

# An engineered nonribosomal peptide synthetase shows opposite amino acid loading and condensation specificity

Aleksa Stanišić,<sup>1</sup> Annika Hüsken,<sup>1</sup> Philipp Stephan,<sup>1</sup> David L. Niquille,<sup>2</sup> Jochen Reinstein,<sup>3</sup> Hajo Kries<sup>1\*</sup>

<sup>1</sup>Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology (HKI) e.V., Beutenbergstr. 11a, 07745 Jena, Germany

<sup>2</sup>Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, 500 Technology Square NE47-140, Cambridge, Massachusetts 02139, USA.

<sup>3</sup>Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany

**KEYWORDS:** NRPS engineering; subdomain swapping; Dynafit; global fit analysis; enzyme kinetics; directed evolution

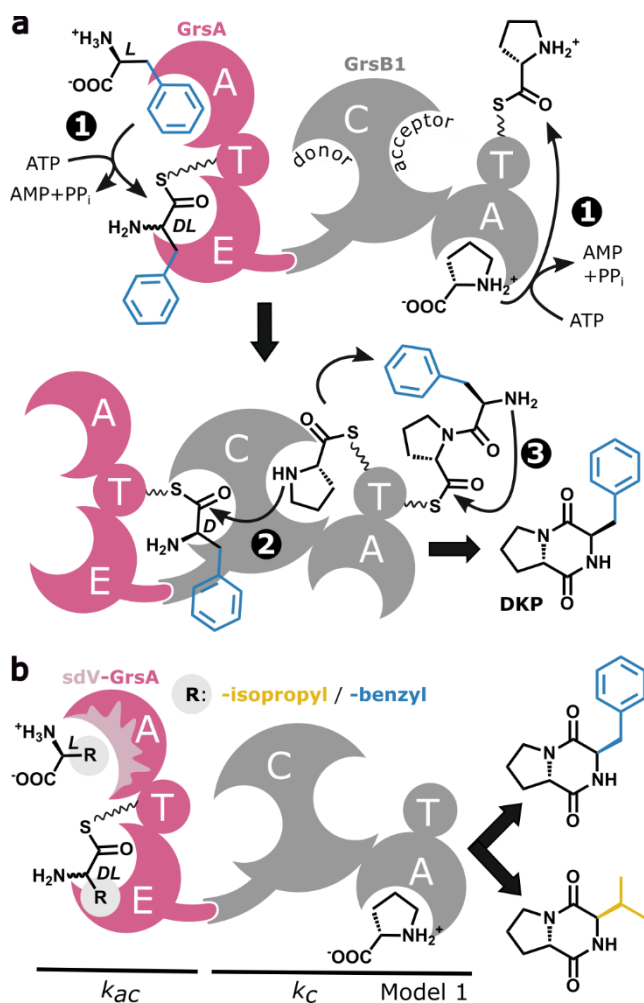
**ABSTRACT:** Engineering of nonribosomal peptide synthetases (NRPS) has faced numerous obstacles despite being an attractive path towards novel bioactive molecules. Specificity filters in the nonribosomal peptide assembly line determine engineering success, but the relative contribution of adenylation (A-) and condensation (C-) domains is under debate. In the engineered, bimodular NRPS sdV-GrsA/GrsB1, the first module is a subdomain-swapped chimera showing substrate promiscuity. On sdV-GrsA and evolved mutants, we have employed kinetic modelling to investigate product specificity under substrate competition. Our model contains one step, in which the A-domain acylates the thiolation (T-) domain, and one condensation step deacylating the T-domain. The simplified model agrees well with experimentally determined acylation preferences and shows that the condensation specificity is mismatched with the engineered acylation specificity. Our model predicts changing product specificity in the course of the reaction due to dynamic T-domain loading, and that A-domain overrules C-domain specificity when T-domain loading reaches a steady-state. Thus, we have established a tool for investigating poorly accessible C-domain specificity through nonlinear kinetic modeling and gained critical insights how the interplay of A- and C-domains determines the product specificity of NRPSSs.

## Introduction

Compared to the revolutionary advances in the development of antibiotics during the second half of the 20<sup>th</sup> century, progress in the field has largely stalled for more than 50 years as the infections caused by multidrug resistant bacterial strains increased worldwide.<sup>1</sup> Nonribosomal peptides (NRPs) have long attracted attention due to their impressive structural diversity and antibiotic activities.<sup>2</sup> Nonribosomal peptide synthetases (NRPSSs) are divided into modules and operate in a linear assembly line fashion, where each module activates, edits and incorporates a single amino acid into the growing peptide chain which is shuttled along the assembly line.<sup>3</sup> In the minimal module needed for elongation, substrate is ATP-activated by the adenylation (A-) domain, tethered to the 5'phosphopantetheine (PPant) arm of the thiolation (T-) domain and condensed by the condensation (C-) domain to the amino acyl or peptidyl intermediate coming from the upstream module. The peptide bond is made in the cleft between the two lobes of the V-shaped C-domain forming two substrate binding sites. From upstream, an acyl-PPant-substrate binds to the C-domain donor site and from the downstream module, the aminoacyl-PPant occupies the acceptor site.<sup>4</sup> Then, the acyl donor is transferred to the acceptor amino group (Figure 1a).

After several elongation steps, terminal thioesterase (TE-) domains release the product. Additional domains frequently tailor products through epimerization, methylation, side-chain cyclization, and other reactions.

Modular structure and straightforward biosynthetic logic of NRPSSs have long inspired engineering attempts aiming at better antibiotics, for instance.<sup>5</sup> NRPS engineering would offer attractive biosynthetic routes towards tailor-made peptides which can be difficult or expensive to produce using standard synthetic or semisynthetic methods. The A-domain acts as a first specificity filter by selecting and activating the substrate before incorporation into peptide by the C-domain.<sup>6</sup> As a consequence, several strategies have been tested to edit the specificity of the A-domain.<sup>7-11</sup> Alternatively, domains and modules have been substituted and reshuffled.<sup>12-15</sup> One key emerging issue is the substrate tolerance of follow-up domains after changing the peptide sequence.<sup>16</sup>



**Figure 1.** a) Acylation of the T-domain of GrsA (pink) and GrsB1 (grey) with *L*-Phe and *L*-Pro, respectively (step 1). *L*-Phe is racemized at the E-domain of GrsA. Condensation of T-domain loaded *D*-Phe and *L*-Pro occurs in the C-domain of GrsB1 (step 2). Spontaneous cyclization at the T-domain of GrsB1 releases *D*-Phe-*L*-Pro diketopiperazine (DKP, step 3). b) Subdomain swapped sdV-GrsA activates two alternative substrates, *L*-Val and *L*-Phe, which are incorporated into DKPs.

It is widely suspected that a second specificity filter at the C-domain is one culprit for non-effective NRPS engineering.<sup>17–19</sup> It was also postulated that the C-domain acceptor site shows more stringent proofreading than the donor site.<sup>20–24</sup> Although the side chain specificity of the C-domain with two enzyme-bound thioesters as substrates is challenging to measure, at least the stereospecificity is well established.<sup>20,23–25</sup> An extreme example of C-domain specificity has been described in glycopeptide antibiotics where the C-domain controls the incorporation of *trans*-modified substrate, despite promiscuous A-domain selection.<sup>26</sup> Considering C-domain specificity adds an additional layer of complexity to NRPS engineering which prompted researchers to preferentially exchange C-A didomains. A novel recombination strategy generating chimeric C-domains has alleviated constraints from A- and C-domain incompatibility when natural NRPS modules were shuffled.<sup>27</sup> However, a specificity code in the C-domain that would be analogous to the powerful A-domain specificity code describing the substrate binding pocket<sup>28,29</sup> remains elusive. Additionally, it

seems that not all C-domains perform stringent proofreading but, in some cases, tolerate alternative substrates well.<sup>30–33</sup> The recently solved crystal structure of a C-domain in complex with the T-domain bound at the acceptor site reveals that a binding pocket which would accommodate the amino acid side-chain is absent.<sup>34</sup> This is in agreement with the outcomes of pyoverdine cluster engineering, which yielded functional chimeras after nonsynonymous A-domain substitutions.<sup>30</sup> Therefore, the importance of C-domain specificity for NRPS engineering is a crucial issue and still under debate.

If C-domains are substrate specific, engineering A-domains alone will result in mismatches and activity losses. By “subdomain-swapping”,<sup>19,35</sup> Kries et al. have generated chimeric initiation module sdV-GrsA<sub>ATE</sub> (subscript: domain architecture; Figure 1b).<sup>8</sup> By minimizing the size of the genetic exchange unit, subdomain swapping constitutes an economic strategy for NRPS specificity transfer. The precursor GrsA, an *L*-Phe activating initiation module from pentamodular gramicidin S synthetase interacts with the excised second module GrsB1<sub>CAT</sub> to generate *D*-Phe-*L*-Pro diketopiperazine (DKP; Figure 1a). The substrate binding A-domain fragment (“subdomain”) from *L*-Val activating GrsB2 was grafted onto GrsA. The resulting chimera, sdV-GrsA, shows designed *L*-Val-preference but also accepts *L*-Phe, and synthesizes Val-Pro and Phe-Pro DKPs with GrsB1 (Figure 1b). Before condensation, sdV-GrsA racemizes loaded amino acids in an epimerization (E-)domain.<sup>25</sup> The natural substrate of the donor site of the GrsB1 C-domain is *D*-Phe, which is condensed with *L*-Pro. Multispecific sdV-GrsA additionally offers *D*-Val for condensation, a noncognate substrate of GrsB1. Dipeptides generated on the T-domain of GrsB1 are released and measured as *D*-Phe-*L*-Pro DKP (DF-DKP; indicating chirality and identity of the first amino acid) and *D*-Val-*L*-Pro DKP (DV-DKP), respectively.

Here we investigate how the mismatch between A- and C-domain specificity affects the designer NRPS sdV-GrsA/GrsB1. Promiscuity of the A-domain in the first module opposed to conserved wild-type specificity of the GrsB1 C-domain creates a unique opportunity to determine the impact of the partial reactions on the overall product preference by non-linear kinetic modelling. Surprisingly, we observed time dependent inversion of product ratios of the engineered NRPS. With a simple model we have extracted rate constants and specificity parameters for amino acid loading and peptide formation from progress curves of sdV-GrsA/GrsB1 and improved variants to illuminate the elusive contribution of C-domains to NRPS specificity.

## Results

**Protein titration.** In the wild type GrsA/GrsB1 system, GrsB1 is clearly rate limiting with a condensation rate of  $1.8 \text{ min}^{-1}$  lagging behind adenylation ( $280 \text{ min}^{-1}$ ),<sup>36</sup> acylation ( $500 \text{ min}^{-1}$ ),<sup>37</sup> and epimerization of *L*-Phe ( $160 \text{ min}^{-1}$ ).<sup>37</sup> Consequently, turnover increases with an excess of the second module, as observed in the closely homologous TycA/TycB1 system.<sup>38</sup> We hypothesized that impaired adenylation in sdV-GrsA might shift this situation. In titration experiments, we investigated the influence of sdV-GrsA/GrsB1 concentration and ratio on the peptide formation rate to find out which module limits the rate. As a standalone initiation module, sdV-GrsA (Val/Phe) interacts with the second module GrsB1 (*L*-Pro) through communication (COM) domains, creating a binary complex. To determine the  $K_d$ , we titrated both enzymes at equimolar concentration (Figure 2a). The titration curve was fitted to a bimolecular binding model to extract an apparent dissociation constant ( $K_{d, \text{app}}$ ) and maximal rate ( $r_{\text{max}}$ , Table 1). The  $K_d$  of 2 to 4  $\mu\text{M}$  measured here is close to that of GrsA and homologous TycB1 (5  $\mu\text{M}$ ) obtained through microscale thermophoresis.<sup>39</sup> The enzyme concentration in subsequent peptide formation assays could not always be saturated because prolonged reactions at concentrations above 2.5  $\mu\text{M}$  showed erratic behaviour, presumably due to instability of sdV-GrsA.

**Table 1. Interaction of sdV-GrsA and GrsB1.\***

Product	$K_{d, \text{app}}$ ( $\mu\text{M}$ )	$r_{\text{max}}$ ( $\text{min}^{-1}$ )
DF-DKP	$4 \pm 1$	$0.032 \pm 0.004$
DV-DKP	$2.0 \pm 0.5$	$0.0088 \pm 0.0008$

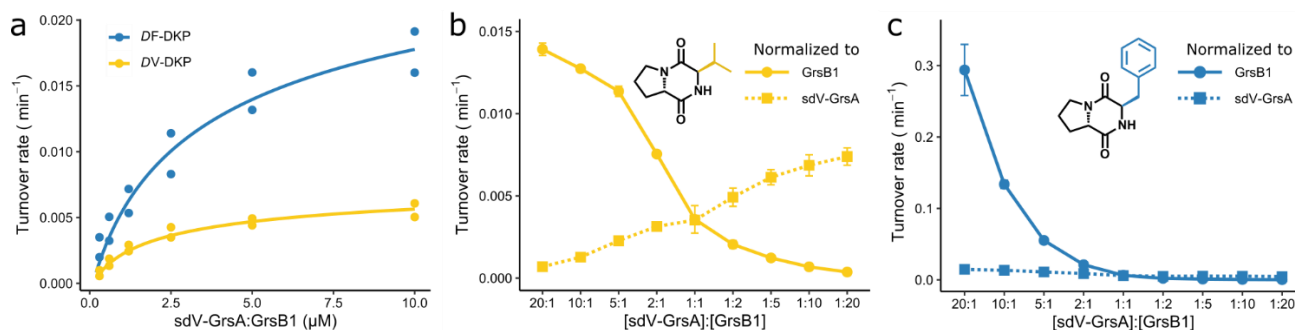
\*Apparent  $K_d$  and maximal rate ( $r_{\text{max}}$ ) were derived from a bimolecular binding model for complex formation of sdV-GrsA and GrsB1 (Figure 2a).

To determine which module limits the peptide formation rate, at fixed 0.5  $\mu\text{M}$  concentration of one module, we titrated the other. Similar maximum rates for DV-DKP formation are reached at a 20-fold excess of sdV-GrsA ( $0.007 \text{ min}^{-1}$ ) or GrsB1 ( $0.014 \text{ min}^{-1}$ ) relative to the less concentrated module (Figure 2b). Hence, both modules seem to

process Val at comparable rates. In contrast, formation of DF-DKP is limited by sdV-GrsA alone, likely because GrsB1 prefers the native donor substrate Phe over Val (Figure 2c).

**sdV-GrsA evolution.** We have previously used directed evolution targeting the A-domain to improve the sdV-GrsA reaction.<sup>40</sup> The resulting variants are characterized here in more detail, because we expected insights into the relationship between A-domain and product formation specificity from subtle differences between the mutants. Mutations were introduced into the subdomain region by reverting amino acid residues to their identities in GrsA, yielding variants enhanced in terms of DV-DKP formation (Supplementary Protocol and Supplementary Figures 3-5). For kinetic profiling in this work, we selected the most active (STAP) and the most selective mutant (MS) by comparing activity and selectivity in a DKP formation assay under *L*-Val/*L*-Phe substrate competition. The selected mutants bear four (STAP) and two (MS) point mutations in a region surrounding the substrate binding pocket of the A-domain (Supplementary Figures 4 and 5). In the DKP formation assay used for screening, the STAP mutant showed a 6-fold increase in activity with slightly lower Val-selectivity (37%) and the MS mutant showed 2-fold higher activity at increased Val-selectivity (91%) compared to sdV-GrsA (54%, Supplementary Figure 6).<sup>40</sup>

**Thermal stability.** Since sdV-GrsA is an unstable, chimeric protein impaired by engineering, we suspected improved structural integrity as a driver of evolutionary improvements. To compare stability of the mutants with sdV-GrsA, we recorded DKP formation at a range of temperatures between 20 and 50 °C (Figure 3a). While sdV-GrsA and the MS mutant have a temperature optimum at 35 °C, the STAP mutant maintains high activity up to 45 °C, suggesting that improved activity in this mutant is due to structural stabilization. To minimize stability issues, we chose 33 °C as a standard temperature for all further measurements. The increased thermal stability of the STAP mutant confirms our prior hypothesis that structural integrity compromised by subdomain swapping can be restored with few mutations.

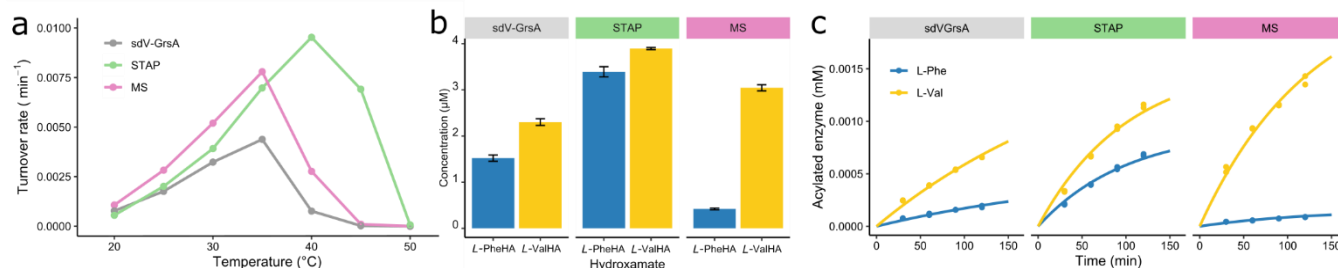


**Figure 2.** a) DV-DKP formation at different equimolar concentrations of sdV-GrsA and GrsB1 (for  $K_{d, \text{app}}$  and  $r_{\text{max}}$  see Table 1). DV-DKP (b) and DF-DKP (c) formation at different ratios of sdV-GrsA and GrsB1. Measured rates are normalized to GrsB1 (circles, full line) or sdV-GrsA (squares, dashed line). All reactions were run for 60 minutes at 33 °C with 1 mM amino acid substrates. Error bars in (b) and (c) indicate the standard deviation from two technical replicates.

**Adenylation and thiolation specificity.** In the directed evolution experiment, mutations were targeted to the A-domain of sdV-GrsA (Supplementary Figures 3-5) and therefore, next to protein stability, expected to mostly affect the rate of the adenylation and thiolation partial reactions catalyzed by this domain. To thoroughly probe the influence of the mutations, we measured saturation kinetics of adenylation, affinity to 5'-O-N-(valyl)sulfamoyladenine (Val-AMS) and Phe-AMS active site inhibitors, substrate specificity profiles of adenylation, and acylation of the T-domain. Saturation kinetics for *L*-Val as a substrate were measured using the MESG/hydroxylamine assay to determine Michaelis-Menten parameters (Table 2 and Supplementary Figure 7).<sup>36</sup> In the stabilized STAP mutant, the adenylation  $k_{cat}$  remains largely unchanged compared to sdV-GrsA (6.0 vs. 8.6 min<sup>-1</sup>) accompanied with a lower  $K_M$  (34 vs. 120 mM). The MS mutant also shows a lower  $K_M$  (51 mM) which is, however, overcompensated by a 10-fold reduction in  $k_{cat}$ . Adenylation of Phe was too slow for the MESG/hydroxylamine assay. To anyway compare preferences for Val and Phe, we determined affinities to the corresponding AMS-type inhibitors which mimic the aminoacyl-AMP intermediate.<sup>6</sup> These affinities were determined with a thermal shift assay<sup>41</sup> which shows two transitions. The first melting temperature ( $T_{m1}$ ) shifts depending on the concentration of AMS inhibitor, while the second ( $T_{m2}$ ) stays almost constant at 57 °C

(Supplementary Figures 8 and 9). The stepwise melting process most likely reflects the multidomain ATE-architecture of the NRPS with a destabilized A-domain ( $T_{m1}$ ) and a stable, native E-domain ( $T_{m2}$ ). From  $T_{m1}$ , we have determined the  $K_d$ 's of the inhibitors which reveal higher affinity for Val-AMS in all enzymes by a factor of 5 (sdV-GrsA), 9 (STAP), and 5 (MS; Table 2). The  $T_{m1}$  of STAP shows a stabilization by 12.1 K relative to sdV-GrsA, in line with the temperature dependence of activity (Figure 3a).

While  $K_d$ 's for AMS inhibitors fail to explain the enhanced DV-DKP formation of the MS mutant, HAMA specificity profiles, which measure hydroxylamine-quenched aminoacyl adenylates, are consistent with this trend (Figure 3b). Both in HAMA and in peptide formation, STAP shows higher activity with almost unchanged specificity, while MS shows higher specificity towards *L*-Val. HAMA, AMS inhibitor binding and the MESG/hydroxylamine adenylation assay inform about the first partial reaction catalysed by the A-domain up to the amino acyl-adenylate. The adenylation partial reaction may behave differently from acylation - comprising both adenylation and thiolation. We measured acylation using Val and Phe, one of which was radioactively labelled with <sup>14</sup>C (Figure 3c and Table 2). Compared to sdV-GrsA (0.028 mM<sup>-1</sup> min<sup>-1</sup>), acylation with Val has been accelerated to 0.062 mM<sup>-1</sup> min<sup>-1</sup> (STAP) and 0.074 mM<sup>-1</sup> min<sup>-1</sup> (MS) in the



**Figure 3.** Evolutionary improvement of sdV-GrsA. a) Thermostability of sdV-GrsA and mutants. DV-DKP formation rate is measured at 5  $\mu$ M enzyme and 1 mM amino acid substrates over 30 min at different temperatures. Error bars indicating the standard deviation from two technical replicates are too small to be visible. b) Adenylation specificity of sdV-GrsA and mutants (5  $\mu$ M) determined with HAMA at 33 °C and 1 mM proteinogenic amino acids (only Phe and Val shown, for full profiles see Supplementary Figure 2). Error bars indicate the standard deviation from three technical replicates. c) Acylation of sdV-GrsA and mutants (2.5  $\mu$ M) with <sup>14</sup>C labelled *L*-Phe and *L*-Val (0.1 mM) under substrate competition. To extract acylation constants ( $k_{acV}$ ,  $k_{acF}$ ), progress data for two technical replicates were fitted to a bimolecular kinetic model with Dynafit (Table 2).

**Table 2. Adenylation, acylation, and inhibitor binding in sdV-GrsA and mutants.**

	Adenylation <sup>§</sup>		Thermal shift assay <sup>§</sup>			HAMA*	Acylation <sup>#</sup>		
	$k_{cat}$ (Val) (min <sup>-1</sup> )	$K_M$ (Val) (mM)	Val-AMS ( $\mu$ M)	Phe-AMS ( $\mu$ M)	$T_{m1}$ (°C)	[ValHA]/ [PheHA]	$k_{acV}$ (mM <sup>-1</sup> min <sup>-1</sup> )	$k_{acF}$ (mM <sup>-1</sup> min <sup>-1</sup> )	$k_{acV}/k_{acF}$
sdV-GrsA	8.6 ± 0.5	120 ± 10	80 ± 10	400 ± 100	33.5 ± 0.4	1.51	0.0281 ± 0.0005	0.0082 ± 0.0005	3.4
STAP	6.03 ± 0.08	34 ± 1	54 ± 5	500 ± 100	45.6 ± 0.3	1.15	0.062 ± 0.002	0.037 ± 0.001	1.7
MS	0.85 ± 0.07	51 ± 9	39 ± 2	200 ± 30	36.3 ± 0.8	7.23	0.074 ± 0.001	0.0050 ± 0.0009	14.8

<sup>§</sup>Kinetic parameters of *L*-Val-adenylation determined with the MESG/NH<sub>2</sub>OH assay. Error margins are obtained from a nonlinear fit with technical duplicates to the Michaelis-Menten equation in R. <sup>§</sup>Dissociation constants of AMS-type inhibitors determined with the thermal shift assay (Supplementary Figures 8 and 9) using a hyperbolic binding model. Error margins are determined from the error of the nonlinear fit. Melting temperatures are given for the first transition ( $T_{m1}$ ) in the absence of inhibitor, with the standard deviation as error margin. Experiments were done with two batches of enzyme in technical triplicates. \*Ratios of hydroxamates (Figure 3b). <sup>#</sup>Experimental acylation rate constants determined with the <sup>14</sup>C assay and ratios for *L*-Val ( $k_{acV}$ ) and *L*-Phe ( $k_{acF}$ , Figure 3c).



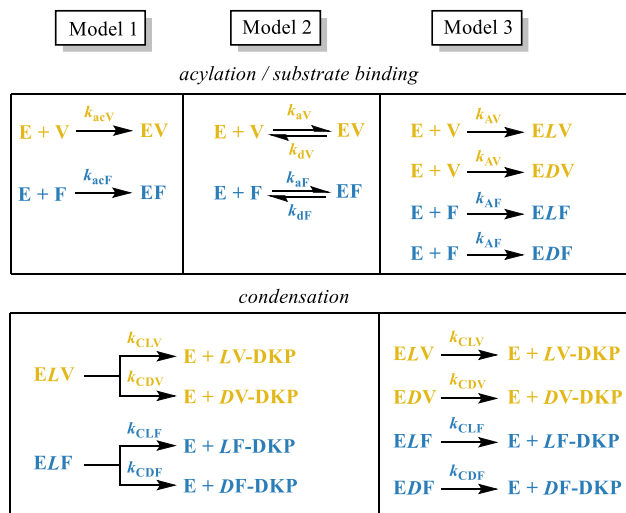
mutants. At the same time, Val/Phe specificity slightly decreased from 3.4-fold in sdV-GrsA to 1.7-fold in the STAP mutant but increased 15-fold in the MS mutant. These values show the same trend as HAMA specificities (Figure 3b). Apparently, the adenylation reaction and not the thiolation reaction is mostly responsible for the differences in activity and specificity of STAP and MS mutants. There is, however, a small trend towards higher Val-specificity at the acylation stage which is most pronounced with the MS mutant and which may indicate a contribution of thiolation to Val-specificity.

**Inversion of product preference.** Compared to adenylation (HAMA) and acylation specificity, peptide formation by sdV-GrsA/GrsB1 in the presence of competing substrates (1:1 *L*-Val and *L*-Phe) shows lower Val-incorporation (Figure 4a). Surprisingly, the incorporation ratio is not even constant over time. Peptide production begins with 3-fold Phe- but ends with slight Val-preference. Substrate depletion cannot account for this effect because substrates are present in large excess. We hypothesized that the inversion might occur due to crosstalk between the A-domain of sdV-GrsA which is weakly Val-specific, and the C-domain of GrsB1 that we assume to have a Phe-specific donor site. Notably, the assembly line architecture of NRPSs allows substrate competition only at the adenylation step, while subsequent steps channel intermediates covalently bound to the NRPS. We explain the inversion of product ratios by variable T-domain loading: the chimeric A-domain loads the T-domain with either Phe or Val and slightly favours Val. However, the C-domain of GrsB1 preferentially consumes *D*-Phe, while *D*-Val stalls on the T-domain. When Phe-loaded sdV-GrsA becomes deacylated by GrsB1, the replacement will more likely be Val than Phe. Hence, the population of Val-loaded sdV-GrsA in the assay grows over time. This scenario explains the slowing of *DF*-DKP and acceleration of *DV*-DKP formation, since the C-domain is increasingly forced to accept Val stalled on the T-domain. The proportion of Val and Phe on the T-domain stays steady once they are loaded and unloaded at the same ratio. Since the loading ratio is determined by A-domain preference, unloading by the C-domain must follow suit. In other words, the A-domain alone determines the product ratio once the steady-state of T-domain loading has been reached.

**Kinetic model of peptide formation.** To test our hypothesis of variable T-domain loading, we have numerically fit reaction progress data with three simplified kinetic models (Scheme 1).<sup>42,43</sup> Progress curves were fit to these models using Dynafit.<sup>44</sup> The three models differ in the equations representing the initial catalytic steps of the NRPS. In model 1, acylation is irreversible, described by the bimolecular rate constants  $k_{acV}$  and  $k_{acF}$ . Model 2 describes reversible association of substrates with the enzyme with binding constant  $K_a = k_a/k_d$ , where binding is assumed to be much faster than the other steps and arbitrarily fixed at a rate of  $10^6 \text{ mM}^{-1} \text{ min}^{-1}$ . Condensation steps ( $k_c$ ) lead from the acyl enzyme intermediate to the peptide product. In models 1 and 2, but not in model 3, the condensation step includes the epimerization step. The Dynafit software numerically integrates the corresponding systems of first-order differential equations (Supplementary Information) and performs least square regression to obtain rate constants. To challenge the model

and increase the reliability of predicted kinetic constants, we recorded time courses at varying *L*-Val and *L*-Phe ratios which were globally fit to the three models. The MS and STAP mutants of sdV-GrsA with altered adenylation properties were tested, too, to confirm the changes in adenylation specificity (Table 2). For simplicity, at first, we assumed that the first (sdV-GrsA) and the second module (GrsB1) act as a functional unit and kinetically modelled them as a single enzyme (E).

### Scheme 1. First generation kinetic models.\*



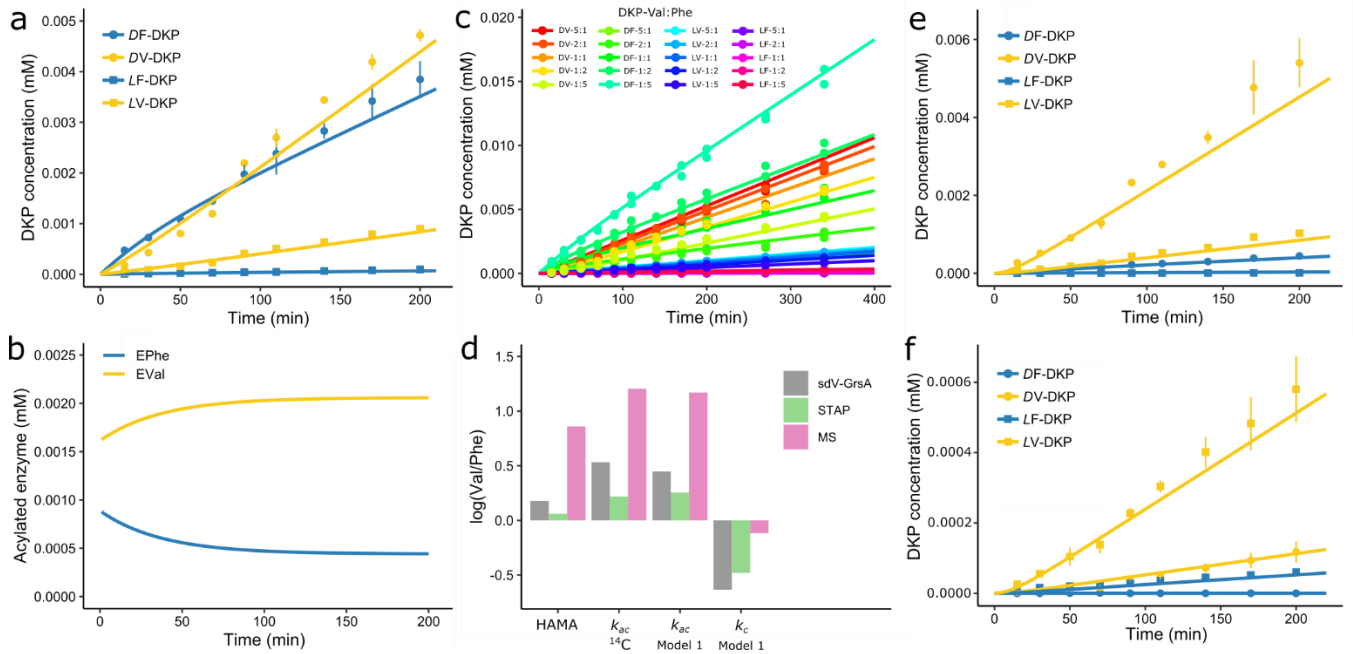
\*E: sdV-GrsA/GrsB1 complex; V: *L*-Val; F: *L*-Phe.

Models 1 and 2, but not model 3, successfully fit the experimental data and yield similar values for condensation constants ( $k_c$ ) in all three enzymes (Figure 4c, Supplementary Table 1). Indeed, the concentration of Val-acylated enzyme increases over time, confirming the hypothesis of dynamic T-domain loading (Figure 4b). Absolute values for the acylation rates in model 1, which most closely reflects our understanding of the NRPS mechanism, failed to be defined. However, the Val/Phe preference ( $k_{acV}/k_{acF}$ ) was calculated from parameters obtained with a Monte Carlo algorithm and aligned well with experimental acylation rates (Figure 4d). Therefore, we created an improved version of model 1 (model 1B, Scheme 2) where we fixed  $k_{ac}$  to the experimentally determined values. Furthermore, the enzyme concentration in the assays falls in the range of the  $K_d$  of sdV-GrsA/GrsB1 (Table 1), which influences the magnitude of  $k_c$ . Therefore, model 1B was amended with equilibria for the interaction of GrsB1 with the acylated and unacylated first module, where the equilibrium constant was fixed to the experimental value.

Model 1B, integrating experimental acylation rates and enzyme dissociation equilibria, successfully fits the data, too (Table 3). As expected,  $k_c$  values are larger compared to model 1, due to incomplete module dimerization which is now accounted for. Experimental acylation constants, which were previously not fully defined, are compatible with the fit. This confirms that the kinetic constants determined with our model are correctly assigned to the mechanistic steps acylation and condensation and the values for the condensation rates are meaningful. In model 1B, sdV-GrsA shows a condensation rate constant for *DF*-DKP for-

mation 16-fold higher than that for *DV*-DKP, possibly reflecting the preference of the donor site for *D*-Phe. However, in case of the STAP and MS mutants, the *DF*-DKP/*DV*-DKP preference is reduced to two to three-fold although the mu-

tants only differ in the A-, not in the C-domain. This discrepancy either indicates an underestimated experimental error in the predicted constants or an intriguing influence of A-domain mutations on the condensation rate.



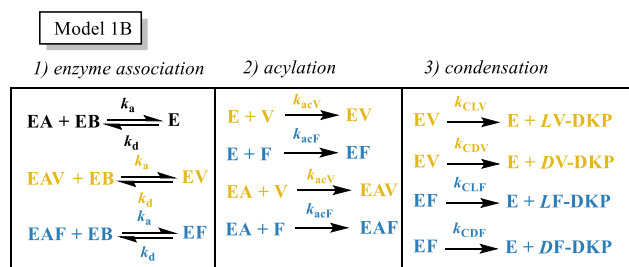
**Figure 4.** a) Peptide formation reaction with STAP at 1 mM competing *L*-Val and *L*-Phe fit with model 1. b) Concentration change of acylated STAP during the reaction course, as predicted by model 1. c) Global fit of STAP progress kinetic data to model 1 (Table 3). d) Selectivity (Val/Phe) comparison between adenylation (HAMA), acylation, and condensation rate constants. e) Time course of the MS mutant and (f) the corresponding E-domain knock-out fit to model 1B (1 mM competing *L*-Val and *L*-Phe). Error bars in a), c), e), and f) indicate the standard deviation from two biological and two technical replicates.

**Table 3. Rate constants fitted with model 1 and 1B.#**

	Constant	sdV-GrSA	STAP	MS
Model 1	$k_{acV}/k_{acF}$	$2.61 \pm 0.01$	$1.800 \pm 0.001$	$14.49 \pm 0.02$
	$k_{CLV} (\text{min}^{-1}) \times 10^3$	$1.3 \pm 0.1$	$2.12 \pm 0.07$	$2.01 \pm 0.06$
	$k_{CDV} (\text{min}^{-1}) \times 10^3$	$6.8 \pm 0.1$	$11.11 \pm 0.08$	$10.5 \pm 0.1$
	$k_{CLF} (\text{min}^{-1}) \times 10^3$	$0.6 \pm 0.4$	$0.6 \pm 0.2$	$0.4 \pm 0.4$
	$k_{CDF} (\text{min}^{-1}) \times 10^3$	$29 \pm 2$	$33.5 \pm 0.5$	$13.8 \pm 0.9$
Model 1B	$K_{d \text{ app}} (\mu\text{M})^*$	2	2	2
	$k_{acV}/k_{acF}^*$	3.4	1.7	14.8
	$k_{CLV} (\text{min}^{-1}) \times 10^3$	$3.7 \pm 0.4$	$5.4 \pm 0.3$	$5.3 \pm 0.2$
	$k_{CDV} (\text{min}^{-1}) \times 10^3$	$19.0 \pm 0.5$	$30.0 \pm 0.4$	$28.1 \pm 0.3$
	$k_{CLF} (\text{min}^{-1}) \times 10^3$	$^{\S}\text{ND}$	$^{\S}\text{ND}$	$5 \pm 2$
	$k_{CDF} (\text{min}^{-1}) \times 10^3$	$300 \pm 50$	$89 \pm 1$	$61 \pm 7$

#The values shown are mean and standard deviation of the output of Dynafit's Monte Carlo algorithm.  $^{\S}$ Individual rate constants  $k_{acV}$  and  $k_{acF}$  could not be obtained but ratios were well defined in the Monte Carlo algorithm. \*These parameters have been fixed to experimental values in model 1B.  $^{\S}\text{ND}$ : Not determined. Standard errors exceed the value more than 100-fold. Values for  $k_{CLF}$  are poorly determined because *LF*-DKP concentrations were close to the limit of detection in all experiments.

## Scheme 2. Kinetic model 1B including enzyme association steps.\*



\*EA: sdV-GrsA and variants; EB: GrsB1. Constants  $k_a/k_d$  and both acylation steps ( $k_{acV}$ ,  $k_{acF}$ ) are fixed to experimentally determined values.

**E-domain inactivation.** Both model 1 and 1B failed to accurately define *L*-Phe condensation constants, because the corresponding *LF*-DKP product was obtained in low yields. Hence, we knocked out the E-domain in the first module, which racemizes amino acids before transfer to the C-domain, to reroute more flow towards the *L*-*L* diastereomers. We inactivated the E-domain by introducing a His753Ala mutation previously shown to abolish epimerization activity,<sup>45</sup> and followed the differences in the progress curves of peptide formation. As expected, DKP products with *D*-*L* configuration are almost abolished when *L*-Val and *L*-Phe are offered as competing substrates as before (Figure 4f). Some *DV*-DKP formation is still detectable, suggesting that the inactivation of the E-domain is not complete or epimerization slowly proceeds via an alternative mechanism. A non-linear fit of the kinetic data with model 1B yielded a similar acylation preference as for the MS mutant (Supplementary Table 2). The condensation rates leading to *LF*-DKP and *LV*-DKP, are now better determined ( $4.4 \pm 0.2$  and  $2.80 \pm 0.02 \times 10^{-3} \text{ min}^{-1}$ , respectively) and show a 10-fold preference for *D*- over *L*-Val in the C-domain.

## Discussion

Transplantation of A-domain specificity is a promising strategy for NRPS engineering, but has been accompanied with losses in activity, for instance in subdomain-swapped sdV-GrsA.<sup>23</sup> A good strategy for restoring activity will be paramount to routinely employing subdomain swapping in the biosynthetic design of natural products. Steric clashes on the surface of the grafted subdomain have been suspected to disturb structural stability and compromise activity. We have shown that testing a small number of reversions to residue identities before swapping can generate significant improvements.<sup>40</sup> In this fashion, substrate specificity has been increased in the MS mutant and the temperature tolerance has been extended by 10 °C in the STAP mutant (Figure 3a). Since subdomain swapping only directly affects a limited number of interface residues, screening of reversion mutations comes at a low cost and may be more generally applicable to chimeric NRPS domains.

We show that mismatched A- and C-domain specificity created through A-domain engineering creates an unexpected change in product specificity over time. GrsB1's preference for the cognate substrate Phe leads to progressive accumulation of Val-loaded sdV-GrsA and the resulting Val-excess eventually overwhelms GrsB1's preference (Figure

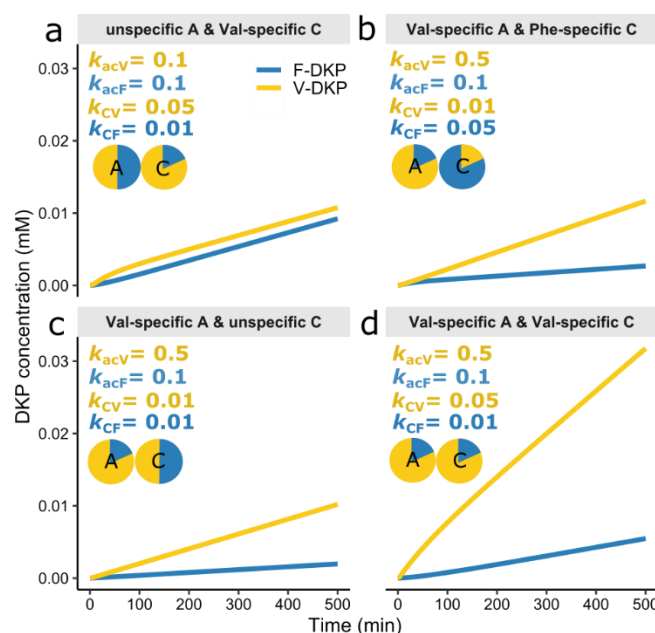
4a). These observations of complex, nonlinear product formation kinetics highlight potential pitfalls in the characterization of engineered NRPSs which may yield contradictory results depending on the exact timing of the assay.

The mismatch between A- and C-domain specificity in sdV-GrsA/GrsB1 offered a unique opportunity to quantify specificity of the condensation step for Val, Phe and their enantiomers. We performed nonlinear kinetic modelling of peptide formation time courses measured at a range of Val/Phe ratios. Given the complexity of the NRPS mechanism, progress data is explained by a conveniently simple model. This model determines acylation rate constants for Val and Phe, and condensation rate constants also for the respective enantiomers. Ratios of acylation rate constants for Val and Phe are clearly reflected in the data and match those experimentally determined with radiolabelled amino acids strikingly well (Figure 4d).

Condensation rate constants are not as uniform as anticipated between mutants having the same C-domain (Table 3). Curiously, it makes a difference for the condensation rate constant which of the marginally different sdV-GrsA variants mutated in the A-domain presents the donor substrate to GrsB1. GrsA/GrsB1 forms the wild-type product *DF*-DKP at a rate of  $> 1 \text{ min}^{-1}$ ,<sup>46</sup> while the sdV-GrsA mutants only reach  $k_{CDF}$  values between 0.06 and  $0.3 \text{ min}^{-1}$ . These differences might indicate an influence of A-domain mutations on a reaction step after T-domain acylation. It is intriguing to speculate that subdomain swapping might have slowed down a conformational change needed to deliver the donor substrate to the C-domain, which is now affected by reversion mutations in the MS and STAP variants.<sup>47</sup>

Strikingly, it follows from our two-step model of NRPS specificity that A-domain dominates C-domain specificity, which is illustrated by simulations of a hypothetical two-module system with tailored acylation and condensation constants (Figure 5). The simulations show that C-domain rate constants matter for the rate of product formation, but not for the specificity. After an initial period of changing product ratios caused by dynamic T-domain loading, the product preference converges to the ratio dictated by the A-domain. While efficient A-domain engineering will overcome C-domain specificity, naturally, condensation can still limit the overall rate. Then, a faster condensation rate directly translates into faster product formation (Figure 5c and d).

The relative importance of A- and C-domain catalysis has decisive implications for choosing the best NRPS engineering strategy. The  $<16$ -fold selectivity for Phe over Val at the C-domain donor site pales compared to five orders of magnitude separating these substrates in terms of  $k_{cat}/K_M$  in a native Phe-A-domain.<sup>48</sup> The only modest differences in condensation rate constants contradict a notion of a previously considered, additional specificity filter.<sup>14,17,23,24</sup> Moreover, the unbalanced loading of the T-domain cancels out C-domain selectivity at the steady-state, so that the A-domain effectively determines the product ratio alone. Hence, even a small preference of the A-domain for Val in sdV-GrsA overrides C-domain preference and *DV*-DKP becomes the main product late in the reaction. Nevertheless, the product yield



**Figure 5.** Hypothetical DKP formation by a dimodular NRPS with different combinations of A- and C-domain specificities. The reaction mechanism of model 1 is used to generate simulated progress curves.

depends on the overall processivity and efficiency of both A- and C-domain catalysis, which is low in sdV-GrsA and both mutants. Small, promiscuity-promoting interventions at the C-domain may suffice to relieve condensation constraints. Keeping A-domains highly functional with minimal structural disturbances will be key for the success of NRPS engineering. The dynamics of T-domain loading are of special importance for engineering in producer strains which express type II thioesterases. These enzymes will presumably remove stalled, noncognate substrates from the T-domain and constantly reset its loading state, thus counteracting the effects of A-domain engineering.

Nonlinear kinetic modelling has been proven here as a useful tool to dissect the complex mechanism of an engineered nonribosomal assembly line synthetase. Our results underline the importance of the A-domain as a gatekeeper and the potential of A-domain engineering as a powerful tool for increasing the diversity of nonribosomal peptides.

## ASSOCIATED CONTENT

The Supporting Information is available free of charge via the Internet at <http://pubs.acs.org>.

Complete experimental procedures and additional data, Figures S1 – S10, Tables S1 – S2, and NMR spectra (PDF).

## AUTHOR INFORMATION

### Corresponding Author

\*Hajo.Kries@Leibniz-HKI.de

### ORCID

Hajo Kries: 0000-0002-4919-2811

Jochen Reinstein: 0000-0003-1497-3084

## Author Contributions

H.K., A.S., A.H., J.R., D.L.N., and P.S. designed the experiments. A.S., A.H., and P.S. performed the experiments. A.S. and A.H. performed mutagenesis, protein production and kinetic measurements. A.S., A.H., J.R., and H.K. analyzed kinetic data. P.S. performed chemical synthesis and thermofluor experiments. A.S., H.K., J.R., and D.L.N. wrote the manuscript with input from all authors.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

We are grateful for financial support by the Fonds der Chemischen Industrie, Daimler und Benz Stiftung, Deutsche Forschungsgemeinschaft and Max Planck Society.

## ABBREVIATIONS

NRPS, nonribosomal peptide synthetase; A-domain, adenylation domain; C-domain, condensation domain; T-domain, thiolation domain; GrsA, gramicidin S synthetase A; GrsB, gramicidin S synthetase B; sdV-GrsA, GrsA with Val-specific subdomain; DKP, diketopiperazine; DV-DKP, D-Val-L-Pro DKP; DF-DKP, D-Phe-L-Pro DKP; LV-DKP, L-Val-L-Pro DKP; LF-DKP, L-Phe-L-Pro DKP; AMS, 5'-O-sulfamoyladenine; HAMA, hydroxamate-based assay for adenylation specificity.

## REFERENCES

1. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
2. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis — Principles and Prospects Reviews. *Angew. Chem. Int. Ed. Engl.* 3770–3822 (2017) doi:10.1002/anie.201609079.
3. Reimer, J. M., Haque, A. S., Tarry, M. J. & Schmeing, T. M. Piecing together nonribosomal peptide synthesis. *Curr. Opin. Struct. Biol.* **49**, 104–113 (2018).
4. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta - Proteins Proteomics* **1865**, 1587–1604 (2017).
5. Brown, A. S., Calcott, M. J., Owen, J. G. & Ackerley, D. F. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228 (2018).
6. Stanišić, A. & Kries, H. Adenylation Domains in Nonribosomal Peptide Engineering. *ChemBioChem* **20**, 1347–1356 (2019).
7. Evans, B. S., Chen, Y., Metcalf, W. W., Zhao, H. & Kelleher, N. L. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. *Chem. Biol.* **18**, 601–607 (2011).
8. Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640–8 (2015).
9. Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* **10**, 282–287 (2018).
10. Thirlway, J. *et al.* Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed. Engl.* **51**, 7181–7184 (2012).
11. Niquille, D. L., Folger, I. B., Basler, S. & Hilvert, D. Biosynthetic Functionalization of Nonribosomal Peptides. *J. Am. Chem. Soc.* **143**, 2736–2740 (2021).
12. Calcott, M. J., Owen, J. G., Lamont, I. L. & Ackerley, D. F. Biosynthesis of novel Pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas*



- aeruginosa. *Appl. Environ. Microbiol.* **80**, 5723–31 (2014).
13. Baltz, R. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. *ACS Synth. Biol.* **3**, 748–58 (2012).
14. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
15. Huang, H.-M., Stephan, P. & Kries, H. Engineering DNA-Templated Nonribosomal Peptide Synthesis. *Cell Chem. Biol.* **28**, 221–227.e7 (2021).
16. Tailhades, J. *et al.* A Chemoenzymatic Approach to the Synthesis of Glycopeptide Antibiotic Analogues. *Angew. Chemie* **132**, 10991–10995 (2020).
17. Bozhüyük, K. A., Micklefield, J. & Wilkinson, B. Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51**, 88–96 (2019).
18. Samel, S. a, Schoenafinger, G., Knappe, T. a, Marahiel, M. a & Essen, L.-O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* **15**, 781–792 (2007).
19. Meyer, S. *et al.* Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. *Cell Chem. Biol.* **23**, 462–471 (2016).
20. Linne, U., Stein, D. B., Mootz, H. D. & Marahiel, M. a. Systematic and quantitative analysis of protein-protein recognition between nonribosomal peptide synthetases investigated in the tyrocidine biosynthetic template. *Biochemistry* **42**, 5114–24 (2003).
21. Marshall, C. G., Burkart, M. D., Keating, T. A. & Walsh, C. T. Heterocycle formation in vibriobactin biosynthesis: Alternative substrate utilization and identification of a condensed intermediate. *Biochemistry* **40**, 10655–10663 (2001).
22. Doekel, S. & Marahiel, M. A. Dipeptide formation on engineered hybrid peptide synthetases. *Chem. Biol.* **7**, 373–84 (2000).
23. Belshaw, P. J. Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science (80-. )*. **284**, 486–489 (1999).
24. Ehmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem. Biol.* **7**, 765–772 (2000).
25. Luo, L. *et al.* Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **41**, 9184–96 (2002).
26. Kaniusaite, M. *et al.* A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. *Chem. Sci.* **10**, 9466–9482 (2019).
27. Bozhüyük, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661 (2019).
28. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
29. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
30. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **11**, 4554 (2020).
31. Ruttenberg, M. A. & Mach, B. Studies on Amino Acid Substitution in the Biosynthesis of the Antibiotic Polypeptide Tyrocidine. *Biochemistry* **5**, 2864–2869 (1966).
32. Peypoux, F. & Michel, G. Controlled biosynthesis of Val7- and Leu7-surfactins. *Appl. Microbiol. Biotechnol.* **36**, 515–517 (1992).
33. Han, J. W. *et al.* Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnol. Lett.* **34**, 1327–1334 (2012).
34. Cryle, M., Steer, D., Goode, R. & Challis, G. Understanding condensation domain selectivity in non-ribosomal peptide biosynthesis: structural characterization of the acceptor bound state. 1–17 (2021).
35. Crüsemann, M., Kohlhaas, C. & Piel, J. Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem. Sci.* **4**, 1041 (2013).
36. Wilson, D. J. & Aldrich, C. C. A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal. Biochem.* **404**, 56–63 (2010).
37. Luo, L. & Walsh, C. T. Kinetic analysis of three activated phenylalanyl intermediates generated by the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **40**, 5329–5337 (2001).
38. Gruenewald, S., Mootz, H. D., Stehmeier, P. & Stachelhaus, T. In vivo production of artificial nonribosomal peptide products in the heterologous host Escherichia coli. *Appl. Environ. Microbiol.* **70**, 3282–3291 (2004).
39. Dehling, E., Rüschenbaum, J., Diecker, J., Dörner, W. & Mootz, H. D. Photo-crosslink analysis in nonribosomal peptide synthetases reveals aberrant gel migration of branched crosslink isomers and spatial proximity between non-neighboring domains. *Chem. Sci.* **11**, 8945–8954 (2020).
40. Stanišić, A., Hüsken, A. & Kries, H. HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes. *Chem. Sci.* **10**, 10395–10399 (2019).
41. Lo, M.-C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **332**, 153–159 (2004).
42. Kuzmič, P. A generalized numerical approach to rapid-equilibrium enzyme kinetics: Application to 17β-HSD. *Mol. Cell. Endocrinol.* **248**, 172–181 (2006).
43. Zhang, D., Kovach, I. M. & Sheehy, J. P. Locating the rate-determining step(s) for three-step hydrolase-catalyzed reactions with dynafit. *Biochim. Biophys. Acta - Proteins Proteomics* **1784**, 827–833 (2008).
44. Kuzmič, P. DynaFit-A Software Package for Enzymology. *Methods Enzymol.* **467**, 247–280 (2009).
45. Stachelhaus, T. & Walsh, C. T. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **39**, 5775–5787 (2000).
46. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for ‘clickable’ amino acids. *Angew Chem Int Ed Engl* **53**, 10105–10108 (2014).
47. Mayerthaler, F. *et al.* Intermediary conformations linked to the directionality of the aminoacylation pathway of nonribosomal peptide synthetases. *RSC Chem. Biol.* (2021) doi:10.1039/D0CB00220H.
48. Villiers, B. R. M. & Hollfelder, F. Mapping the Limits of Substrate Specificity of the Adenylation Domain of TycA. *Chembiochem* **10**, 671–682 (2009).

