term objective of creating nonribosomal peptides at will. The directed evolution approach used here demonstrates how a weak, promiscuous activity can be used as a springboard for generating A-domains with altered specificity and substantial rate enhancements. This work also shows that a library approach can usefully complement previous rational approaches and give insight into the molecular recognition mechanisms involved in substrate recognition. The challenging goal of producing new therapeutically useful bioactive peptides by fermentation, potentially containing unnatural amino acids, will clearly have to involve contributions from both evolutionary and rational strategies.

EXPERIMENTAL PROCEDURES

Mutagenesis

The eight nonconserved binding pocket residues of the TycA A-domain (Figure 1B) were successively mutated with the degenerate codon NNS (32 possible codons for the 20 amino acids) by the QuikChange method (Stratagene). The sequence of the mutagenic oligonucleotides (Operon) used at the different rounds of SM is shown in Supplemental Information. For each of the eight QuikChange reactions, the PCR (50 μl) contained DNA template (30 ng; pSU18-*tycA-PheATE-His* (Grunewald et al., 2004) for the first round of SM, pSU18-*tycA-PheATE-His* A301C/C331I for the second round of SM, and pSU18-*tycA-PheATE-His* A301C/C331I/W239M for the third round of SM), dNTPs (0.2 mM each), the forward and reverse mutagenic oligonucleotides (0.24 μM each; Supplemental Experimental Procedures) and *PfuTurbo* DNA polymerase (2.5 U) in cloned *Pfu* DNA polymerase reaction buffer (Stratagene). The following cycling protocol was used: initial denaturation for 30 s at 95°C; 18 cycles of denaturation for 30 s at 95°C, annealing for 1 min at 55°C and polymerization for 7 min at 68°C. Each of the eight PCR products (50 μl) was incubated (1 hr, 37°C) with DpnI

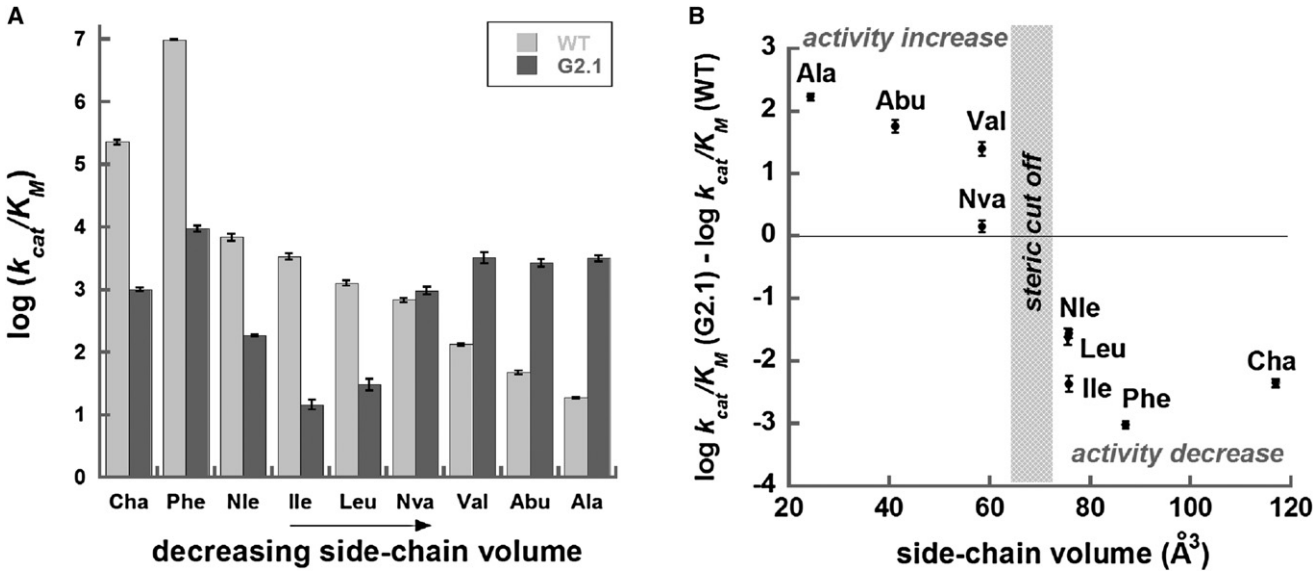

Figure 5. Analysis of the Relationship of Catalytic Efficiency (log k_{cat}/K_M) and Substrate Size before and after Directed Evolution Reveals Tightening of a Steric Cutoff Limit

(A) Overview of the catalytic efficiencies (log k_{cat}/K_M) of WT and mutant G2.1 toward a range of hydrophobic substrates with different side-chain volumes.

(B) Comparison of the gain or loss of activity for mutant G2.1 (in comparison to WT) as a function of substrate side-chain volume. A clear dichotomy between small substrates that are accepted more efficiently and larger substrates that suffer activity loss as a consequence of directed evolution.

The catalytic efficiency of the TycA variants was measured by ATP/PP_i-exchange assay (Villiers and Hollfelder, 2009). The kinetic parameters are shown in Table 2. The structures of the substrates are shown in Figure 2: β-cyclohexyl-L-alanine (Cha), L-Phe (Phe), L-norleucine (Nle), L-Ile (Ile), L-Leu (Leu), L-norvaline (Nva), L-Val (Val), L-2-aminobutyric acid (Abu), L-Ala (Ala). See also Table S3.

restriction enzyme (1 μl, 20 U; NEB). XL1-Blue cells (100 μl) were transformed by heat shock (42°C) with the DpnI digestion product (10 μl) and plated on 2 × YT agar plates containing chloramphenicol (20 μg/ml). After overnight incubation (37°C), the thousands of colonies obtained on each plate were resuspended in 2 × YT medium (5 ml) containing chloramphenicol (20 μg/ml). Plasmid DNA was prepared from each of the eight cell resuspensions with the QIAprep Spin Miniprep Kit (QIAGEN), and the concentration was measured by absorbance (260 nm). The equal representation of the expected nucleotides at the degenerate codons was confirmed by sequencing using the oligonucleotide indicated in Supplemental Experimental Procedures. Eight individual clones were sequenced per library and showed a negligible error rate and no apparent bias, suggesting full coverage of diversity.

After the first round of SM, the selected mutations A301C and C331I were recombined by introducing the mutation A301C in the mutant plasmid pSU18-*tycA-PheATE-His* C331I with the mutagenic oligonucleotides listed in Supplemental Experimental Procedures and the QuikChange method described above. After the second round of SM the mutations I330V and W239M were recombined similarly, by introducing I330V in the mutant plasmid pSU18-*tycA-PheATE-His* A301C/C331I/W239M with the mutagenic oligonucleotides indicated in Supplemental Experimental Procedures.

Production of the Purified SM Libraries in 96-Well Format

For each of the eight SM libraries, BL21(DE3) cells (100 μl) were transformed by heat shock (42°C) with plasmid DNA (approximately 150 ng), and plated on 2 × YT agar plates containing chloramphenicol (20 μg/ml). For each of the eight SM libraries, 45 colonies were picked and grown overnight (750 rpm, 37°C) in a 96-well plate (Nunclon, Nunc) in 2 × YT medium (200 μl) containing chloramphenicol (20 μg/ml). The incubator (Titramax 1000, Heidolph) contained a water-filled box to limit evaporation and maintain the same level of humidity along the cell growth. The plates were sealed with an adhesive seal (Nunc) and a small hole was pierced in each well with a needle to allow homogeneous oxygenation and growth. Each 96-well plate included six wells dedicated to control colonies: five containing the SM template and one containing the empty plasmid (pSU18). Thus, four 96-well plates were required per round of SM.

An aliquot of the overnight culture (100 μl) was added with a multichannel pipette (Rainin) to 2 × YT medium (900 μl) containing chloramphenicol (20 μg/ml) and IPTG (1.1 mM) in a deep 96-well plate (2.2 ml deep well plate, ABgene). The plates were sealed with a heat seal (ABgene) and a heat sealer (Eppendorf). For the same reasons as stated above, a small hole was pierced in each well with a needle. The protein expression was carried out (24 hr, 1350 rpm, 20°C) in the plate incubator stated above, containing a water-filled box.

At this step of the procedure, glycerol 50% (32 μl) was added to each well of the 96-well plate containing the rest of overnight cell culture (approximately 75 μl). The plates were sealed with an adhesive seal (Nunc) and stored at −80°C to conserve the genetic information of the library clones.

After protein expression, the deep 96-well plates were centrifuged (10 min, 4000 rpm, 4°C) and the supernatant was discarded by inverting the plate. Each cell pellet was lysed (30 min, 1350 rpm, 25°C) with a lysing mixture (135 μl Milli-Q H_2O; 15 μl 10× BugBuster reagent, Novagen; 0.15 μl lysonase, Novagen). After the cell lysis, equilibration buffer (850 μl; 50 mM NaH_2PO_4, 300 mM NaCl, pH 8) was added to the cell lysate (approximately 150 μl) and the deep 96-well plates were stored at 4°C for a few hours until the protein purification step.

After centrifugation (10 min, 4000 rpm, 4°C), the supernatants containing soluble protein were transferred (950 μl) with a multichannel pipette and purified in a 96-well plate containing a cobalt-based hexahistidine tag affinity resin (TALON, Clontech), according to the manufacturer's instructions and by using a vacuum manifold (Qiavac 96, QIAGEN). The elution of hexahistidine-tagged protein was initiated with imidazole elution buffer (300 μl; 50 mM NaH_2PO_4, 300 mM NaCl, 150 mM imidazole [pH 7]). The resin was used twice before being regenerated as recommended by the manufacturer with a cobalt solution (50 mM $CoCl_2$). The purified protein solutions could be stored for a few months at 4°C in the elution buffer without significant loss of enzymatic activity.

The relative amount of protein obtained in each well after purification (soluble protein) was determined by Bradford assay. An aliquot (30 μl) of the purified protein solution was added to Bradford reagent (200 μl) in a 96-well plate (Nunc). After a short incubation (approximately 10 min, 900 rpm, 25°C), the OD_{595} was measured with a spectrophotometer (SpectraMax M5, Molecular Devices).

Library Screening by ATP/PP_i-Exchange Assay in 96-Well Format

The reaction (60 μl; 10 μl purified enzyme solution, 10 mM $MgCl_2$, 2 mM ATP, 0.1 mM $Na_4P_2O_7$, 1.3 nCi/μl [^{32}P]PP_i, 15 mM L-Thr for the first round of SM or 2.5 mM L-Thr for the second round of SM or 1.5 mM L-Thr for the third round of SM to guarantee a screening in k_{cat}/K_M conditions) was carried out in assay buffer (50 mM HEPES, 100 mM NaCl, 1 mM EDTA [pH 8]) and initiated by adding the purified enzyme solution (10 μl) to the well of a Black & White 96-well plate (Perkin-Elmer) containing the pre-reaction mixture (50 μl). The reaction was started with 1 min intervals for each column and stopped with Stop Mix (60 μl; 150 mM $Na_4P_2O_7$, 840 mM perchloric acid, 3.6% (w/v) charcoal) with the same intervals, allowing for a 30 min reaction time. Each plate contained one negative control, i.e., the well corresponding to the cells transformed with the empty plasmid (pSU18), four positive controls, i.e., the wells corresponding to the cells transformed with the template plasmid and tested with L-Thr, and one saturation signal control, i.e., a well corresponding to the cells transformed with the template plasmid and tested with the native substrate L-Phe (1 mM).

To remove the excess of [^{32}P]PP_i and measure the amount of [^{32}P]ATP formed during the reaction and adsorbed to charcoal, three washing steps were carried out. Water (130 μl) was added to each well and the plate was centrifuged (3 min, 4000 rpm, 20°C). The fluid containing the excess of [^{32}P]PP_i was removed by inverting the plate. Two additional washing steps were carried out similarly (each with 220 μl water). After the washing, charcoal was resuspended in scintillation fluid (150 μl; Optiphase Supermix, Perkin-Elmer) and the radioactivity was measured after 24 hr of incubation by reading in top mode for one minute per well in a MicroBeta 96-well plate liquid scintillation counter (Perkin-Elmer).

Expression and Purification of the TycA Variants

The production of the hexahistidine-tagged TycA (PheATE-His) variants was performed in BL21(DE3) cells bearing the variants of the plasmid pSU18-*tycA-PheATE-His* (Gruenewald et al., 2004). In a baffled flask (2 liters), 2 × YT medium (500 ml) containing chloramphenicol (20 μg/ml) was inoculated (1:100) with an overnight culture made from a fresh colony of the above strains and grown at 37°C to OD_{600} approximately 0.6. IPTG was added (final concentration of 0.5 mM) and the culture grown for another 4 hr at 30°C. The cells were pelleted (30 min, 4000 rpm, 4°C), resuspended in binding buffer (0.5 M NaCl, 20 mM Tris-HCl, 5 mM imidazole, pH 7.8) and stored at −20°C if necessary. For cell lysis DNaseI (NEB, 0.2 U per ml) was added, the cells were broken using an emulsifier, centrifuged (1 hr, 18,000 rpm, 4°C) and filtered through a filter (0.45 μm). The protein was affinity-purified on a HisTrap FF column on an ÄKTA chromatography system (GE Healthcare) using binding buffer with linearly increasing imidazole concentration (5–250 mM). The imidazole buffer was exchanged against assay buffer by gel filtration (PD-10 column, GE Healthcare) and the pure protein was stored in aliquots at −20°C. Protein concentration was determined by Bradford assay and by measuring the absorbance at 280 nm (for each TycA variant ε_{280} was calculated with ProtParam, an ExPASy tool http://www.expasy.ch/tools/protparam.html).

Kinetic Characterization of the TycA Variants by ATP/PP_i-Exchange Assay in 96-Well Format

The reactions (60 μl) were carried out in the assay buffer (final concentrations: 10 mM $MgCl_2$, 0.7 nCi/μl [^{32}P]PP_i, 2 mM ATP, 0.1 mM $Na_4P_2O_7$; see Supplemental Experimental Procedures for enzyme concentration, substrate concentration, and reaction times). A concentrated solution (3×; 20 μl) of enzyme (containing also $MgCl_2$ and [^{32}P]PP_i in the assay buffer) was used to start the reaction with a multichannel pipette in microtubes (1.2 ml, Greiner Bio One) by adding it to a prereaction mixture (1.5×; 40 μl) containing the substrate, ATP and PP_i in the assay buffer. Enzyme concentration, substrate concentration and reaction times were optimized for every substrate and TycA variant. The reaction was stopped at the different time points by transferring the reaction mixture (60 μl) to Stop Mix (60 μl) in a Black & White plate (Perkin-Elmer). Charcoal was washed and resuspended in scintillation fluid in the same way as described above but radioactivity was measured after 48 hr of incubation. The amount of radiolabeled ATP formed (adsorbed to charcoal) was calculated from counts per minute (CPM) values, corrected for the counts measured in the absence of amino acid substrate. The Michaelis-Menten parameters K_M and k_{cat} were calculated from the initial velocity of time curves with different substrate concentrations by nonlinear least-squares fit to the Michaelis-Menten equation using Kaleidagraph (Villiers and Hollfelder, 2009). For some substrate/TycA variant combinations (L-Thr/WT, L-Ala/WT, L-norleucine/G2.1) (Figure S1; Table 1) no saturation could be observed so the catalytic efficiency (k_{cat}/K_M) was calculated by linear regression for better reliability.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, five tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.chembiol.2011.06.014.

ACKNOWLEDGMENTS

This research was funded by the MRC and a studentship from the EU Early-stage Training Site ChemBioCam to B.V. F.H. is an ERC Starting Investigator. The authors thank several colleagues for comments on the manuscript.

Received: April 22, 2011
Revised: June 18, 2011
Accepted: June 22, 2011
Published: October 27, 2011

REFERENCES

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195–201.

Babtie, A., Tokuriki, N., and Hollfelder, F. (2010). What makes an enzyme promiscuous? Curr. Opin. Chem. Biol. *14*, 200–207.

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.

Bone, R., Silen, J.L., and Agard, D.A. (1989). Structural plasticity broadens the specificity of an engineered protease. Nature *339*, 191–195.

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.

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.

Chen, C.Y., Georgiev, I., Anderson, A.C., and Donald, B.R. (2009). Computational structure-based redesign of enzyme activity. Proc. Natl. Acad. Sci. USA *106*, 3764–3769.

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.

Doekel, S., Coëffet-Le Gal, M.F., Gu, J.Q., Chu, M., Baltz, R.H., and Brian, P. (2008). Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in Streptomyces roseosporus. Microbiology *154*, 2872–2880.

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.

Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides1. Annu. Rev. Microbiol. *58*, 453–488.

Firth, A.E., and Patrick, W.M. (2005). Statistics of protein library construction. Bioinformatics *21*, 3314–3315.

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.

Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T., and Wagner, G. (2008). Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. Nature *454*, 903–906.

Grünewald, J., and Marahiel, M.A. (2006). Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. Microbiol. Mol. Biol. Rev. *70*, 121–146.

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.

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.

Hawe, A., Sutter, M., and Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. Pharm. Res. *25*, 1487–1499.

Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. *79*, 471–505.

Khersonsky, O., Roodveldt, C., and Tawfik, D.S. (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. Curr. Opin. Chem. Biol. *10*, 498–508.

Lee, S.G., and Lipmann, F. (1975). Tyrocidine synthetase system. Methods Enzymol. *43*, 585–602.

Luo, L., Burkart, M.D., Stachelhaus, T., and Walsh, C.T. (2001). Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. J. Am. Chem. Soc. *123*, 11208–11218.

Meier, J.L., and Burkart, M.D. (2009). The chemical biology of modular biosynthetic enzymes. Chem. Soc. Rev. *38*, 2012–2045.

Miller, D.W., and Agard, D.A. (1999). Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease. J. Mol. Biol. *286*, 267–278.

Miyazaki, K., and Arnold, F.H. (1999). Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. J. Mol. Evol. *49*, 716–720.

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.

Mootz, H.D., and Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J. Bacteriol. *179*, 6843–6850.

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.

Nguyen, K.T., Ritz, D., Gu, J.Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R.H. (2006). Combinatorial biosynthesis of novel antibiotics related to daptomycin. Proc. Natl. Acad. Sci. USA *103*, 17462–17467.

O'Brien, P.J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. Chem. Biol. *6*, R91–R105.

Orr, H.A. (2005). The genetic theory of adaptation: a brief history. Nat. Rev. Genet. *6*, 119–127.

Otten, L.G., Schaffer, M.L., Villiers, B.R., 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.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. *33*, 5799–5808.

Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., and Essen, L.O. (2007). Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. Structure *15*, 781–792.

Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. *105*, 715–738.

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

Stevens, B.W., Lilien, R.H., Georgiev, I., Donald, B.R., and Anderson, A.C. (2006). Redesigning the PheA domain of gramicidin synthetase leads to a new understanding of the enzyme's mechanism and selectivity. Biochemistry *45*, 15495–15504.

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.

Tatusova, T.A., and Madden, T.L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. *174*, 247–250.

Tokuriki, N., and Tawfik, D.S. (2009). Protein dynamism and evolvability. Science *324*, 203–207.

Toscano, M.D., Woycechowsky, K.J., and Hilvert, D. (2007). Minimalist active-site redesign: teaching old enzymes new tricks. Angew. Chem. Int. Ed. Engl. *46*, 3212–3236.

Uguru, G.C., Milne, C., Borg, M., Flett, F., Smith, C.P., and Micklefield, J. (2004). Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. J. Am. Chem. Soc. *126*, 5032–5033.

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

Walsh, C.T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science *303*, 1805–1810.