MECHANISTIC ANALYSIS OF NONRIBOSOMAL PEPTIDE SYNTHETASES

Aleksa Stanisic

in Partial Fulfilment of the Requirements for the Degree of

"Doctor of Philosophy" (PhD)

Submitted to the Council of the Faculty of Biological Sciences of Friedrich Schiller University Jena

by Aleksa Stanišić, Master of Science

born on 29th April 1991 in Bar, Montenegro

This thesis was performed at the Leibniz Institute for Natural Product Research and Infection Biology e.V. – Hans-Knöll-Institute, Jena in the Department of Biosynthetic Design of Natural Products.
This work was financed by the 441781663 Project (Deutsche Forschungsgemeinschaft, DFG), a fellowship from the Daimler und Benz Foundation, the Fonds der Chemischen Industrie (FCI) and supported by the Graduate School Jena: International Leibniz Research School (ILRS).

Date of defense
08.11.2021
Thesis reviewers
Prof. Dr. Dirk Hoffmeister (Friedrich Schiller University, Jena, Germany)
Prof. Dr. Christian Jogler (Friedrich Schiller University, Jena, Germany)
Prof. Dr. Barrie Wilkinson (University of East Anglia, Norwich, United Kingdom)

“Да ми је да доживим дан кад Алекса школу заврши...”

Тетка Славка, 1934 - 2020

“I wish to live to see the day when Aleksa is finished with school...”

Aunt Slavka, 1934 - 2020

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of the exceptional people in the Biosynthetic Design of Natural Products group. First and foremost, I am indebted to my supervisor Hajo for showing up at the right moment and giving me the opportunity to work on such weird and exciting enzymes like NRPSs. I am especially grateful for your patience, understanding, swift feedback and continuous approachability over the years. It's been an incredible journey, starting from our honeymoon LC-MS shopping trip to Duisburg, mid-life apartment change in summer 2018, culmination with a luxurious Mexican fiesta and ending with publications. There is nothing more I can expect from a supervisor and I am fully aware of the privilege of having you as one.

To all the members of BDN group, I thank for the pleasant and relaxed atmosphere necessary to compensate for the inevitable frustration accompanying our PhD lives. The contagious vitality of Farzaneh conveyed through her regular, interruptive office break-ins. Philipp's humor, darker than his 15 daily cups of coffee. A gentle and always considerate demeanor of Hsin Mei in the face of suffering. Reliability and kindness Sayan brought with him from warm Kolkata (alongside a few ghosts). Daniela, who revolutionized our understanding of memes, while Oliver regularly summoned the spirit of existential ponderance into our hoods and corridors. I was fortunate enough to supervise and work alongside wonderful students: Shobana, who taught me the secrets of ghee; Annika, with her lightheartedness and proclivity for art; Piera, in front of whom I could pretend to speak Italian and Ulrich, who was brave enough to prank me. You were a reminder of the joy of teaching and without your help, this work would not have been possible. Although I could spend only a limited amount of time with new lab members Max and Huiyun, it was enough to stay reassured that Hajo is left in good hands. Considering the size of our group, we would be rather lonely if it weren't for the groups of Christine and Florian with whom we regularly exchanged goods, services and expertise. Special thanks to Luka, with whom I could speak serb..khm..ocroatian, Melanie, Elena, Francois, René, Nina, Chia-Chi, David, Hector, Daniel, Jan, Huijuan, Jamshid.

I would like to thank to our collaborators Felix, Jacob, Carl, Andrei, Prof. Hertweck, Prof. Sasso, Dr. Figge, Dr. Löffler, for providing me the opportunity to work on different and exciting projects as well as to the Prof. Dirk Hoffmeister as a part of my thesis committee and to the reviewers of this thesis. I am also grateful to Prof. Jonathan Gershenzon and Dr. Grit Kunert from MPI-CE for their understanding, support and respect for my decision to change the PhD project.

I would like to thank Mikhaila Peterson for introducing me to her bizzare, but extremely effective “lion diet”, consisting exclusively of ruminant meat, which profoundly transformed my life by completely recovering my crippled mental health.

Last, but far from the least, I give an enormous thanks to my friends. My two chicas, Marina and Sofia who were the main pillars of my life in Jena. Thank you for being my family and my biggest treasure here. Mari, na početku te nisam baš nešto gotivio, ali godinu po godinu, navikne čovjek, bolje išta nego ništa, zavolio sam te iako si đubretara jedna. Sofki, our direkt walks will enter the tourist guides of Jena. Gracias por darme la vida and....1....2....3....eh! Special thanks goes to another Lobeda veteran, my flatmate Vesko, for the familliar atmosphere in our flat and his no-bullshit approach to life. A big shout-out goes to Anushka for making us company and for intiating me into the Russian pop music. I give thanks to my Berlin squad, Nina and Kaća, and to Ivan and Saša for maintaining my Eurosong memory ever fresh. I am also grateful to Vera, Marinita, Osama, Asya, Fadi, Sandra, Falk, Silke, Maria, Franzi, Carlos, Elena, Dragan, Stefan, Agnes, Jan, Uwe, Joel, Kief for all the memorable moments over the years.

To Milica and Džejlana, close friends who are very far, I stay forever indebted and confident that we will continue to fight the distances. I am especially obliged to Ljubica, for her soulful mountain voice. I am also grateful for another Milica for her continuous presence. Neda, Vanja, Vlado, Nataša, Bojana, thank you for making my every return home feel like I’ve just left. Finally, I thank my family: parents and grandparents for their many sacrifices, unceasing support and for reminding me to always discern important from the trivial. At the very end, my heart goes to my sister Ana, my fiercest and most faithful ally.

CONTENTS

SUMMARY

Considering the ongoing rise of the multidrug-resistant bacterial infections, it is essential to expand the available repertoire of therapeutic agents. Microbial natural products are an indispensable source of novel activities and continue to serve as our main provider of antibiotics and chemotherapeutics. However, natural microbial compounds typically require additional modifications introduced by semisynthesis to optimize them for human use. Additionally, a growing need for drugs with novel mechanism of action requires screening of libraries containing diverse chemical structures. This fuels the interest for repurposing natural biosynthetic systems to generate tailored structures or diversify the existing ones.

Nonribosomal peptides are among the most widespread natural products in bacteria and fungi. Their importance is best illustrated by their complexity and the amounts of resources dedicated to building the underlying biosynthetic machineries nonribosomal peptide synthetases (NRPS). These gigantic, multidomain enzymes synthesize peptides by linking individual amino acid units in an assembly line fashion. Each substrate is activated by a dedicated adenylation (A-) domain. Intermediates are shuttled along the assembly line by thiolation (T-) domains and the peptide bond is formed in the condensation (C-) domain. Unlike the ribosomal system, NRPSs incorporate over 500 different monomers resulting in a myriad of peptide structures. Their modular structure, acceptance of alternative substrates and tailoring capacity make them an ideal target for engineering. However, the majority of interventions is plagued with low product titres.

One of the main engineering bottlenecks is the lack of a straightforward assay for determining the substrate specificity of the A-domain. Here, I have developed a specificity assay (HAMA) which enables the determination of a complete specificity profile in a single enzymatic reaction. HAMA is based on the specific detection of amino acid hydroxamates formed by quenching of aminoacyl adenylates with hydroxylamine in the active site of the A-domain. This enables the assaying of multiple substrates at once, thus mimicking competition conditions present in the natural context. This makes a significant step forward from experimentally demanding assays traditionally used for this purpose. The amounts of hydroxamates reflect the specificity constants of corresponding substrates, which was demonstrated against a panel of previously characterized A-domains of known specificity. Additionally, HAMA proved its utility in the characterization of novel fungal A-domains from *Mortierella alpina*.

HAMA offers an unprecedented opportunity for exploring the substrate promiscuity of A-domains. After adapting the experimental conditions for a microtiter plate screening format, I determined the functional landscape of the A-domain in SrfA-

C, the terminal module of surfactin synthetase. First, I developed VSA, an ancestor-like version of SrfAC with relaxed substrate specificity and preserved catalytic efficiency. Second, 15 residues in the binding pocket were individually fully randomized and HAMA profiles determined for each variant. Thus, I obtained the first experimental quantification of the relative contribution of each position and residue to the activity and specificity of the A-domain. I show that A-domains are inherently flexible in terms of substrate selection and that only a few mutations at defined positions can be sufficient to dramatically change the specificity profile. Notably, I show that A-domain mutagenesis need not be accompanied with large activity losses, which have often suffered in previous engineering attempts. This provides a fresh perspective on the directed evolution of A-domains towards new activities.

A functional A-domain is one prerequisite for successful NRPS engineering. However, formation of the modified peptide can be hindered by stalling of the intermediates at the downstream catalytic steps. Of particular importance is the question of a second specificity filter in the C-domain which was suggested to be the additional culprit for low product titres. To probe the relationship between A- and C-domain specificity, we take advantage of a two module system (sdV-GrsA:GrsB1) where the chimeric A-domain of sdV-GrsA shows conflicting specificity with the GrsB1 C-domain. We demonstrate that the A-domain is able to overrule C-domain specificity and dictate the identity of the final product while the C-domain may constrain the product formation rate.

Six decades of NRPS research have resulted in several remarkable tailoring successes. However, the lack of mechanistic understanding of the inner workings of NRPSs has prevented the development of a general workflow which would reliably generate functional enzymes and new drugs. Aspiring to alleviate these obstacles, this thesis offers critical insights into adenylation and the interplay with condensation, two fundamental NRPS reactions.

ZUSAMMENFASSUNG

In Anbetracht der Zunahme von Infektionen mit multiresistenten Bakterien ist es notwendig das Repertoire an verfügbaren Wirkstoffen zu erweitern. Mikrobielle Naturstoffe sind eine nicht weg zu denkende Quelle an neuen Bioaktivitäten und dienen weiterhin als bedeutende Ausgangspunkte für Antibiotika und Chemotherapeutika. Allerdings benötigen neuartige mikrobielle Verbindungen meist zusätzliche Modifikationen, welche durch Semisynthese eingeführt werden können, um sie für Anwendungen am Menschen zu optimieren. Des Weiteren bedingt die wachsende Nachfrage nach Verbindungen mit neuartigen Wirkmechanismen die Entwicklung von Screeningverfahren zur Analyse von Substanzbibliotheken, die diverse chemische Strukturen enthalten. Dies alles treibt das steigende Interesse am Umnutzen von natürlichen Biosynthesesystemen an, um maßgeschneiderte chemische Strukturen zu erschaffen oder existierende anzupassen.

Nicht-ribosomale Peptide gehören zu den am weitest verbreiteten Naturstoffen in Bakterien und Pilzen. Ihre Bedeutung zeigt sich am besten in ihrer Komplexität und der Menge an Ressourcen, die einzig für die Bereitstellung der ihnen zugrunde liegenden biosynthetischen Maschinerie - Nicht-ribosomale Peptidsynthetasen (NRPS) - aufgebracht wird. Hierbei handelt es sich um gigantische Multidomänenzymkomplexe, die Peptide synthetisieren, in dem sie individuelle Aminosäuren im Stil eines Fließbandes miteinander verknüpfen. Jedes Substrat wird von einer spezifischen Adenylierungs-(A-)-Domäne aktiviert. Intermediate werden entlang des Enzymkomplexes über Thiolations-(T-)-Domänen weitergereicht und die Peptidbindung durch Kondensations-(C-)-Domänen gebildet. Im Gegensatz zur ribosomalen Peptidsynthese verwenden NRPSs über 500 verschiedene Bausteine, was in unzähligen Peptidstrukturen resultiert. Ihre modulare Struktur, die Akzeptanz für alternative Substrate und ihre Anpassungsfähigkeit macht NRPS zu idealen Zielen für Protein Engineering. Allerdings sind die meisten dieser Eingriffe durch niedrige Produkttiter gekennzeichnet.

Einer der bedeutendsten Flaschenhälse während des Engineerings ist der Mangel an direkten Assays zur Bestimmung der Substratspezifität von A-Domänen. In dieser Arbeit habe ich einen Spezifitätsassay (HAMA) entwickelt, der Bestimmung eines kompletten Spezifizitätsprofils in einer einzigen enzymatischen Reaktion ermöglicht. HAMA basiert auf der spezifischen Detektion von Aminosäurehydroxamaten, die durch Quenchen von Aminosäureadenylaten mit Hydroxylamin im aktiven Zentrum von A-Domänen gebildet werden. Dies ermöglicht die gleichzeitige Analyse multipler Substrate und bildet derart die Substratkompetition im natürlichen Kontext ab. Hierbei handelt es sich um einen bedeutenden Schritt vorwärts verglichen mit experimentell anstrengenden Assays, die bisher zu diesem Zweck durchgeführt wurden. Die Menge an gebildeten

Hydroxamaten bildet die Spezifitätskonstanten von entsprechenden Substraten ab. Dies wurde durch Analyse von zuvor charakterisierten A-Domänen und dem Vergleich mit für sie publizierten Daten bestätigt. Zusätzlich konnte HAMA seinen Nutzen in der Charakterisierung von neuartigen pilzlichen A-Domänen aus Mortierella alpine nachgewiesen.

HAMA bietet eine beispiellose Möglichkeit zur Untersuchung der Substratpromiskuität von A-Domänen. Nach Anpassung der experimentellen Bedingungen an ein Mikrotiterplatten-Screening-Format konnte ich die funktionellen Gegebenheiten in SrfA-C, dem terminalen Modul von Surfactin Synthetase C, bestimmen. Zuerst erschuf ich mit VSA eine prototypartige Version von SrfAC, welche eine entspanntere Substratspezifität bei gleichzeitig erhaltener katalytischer Effizienz zeigte. Anschließend wurden 15 Positionen in der Bindetasche individuell komplett randomisiert und HAMA Profile für jede Variante gemessen. Auf diese Art und Weise erhielt ich die ersten experimentellen Quantifizierungen des relativen Beitrags jeder einzelnen Position und dort jeder einzelnen Aminosäure zu Aktivität und Spezifität der A-Domäne. Ich konnte zeigen, dass A-Domänen von Natur aus flexibel in Bezug auf Substratselektion sind und nur ein paar Mutationen an klar definierten Positionen ausreichen, um das jeweilige Spezifitätsprofil drastisch zu verändern. Interessanterweise konnte ich zeigen, dass Mutagenese von A-Domänen nicht zwingend mit einem starken Verlust der Aktivität, welche unter vorherigen Engineering Ansätzen litt, einhergehen muss. Dies eröffnet einen neuen Blickwinkel auf die gerichtete Evolution von A-Domänen hin zu neuen Aktivitäten.

Eine funktionelle A-Domäne ist die Voraussetzung für erfolgreiches NRPS Engineering. Die Bildung des modifizierten Peptides kann jedoch durch Festsitzen der Intermediate während nachfolgender katalytischer Schritte behindert werden. Von besonderer Bedeutung ist die Frage nach einem sekundären Spezifitätsfilter in der C-Domäne, welcher als zusätzlicher Verdächtiger für niedrige Produkttiter vermutet wird. Um die Beziehung zwischen A- und C-Domänenspezifität zu adressieren, haben wir ein Zwei-Modul-System (sdV-GrsA:GrsB1) ausgenutzt, in welchem die chimäre A-Domäne von sdVGrsA widersprüchliche Spezifität zur C-Domäne von GrsB1 zeigt. Wir konnten zeigen, dass die A-Domäne in der Lage ist die Spezifität der C-Domäne zu überstimmen und die Identität des finalen Produktes vorzugeben. Allerdings könnte die C-Domäne die Bildungsrate des Produktes einschränken.

Sechs Jahrzehnte an NRPS Forschung resultierten in mehreren bemerkenswerten Erfolgen im gezielten Anpassen von NRPSs. Der Mangel an mechanistischem Verständnis der genauen Funktionsweise von NRPSs hat allerdings die Entwicklung eines generellen Arbeitsablaufs zur verlässlichen Erzeugung von funktionellen Enzymen und neuen Wirkstoffen verhindert. Bestrebt diese Hindernisse abzumildern, biete diese

Doktorarbeit kritische Einblicke in die Adenylierungsreaktion und deren Zusammenspiel mit der Kondensationsreaktion – zwei grundlegenden NRPS-Reaktionen.

1 INTRODUCTION

1.1 The race with infectious diseases

The formation of agricultural human communities and the rise of civilizations 10 000 years ago did not happen without trade-offs. Egalitarian hunter-gatherer societies gave way to the hierarchical communities of peasants, warriors, priests and kings living in tight proximity with newly domesticated animals in settlements of problematic hygienic conditions. Cities were a perfect breeding ground for the rise of *zoonoses*, animal infectious agents which acquired the ability to infect humans.[1] It is considered that all today's 'childhood diseases' originated in this way and continue to coevolve together with their human hosts. This process is still ongoing, as exemplified by this thesis being written during the pandemic of Covid-19 virus, presumably acquired from bats.[2]

The rise of wealth and material possessions, however freed time for studying the environment and eventually discovering ways to tackle these issues. In 1904 Paul Ehrlich developed arsenic-based salvarsan, the first antimicrobial compound effective against syphilis. After 1928, with the discovery of penicillin G by Alexander Fleming, it became obvious that antibiotics have the power to reshape human mankind. It began The Golden Age of antibiotics which peaked in the 1950s and 1960s with the development of penicillins, tetracyclins, aminoglycosides and quinolones.[3] It soon became evident that bacteria have the ability to evade and develop resistance to all of these compounds classes and an arms race for the development of new drugs began.

1.1.1 Antibiotic resistance

The arms race for new antibiotics continues to this day, as we have entered the 'post-antibiotic era'. One of the biggest challenges of modern medicine is the treatment of infectious diseases due to the worldwide rise of multiple drug resistant bacterial strains (MDR). While this process also happens naturally with the exposure of bacteria to the toxic agents, it is strongly facilitated by extensive misuse of antibiotics. For decades, this rise in MDRs was not accompanied with comparable breakthroughs in the development of antimicrobial agents.[4] While Gram positive infections cause a significant burden on the healthcare system, a large panel of efficient antibiotics is available which provides some leeway for treatment. In contrast, Gram negative infections are estimated to pose a five times higher clinical burden due to the limited effective treatment options, quinolones being the last novel antibiotic class developed for this purpose in the last 60 years.[5] The

presence of an additional lipopolysaccharide envelope around the cell membrane of Gram negative bacteria creates a barrier for the penetration of antimicrobial compounds into the cell. Current procedures for treating Gram negative MDR infections involve a combination of new generation carbapenems or cephalosporins with inhibitors of carbapenemases or beta lactamases, resistance-conferring enzymes employed by bacteria. The era of broad-spectrum antibiotics against Gram negative bacteria is over as new treatment protocols require tailoring to the specific infection and the patient. As resistance continues to develop, clinicians are forced to resort to older antibiotics such as colistin, tigecyclin and fosfomycin with problematic safety profiles or pharmacokinetic properties.[6]

1.1.2 Routes towards new antibiotics

As pharmaceutical industry has largely abandoned the search for new antibiotics, this task falls on the shoulders of academic research. Traditional antibiotic discovery approaches based on the isolation of bacterial strains from the natural environment are failing to provide novel lead structures.[7,8] A limited fraction of environmental bacteria is culturable in the laboratory setting. This pool of bacterial taxa has been largely exhausted and results in high rediscovery rates. An alternative approach is targeted drug-design for generating synthetic compound libraries. However, this process does not address the problem of permeability or bioavailability due to the unique properties necessary to penetrate cell envelopes of Gram negative bacteria.[9] Selection pressure in nature optimizes antimicrobials to highly efficient structures able to both penetrate the bacterial cell wall as well as bind to the target with high affinity which is a challenge for de novo design.[10] Novel sequencing techniques revealed an opportunity to access the microbial dark space: whole genomes and biosynthetic gene clusters from unculturable organisms are becoming available and powerful sequence-based algorithms made it possible even to deduce natural product structure from genetic information.[11,12] It has been revealed that even the most extensively studied *Streptomyces* and *Actinomyces* strains contain silent biosynthetic gene clusters (BGCs), inactive under standard cultivation conditions. Additionally, isolating and sequencing genetic material directly from environmental samples allows building large metagenomic libraries with previously unknown BGCs encoding specialized metabolites.[13]

The majority of naturally occurring antimicrobial secondary metabolites belong to the classes of nonribosomal peptides (NRPs) and polyketides (PKs). These structures have been adapted for human use and make up a significant fraction of antibiotics (β-lactams, macrolides, tetracyclins). Beside the discovery of novel structures, there is a growing interest in repurposing the old antibiotics which did not enter widespread use due to undesirable toxicity profiles, pharmacokinetic properties or limited supply. Prominent examples are linezolid and tedizolid, two resurrected members of the

oxazolidinone class of antibiotics, which were discovered in the 1980s and abandoned shortly afterwards due to liver toxicity.[14,15]As the search for novel antibiotics expands, natural products are likely to remain the major innovative lead structures.

1.2 Nonribosomal peptide synthetases

1.2.1 Nonribosomal peptides (NRPs)

Peptide natural products are one of the most diverse compound classes playing a pivotal role in the drug discovery.[16] A panel of 20 proteinogenic amino acid building blocks utilized by the ribosome is diverse enough to generate proteins with myriad of roles and activities. Secondary metabolism draws on this pool and expands the canonical code in order to generate structurally unique compounds fulfilling specific biological roles. This is achieved via two main routes, resulting in two classes of natural products: ribosomally synthesized and post-translationally modified peptides (RiPPs) and nonribosomal peptides (NRPs) synthesized by a distinct family of bacterial and fungal enzymes named nonribosomal peptide synthetases (NRPSs). Additional ribosome-independent pathways exist such as tRNA-independent acyl-AMP-ligases, ATP-grasp-ligases, tRNA-dependent cyclodipeptide synthases and Fem-like ligases.[17]

While the importance of secondary metabolites for producer organisms is not always clear, their potential for clinical application is obvious. The chemotherapeutic class of antimicrobial, anticancer and immunosupressant clinical drugs is rich with structures of nonribosomal origin such as penicillin, chloramphenicol, vancomycin, cyclosporine, teicoplanin, bleomycin etc.[18] The secret for the success of NRPs lies in their unique chemical structures. Despite their enormous diversity, several patterns can be drawn from this family of natural products. Noncanonical building blocks are frequently introduced, such as amino acids decorated with hydroxyl and methyl groups as well as heterocycles and halogens.[18] D- and β-amino acids are particularly prevalent in NRPs, adding additional conformational constraints and rendering structures resistant to the degradation by proteases. The majority of NRPs are further rigidified by cyclization. Hybridization with other metabolic pathways is fairly common, as found in polymyxins with fatty acids at the N terminus or virginiamycin bearing a polyketide fragment.[19,20] The incorporation of noncanonical building blocks in NRPs can be achieved by the direct activation of nonproteinogenic substrates.[21] Alternatively, primary metabolites are activated and edited in specialized compartments within the assembly line. Some of the most complex, crosslinked glycopeptide scaffolds are edited by standalone tailoring enzymes after the peptide is released from the NRPS.[22] The drug development of NRPs faces challenges when sourcing them from the natural environment, exploiting their diversity in laboratory setting and producing them on a large scale by chemical synthesis.

1.2.2 NRPS mechanism

While ribosomal peptide synthesis employs diverse mRNA templates and a single catalytic center, the synthesis of NRPs occurs on complex, modular enzymes whose structural organization itself dictates the identity of the peptide natural product. These are often enormous enzymes encoded on single or multiple polypeptide chains. Current bacterial record keeper is kolossin A synthetase (with a mass of 1.8 MDa) consisting of 15 modules.[23] The biological importance of NRPs is best reflected in the fact that metabolic resources are invested into the synthesis of these gigantic proteins. NRPSs are widely distributed among prokaryotes. In bacteria, they are especially enriched within *Proteobacteria*, *Actinobacteria*, *Firmicutes* and *Cyanobacteria*, while those of fungal origin are concentrated in *Ascomycota*.[24] Classical NRPSs belong to the type I, organized in large, multimodular enzymes, imagined as an assembly line divided into distinct modules where each module activates and incorporates a single substrate into the growing peptide chain. By following this collinearity rule, it is possible to deduce the number of amino acids in the natural product from the number of NRPS modules. Alternative NRPS architectures are found in iterative systems, which contain a small number of modules undergoing repeated use to build the final product by concatenating the same peptide fragment multiple times.[25] Other NRPS scaffolds with unusual module organization not following the collinearity rule are gathered in non-linear group, typically catalysing internal cyclizations or branch-point synthesis.

NRPS modules are further split into individual catalytic units – domains of which four are essential for a fully functional NRPS: 1) adenylation (A-) domains for the activation of the amino acid substrate, 2) catalytically inactive thiolation (T-) domains for shuttling substrates and intermediates along the assembly line, 3) condensation (C-) domains for generating the peptide bond and 4) thioesterase (Te-) domains for releasing the final product. Initial efforts to heterologously produce and assay NRPS proteins yielded inactive enzymes which prevented any progress in NRPS research. The first breakthrough in NRPS enzymology came with the discovery of broadly specific, *in trans* acting phosphopantetheine transferases (PPTase) necessary for the attachment of a 4'-phosphopantetheine (Ppant) prosthetic group to the T-domain, thus converting the enzyme from inactive *apo* to the functional *holo* form.[26] In a typical NRPS system, amino acid substrate is selected by the A-domain and activated with ATP·Mg^{2+}. Resulting aminoacyl adenylate is attacked by a free thiol of the Ppant arm of the T-domain and loaded in the form of a thioester. The loaded T-domain is a central feature of NRP synthesis, as it needs to interact with several domains within the module as well as shuttle intermediates between the modules. The tethered amino acid thioester is then passed to the C-domain where it is coupled to the donor amino group from the upstream peptidyl- or aminoacyl intermediate, thus generating the peptide. Initiation modules (A-T) typically

lack a C-domain since the donor substrate is not available. Elongation modules (C-A-T) extend and transfer the growing peptide chain to the termination module (C-A-T-Te) ending with the Te-domain which releases the final product by hydrolysis or intramolecular cyclization. Additional modifications of loaded substrates and intermediates are achieved through editing domains embedded in the NRPS scaffold such as epimerization (E-), formylation (F-), methylation (M-), oxidation (Ox-), reduction (R-) and heterocyclization (Cy-) domains.

All NRPS modules can be located on a single protein but it is more common to have them split into several polypeptide chains. At split sites, these proteins typically carry additional 20-30 residue long communication (COM) domain pairs at the C terminus of the first and the N terminus of the second module, which enables productive interaction.[27] The interplay between individual domains and modules and orchestration of all steps to maintain the continuous flow of intermediates point to a complex sequence of protein-protein interactions and a highly dynamic structure. This flexibility was responsible for the difficulties faced during structural characterization of NRPSs. However, X-ray crystal structures of several individual domains were successfully solved and a series of recent breakthroughs provided insight into the workings of entire modules and multimodular NRPSs.[28–33]

Entering the assembly line

The A-domain provides the gateway for substrates to enter the NRPS assembly line. In a sequence of conformational changes, two half-reactions occur within the A-domain. First, in the adenylation reaction, substrate is coupled with ATP to generate the reactive aminoacyl adenylate. Second, in the thiolation reaction, adenylate is attacked by a nucleophilic thiol group of the PPant arm of the T-domain forming a thioester bound aminoacyl-T domain. The A-domain (ca. 60 kDa) belongs to the ANL (**a**cyl-CoA synthetases, **N**RPS adenylation domains and **L**uciferase enzymes) superfamily of adenylating enzymes.[34] A-domains generally remain active when excised from their native NRPS context, which was employed to record several crystal structures trapped in different conformations enabling a detailed clarification of their mechanism of action.[28,29,32,35] The core of the A domain is a topologically conserved fragment with a flavodoxin-like fold[36] made of five-stranded β-sheets between the two α-helices, comprising the binding pocket for the side chain of the substrate.

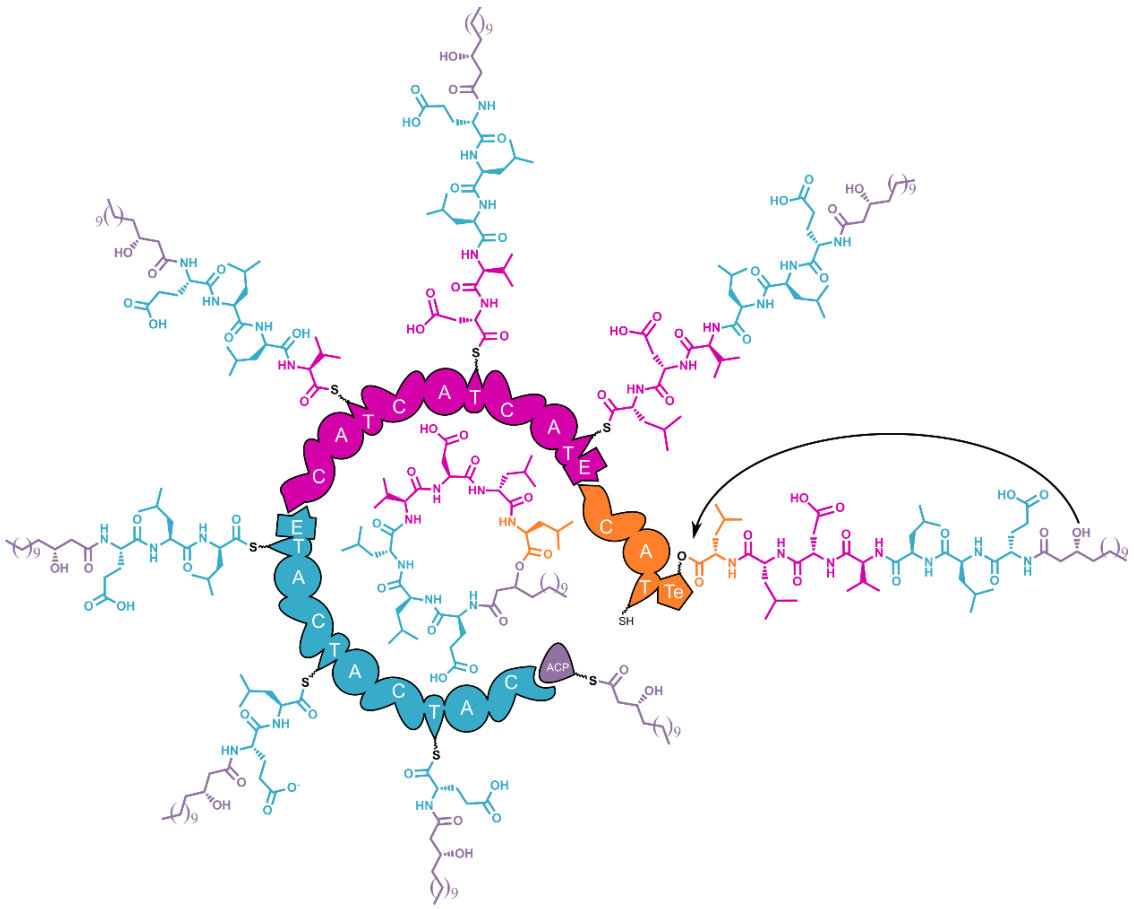


Figure 1. Biosynthesis of surfactin. Surfactin synthetase encompasses seven modules split into three polypeptide chains. The acyl carrier protein (ACP) initiates biosynthesis by providing the fatty acid for condensation. The terminal Te-domain releases the cyclic depsipeptide through intramolecular cyclization.

The A-domain consists of two subdomains: a larger 50 kDa N-terminal core (A_{core}) and a smaller, flexible 10 kDa C-terminal subdomain (A_{sub}).[29] Before substrates are bound, the A-domain adopts an open conformation with A_{sub} turned away, leaving A_{core} exposed to allow binding of amino acid and ATP·Mg^{2+}. Two highly conserved residues are essential for the positioning and reactivity of α-amino acid carboxylate and amino group in the binding pocket: Asp235 in A_{core} at the entrance to the binding pocket and Lys517 in a loop of A_{sub} pointing towards the active site (PheA numbering, PDB: 1AMU).[37] The side chain of the amino acid is positioned below Asp235 in the binding pocket between an α-helix and β-sheet. The adenylate binding cleft (TSGTTGNPKG) is highly conserved within adenylating enzymes, with an extensive network of electrostatic and hydrogen bonds responsible for the tight binding of ATP·Mg^{2+}. Upon substrate binding, A_{sub} covers the active site, bringing Lys517 in contact with substrates and forming the catalytically active closed conformation (Figure 2). The invariable lysine residue stabilises the negatively charged transition state making it a key component of the active site. In contrast, Asp235 can be mutated to accommodate substrates which do not contain α-amino groups such as β-amino acids, α-hydroxy acids, α-keto acids or aminobenzoic acids.[18] Pyrophosphate is released while the highly reactive adenylate

intermediate is protected from hydrolysis by staying tightly bound in the active site of the enzyme until it is loaded on the *holo* T-domain. Once the adenylate is generated, A_{sub} rotates to allow the binding of the T-domain PPant arm which loads the amino acid, releasing the AMP and restoring the open A-domain conformation for the following catalytic cycle.

A breakthrough in NRPS research came with the discovery of specificity conferring residues in the A-domain binding pocket. This was enabled by solving the crystal structure of an A-domain from GrsA, first module of gramicidin S cluster in complex with L-Phe and AMP (1AMU).[37] When the binding pocket residues were pinpointed, sequence analysis of A-domains with different specificities identified eight key positions whose combination provides substrate-specific signatures, termed the 'nonribosomal code'. Out of the eight identified positions, two are defined as 'wobble' residues with lower degree of conservation while the remaining six are highly conserved withing the same substrate group. Binding pockets for nonpolar substrates are generally less conserved, while pockets activating polar substrates typically contain one or more polar residues.[38,39] As more A-domain sequences became available over the years, the specificity conferring code was refined and expanded to encompass the second shell residues using more sophisticated algorithms.[40,41] This information was used to develop sequence-based predictors able to deduce the identity of the final NRPS product from protein sequence data.[42–45]

In addition to NRPS genes, natural product BGCs can encode additional auxiliary MbtH-like proteins (MLPs) named after MbtH in mycobactin cluster in *Mycobacterium tuberculosis H37Rv*.[46,47] They have been shown to copurify with A-domains and, in some cases are essential for A-domain activity and solubility. Gulick et al. solved the first crystal structure of two MLPs bound to the A-domain of EntF from enterobactin synthetase, identifying the signature sequence for MLP-A-domain interaction.[30] However, it has been found that the majority of A-domains contain this motif regardless of their MLP dependence. Despite being essential in some NRPS systems, purpose and mechanism of MLPs remain elusive.

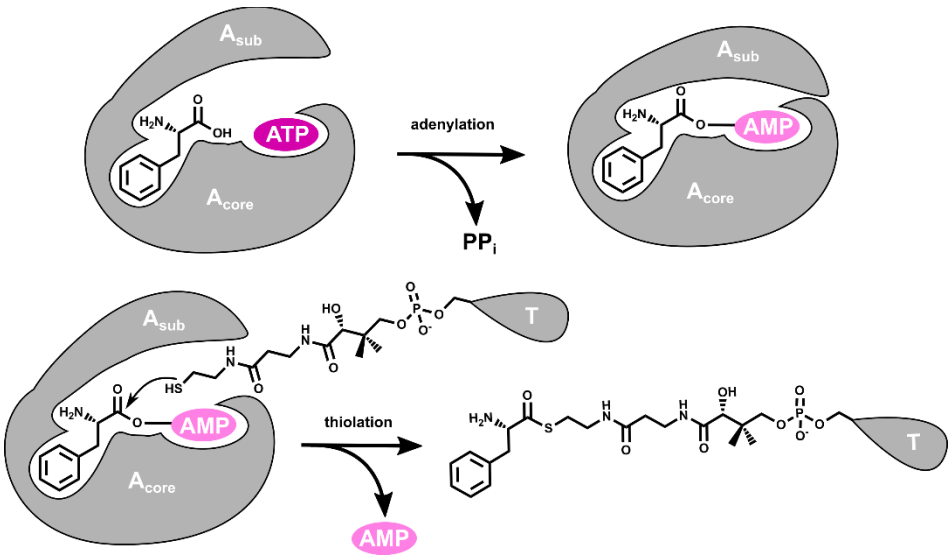


Figure 2. Reactions of the A-domain. In the first adenylation half-reaction, A-domain binds the amino acid and ATP to catalyse the formation of aminoacyl adenylate which stays tightly bound in the active site. In the second thiolation half-reaction, free thiol from PPant arm of the adjacent T-domain attacks the adenylate and tethers the aminoacyl residue as a thioester.

The supply chain of the intermediates

A central role in the NRPS system is played by the small 10 kDa T-domain. T-domains with a flexible PPant extension arm shuttle the aminoacyl- and peptidyl thioester intermediates between domains and modules. Devoid of catalytic activity, T-domains are four helix bundle proteins with a highly conserved GxxS motif used for posttranslational modification by PPTases. Furnishing of adjacent domains with corresponding substrates requires traversing large distances which led to a view that T-domain is highly flexible "swinging arm", undergoing significant conformational changes to interact with all partner domains, while the rest of the NRPS remains relatively rigid around the scaffold built by A-C interfaces.[32] This view has been challenged in a study visualizing complete, multimodular NRPS in different conformations, indicating that the whole assembly line undergoes large conformational shifts.[33] Not only that the T-domain may be relatively inflexible, but it seems that A_{sub} movement during the A-domain cycle is a main driver of T-domain interactions.[29] Thiolation is fast[48] and believed not to contribute to substrate selection. Since the T-domain interacts with several partner domains, protein-protein interaction is an issue. Some specificity has been observed on T-domains from elongation modules transplanted in the termination module which, as a consequence lost the ability to interact with the Te domain.[49] This question is particularly relevant in the context of the C-domain binding, as T-C interaction is essential for the directionality of NRP synthesis.

Connecting the building blocks

Monomers activated in the A-domain are joined together by the C-domain located at the N terminus of each elongation module. This 50 kDa enzyme from CAT (chloramphenicol

acetyltransferase) superfamily couples α-amino group of the amino acid tethered to the T-domain with aminoacyl or peptidyl intermediate of the upstream module.[50] The C-domain is a pseudodimer consisting of two lobes connected with a flexible hinge in a V-shape. Two substrates bound to the PPant arm of the upstream and downstream T-domain meet in the cleft between the lobes where the active site (HHxxxDG motif) is located.[51] Upon peptide bond formation, the resulting peptidyl residue domain is translocated to the downstream module. C-domains have been observed in two conformations based on the distance between the lobes, but it is unclear how relevant these states are for catalysis.[52] It was initially suggested that the second histidine acts as a base, deprotonating the α-amino group of the acceptor aminoacyl-S-T domain for the nucleophilic attack on the donor peptidyl-S-T domain thioester.[53,54] However, mutation of the second histidine does not always lead to a complete loss of peptide formation which casts doubt on the proposed mechanism.[50,55,56] Samel et al. proposed an alternative hypothesis in which catalytic His, acts by stabilizing the tetrahedral reaction intermediate.[57] As the pKa of the α-amino group has been estimated to be ~7, it is likely that active deprotonation is not necessary for the nucleophilic attack in which case active site residues would act mostly through substrate positioning.[58] In addition to the conserved motif, the V shaped cleft forms a solvent channel providing access for binding of the PPant loaded intermediates: the acceptor site binds aminoacyl-S-PPant and the donor site binds aminoacyl- or peptidyl-S-PPant from the upstream module.[50]

The lack of a straightforward and robust assay for the condensation reaction has prevented detailed C-domain substrate specificity investigation. The acceptor site was considered to show strict stereo- and side chain specificity, while the donor site exerts predominantly stereoselectivity.[59,60] Stereoselectivity is further corroborated with phylogenetic analyses which can distinguish four different C-domain classes depending on the chirality of the substrates: starter, $^{L}C_{L}$, $^{D}C_{L}$, and $^{L}C_{D}$.[61,62] However, when it comes to side chain specificity, both sequence analysis and crystal structures failed to identify distinct binding sites and a specificity code analogous to that in the A-domain.[63] Additionally, tolerance towards different substrates seems to vary between C-domains.[64–67] C-domains may also influence the specificity of adjacent A-domains, presumably due to the extensive protein-protein interaction between these domains.[68] This was demonstrated on cyanobacterial microcystins whose A-domains show high promiscuity when assayed as AT constructs, while the inclusion of a native C-domain increases the specificity of the A-domain.[69,70]

It is generally accepted that condensation reactions in C-domains limit the overall rate of NRP biosynthesis, being approximately 100-fold slower than respective adenylation and thiolation reactions.[59,71] An important question is how the activities of different modules are concerted and what mechanisms prevent initiation at internal modules, which would result in incomplete products. One could assume that strict

selectivity of C-domain donor sites would prevent condensation of incomplete peptidyl intermediates, however, the limited data available point towards relaxed specificity of C-domain donor sites.[59] Misinitiated peptidyl intermediates might stall at the T-domains, making them prone to hydrolysis by type II thioesterases.[72] Recent advances in structural biology of C-domains are beginning to clarify the picture. Aldrich et al. synthetised non-hydrolysable ketone and α, α-difluoroketone derivatives of pantheteine probes to stabilize the ternary complex formed during reaction between the donor and acceptor substrates bound to their cognate T-domains in enterobactin synthetase.[73] First crystal structure of a C-domain in complex with aminoacyl-T domain acceptor substrate from Cryle group revealed the absence of a distinct binding pocket for the amino acid side chain.[63] Additionally, the gating mechanism of $^{L}C_{L}$ domains is conferred by Arg2577 which repels the unmodified PPant arm thus preventing the binding of unloaded T-domains. Further development of strong mechanism-based inhibitors which would trap the C-domain in a relevant conformation with T-domains is essential to provide insight into the C-domain mechanism and substrate selection.

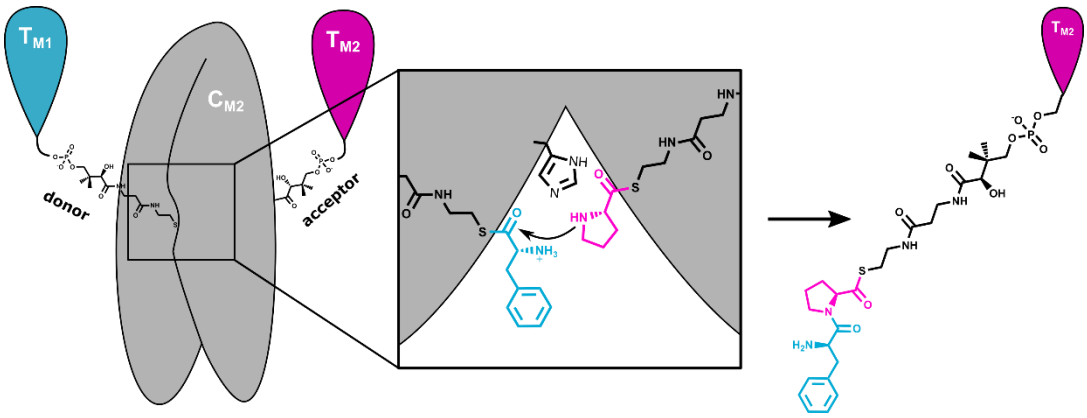


Figure 3. Condensation reaction. The C-domain binds two acylated PPant-T-domains from the upstream and the downstream module. The peptide bond is formed in the cleft between the two lobes of the C-domain by the attack of the acceptor amino group on the donor thioester. Formed peptidyl-PPant-T-domain is released for further processing by the downstream module.

Releasing the product

Thioesterase domains are located at the termination module of the assembly line, and are essential for the release of the mature peptide product, thus enabling the continuous operation of the whole NRPS machinery. Te-domains (~30 kDa) belong to the α/β hydrolase superfamily with a conserved catalytic Ser-Asp-His triad. The release of the mature product is achieved in a two-step process: peptidyl intermediate is transferred from the T-domain to the activated Ser residue in the Te domain, forming an O-peptidyl intermediate at the Te-domain. The activated ester bond is hydrolyzed by a water molecule resulting in the release of the linear peptide or, more frequently, a nucleophilic group from within the peptide yields the macrocyclic product.[74] The shape of the final product depends on the nucleophile used for cyclization. Cyclization with N-terminal amines forms cyclic macrolactams while attack by internal Lys or Orn side chain yields

branched structures. Depsipeptides are formed by utilizing side chain hydroxyl groups or α-hydroxy building blocks as nucleophiles for cyclization. This process is typically highly specific for each system although the Te-domain is too small to accommodate the full peptide intermediate. The cyclization process is governed by conformational positioning of the linear peptide intermediate and the Te-domain mediated specificity towards ring size or amino acid residues.[74] The ability of Te domains to catalyse macrocyclizations of supplied peptidyl substrates has been exploited for organic synthesis where large dilutions are typically required to favour conditions for intramolecular cyclization.[75]

Abovementioned Te-domains embedded at the termination modules of NRPSs belong to the type I thioesterase family. Type II thioesterases are standalone enzymes performing *in trans* proofreading by freeing up incorrectly loaded T-domains which can block the assembly line. Type II thioesterases are able to hydrolyze stalled intermediates and restore the function of the biosynthetic system.[72,76]

Introducing diversity

A-domains have evolved to activate hundreds of nonproteinogenic building blocks, thereby diversifying the peptide product.[77,78] In addition to that, a frequent feature of NRPS systems is the presence of additional, tailoring domains for *in cis* modifications of the peptide. A remarkable subgroup of A-domains termed "interrupted A-domains" contains additional methyltransferase, ketoreductase, oxidase or monooxygenase domains embedded within the A-domain, most frequently at the hinge connecting A_{core} with A_{sub}.[79] The most frequently occurring type has integrated methyltransferase domains catalyzing S-adenosyl methionine (SAM) dependent methylation of the peptide backbone.

C-domains also perform several diversity enhancing functions within NRPS systems. Hybrid polyketide-nonribosomal peptide systems (PKS/NRPS) contain C-domains which can condense polyketide intermediates with the nonribosomal scaffold.[80] Biosynthetic clusters for lipopeptide antibiotics contain a starter C domain which condenses fatty acid CoA thioester.[81] C-domains can also catalyze ester bond formation, as is the case with depsipeptides.[82] Te-domains are often lacking in fungal NRPSs and instead, terminal C-domains catalyse the attack of the α-amino group of the first amino acid in the linear peptide on the thioester, thus generating the cyclic product.[83,84] Moreover, C-domains can play a key role for the recruitment of *trans* acting editing enzymes and the control of incorporation of modified substrates in glycopeptide antibiotics.[85]

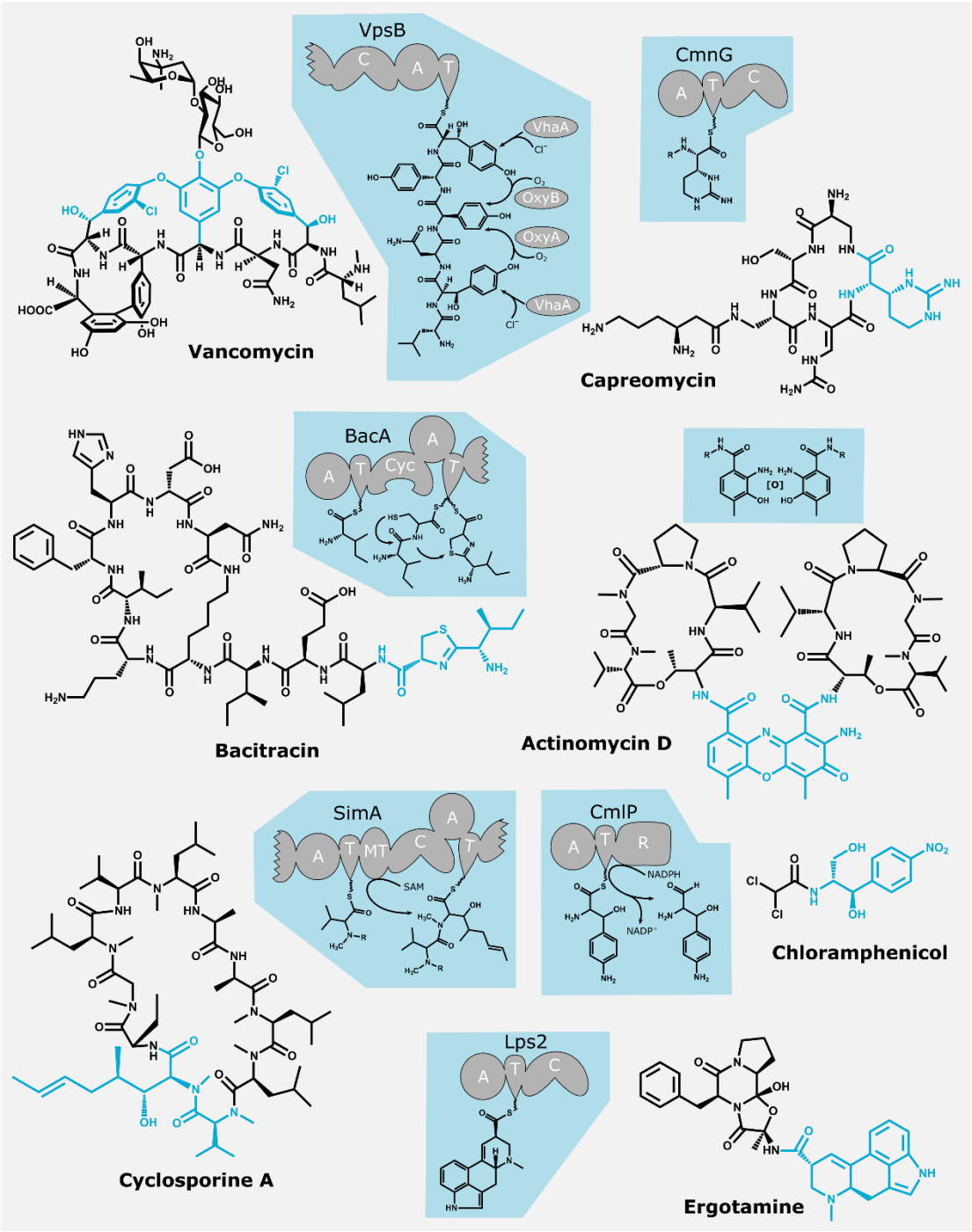


Figure 4. Natural nonribosomal peptides with the biosynthetic origins of selected moieties. Cyc, cyclization domain; MT, metyltransferase domain; OxyA&B, *in trans* acting oxygenases; R, reductase domain; VhaA, *in trans* acting halogenase.

Epimerization and cyclization domains

Both E- and Cy-domain are repurposed C-domains. Cy domains introduce heterocycles into the peptide, most commonly five-membered thiazoline, oxazoline and methyloxazoline. They are characterized by a highly conserved DxxxxD motif catalyzing the two separate reactions: peptide formation between aminoacyl- or peptidyl-S-T domain and cyclodehydration of thiol/hydroxyl of the side chain of condensed serinyl/threonyl/cysteinyl residue.[86]

One of the typical characteristics of NRPs is the presence of D amino acids. While some A-domains can directly activate D substrates, this modification is more often achieved by internal E or C/E domains which catalyse racemization of peptidyl-S-T domain intermediates. Although C- and E-domains share less than 20% homology, they are structurally similar and share the same active site motif.[62] The reaction mechanism of E-domains has been well studied and consists in the abstraction of the Cα proton by the second histidine in the His-motif and consequent reprotonation yielding racemic mixture which can be slightly biased towards the D-isomer.[87,88] Stereoselectivity of the donor site of the downstream C-domain thus ensures the incorporation of D substrate into the peptide.

1.2.3 Studying NRPSs in the laboratory

Heterologous expression of complete, multimodular NRPSs for *in vitro* analysis is often a daunting task which is why domains and modules are typically excised and assayed as standalone constructs. Additionally, the investigation of individual catalytic steps requires them to be isolated. The choice of the cut sites is often arbitrarily made according to the location of the conserved domain motifs. The rationale is that interdomain linker regions do not play a significant role in interdomain communication. Recent studies of linker regions indicate that they may be more important than previously considered raising caution when assaying NRPSs in vitro and planning engineering projects.[89–92]

Measuring adenylation and thiolation

Here, a brief overview of assays employed for measuring adenylation activity is provided while a more comprehensive review is laid out in Manuscript I. The majority of assays for adenylating enzymes measure the activity indirectly, through the pyrophosphate released during the adenylation reaction. However, the development of assays for probing the adenylation half-reaction of the A-domain has been hampered due to the low turnover rate caused by tight binding of the aminoacyl adenylate intermediate to the A-domain.[93] In the absence of a functional T-domain or downstream modules, the reaction is halted after the first cycle while the residual activity detected under these conditions is the leakage rate from the slow dissociation of aminoacyl adenylate.[94] Two workarounds overcame this issue:

Pyrophosphate exchange radioassay

To obviate the need for the T-domain as nucleophile, a pyrophosphate exchange radioassay based on the reversibility of adenylation half-reaction was developed.[95–97] By adding a ^{32}P-radiolabelled pyrophosphate *in vitro,* reaction equilibrium is shifted towards the ATP synthesis yielding ^{32}P-ATP which can be adsorbed on active charcoal and measured by liquid scintillation counting. In this manner, the adenylation rate is indirectly

determined through quantitation of the trapped ^{32}P-ATP. The pyrophosphate exchange assay has been a gold standard for measuring A-domain activity and has been adapted for use in measuring saturation kinetics, specificity profiling and microtiter plate screening experiments.[69,98,99] However, this assay suffers certain limitations. Experimental handling is tedious, requiring the use of expensive and short lived $^{32}PP_i$ and several washing steps which increases the technical error. Additionally, assays must be conducted with a single substrate per reaction with long incubation times which can yield falsely promiscuous specificity profiles.

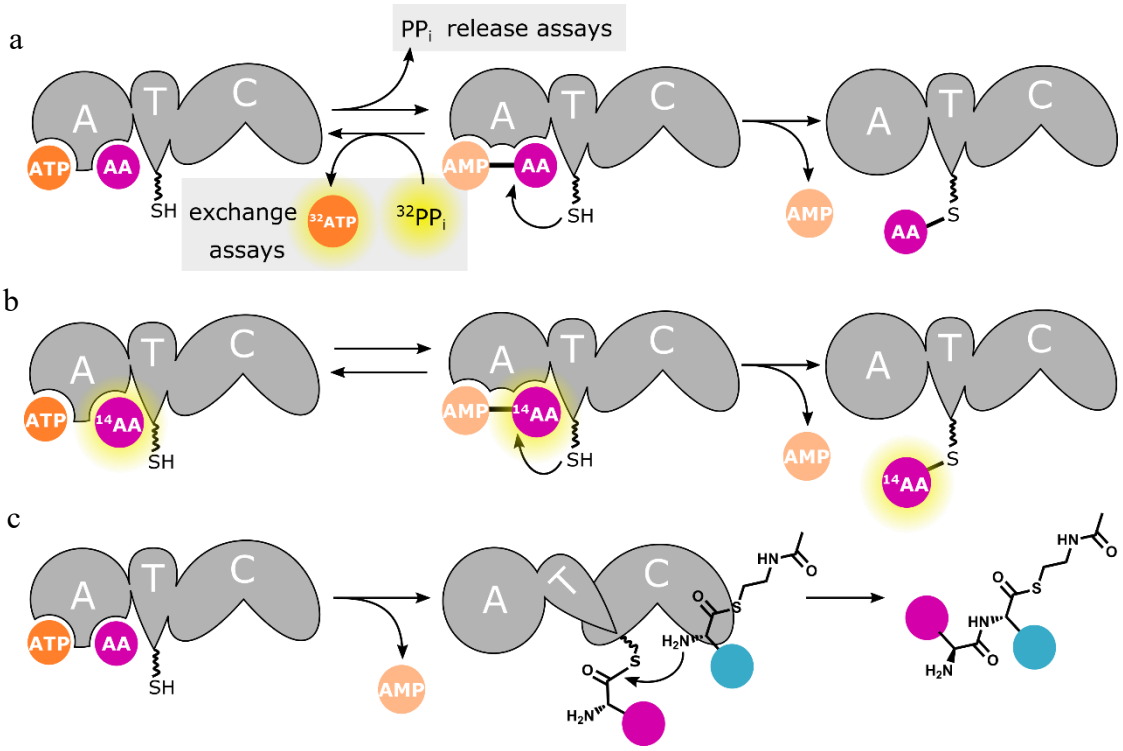


Figure 5. Overview of assays used for measuring NRPS activity. a) Adenylation is measured based on the detection of released pyrophosphate (release assays) or the incorporation of radiolabelled pyrophosphate into ^{32}ATP (exchange assays). b) Acylation is measured by detection of radioactivity trapped on the enzyme after incubation with ^{14}C-labeled substrate and precipitation with trichloroacetic acid. c) Condensation is measured by use of aminoacyl N-acetylcysteamine thioesters as a surrogate substrate for the C-domain and the detection of resulting SNAC-peptide.

Pyrophosphate release assays

A standard method for measuring adenylation is through the detection of released pyrophosphate (PP_i) or by enzymatic hydrolysis to phosphate (P_i). Several platforms have been developed for spectrophotometric detection by generating molybdate and malachite green complexes,[100–103] coupling to NADH oxidation through accessory enzymes[104] and generating chromogenic substrate through phosphorylation of guanosine analogue 2-amino-6-mercapto-7-methylpurine (MesG).[105–108] When applied to the A-domains without a suitable quencher which would release the bound adenylate, these assays are effectively measuring the leakage rate rather than adenylation and present a risk of obtaining misleading results. Addition of excess *holo* T-domain to the reaction restores the turnover, however that is usually not feasible in a standard experimental setting.[104]

Aldrich et al. have successfully utilised hydroxylamine as an alternative quencher and adapted the MesG assay for the continuous detection of A-domain activity.[109] Hydroxylamine is relatively inert towards enzymes and small enough to diffuse into the active pocket and react with the adenylate forming amino acid hydroxamate and AMP, thus enabling the adenylation turnover. This inspired the development of hydroxylamine-based specificity assay in Manuscript II.

Inhibitors of A-domains

Strong interaction between the A-domain and aminoacyl adenylate inspired the development of potent mechanism-based inhibitors. Reactive mixed anhydride group of AMP is exchanged with a similar but stable sulfonamide group resulting in 5′-O-sulfamoyladenosine amino acid derivatives.[110,111] Similar vinylsulfonamide inhibitors with an additional Michael acceptor group were used to trap the T-domain in a thioesterification stage and facilitate crystallization of A-T complexes, thus delineating A-T interactions during the thiolation step.[112,113]

Acylation radioassay

The thiolation half-reaction can be accessed by measuring the acylation rate, encompassing both half-reactions of A- and T-domain pairs.[48,88] Conversion to the *holo* form is an essential prerequisite necessary for the NRPS to be acylated. In experimental settings, this is achieved by preincubating the NRPS construct with coenzyme A and the nonselective PPTase Sfp or expressing NRPS constructs directly in strains bearing integrated nonspecific PPTases.[114] In a following step, enzyme is incubated with ^{14}C labelled substrate and the reaction is quenched at different time points by precipitating the protein. The amount of activity trapped on the protein indicates loading of the substrate on the T-domain. In wild type systems, both adenylation (200 min^{-1})[109] and thiolation (500 min^{-1})[48] half reactions are typically fast, with adenylation being rate limiting. Since the acylation radioassay is discontinuous, probing the quick acylation reaction in wild type systems requires laborious stopped-flow kinetic measurements.

Measuring peptide formation

Compared to the well investigated A-domains, C-domains are experimentally more difficult to assess. This is due to the unique nature of their substrates: aminoacyl-S-T domains or peptidyl-S-T domains which are not easily supplied *in vitro*. Initial investigations of donor- and acceptor site activity were conducted by using aminoacyl-CoA and peptidyl-CoA probes for a C-domain of the tyrocidine synthetase.[59,60] To mimic the terminal part of the PPant moiety in aminoacyl-S-T domain substrates, Walsh and his group utilized aminoacyl N-acetylcysteamine thioesters (SNACs) to characterize substrate- and stereoselectivity of C domains in the enterobactin and tyrocidine synthetase.[61,115] However, SNACs require synthesis and have a propensity towards

nonenzymatic hydrolysis. The development of a straightforward assay for probing C-domain remains a key bottleneck for a more complete picture of its activity.

1.2.4 Repurposing NRPSs

Since their discovery, their structure of distinct and seemingly autonomous units inspired researchers to adapt NRPSs as a source of tailored peptides. The diversity of naturally occurring NRP scaffolds and wealth of different module combinations in NRPS BGCs indicate the potential to adopt custom rearrangements. However, this field of research has been plagued with losses in activity of engineered constructs resulting in decreased yields of peptide products. Nevertheless, recent insights into the inner workings of NRPSs enabled a more informed choice of recombination points. Additionally, directed evolution has been used to hone the activity of impaired chimeras, resulting in a number of remarkable engineering successes.

Targeting specificity gateway

Acting as a main gatekeeper, the A-domain was the first target for introducing novelty in NRPs. A landmark study from Mohammad Marahiel's group substituted the whole A-domain within SrfAC, a termination module in surfactin A synthetase.[116] Three bacterial and two fungal A-domains with a wide range of specificities were transplanted in place of the Leu-specific A-domain of SrfAC, generating modified surfactin A variants. Although peptides were produced at low levels, this was the first evidence that rational alteration of NRPS activity is, in principle possible. The discovery of A-domain specificity code enabled a less disruptive approach by targeting binding pocket residues via site directed mutagenesis. The hope was that they can be utilized to change the specificity without introducing large structural disturbances by exchanging complete domains. However, only conservative changes were achievable, initially switching specificity between structurally similar substrates such as L-Asp to L-Asn and L-Glu to L-Gln.[38,117,118] Naturally promiscuous A-domains from fusaricidin and anabaenopeptin clusters showed less resistance to shifting specificity when their specificity codes were altered.[66,119]

Exchanging the specificity code between different A-domains proved to be too simplistic for generally achieving significant specificity switches. Natural recombination events in hormaomycin cluster inspired the strategy of "subdomain swapping" based on the exchange of a structurally distinct fragment of the A-domain enclosing the specificity conferring residues as well as second shell residues.[99,120] Subdomain exchange includes the whole substrate binding pocket while maintaining native interactions between the domains. However, success was limited to the exchanges of subdomains from the same biosynthetic cluster indicating the importance of homology of A-domains. This was further corroborated by evolutionary study of natural product BGCs showing that they

evolve by a mixture of gene diversification and concerted evolution in distinct families, thus preventing straightforward recombination of distantly related clusters.[121]

Directed evolution is a powerful tool for generating new functionalities and improving the efficiency of engineered enzymes.[122–124] Iterative rounds of mutagenesis can accumulate beneficial mutations by enhancing selection processes that spontaneously occur in nature. In order to replicate this process in the laboratory, it is essential to develop a robust, high-throughput assay for screening of the desired activity. Liu et al. introduced noncognate A-domains into the andrimid cluster and employed mutagenic PCR to generate mutant libraries, followed by activity screening on selection plates.[125] Three rounds of mutagenesis were enough to improve the antibiotic production of synonymous A-substitution variants up to near-wild type levels. In a following study, a high-throughput LC-MS/MS assay was used to screen a library of 14 000 clones and identify four andrimid analogs with improved antibiotic activity.[126]

By screening a single mutant library of specificity conferring residues, Kries et al. generated a W239S mutant of L-Phe specific GrsA, which activates propargyl-L-Tyr.[127] The exchange of a bulky tryptophan residue at the bottom of the binding pocket opened additional space for the propargyl chain which enabled the incorporation of the "click" amino acid into peptide at wild type rates. In another A-domain engineering feat, Niquille et al. adapted yeast surface display (YSD) and fluorescence activated cell sorting (FACS) to change the specificity of TycA in tyrocidine synthetase from α-Phe to β-Phe without losing catalytic efficiency.[128] By combining rational targeted deletion of a single residue and saturation mutagenesis of four specificity conferring positions, novel TycA variant was generated with a remarkable 40 000-fold specificity switch and the production of modified peptide at wild type rate.

Shuffling domains and modules

Initial failures with A-domain engineering raised an issue of additional specificity filters at downstream domains. Acceptor and donor sites are considered to be optimized for wild type aminoacyl- and peptidyl-intermediates which would prevent condensation of noncognate substrates. Initial experiments with artificially loaded modules[59] and aminoacyl-SNACs[115] established stereoselectivity at both C-domain binding sites for the loaded T-domain. Additionally, a structure of SrfAC, a complete NRPS termination module, revealed a large interface between the A and the C-domain.[32] It was suggested that manipulating A-domains in isolation would disturb this interaction and further impair the function of engineered constructs. Hence, the focus was shifted towards maintaining C-A pairs by transplanting entire C-A or C-A-T units.[90,91,129] A comprehensive NRPS engineering project was conducted by Richard Baltz and co-workers at Cubist Pharmaceuticals on the daptomycin biosynthetic cluster. By using a combination of C-A didomain and module recombination from closely related clusters as well as deletions of

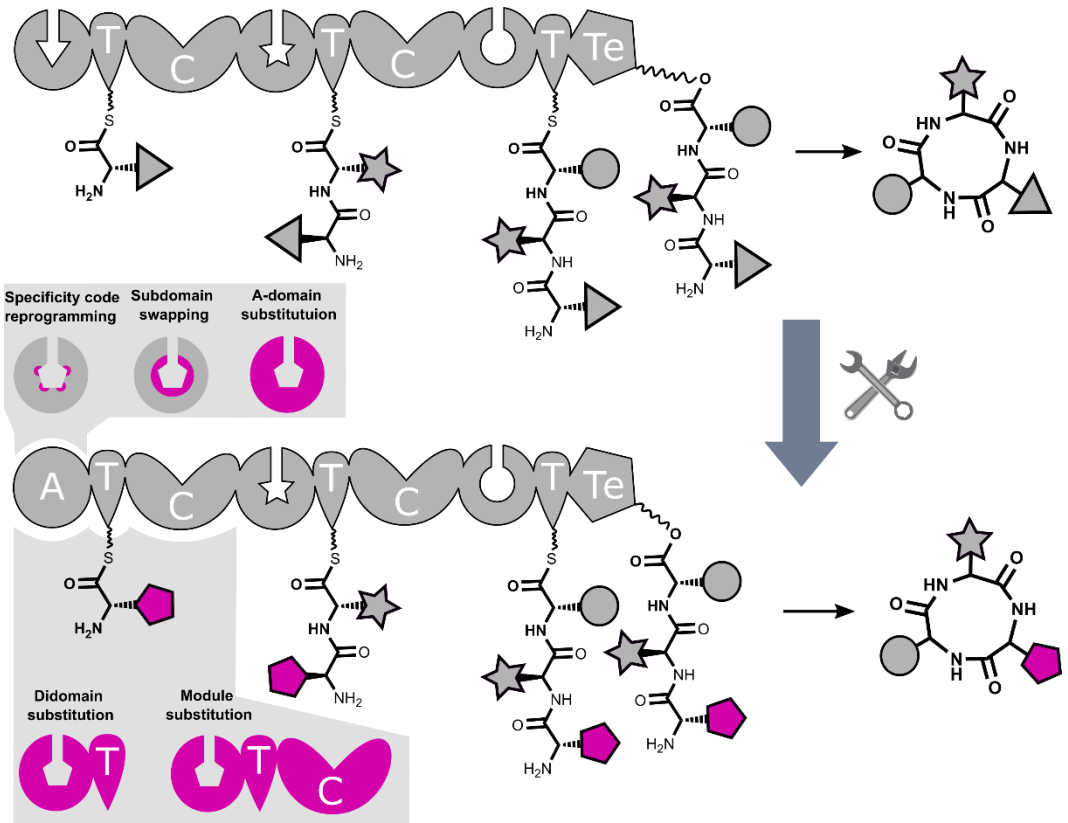


Figure 6. Approaches to NRPS engineering. The incorporation of a new substrate can be achieved by A-domain editing or exchange of entire domains and modules.

tailoring enzymes, more than 40 daptomycin analogues were generated.[130–133] Although some compounds showed improved physicochemical properties, none surpassed daptomycin in terms of antimicrobial activity. Nevertheless, the majority of generated constructs was inactive and those where products were detectable showed highly variable yields confirming that combinatorial NRPS engineering faces serious challenges.

Bode et al. used alternative exchange units comprising A-T-C domains based on the identified cutting site at the linker region between C and A-domain.[90] By analysing the sequences of C-A linkers, they identified a conserved, flexible loop expected to be more susceptible to changes. A panel of exchange units was generated from related bacterial taxa and assembled de novo to yield naturally occurring peptides including several novel structures. While this approach still requires matching of the acceptor site specificity of the exchange unit with the downstream module specificity, it is an important step forward due to the high number of effective recombinations. In a following study, an alternative cutting site at the hinge connecting the two lobes of the C-domain was employed to generate exchange units with (presumably) relaxed specificity of the C-domain, obviating the need to match the exchange unit with the downstream module, thus reducing the number of necessary building blocks.[91] By lifting the requirement to match specificities of the adjacent modules, a series of tailored peptides was generated. Despite some even surpassing the wild type titres, the majority of products still suffered low

production rates. Moreover, the success was more likely in homologous units, again substantiating the evolutionary constraint.

A contrasting perspective came from the work of Ackerley et al. indicating that the C-domain specificity barrier may not be responsible for challenges in NRPS engineering.[92] By shuffling three variable regions of the C-domain, an alternative cut site within the C-A linker region was identified enabling functional A-domain substitutions. Interestingly, their previous work on pyoverdins aligned with the C-domain specificity filter hypothesis resulting in effective synonymous and ineffective nonsynonymous substitutions.[134,135] Remarkably, just by using a different recombination boundary for swapping A-domains in the same system they achieved six successful nonsynonymous substitutions resulting in active constructs producing pyoverdines at high titres.

While A-domain engineering achieved some notable successes, the shuffling of domains and modules is largely influenced by conflicting hypotheses on the stringency of the C-domain selectivity filter, the question here explored in Manuscript IV. The most striking example of C-domain selectivity are glycopeptide antibiotic NRPSs where the C-domain controls the incorporation of *trans*-modified substrate, despite promiscuous A-domain selection.[136] On the contrary, other C-domains show relaxed specificity.[64–66] Additionally, the suggestion that A-C pairs need to be preserved is not supported by phylogenetic evidence of NRPS clusters which evolved by complete or partial A-domain substitution.[137–140] Particularly compelling are microcystins which diversify through A-domain substitutions, paradoxically contrasting the *in vitro* findings of a C-domain gatekeeping role in microcystin synthetases.[69] A possible explanation for these discrepancies is our insufficient understanding of linker regions in NRPSs. Most engineering studies to date have used arbitrary cut sites at linkers between domains and modules without systematically investigating the consequences of the cut site location. It has since been suggested that interdomain linkers are specific for a particular pair of substrates activated by two modules, and their role should be taken into account.[89] Phylogenetic analysis of A-domains additionally identified recombination hotspots within the A-domain and at the interface between the A- and C-domain which were conserved among three bacterial taxa.[92] The possibility remains that stringency of C-domain selectivity is adapted for every system and cannot be generalized. It remains to be seen whether recent successes with the exchange of NRPS units are isolated examples of C-domains with relaxed specificity, or generally applicable approaches.

1.3 Aims of this thesis

Rational modification of NRPS activity demands detailed knowledge of principles and mechanisms behind substrate selection and individual catalytic steps. This thesis aims to address the issues of substrate selection by the A-domain and the two most prominent kinetic bottlenecks at the adenylation and condensation steps.

In Manuscript I, the importance of the A-domain for NRPS engineering and the assays developed to measure A-domain activity will be reviewed. To alleviate current obstacles for determining A-domain specificity, an improved, hydroxamate-based specificity assay (HAMA) is envisioned (Manuscript II). Specificity profiles will be recorded under competition conditions and analysed in a multiplexed fashion by LC-MS/MS to reduce the required experimental effort and yield more meaningful specificity profiles better reflecting intracellular conditions. Specificity assays will be used to characterize novel A-domains (Manuscripts V, VI and VII) and to exploit the full biosynthetic potential of naturally promiscuous, fungal A-domains.

Next, HAMA will be implemented for the first comprehensive investigation of the specificity limits of the A-domain (Manuscript III) – a longstanding problem of NRPS enzymology. First, a promiscuous, ancestor-like version of an A-domain will be developed, assuming that promiscuous enzymes serve as evolutionary branch points from which multiple pathways lead to different specificities and functions. Second, libraries of single mutants at the binding pocket residues will be exhaustively screened with HAMA to determine the functional landscape of the A-domain. Thus, we will quantify the contribution of all binding pocket residues to specificity and reveal the evolutionary distance between A-domains with different specificities.

In addition to the A-domain, it is widely considered that a second C-domain specificity filter may hinder efficient peptide production. To investigate the dynamics between the A- and the C-domain specificity, a chimeric, dimodular NRPS system with impaired catalytic efficiency will be used. Aiming to identify kinetic bottlenecks at specific catalytic steps, experimental determination of individual rate constants coupled with kinetic modelling will be used (Manuscript IV).

Taken together, this thesis offers valuable insights into the mechanisms governing substrate selection and the flow of intermediates in NRPSs. These advances make an important contribution to a more universal and reliable NRPS tailoring methodology.

2 MANUSCRIPT I

Adenylation Domains in Nonribosomal Peptide Engineering

Aleksa Stanišić and Hajo Kries

Published manuscript: *ChemBioChem* **20**, 1347–1356 (2019).

doi: 10.1002/cbic.201800750

Summary:

Adenylation domains are first specificity filters for the incorporation of the substrate into the peptide by nonribosomal peptide synthetases, thus holding a central role in all engineering strategies. This review focusses on the assays for adenylation activity as well as methods for changing A-domain specificity. Strengths and pitfalls of different approaches are discussed aiming to provide an overview of methods used to probe and engineer the A-domain.

The candidate is

☒ First author ☐ Second author ☐ Corresponding author ☐ Coauthor

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
AS	70 %			70 %	
HK	30 %			30 %	

ChemPubSoc Europe
DOI: 10.1002/cbic.201800750
Check for updates
CHEMBIOCHEM Minireviews

Adenylation Domains in Nonribosomal Peptide Engineering

Aleksa Stanišić and Hajo Kries*[a]

Dedicated to Donald Hilvert on the occasion of his birthday

Nonribosomal peptides are a prolific source of bioactive molecules biosynthesized on large, modular assembly line synthetases. Synthetic biologists seek to obtain tailored peptides with tuned or novel bioactivities by engineering modules and domains of these nonribosomal peptide synthetases. The activation step catalyzed by adenylation domains primarily selects which amino acids are incorporated into nonribosomal peptides. Here, we review experimental protocols for probing the adenylation reaction that are applicable in natural product discovery and engineering. Several alternatives to the established pyrophosphate exchange assay will be compared and potential pitfalls pointed out. Binding pocket mutagenesis of adenylation domains has been successfully conducted to adjust substrate preferences. Novel screening methods relying on yeast surface display, for instance, search a larger sequence space for improved mutants and thus allow more substantial changes in peptide structure.

1. Introduction

Myriad structures and functions emerge when amino acids are connected into peptides and proteins. In secondary metabolism, ribosome-dependent and -independent pathways produce cyclic and highly modified compounds beyond the linear polymers of canonical amino acids well-known from typical proteins. Posttranslational modifications expand the structural diversity of ribosomal peptides, giving rise to the natural product class of RiPPs.[1] In contrast to the ribosome, nonribosomal peptide synthetases are inherently able to activate nonproteinogenic substrates and can have several built-in editing domains. The importance of nonribosomal peptides (NRPs) is underscored by their clinical application; for instance, the life-saving molecules daptomycin, vancomycin, cyclosporine, penicillin and cephalosporin are of nonribosomal origin.[2] Nonribosomal peptide synthetase (NRPS) refers to a specific protein family and is not a general term for peptide synthesis outside of the ribosome. Other, more exotic ribosome independent pathways create peptides employing tRNA-independent acyl-AMP-ligases and ATP-grasp-ligases or tRNA-dependent cyclodipeptide synthases and Fem-like ligases.[3]

Regardless of the pathway, the first step in peptide biosynthesis is always the activation of the unreactive carboxylate group of an amino acid by phosphorylation or, as in aminoacyl tRNA synthetases (aaRS) and NRPSs, by adenylation.[4] Similar activation reactions have been honed to perfection in the synthetic organic chemistry laboratory with the advance of solid-phase peptide synthesis over the last 60 years. However, current chemosynthetic methodologies face high costs when generating production scale amounts of cyclic and highly modified peptide structures. Therefore, bioengineering is emerging as a means to repurpose natural machineries for this task.[2,5,6]

1.1. Nonribosomal peptide synthesis

Most NRPSs are enormous enzymes, often hundreds of kilodaltons in size, in which multiple catalytic units (domains) are encoded on one or few polypeptide chains.[2] Domains are grouped in modules, each incorporating one amino acid into the peptide (Figure 1 A). As the growing peptide chain is shuttled from the first to the last module, order and identity of the modules and domains determine the sequence of the peptide. The convenient collinear arrangement of NRPS genes, NRPS proteins and peptide products, that is generally observed, greatly simplifies the analysis of NRPS clusters.

A minimal NRPS elongation module consists of an adenylation (A), a thiolation (T) and a condensation (C) domain. The adenylation domain (≈600 residues) operates as a gatekeeper, controlling which substrate enters the assembly line and becomes attached to the adjacent T domain. Beside the 20 proteinogenic amino acids, the substrate scope of A domains includes a wide range of nonproteinogenic amino acids, fatty acids and α-hydroxy acids.[7] A domains are associated with small (≈100 residues) T domains with long and floppy 4′-phosphopantetheine (Ppant) prosthetic groups attached to a conserved Ser residue. The T domain serves as a flexible, cata-

[a] A. Stanišić, Dr. H. Kries
Independent Junior Research Group
Biosynthetic Design of Natural Products
Leibniz Institute for Natural Product Research and Infection Biology
Hans Knöll Institute (HKI Jena), Beutenbergstrasse 11a
07745 Jena (Germany)
E-mail: hajo.kries@leibniz-hki.de
The ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201800750.

This article is part of the young researchers' issue ChemBioTalents. To view the complete issue, visit http://chembiochem.org/chembiotalents.

ChemPubSoc Europe
CHEMBIOCHEM Minireviews

lytically inert shuttle and transfers tethered substrates between domains and modules in the NRPS biosynthetic machinery. The first C domain forms the peptide bond between the amino acid thioester attached to the initiation module and the amino acid attached to the first elongation module. Downstream C domains elongate the growing peptidyl chain with aminoacyl groups furnished by the respective AT didomains. In addition to the standard CAT module organization, modifying domains such as epimerization, cyclization or methyltransferase domains can enrich structural diversity. Thioesterase domains at the end of the assembly line catalyze release of the mature product by hydrolysis or cyclization.

1.2. Importance of adenylation domains

Reminiscent of a robot arm with multiple tools attached, the AT didomain undergoes complex rotations and conformational changes to load selected amino acids onto the T domain (Figure 1 B).[8] Two half-reactions occur in A domains: activation of the amino acid with adenosine-5′-triphosphate (ATP) to form amino acyl adenylate (adenylation), and transfer of the amino acyl residue to the Ppant arm of the adjacent T domain (thioesterification). Several crystal structures of adenylation domains trapped in different conformations have been solved to shed light on this intricate catalytic process.[9–13] The strictly invariant, positively charged Lys517 residue (numbering of GrsA-A structure)[14] compensates the charge and brings amino acid carboxylate and ATP-α-phosphate into proximity, providing a crucial catalytic driving force for the adenylation reaction (Figure 1 C). Conserved Asp235 is positioned to interact with the amino group of α and β amino acids but is unsurprisingly absent in A domains activating aromatic and aliphatic acids. A large network of hydrogen bonds consisting of polar residues enables tight binding of the pentavalent transition state.[14]

The catalytic process is accompanied by large conformational changes of the small, C-terminal subdomain (A_{sub}) which moves relative to the large, N-terminal core (A_{core}; Figure 1 B).[12] Initially, the A domain adopts an open conformation in which A_{sub} is oriented away from the active site, allowing binding of amino acid and ATP. In the following step, the A domain adopts a catalytically active conformation in which A_{sub} is closed upon the active site. After the amino acyl adenylate has been generated, the A domain adopts a third conformation in which A_{sub} rotates by 140°, opening space for the Ppant arm of the T domain. Then, the A domain loads the amino acid onto the T domain and returns to the initial open state.

A bedrock of NRPS research is the knowledge of specificity-determining residues in the amino acid binding pocket which allows guessing natural product structures from NRPS genes. The first crystal structure of an A domain extracted from the initiation module GrsA of the gramicidin S cluster, in complex

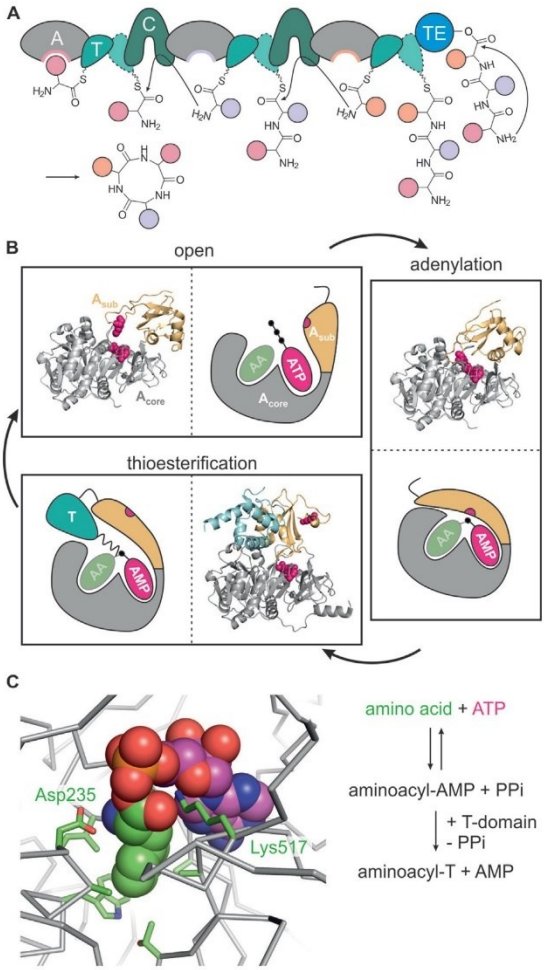


Figure 1. A) A hypothetical NRPS assembles a cyclic tripeptide. The role of the T domain as flexible carrier domain is emphasized by showing it in two positions. B) During adenylation and thioesterification, the A domain goes through a series of conformational states which are exemplified here by X-ray structures of linear gramicidin synthetase LgrA (PDB IDs: 5ES5 and 5ES8[12]). Similar conformational changes have been observed in a number of structures.[8] The AMP ligand modeled from PDB ID: 1AMU[14] and the catalytic lysine are shown as spheres in pink to indicate the position of the active site and the relative movement of A_{sub}. The formylation domain is omitted for clarity. The amino acid is greyed out in the cartoon representation because it is not shown in the crystal structures. C) The specificity code residues in the GrsA-A crystal structure (PDB ID: 1AMU[14]) are shown as green sticks.

Hajo Kries obtained his doctoral degree from ETH Zurich (Switzerland) for his work in the laboratory of Donald Hilvert on enzyme design and engineering. After postdoctoral studies in Sarah E. O'Connor's laboratory at the John Innes Centre (Norwich, UK), he became junior group leader at the Hans Knöll Institute in Jena, Germany. He is currently investigating engineering methods for nonribosomal peptide synthetases.

ChemPubSoc Europe
CHEMBIOCHEM Minireviews

with L-Phe and AMP,[14] revealed eight key residues in the active site that determine substrate specificity (Figure 1 C).[15,16] This insight enabled the development of sequence-based predictors of A-domain specificity. Later, the eight-residue code was amended with second and third shell residues to improve prediction accuracy.[17,18] These tools deduce the substrate or substrate type of A domains with a reasonable accuracy based on the NRPS protein sequence and have aided discovery and structural elucidation of peptides.[19–21]

Here, we review NRPS engineering with a focus on the nonribosomal adenylation (A) domain which provides not the only, but an important route to tailored nonribosomal peptides. Excellent reviews have been published elsewhere in recent years about other aspects of NRPS enzymology and various engineering strategies.[2,22–24] First, we will compare methods to investigate the adenylation domain. Second, we will highlight recent successes in NRPS engineering centered on the A domain.

2. Kinetic Profiling

Given the control that A domains exert over NRPS specificity, it is essential for NRPS discovery and engineering to have suitable assays for dissecting their activity. Over the years, a wide array of assays has been developed for this purpose. Adenylation assays are not only pivotal for engineering NRPSs but there is also a large overlap with assays for aaRSs involved in ribosomal synthesis. Although phylogenetically unrelated, nonribosomal A domains and ribosomal aaRSs catalyze conceptually the same reaction with different acyl group acceptors—T domain and tRNA, respectively. Notably, aaRS engineering has unlocked miscellaneous building blocks for expanding the genetic code with non-canonical amino acids.[25]

The A domain has to orchestrate a complex series of reactions and protect reactive intermediates to avoid formation of side products (Figure 2 A).[26] During adenylation, a pyrophosphate is released while the adenylate is retained in the active site and adenosine-5′-monophosphate (AMP) is subsequently released during substrate transfer to the T domain. Several off-pathway reactions can occur: the amino acyl adenylate intermediate can diffuse into solution, hydrolyze, or react with ATP to free amino acid and adenosine-5′-diphosphate (ADP).[26] Due to the multistep nature of the adenylation reaction and the multitude of side-reactions, several assays are required to gain a complete understanding of A domain reactivity.

2.1. Measuring the adenylation half reaction

In order to examine the kinetics of an individual A domain it must be separated from other A domains in the NRPS cluster and purified for in vitro assays—not always an easy task. Only few standalone A domains act on separate partner T domains in trans.[27,28] Either way, the A domain needs the partner T domain to complete the catalytic cycle or the amino acyl adenylate remains tightly bound to the A domain. Walsh and coworkers have shown that adenylate dissociation is two to three orders of magnitude slower than acyl transfer to the

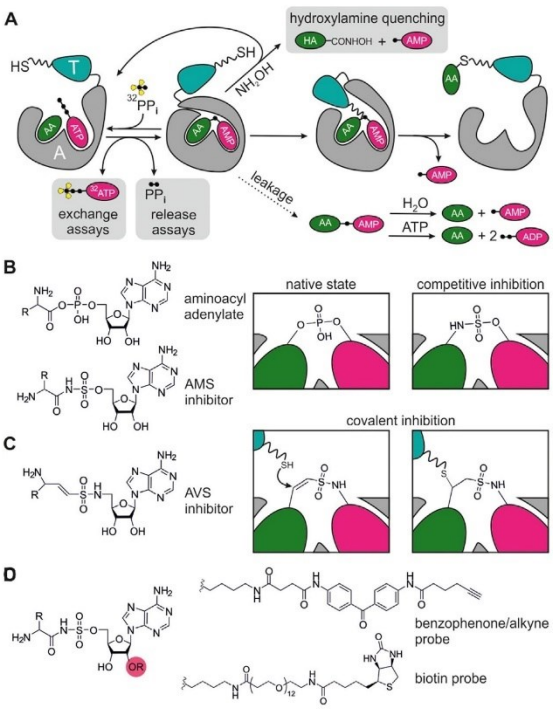


Figure 2. A) Kinetic profiling of the A domain. B) Competitive inhibitors, C) a covalent inhibitor, and D) probes for the A domain.

native T domain.[29–31] Consequently, in A domains excised from the multidomain synthetase, the absence of downstream modules stalls catalysis and greatly diminishes turnover. Under these conditions, turnover relies solely on leakage and hydrolysis of the adenylate intermediate. Therefore, the presence of cognate T domains or alternative nucleophiles has turned out to be crucial for quantifying adenylation activity. Under substrate saturating conditions, thioesterification is completed within minutes which shows as an initial adenylation burst.[26] After exhaustive acylation of all available T domains, the reaction stalls, prohibiting steady-state kinetic measurements of the most interesting A domain activity. A set of alternative solutions to measure the relevant kinetic constants has been developed.

2.2. Pyrophosphate exchange

The laborious pyrophosphate exchange assay has been the gold standard for the precise analysis of adenylation domains and aaRSs in the last 50 years. The reversibility of the adenylation half-reaction is exploited in order to obviate the need for a downstream acyl acceptor.[32–34] In the cell, the approximately isoenergetic adenylation reaction is dragged to the product side by the rapid, enzymatic hydrolysis of the pyrophosphate product (Figure 2 A).[35] An excess of ^{32}P-radiolabeled pyrophosphate ($^{32}PP_i$) added to the reaction mixture in vitro tips the

ChemPubSoc Europe

CHEMBIOCHEM Minireviews

scale to the other side: the backwards reaction is accelerated, yielding radiolabeled ATP which can be quantitatively adsorbed on active charcoal. After washing the charcoal to remove free $^{32}PP_i$, radioactivity is measured in liquid scintillator and the concentration of ^{32}P-ATP formed is calculated in comparison with a reaction run to complete exchange. It is assumed that the exchange velocity strongly correlates with the adenylation velocity. For calculating Michaelis–Menten constants, reactions have to be quenched at different time points in order to determine initial velocities.[26,36]

Pyrophosphate exchange is also employed for specificity profiling of adenylation domains, in which case the enzyme is incubated with a range of amino acids, usually with one arbitrarily chosen time point and substrate concentration for each.[37–39] Since specificity profiles are, due to the work effort, usually not conducted under conditions where the specificity constant k_{cat}/K_M could be measured, such profiles must be taken with a grain of salt. The pyrophosphate exchange assay has also been adapted to a 96-well plate format to increase throughput for screening experiments.[38]

The pyrophosphate exchange assay has several drawbacks. It requires large amounts of expensive, short-lived $^{32}PP_i$ and the procedure itself is technically demanding requiring several washing steps and handling of radioactivity. Bachmann et al. have described a more convenient, nonradioactive version of the exchange assay employing stable γ-^{18}O-labeled ATP and following its exchange with unlabeled PP_i by mass spectrometry.[40] A common drawback of all exchange assays is that they are discontinuous, inflating the workload necessary for recording a full kinetic time course. Also, the exchange rate is an indirect measure of adenylation and not always biologically meaningful. For instance, NRPS-independent siderophore synthetases activating benzoic acids do not exchange, but tightly bind pyrophosphate.[41]

2.3. Pyrophosphate release

PP_i released during adenylation can be measured spectrophotometrically directly or after hydrolysis to phosphate. These spectroscopic assays can be conducted in 96 or 384-well format. In general, PP_i release assays are not the first choice for A-domain engineering projects: In the absence of acyl acceptors and after the initial burst, pyrophosphate release depends on leakage of adenylate from the binding pocket, which should be slower for better substrates. Hence, pyrophosphate release may in some cases *inversely* correlate with the native substrate preference of the adenylation enzyme. However, by adding artificial acyl acceptors, pyrophosphate release assays have been developed into convenient, continuous assays for adenylation activity.

Discontinuous assays

Pyrophosphate can be detected directly through the colored molybdopyrophosphate complex, however this method suffers from poor sensitivity.[42,43] Alternatively, pyrophosphatase converts PP_i into orthophosphate which forms a colored complex with molybdate and malachite green, which is feasible in a microplate format.[44,45] While K_M values from colorimetric PP_i release assays are similar to the K_M values from PP_i exchange assays, turnover rates are orders of magnitude slower (Table 1), corroborating that turnover observed in PP_i release in the absence of nucleophiles reflects leakage from the active site.

Table 1. Variation in kinetic constants reported for homologous enzymes GrsA and TycA.

Assay type	Enzyme variant (Phe chirality)	K_M [μM][a]	k_{cat} [min^{-1}]	k_{cat}/K_M [μM^{-1} min^{-1}]	Ref.
PPi exchange	GrsA (L)	100	690	6.9	[46]
	GrsA-A (D)	110	480	4.5	[47]
	GrsA-AT (D)	160	240	1.5	[47]
	GrsA-A (L)	12	300	25	[48]
	GrsA (L)	70	690	9.9	[49]
	GrsA (D)	70	720	10	[49]
	TycA (L)	12	120	9.9	[36]
PPi release	GrsA (L)	30	0.06	0.002	[49]
	GrsA-A (L)	0.6	0.08	0.13	[50]
MESG/NH_2OH	GrsA-A (D)	70	280	4.1	[47]
	GrsA-AT (D)	100	94	0.9	[47]

[a] Michaelis constant for the Phe substrate.

Continuous assays

Aiming to provide a continuous adenylation assay, a spectroscopic, NADH-based format has been developed.[51] PP_i production was coupled to NADH oxidation over four accessory enzymes (fructose-6-phosphate kinase, aldolase, triosephosphate isomerase, glycerophosphate dehydrogenase) and measured by the drop in NADH absorbance. However, high background activity was observed, possibly derived from traces of highly active NADH dehydrogenases copurified with the protein. This background activity could be reduced by additional protein purification steps. The turnover rate for D-alanyl carrier protein ligase, a standalone A domain, could be enhanced 300-fold by adding the cognate in trans T domain in large excess. Notably, NADH oxidation has also been used as a readout for AMP formation in the study of acyl-CoA ligases which are not integrated into an assembly line and generate AMP directly.[52]

In another continuous adenylation assay format, three auxiliary enzymes are used to couple pyrophosphate production to consumption of a chromogenic substrate. In this process, inorganic pyrophosphatase hydrolyzes pyrophosphate to phosphate. Catalyzed by purine nucleoside phosphorylase, the phosphate expels the nucleobase from guanosine analogue 2-amino-6-mercapto-7-methylpurine (MESG), which increases absorbance at 360 nm.[53–56] In a "rapid quench" study of GrsA this method was employed to measure single turnover catalysis.[57]

2.4. Hydroxylamine quenching

In PP_i release assays, addition of quenchers can prevent stalling of the reaction due to product inhibition. Hydroxylamine has already been used in adenylation assays with aaRSs in the

ChemPubSoc Europe
CHEMBIOCHEM Minireviews

1960s in order to quench aminoacyl adenylates and generate hydroxamic acids. These hydroxamic acids were detected spectrophotometrically as colored Fe^{3+} complexes.[32] With this assay, substrate profiling of A domains unable to release PP_i has been performed.[41] Interestingly, it was shown that the rate of hydroxamate formation correlates with other nonenzymatic quenching reactions, when L-Pro is employed as a nucleophile instead of hydroxylamine, for instance.[58] Nonenzymatic quenching of adenylates with various amine nucleophiles could have synthetic applications.[59,60] The main problem with Fe^{3+}-based detection of hydroxamates is poor sensitivity in the high micromolar range[58] and instability of the complex. Additionally, not all hydroxamates form spectroscopically detectable complexes with ferric iron, and the reaction cannot be monitored continuously.

The Aldrich group has combined the continuous MESG phosphate detection system with hydroxylamine as an alternative nucleophile, thus providing a convenient alternative to PP_i exchange.[47] For standalone A domains, hydroxylamine increases the turnover rate 18-fold above the rate of adenylate leakage of the native substrate. For non-native substrates, k_{cat} values were paradoxically higher than the corresponding values for the native substrates, which was ascribed to the high leakage rate of poorly bound adenylates. Despite these discrepancies, specificity constants (k_{cat}/K_M) are in good agreement with the PP_i exchange assay (Table 1). With this assay, kinetic parameters could be determined for the fatty acid adenylating enzyme FadD28 which does not catalyze PP_i exchange.

2.5. Inhibitors and affinity probes

Potent, mechanism-based inhibitors are important tools in enzymology because inhibitor binding is informative about transition state stabilization,[61] even more so when structures of the enzyme–ligand complex can be analyzed.[62] In directed evolution experiments,[63] mechanism-based inhibitors can pull active enzymes out of mutant libraries. The design of mechanism-based A-domain inhibitors exploits the high affinity of A domains towards the aminoacyl adenylate intermediate (Figure 2B and C). Inhibitors derived from the non-hydrolysable AMP analogue 5′-*O*-sulfamoyladenosine (AMS) are good mimics of this intermediate.[64,65] These aminoacyl AMS inhibitors bind strongly to the A domain with inhibition constants in the low nanomolar range, which makes them convenient crystallographic ligands freezing the A domain in a pose relevant for the adenylation reaction.[22,66] Related AVS inhibitors (Figure 2C) with an additional Michael acceptor group can establish a covalent link to the T domain, when the A–T interaction is of interest.[28,67] Medical applications are conceivable, because some infectious bacteria and fungi use nonribosomal peptides as virulence factors.[68,69] Selected derivatives of salicyl-AMS, for instance, show potent antitubercular activity due to inhibition of MbtA, the A domain involved in the biosynthesis of mycobactin siderophores.[69]

The Kakeya and Ishikawa laboratories have pioneered the design of inhibitor-based affinity probes for the enrichment and identification of NRPS modules in bacterial proteomes by LC-MS/MS analysis (Figure 2D). Besides the discovery and characterization of expressed NRPSs, chemical probes could be applied in monitoring the expression of NRPSs and used to optimize culture conditions. A biotinylated version of L-Phe-AMS was synthesized and used to pull down and monitor the expression of the L-Phe specific GrsA module from lysate of the native producer strain.[70] In a similar fashion, recombinant His_6-tagged A domains binding to an inhibitor coated surface could be detected and distinguished by their specificity in an ELISA format.[71] When photoreactive benzophenones were integrated into the inhibitor constructs, sensitive detection of NRPS domains and intact multimodular proteins was possible in proteomes by in-gel fluorescence imaging.[71–74]

3. Engineering

Alarmed by the rise in antimicrobial resistance and the resulting need for novel compounds, researchers have invested intense efforts over the last decades to discover nonribosomal routes toward tailored antibiotics. Despite major advances in understanding NRPS enzymology, success stories are still rare. This review focusses on strategies targeting the A domain, for which the assays discussed in Section 2 are setting the stage. A strong impetus for NRPS engineering has come in recent years from efforts aiming to increase the throughput of these adenylation assays to microtiter plate format or more. But we do not imply that A-domain engineering alone can unleash the full potential of NRPSs. Contributions of other domains to NRPS specificity often block mutasynthetic incorporation of novel building blocks. Moreover, novel recombination strategies on the module level promise to unlock unprecedented combinatorial freedom.[75]

3.1. Domain swapping

Inspired by the modular NRPS structure divided into distinct catalytic units, the initial, rational approach to NRPS engineering has been the exchange of domains (Figure 3A) and modules.[76] A domains were targeted first due to their role in substrate recognition. A landmark success was the substitution of the A domain in the surfactin operon of *Bacillus subtilis*.[77] In place of the Leu-specific SrfA-C A domain, three bacterial (Phe-, Orn-, Leu-) and two fungal (Cys- and Val-) A domains were incorporated yielding five modified surfactin variants produced in the native host, albeit at very low titres.[78] N-Methylation of Val has been engineered into the actinomycin cluster where a Val-specific adenylation domain was replaced with a synonymous domain from the same cluster carrying an additional methyltransferase domain.[79] The fascinating architecture of methylating A domains, where the methyltransferase interrupts the A domain, has also been artificially recreated by integrating methyltransferases.[80] Results of these engineering studies indicate that domain swapping is in principle possible, but unaccounted obstacles, such as intermodular communications and downstream specificity filters, can compromise engineering efficiency.

ChemPubSoc Europe
CHEMBIOCHEM Minireviews

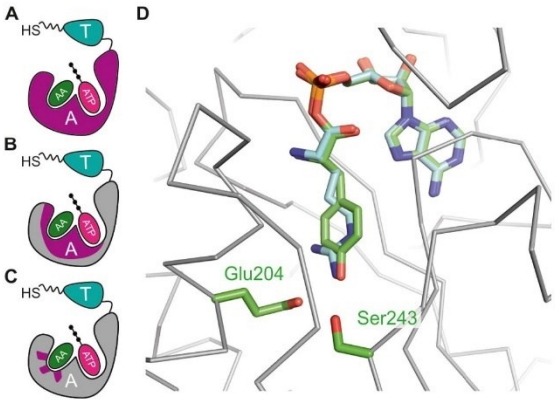


Figure 3. A) Domain swapping, B) subdomain swapping, and C) binding pocket mutagenesis are applied for reengineering NRPS specificity. D) The bispecific binding pocket of ApnA-A_1 (PDB ID: 4D56)[81] in complex with the ligands Arg-AMS (cyan) and Tyr-AMS (green). Mutations E204G/S243E and S243H promote activation of 4-azido-Phe.

3.2. Subdomain swapping

A domains contain a distinct and compact core structure, the binding subdomain, which encompasses the specificity code residues (Figure 3 B). In the evolution of the hormaomycin cluster, duplication and recombination of such binding subdomains presumably transferred substrate specificity.[82] This type of exchange, after being observed in nature, has also been exploited for engineering purposes; with the smaller exchange unit, better maintenance of intermodular communication was expected. Piel and co-workers have swapped subdomains to generate a series of constructs derived from the hormaomycin cluster with subdomains from the same cluster as well as from an unrelated *Streptomyces coelicolor* cluster. The specificities of chimeric enzymes matched those of the donor A domains, and adenylation turnover was preserved. However, constructs with inserts from distantly related clusters were inactive.[82] Following a similar concept, we transplanted a diverse scope of subdomains with varying specificities into Phe-specific GrsA. Chimeric enzymes with various subdomains showed adenylation activity and a Val-specific one was able to participate in dipeptide formation.[39] Again, the most active construct was spliced together with exchange partners from the same cluster—the gramicidin S NRPS—possibly indicating higher success rates with more closely related A domains.

3.3. Specificity code mutagenesis

When signature sequences in the A domain dictate the NRP sequence, it should be possible to deduce mutations leading to custom peptides (Figure 3 C).[15,16] In the work of the Marahiel group, specificity code mutations were introduced into the A domains from the gramicidin S and surfactin synthetases. Complete specificity switches were achieved with conservative substrate changes (L-Glu to L-Gln and L-Asp to L-Asn). Despite activity losses, a modified surfactin variant containing L-Asn could be produced in vivo.[15,83] Using a similar strategy, the Micklefield group altered the specificity of an A domain of the calcium-dependent antibiotic synthetase CdaPS3 to incorporate methyl-glutamine and glutamine instead of methyl-glutamate and glutamate. The antibiotic with the desired methyl-glutamine residue was successfully produced in *S. coelicolor*, constituting the first example of non-proteinogenic amino acid incorporation into a nonribosomal peptide by active site modification.[84]

Bispecific A domains

A domains with relaxed specificity could be gateways to NRPS diversification in both evolution and engineering.[85] In recent work, it has been demonstrated that binding pocket mutations can either enhance promiscuity[37] or skew specificity towards the desired substrate.[81,86] The third A domain in the fusaricidin cluster activates various nonpolar amino acids yielding a mixture of at least twelve fusaricidin analogues. As the L-Phe-fusaricidin variant possesses improved antimicrobial activity, the specificity code of the corresponding domain has been mutated to enhance its production. Indeed, it was possible to increase the specificity towards L-Phe and thereby increase the fraction of the desired L-Phe containing antibiotic in vivo by a factor of three, while the total yield of all fusaricidins was maintained.[86]

For an unusual, bispecific cyanobacterial A domain of an anabaenopeptin synthetase (ApnA-A_1), which naturally activates Arg and Tyr, a structural overlay of Tyr-AMP and Arg-AMP complexes has revealed highly similar orientations of the structurally dissimilar side-chains, explaining the bi-specificity of this enzyme (Figure 3 D).[81] Informed by the structure, Kaljunen et al. have performed site directed saturation mutagenesis on the binding pocket and thus created variants monospecific for either Arg, Tyr or Trp. These mutants preserve, or even surpass wild type adenylation rates, as in the case of a Trp-specific mutant. In this work, both a screening for Fe^{3+}/hydroxamate formation (Section 2.4.) and a kinetic characterization by PP_i exchange (Section 2.2.) were performed.

"Click" residues

Peptide variants are particularly valuable when they carry bioorthogonal handles, for example, azides and alkynes, for selective diversification through chemical conjugation.[48,81] Interestingly, ApnA-A_1 mutants also activate 4-azidophenylalanine (Figure 3 D), which has potential for post-biosynthetic conjugation by click chemistry.[81] In a recently discovered NRP producing fungus, incorporation of click residues has been achieved by feeding the precursor amino acids and exploiting the natural promiscuity of an A domain for L-Phe and O-homoallenyl-Tyr.[87] Nevertheless, selective and efficient incorporation into peptide is challenging. In the A-domain binding pocket of GrsA, we have constructed a single mutant library of eight specificity conferring residues and screened for PP_i exchange in 96-well plates, yielding a point mutation with new specificity. The W239S mutation switches substrate selectivity towards the

ChemPubSoc Europe
CHEMBIOCHEM Minireviews

nonproteinogenic amino acid *O*-propargyl-L-Tyr, while preserving wild-type-like catalytic efficiency.[48] With our "mutasynthetic" approach, formation of a propargylated cyclic dipeptide in vivo proceeded selectively and at wild-type-like rates.

3.4. Directed evolution and screening

Given the success of rationally guided specificity code mutagenesis, even stronger impacts on A-domain specificity would be expected from larger mutational screens in the context of directed evolution experiments. Directed evolution in the laboratory is a powerful approach for optimizing and improving the activity of proteins by mimicking the selection processes that drive natural evolution.[63,88,89] It is based on the generation of mutant libraries which are screened for enhancements of a certain phenotype. Small improvements accumulate over iterative cycles of mutagenesis and screening until the desired activity, stability or selectivity of an enzyme has been reached. A plethora of gene diversification techniques is available for directed enzyme evolution. However, especially in the directed evolution of large and complex NRPSs, a major challenge is to find a powerful screening assay which can sieve through large enough libraries to find the proverbial needle in the haystack.

Fixing chimeric A domains

In their work on the enterobactin and andrimid synthetases, Liu and co-workers were first to demonstrate that the activity of chimeric NRPSs can be improved to near wild-type levels by several rounds of directed evolution.[90] Noncognate A domains from *Streptomyces* and *Bacillus* NRPSs were introduced into the andrimid cluster, mutagenic PCR was employed to generate small mutant libraries (10^3–10^4 clones) and production of andrimid was detected in an inhibition zone screen. After three rounds of screening, antibiotic production with a synonymous A-domain substitution could be improved elevenfold to near wild type levels. In a related experiment, a Val to Ile variant of andrimid with enhanced antibiotic activity was identified. Evolved clones had four to nine amino acid substitutions scattered throughout the A domain which confirmed that residues distant from the binding pocket should not be neglected.

Changing specificity

Going the long way from large L-Phe to small substrates like L-Ala and L-Thr in TycA, for instance, is more demanding because they are disfavored by five- to six-orders of magnitude in catalytic efficiency (k_{cat}/K_M).[36] To this end, iterative saturation mutagenesis has been performed on eight binding pocket residues of the TycA-A domain.[85,91] After two cycles of screening for PP_i exchange and recombination of beneficial mutations, a mutant was isolated with 10^5-fold improved selectivity for L-Ala, to a large extent due to a 10^3-fold decrease for L-Phe.[85] Nevertheless, these significant improvements were not enough to match TycA's high level of activity and selectivity for L-Phe.

Screening for novel peptides

Inhibition zone screens can detect antibiotic activity,[90] but they cannot tell apart structural variations of the inhibitory compound. Moreover, mutated NRPSs making new peptides will typically suffer from low production titers, possibly below the minimum inhibitory concentration. To resolve structures, a highly sensitive LC-MS/MS screening method employing a Fourier-transform ion cyclotron resonance detector was used which identified alternative andrimid derivatives produced in the native host *Pantoea agglomerans* (Figure 4A).[92] Based on the predictable peptide fragmentation, conserved and variable andrimid fragments were used to pinpoint novel derivatives in a defined mass range. Pooling of 96 samples in one LC-MS/MS run boosted the analytical throughput to 14000 clones—enough to screen a library with three binding pocket residues mutated to saturation. Four active mutants of A domain AdmK

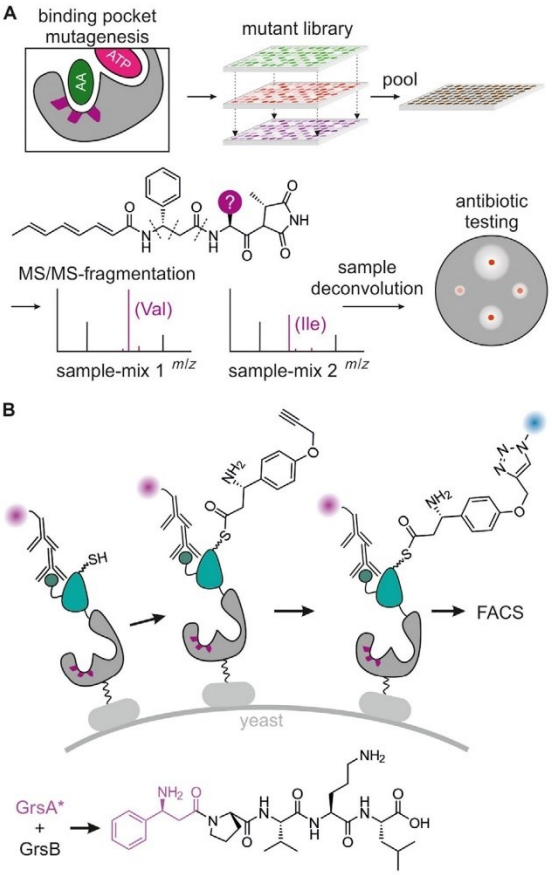


Figure 4. A-domain engineering. A) Screening for structural diversity in a library of A-domain mutants gives rise to andrimid analogues.[92] B) Fluorescence activated sorting of NRPS libraries displayed on yeast surface allows screening of millions of mutants. Mutants were retrieved that enabled synthesis of a backbone modified peptide related to gramicidin S.[66]

were isolated showing the desired specificity towards alternative substrates. For the previously known Ile-andrimid analogue[90] with improved antibiotic activity, the volumetric yield was wild type-like (470 mg L^{-1}), while L-Phe and L-Ala analogues were approximately 1000-fold reduced. Again, none of the mutants showed the same strict selectivity as the wild type for a single substrate.

Yeast surface display

The aforementioned inhibition zone screens[90] and LC-MS assays[92] are limited to library sizes on the order of 10^4 mutants, which is barely enough to cover a library with three positions mutated to saturation ($>20^3$ members). A powerful method for screening larger libraries with more than 10^6 members—yeast cell surface display (YSD)[93]—has recently been adopted by NRPS engineers. Already well established for the development of antibodies and growth factors, YSD is based on the N-terminal fusion of the target protein with Aga2p, which expresses on the yeast surface. A fluorescent label indicating the desired activity is evaluated by fluorescence-activated cell sorting (FACS). For instance, YSD has been employed to screen for mutants of DhbE, a 2,3-dihydroxybenzoic acid specific A domain from the bacillibactin cluster.[94] Biotin-linked acyl-AMS analogues of the noncognate substrates 3-hydroxybenzoic acid and 2-aminobenzoic acid were used to label strongly binding mutants and sort them by FACS. Four positions in the binding pocket were randomized and a library of 5×10^6 clones was screened. The purified mutants were indeed able to bind noncognate substrates more strongly than the wild type but unfortunately, the turnover rate k_{cat} did not follow the same trend.

The Hilvert group has harnessed YSD to switch the specificity of TycA from α-Phe to β-Phe and to make peptide carrying this useful, stability conferring backbone modification (Figure 4B).[66] Inspired by the crystal structure of VinN,[95] a β-amino acid specific A domain, they designed a library containing a deletion and four randomized positions at the binding site. Enabled by the previously discovered W239S mutation,[48] a propargyloxy group was installed on the substrate to allow fluorescent labeling of active AT domains on the yeast surface. By using FACS, a mutant was isolated that strongly preferred β-Phe (40 000-fold specificity switch) and retained high catalytic efficiency. Importantly, the engineered A domain promoted efficient incorporation of β-Phe into a pentapeptide, yielding 120 mg L^{-1} in an *Escherichia coli* culture. The beauty of this assay lies in the direct detection of the relevant amino acyl intermediate, rigorously testing adenylation and thioesterification at the same time. It is evident that a carefully adjusted screening platform is vital to truly capture the desired activity.

4. Discussion and Outlook

Initial enthusiasm about the intriguing genetics of modular biosynthetic pathways has been dampened by various struggles to "harness the biosynthetic code"[76] in practice. The specificity code is an indispensable, predictive tool in natural product discovery. However, rational active site engineering was limited to conservative specificity changes if inspired solely by this code. It seems that the success rate can be remarkably enhanced by covering larger sequence space in mutant screens or directed evolution experiments.[48,81,90,92] After the development of several medium to high throughput screening methods, NRPS engineering again seems a promising route towards novel peptide structures.

NRPSs are complex, dynamic multienzymes and our understanding of their inner workings remains incomplete so that NRPS engineers have learned to endure a high failure rate. We believe that a good comprehension of the kinetics and the orchestration of catalytic events facilitates troubleshooting and enhances engineering success rates. Adequate activity assays for A domains are essential throughout the design process, for gaining mechanistic understanding, performing meaningful screens, diagnosing failures or analyzing hits. In natural product discovery, adenylation assays are vital for validating biosynthetic proposals.

4.1. A guide to adenylation assays

The range of published values for the catalytic constants of the close homologs GrsA and TycA illustrates the difficulties of NRPS kinetic measurements (Table 1) and the fundamentally different purview of the various assay formats discussed in this review. Reported k_{cat}, K_M and k_{cat}/K_M values are vastly different, not only between different assay formats. Some of this variation is owed to the complexity of the system and the numerous variables affecting activity. What should be the best practice for reporting adenylation kinetics? As a rule of thumb, only adenylation assays that are not limited by leakage of the adenylate should be considered for judging the specificity of an A domain. This requirement has been met by discontinuous pyrophosphate exchange assays using $^{32}PP_i$ or γ-^{18}O-ATP, the continuous hydroxylamine/MESG assay[47] and with release assays, but only when natural downstream acceptors were present.[51] Other combinations of hydroxylamine quenching with continuous detection methods for AMP or PP_i are conceivable.[10]

The hydroxylamine/MESG assay holds great promise but more data are needed to ascertain its general applicability. Turnover rates seem slightly lower than in PP_i exchange (Table 1) and possibly, some A domains might be negatively affected by high hydroxylamine concentrations. Also, influence of the amino acid structure on the hydroxylamine quenching rate cannot be excluded. As with other assays, attention must be payed to background from phosphate or amino acid contaminants often present in protein preparations. These issues notwithstanding, continuous data collection and avoidance of expensive isotopes and laborious sample workup are important advantages over PP_i exchange assays for routine kinetic measurements and specificity profiling. However, adenylation assays capture but one step in a long series of events which may all contribute to reaction rate and substrate specificity of the full NRPS. Multiple assays focusing on different mechanistic aspects must be applied to disentangle all these steps.

ChemPubSoc Europe

CHEMBIOCHEM Minireviews

4.2. Future directions

No single approach alone will suffice to create tailored peptide drugs with novel bioactivities. Successful A-domain engineering does not guarantee proficient peptide production, since the processivity of NRPSs depends on numerous factors in and outside of the A domain. A-domain engineering should be combined with novel strategies for module exchange in the future[75] and more insights on C-domain specificity and engineering should be pursued. What comes next? Screening methods should directly address the target and if we want new antibiotics, we should learn how to screen for them in libraries of NRPS clusters that are as diverse as possible and as distant as possible from their natural counterparts. A miniaturization of inhibition zone screens to nanoliter droplets, currently under development in several laboratories,[96,97] could become a cornerstone for building efficient NRPS engineering pipelines to reach that formidable goal.

Acknowledgements

This work was kindly supported by a fellowship of the Daimler und Benz foundation (H.K.). We are grateful to Luka Raguž for helpful comments and discussions.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: adenylation domains · biocatalysis · directed evolution · enzyme engineering · nonribosomal peptide

[1] M. A. Ortega, W. A. van der Donk, *Cell Chem. Biol.* **2016**, *23*, 31–44.
[2] R. D. Süssmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770–3821; *Angew. Chem.* **2017**, *129*, 3824–3878.
[3] T. W. Giessen, M. A. Marahiel, *FEBS Lett.* **2012**, *586*, 2065–2075.
[4] S. Schmelz, J. H. Naismith, *Curr. Opin. Struct. Biol.* **2009**, *19*, 666–671.
[5] A. S. Brown, M. J. Calcott, J. G. Owen, D. F. Ackerley, *Nat. Prod. Rep.* **2018**, *35*, 1210–1228.
[6] H. Kries, *J. Pept. Sci.* **2016**, *22*, 564–570.
[7] C. T. Walsh, R. V. O'Brien, C. Khosla, *Angew. Chem. Int. Ed.* **2013**, *52*, 7098–7124; *Angew. Chem.* **2013**, *125*, 7238–7265.
[8] T. Kittilä, A. Mollo, L. K. Charkoudian, M. J. Cryle, *Angew. Chem. Int. Ed.* **2016**, *55*, 9834–9840; *Angew. Chem.* **2016**, *128*, 9988–9995.
[9] X.-F. Tan, Y.-N. Dai, K. Zhou, Y.-L. Jiang, Y.-M. Ren, Y. Chen, C.-Z. Zhou, *Acta Crystallogr. Sect. D* **2015**, *71*, 873–881.
[10] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. Leigh Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235–238.
[11] B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich, A. M. Gulick, *J. Biol. Chem.* **2016**, *291*, 22559–22571.
[12] J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing, *Nature* **2016**, *529*, 239–242.
[13] M. J. Tarry, A. S. Haque, K. H. Bui, T. M. Schmeing, *Structure* **2017**, *25*, 783–793.e4.
[14] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174–4183.
[15] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493–505.
[16] G. L. Challis, J. Ravel, C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211–224.
[17] C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben, D. H. Huson, *Nucleic Acids Res.* **2005**, *33*, 5799–5808.
[18] M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362–W367.
[19] K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. de Los Santos, H. U. Kim, M. Nave, et al., *Nucleic Acids Res.* **2017**, *45*, W36–W41.
[20] M. A. Skinnider, C. A. Dejong, P. N. Rees, C. W. Johnston, H. Li, A. L. H. Webster, M. A. Wyatt, N. A. Magarvey, *Nucleic Acids Res.* **2015**, *43*, 9645–9662.
[21] M. H. T. Li, P. M. U. Ung, J. Zajkowski, S. Garneau-Tsodikova, D. H. Sherman, *BMC Bioinform.* **2009**, *10*, 185.
[22] A. M. Gulick, C. C. Aldrich, *Nat. Prod. Rep.* **2018**, *35*, 1156–1184.
[23] T. Izoré, M. J. Cryle, *Nat. Prod. Rep.* **2018**, *35*, 1120–1139.
[24] M. Winn, J. K. Fyans, Y. Zhuo, J. Micklefield, *Nat. Prod. Rep.* **2016**, *33*, 317–347.
[25] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
[26] X. Sun, H. Li, J. Alfermann, H. D. Mootz, H. Yang, *Biochemistry* **2014**, *53*, 7983–7689.
[27] G. H. Hur, C. R. Vickery, M. D. Burkart, *Nat. Prod. Rep.* **2012**, *29*, 1074–1098.
[28] J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, *Chem. Biol.* **2012**, *19*, 188–198.
[29] T. A. Keating, Z. Suo, D. E. Ehmann, C. T. Walsh, *Biochemistry* **2000**, *39*, 2297–2306.
[30] D. E. Ehmann, C. A. Shaw-Reid, H. C. Losey, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2509–2514.
[31] F. Rusnak, W. S. Faraci, C. T. Walsh, *Biochemistry* **1989**, *28*, 6827–6835.
[32] M. P. Stulberg, G. D. Novelli, *Methods Enzymol.* **1962**, *5*, 703–707.
[33] B. S. Hansen, *Anal. Biochem.* **1980**, *109*, 12–17.
[34] U. Linne, M. A. Marahiel, *Methods Enzymol.* **2004**, *388*, 293–315.
[35] A. R. Fersht, J. S. Ashford, C. J. Bruton, R. Jakes, G. L. E. Koch, B. S. Hartley, *Biochemistry* **1975**, *14*, 1–4.
[36] B. R. M. Villiers, F. Hollfelder, *ChemBioChem* **2009**, *10*, 671–682.
[37] S. Meyer, J. C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth, E. Dittmann, *Cell Chem. Biol.* **2016**, *23*, 462–471.
[38] L. G. Otten, M. L. Schaffer, B. R. M. Villiers, T. Stachelhaus, F. Hollfelder, *Biotechnol. J.* **2007**, *2*, 232–240.
[39] H. Kries, D. L. Niquille, D. Hilvert, *Chem. Biol.* **2015**, *22*, 640–648.
[40] V. V. Phelan, Y. Du, J. A. McLean, B. O. Bachmann, *Chem. Biol.* **2009**, *16*, 473–478.
[41] N. Kadi, G. L. Challis, *Methods Enzymol.* **2009**, *458*, 431–457.
[42] H. Katano, H. Watanabe, M. Takakuwa, C. Maruyama, Y. Hamano, *Anal. Sci.* **2013**, *29*, 1095–1098.
[43] H. Katano, R. Tanaka, C. Maruyama, Y. Hamano, *Anal. Biochem.* **2012**, *421*, 308–312.
[44] T. P. Geladopoulos, T. G. Sotiroudis, A. E. Evangelopoulos, *Anal. Biochem.* **1991**, *192*, 112–116.
[45] T. J. McQuade, A. D. Shallop, A. Sheoran, J. E. Delproposto, O. V. Tsodikov, S. Garneau-Tsodikova, *Anal. Biochem.* **2009**, *386*, 244–250.
[46] T. Stachelhaus, M. A. Marahiel, *J. Biol. Chem.* **1995**, *270*, 6163–6169.
[47] D. J. Wilson, C. C. Aldrich, *Anal. Biochem.* **2010**, *404*, 56–63.
[48] H. Kries, R. Wachtel, A. Pabst, B. Wanner, D. Niquille, D. Hilvert, *Angew. Chem. Int. Ed.* **2014**, *53*, 10105–10108; *Angew. Chem.* **2014**, *126*, 10269–10272.
[49] L. Luo, M. D. Burkart, T. Stachelhaus, C. T. Walsh, *J. Am. Chem. Soc.* **2001**, *123*, 11208–11218.
[50] B. W. Stevens, R. H. Lilien, I. Georgiev, B. R. Donald, A. C. Anderson, *Biochemistry* **2006**, *45*, 15495–15504.
[51] T. Kittilä, M. Schoppet, M. J. Cryle, *ChemBioChem* **2016**, *17*, 576–584.
[52] A. R. Horswill, J. C. Escalante-Semerena, *Biochemistry* **2002**, *41*, 2379–2387.
[53] M. R. Webb, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4884–4887.
[54] R. H. Upson, R. P. Haugland, M. N. Malekzadeh, R. P. Haugland, *Anal. Biochem.* **1996**, *243*, 41–45.
[55] A. E. Nixon, J. L. Hunter, G. Bonifacio, J. F. Eccleston, M. R. Webb, *Anal. Biochem.* **1998**, *265*, 299–307.
[56] J. Cheng, V. Farutin, Z. Wu, G. Jacob-Mosier, B. Riley, R. Hakimi, E. H. Cordes, *Bioorg. Chem.* **1999**, *27*, 307–325.
[57] L. Luo, C. T. Walsh, *Biochemistry* **2001**, *40*, 5329–5337.
[58] R. Hara, R. Suzuki, K. Kino, *Anal. Biochem.* **2015**, *477*, 89–91.
[59] R. Hara, K. Hirai, S. Suzuki, K. Kino, *Sci. Rep.* **2018**, *8*, 2950.

[60] M. Petchey, A. Cuetos, B. Rowlinson, S. Dannevald, A. Frese, P. W. Sutton, S. Lovelock, R. C. Lloyd, I. J. S. Fairlamb, G. Grogan, *Angew. Chem. Int. Ed.* **2018**, *57*, 11584 – 11588; *Angew. Chem.* **2018**, *130*, 11758 – 11762.
[61] J. Kurz, *J. Am. Chem. Soc.* **1963**, *85*, 987 – 991.
[62] M. M. Mader, P. A. Bartlett, *Chem. Rev.* **1997**, *97*, 1281 – 1302.
[63] C. Zeymer, D. Hilvert, *Annu. Rev. Biochem.* **2018**, *87*, 131 – 157.
[64] S. Bernier, P.-M. M. Akochy, J. Lapointe, R. Chênevert, *Bioorg. Med. Chem.* **2005**, *13*, 69 – 75.
[65] R. Finking, A. Neumüller, J. Solsbacher, D. Konz, G. Kretzschmar, M. Schweitzer, T. Krumm, M. A. Marahiel, *ChemBioChem* **2003**, *4*, 903 – 906.
[66] D. L. Niquille, D. A. Hansen, T. Mori, D. Fercher, H. Kries, D. Hilvert, *Nat. Chem.* **2018**, *10*, 282 – 287.
[67] C. A. Mitchell, C. Shi, C. C. Aldrich, A. M. Gulick, *Biochemistry* **2012**, *51*, 3252 – 3263.
[68] T. D. Davis, P. Mohandas, M. I. Chiriac, G. V. Bythrow, L. E. N. Quadri, D. S. Tan, *Bioorg. Med. Chem. Lett.* **2016**, *26*, 5340 – 5345.
[69] C. Qiao, A. Gupte, H. I. Boshoff, D. J. Wilson, E. M. Bennett, R. V. Somu, C. E. Barry, C. C. Aldrich, *J. Med. Chem.* **2007**, *50*, 6080 – 6094.
[70] F. Ishikawa, H. Kakeya, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 865 – 869.
[71] F. Ishikawa, K. Miyamoto, S. Konno, S. Kasai, H. Kakeya, *ACS Chem. Biol.* **2015**, *10*, 2816 – 2826.
[72] F. Ishikawa, T. Suzuki, N. Dohmae, H. Kakeya, *ChemBioChem* **2015**, *16*, 2590 – 2594.
[73] S. Kasai, S. Konno, F. Ishikawa, H. Kakeya, *Chem. Commun.* **2015**, *51*, 15764 – 15767.
[74] S. Konno, F. Ishikawa, T. Suzuki, N. Dohmae, M. D. Burkart, H. Kakeya, *Chem. Commun.* **2015**, *51*, 2262 – 2265.
[75] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat. Chem.* **2017**, *10*, 275 – 281.
[76] D. E. Cane, C. T. Walsh, C. Khosla, *Science* **1998**, *282*, 63 – 68.
[77] T. Stachelhaus, A. Schneider, M. A. Marahiel, *Science* **1995**, *269*, 69 – 72.
[78] A. Schneider, T. Stachelhaus, M. A. Marahiel, *Mol. Gen. Genet.* **1998**, *257*, 308 – 318.
[79] F. Schauwecker, F. Pfennig, N. Grammel, U. Keller, *Chem. Biol.* **2000**, *7*, 287 – 297.
[80] T. A. Lundy, S. Mori, S. Garneau-Tsodikova, *ACS Synth. Biol.* **2018**, *7*, 399 – 404.
[81] H. Kaljunen, S. H. H. Schiefelbein, D. Stummer, S. Kozak, R. Meijers, G. Christiansen, A. Rentmeister, *Angew. Chem. Int. Ed.* **2015**, *54*, 8833 – 8836; *Angew. Chem.* **2015**, *127*, 8957 – 8961.
[82] M. Crüsemann, C. Kohlhaas, J. Piel, *Chem. Sci.* **2013**, *4*, 1041 – 1045.
[83] K. Eppelmann, T. Stachelhaus, M. A. Marahiel, *Biochemistry* **2002**, *41*, 9718 – 9726.
[84] J. Thirlway, R. Lewis, L. Nunns, M. Al Nakeeb, M. Styles, A.-W. Struck, C. P. Smith, J. Micklefield, *Angew. Chem. Int. Ed.* **2012**, *51*, 7181 – 7184; *Angew. Chem.* **2012**, *124*, 7293 – 7296.
[85] B. Villiers, F. Hollfelder, *Chem. Biol.* **2011**, *18*, 1290 – 1299.
[86] J. W. Han, E. Y. Kim, J. M. Lee, Y. S. Kim, E. Bang, B. S. Kim, *Biotechnol. Lett.* **2012**, *34*, 1327 – 1334.
[87] H. Guo, A. Schmidt, P. Stephan, L. Raguž, D. Braga, M. Kaiser, H.-M. Dahse, C. Weigel, G. Lackner, C. Beemelmanns, *ChemBioChem* **2018**, *19*, 2307 – 2311.
[88] M. S. Packer, D. R. Liu, *Nat. Rev. Genet.* **2015**, *16*, 379 – 394.
[89] F. H. Arnold, *Angew. Chem. Int. Ed.* **2018**, *57*, 4143 – 4148; *Angew. Chem.* **2018**, *130*, 4212 – 4218.
[90] M. A. Fischbach, J. R. Lai, E. D. Roche, C. T. Walsh, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11951 – 11956.
[91] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891 – 903.
[92] B. S. Evans, Y. Chen, W. W. Metcalf, H. Zhao, N. L. Kelleher, *Chem. Biol.* **2011**, *18*, 601 – 607.
[93] E. T. Boder, K. D. Wittrup, *Nat. Biotechnol.* **1997**, *15*, 553 – 557.
[94] K. Zhang, K. M. Nelson, K. Bhuripanyo, K. D. Grimes, B. Zhao, C. C. Aldrich, J. Yin, *Chem. Biol.* **2013**, *20*, 92 – 101.
[95] A. Miyanaga, J. Cieślak, Y. Shinohara, F. Kudo, T. Eguchi, *J. Biol. Chem.* **2014**, *289*, 31448 – 31457.
[96] L. Mahler, K. Wink, R. J. Beulig, K. Scherlach, M. Tovar, E. Zang, K. Martin, C. Hertweck, D. Belder, M. Roth, *Sci. Rep.* **2018**, *8*, 13087.
[97] S. S. Terekhov, I. V. Smirnov, A. V. Stepanova, T. V. Bobik, Y. A. Mokrushina, N. A. Ponomarenko, A. A. Belogurov, M. P. Rubtsova, O. V. Kartseva, M. O. Gomzikova, et al., *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2550 – 2555.

Manuscript received: November 30, 2018
Accepted manuscript online: January 10, 2019
Version of record online: March 13, 2019

3 MANUSCRIPT II

HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes

Aleksa Stanišić, Annika Hüsken and Hajo Kries

Published manuscript: *Chem. Sci.* **10**, 10395–10399 (2019).

doi: 10.1039/C9SC04222A

Summary:

Adenylation domains are one of the main targets for engineering of nonribosomal peptide synthetases. Conventional adenylation assays based on pyrophosphate release or exchange are cumbersome and can measure only a single substrate at a time. Here, we describe a new adenylation assay based on the detection of formed hydroxamates of corresponding substrates after quenching with hydroxylamine. HAMA offers a quick and straightforward way to determine complete specificity profiles in competition conditions from a single reaction. This simplified procedure will facilitate A-domain characterization and NRPS engineering.

The candidate is

☒ First author ☐ Second author ☐ Corresponding author ☐ Coauthor

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
AS	70 %	80 %	90 %	80 %	
AH		10 %	10 %	10 %	
HK	30 %	10 %		10 %	

Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Sci.*, 2019, **10**, 10395

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 22nd August 2019
Accepted 13th September 2019

DOI: 10.1039/c9sc04222a

rsc.li/chemical-science

HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes†

Aleksa Stanišić, Annika Hüsken and Hajo Kries *

Adenylation enzymes selecting substrates for ribosomal and nonribosomal protein and peptide biosynthesis have been popular targets of enzyme engineering. Previous standard assays for adenylation specificity have been cumbersome and failed to reflect the competition conditions inside a cell because they measure substrates one at a time. We have developed an adenylation assay based on hydroxamate quenching and LC-MS/MS detection of hydroxamate products testing dozens of competing amino acid substrates in parallel. Streamlined specificity profiling of adenylation enzymes will facilitate engineering and directed evolution of ribosomal and nonribosomal peptide synthesis.

Adenylate forming enzymes control the substrate selection process in ribosomal and nonribosomal peptide synthesis (Fig. 1A and B).[1–3] Hence, aminoacyl-tRNA synthetases (aaRSs) from ribosomal protein synthesis and nonribosomal adenylation (A) domains have been extensively engineered[4–6] in order to change their specificity for incorporation of alternative substrates. Non-natural building blocks have been of particular interest for enriching the functional spectrum of peptides and proteins, for instance with handles for bio-orthogonal cross-linking.[7,8] While aaRSs function as standalone enzymes, A domains are embedded in a large biosynthetic scaffold together with thiolation (T), condensation (C), thioesterase (Te) and a number of editing domains. Outside these cellular peptide synthesis machineries, adenylating enzymes have recently found application as biocatalysts for amide bond formation, one of the most important reactions in pharmaceutical chemistry.[9,10]

Although the importance of adenylating enzymes has long been recognized, labour intensive specificity profiling hinders engineering efforts.[11] None of the available assays can detect adenylation activity under substrate competition which is critical for intracellular reactions. The widely used pyrophosphate (PP_i) exchange assay sensitively detects incorporation of radioactive ^{32}P-PP_i in the reverse reaction but handling of radioactivity and laborious sample work-up are major drawbacks.[11–13] Non-radioactive but still discontinuous is the mass-spectrometric detection of exchange between PP_i and γ-$^{18}O_4$-ATP.[14] Release assays detecting the liberation of PP_i in the forward reaction allow continuous data collection and are more convenient in terms of instrumentation and handling but suffer from low sensitivity and strong product inhibition.[15–18] In order to overcome inhibition by tightly binding aminoacyl-adenylates in release assays, hydroxylamine has been added as a quencher[19,20] and resulting hydroxamates have been detected as iron complexes.[21]

Independent Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans Knöll Institute (HKI Jena), Beutenbergstr. 11a, 07745 Jena, Germany. E-mail: hajo.kries@leibniz-hki.de

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9sc04222a

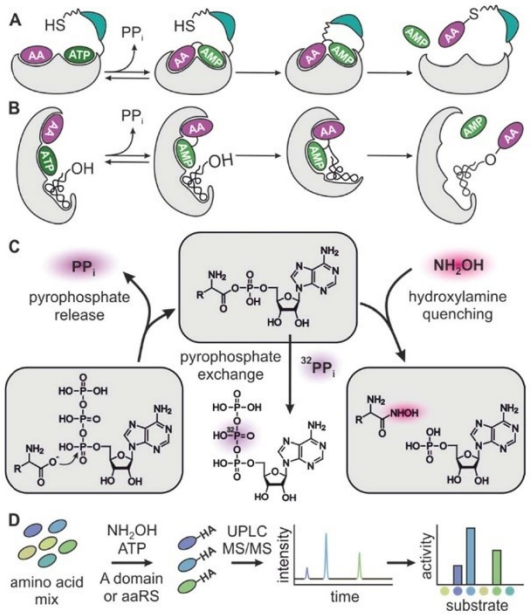


Fig. 1 (A) Adenylation domains and (B) aaRSs activate amino acids (AAs) for peptide synthesis. (C) Aminoacyl hydroxamate (HA) release after quenching of aminoacyl adenylates provides an alternative assay reaction to PP_i release and PP_i exchange. (D) Hydroxamates are quantified by UPLC-MS/MS to obtain full specificity profiles of adenylation enzymes in a single chromatographic run.

A salient feature of many nonribosomal adenylation domains is a high degree of substrate promiscuity which can serve as a springboard for natural enzyme evolution[22] and engineering in the laboratory.[23–28] Given the availability of vast numbers of structurally related amino acid substrates, adenylation enzymes are highly suitable for studying enzyme promiscuity.[29] However, specificity profiles of adenylating enzymes must be measured one substrate at a time[24,27,30] which poorly reflects the situation in a cell where numerous amino acids and carboxylic acids compete for the active site. Consequently, the natural promiscuity of adenylate forming enzymes is imperfectly approximated unless specificity constants (k_{cat}/K_M) are determined for each substrate in saturation kinetics.[31] It follows from an extension of Michaelis–Menten kinetics to competition conditions that product formation rates are proportional to the k_{cat}/K_M of the respective substrate times the substrate concentration (ESI eqn (1)–(3)†).[32] It is hence possible to predict the performance of an adenylation enzyme inside the cell from k_{cat}/K_M values and intracellular substrate concentrations.

Here, we present an adenylation assay that allows deduction of k_{cat}/K_M ratios from hydroxamate product concentrations after reacting a mixture of substrates in the presence of hydroxylamine. In the multiplexed hydroxamate assay (HAMA), product mixtures are disentangled *via* highly specific and sensitive LC-MS/MS. Since the experimental effort is minimal for determining a full specificity profile including dozens of substrates, this assay has great potential for exploring and evolving the promiscuity of adenylate forming enzymes.

Results and discussion

In order to test whether formation of hydroxamates in an adenylation reaction would yield meaningful specificity constants (k_{cat}/K_M), we used Phe specific TycA, the first module from tyrocidine synthetase, as a reference. In a thorough kinetic analysis based on PP_i exchange, full Michaelis–Menten kinetics have previously been determined for a range of substrates.[31] First, we confirmed that TycA would be stable in the presence of the hydroxylamine concentration (150 mM) required for efficient quenching.[33] TycA maintained full activity after treatment with hydroxylamine for up to one hour (ESI Fig. S1†). By using the established MesG/hydroxylamine assay which also relies on hydroxylamine quenching but detects released PP_i,[20] kinetic constants were determined for the substrates L-Phe, D-Phe, L-Trp, L-Tyr, L-Leu, L-Met, and L-Val (ESI Fig. S2†). The k_{cat}/K_M's for these substrates cover a range of five orders of magnitude. MesG/hydroxylamine and PP_i exchange assay yielded overall consistent results (ESI Table S1 and ESI Fig. S3†). Deviations could arise from intrinsic chemical preferences of the assay reactions – attack of hydroxylamine on the aminoacyl adenylate in one case and of pyrophosphate in the other. We conclude that hydroxamate formation is an informative parameter for the characterization of adenylation reactions.

A prerequisite of hydroxamate quantification by LC-MS/MS are standards for optimizing and calibrating hydroxamate detection. Hydroxamates of proteinogenic and three nonproteinogenic amino acids (phenyl-glycine, β-phenylalanine and pipecolic acid) were synthesized from corresponding methyl esters by treatment with hydroxylamine. Asparagine, glutamine and ornithine hydroxamates could not be obtained, presumably due to intramolecular cyclization and instability. Hydroxamate standards are available from the authors upon reasonable request. Hydrophilic interaction chromatography on a BEH-amide phase provided efficient separation of the highly polar compounds which showed little retention on reversed phase columns. Specific fragmentations of hydroxamates were detected by ESI-MS/MS on a Xevo TQ-S micro (Waters), with limits of quantitation in the range of 3 to 400 nM and a dynamic range of at least three orders of magnitude (ESI Table S2†). Detection of serine hydroxamate was hampered by isobaric, coeluting compounds present in the assay mixture. Ile and Leu hydroxamates coelute but were differentiated by addition of deuterium labelled Leu-d7. Similarly, addition of deuterium labelled isotopes allowed mass-differentiation of enantiomeric pairs of Phe and Val.

For HAMA, we performed reactions similar to the MesG/hydroxylamine assay but in the presence of a 1 mM substrate mixture containing all amino acids of interest. To avoid substrate depletion, reactions were run up to 10% conversion of the most active substrate. Under these conditions, the amounts of hydroxamates determined by LC-MS/MS should be proportional to the corresponding specificity constant k_{cat}/K_M of the amino acid substrate. Since the activity of TycA is *ca.* 10^3-fold larger for L-Phe ($k_{cat}/K_M = 1600\ mM^{-1}\ min^{-1}$; ESI Table S1†) than for the best alternative substrates, a second reaction without L-Phe was conducted with longer reaction time to bring the less active substrates into the quantifiable range. The hydroxamate profile obtained from two reactions and two chromatographic runs yielded results consistent with full saturation kinetics recorded with the MesG/hydroxylamine assay (Fig. 2A). While the detailed kinetic analysis with PP_i exchange and release assays is tedious and time consuming, a complete specificity profile with HAMA can be completed in less than one hour.

Based on the subtle, two atom difference between amino acids and amino acid hydroxamates, we hypothesized that hydroxamates would act as competitive inhibitors of adenylation. Indeed, we found that L-Phe hydroxamate (PheHA) is a weak competitive inhibitor of TycA with an inhibition constant ($K_i = 30$ μM) similar to the Michaelis constant ($K_M = 20$ μM) of L-Phe (Fig. 2B and ESI Fig. S2†). Notably, competitive inhibition is not expected to skew specificity profiles, since the preference of the free enzyme remains unaltered. Formation of the enzyme–inhibitor complex only diminishes the free enzyme concentration, and hence, the overall rate. As predicted, ratios of hydroxamates remained constant over the course of the reaction (ESI Fig. S4†). Determination of k_{cat} values for individual substrates is not the purpose of HAMA but if necessary, recording the time course of hydroxamate formation (ESI Fig. S5†) will ensure linear, initial velocity conditions without inhibition.

After validating HAMA on TycA, we proceeded to demonstrate its general applicability with a panel of NRPS modules encoding various known specificities. For this purpose, four

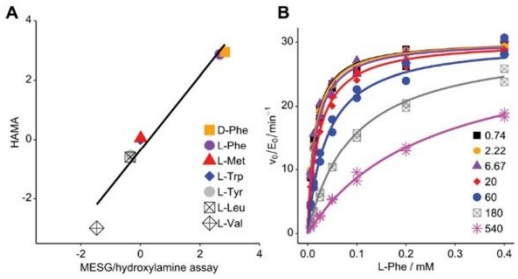


Fig. 2 (A) Rapid LC-MS/MS quantification of hydroxamates formed by Phe-activating NRPS module TycA in a competition reaction yields specificity data equivalent to saturation kinetics recorded with the MESG/hydroxylamine assay (ESI Table S1†; $R^2 = 0.959$; slope = 1.26 ± 0.12). Logarithms of hydroxamate concentrations obtained by HAMA are plotted against $\log(k_{cat}/K_M$ * mM min) values obtained with the MesG/hydroxylamine assay. All activities are relative to Trp. Reported activity for L-Ile (3.4 mM^{-1} min^{-1})[29] could not be detected in either assay format. (B) PheHA is a weak competitive inhibitor of TycA (K_i = 30.3 ± 1.4 μM). The inset shows PheHA concentrations in μM. Each point was measured as technical duplicate.

modules from the gramicidin S cluster ($GrsB1_{CAT}$ [L-Pro], $GrsB2_{CAT}$ [L-Val], $GrsB3_{CAT}$ [L-Orn or L-Lys], $GrsB4_{CATTe}$ [L-Leu])[34] and three modules from the surfactin A cluster (SrfA-$A1_{CAT}$ [L-Glu], SrfA-$B2_{CAT}$ [L-Asp], SrfA-C_{CATTe} [L-Leu])[35] were expressed in *Escherichia coli*, purified *via* nickel affinity chromatography and assayed with a mixture of 1 mM proteinogenic amino acids (Fig. 3A). The published specificities were correctly identified by HAMA while promiscuous side activities were minimal. Additionally, one uncharacterised AT domain from the jessenipeptin biosynthetic cluster from *Pseudomonas aeruginosa* QS1027,[36] supposed to activate L-Thr ($JesA1_{AT}$), was expressed and assayed, and the specificity profile clearly matched the expectation (Fig. 3A).

In ribosomal protein synthesis, aaRSs are key enzymes which have been thoroughly engineered aiming for expansion of the genetic code with unnatural amino acids.[3,4] To demonstrate the potential of HAMA for the analysis of this enzyme class, we expressed and profiled three aaRSs from *E. coli* (MetG, LeuS, HisS; Fig. 3B). Again, all three specificities were correctly identified and almost no side-activities detected, as expected for highly proficient enzymes from primary metabolism.

Production of large NRPS proteins in a pure form is notoriously difficult, but a meaningful specificity profile was also obtained by HAMA using an enzyme preparation (NRPS module JesA1) of low purity (Fig. 3A and ESI Fig. S6†). The small fraction of expressed enzyme generated enough hydroxamate to deduce specificity and the large quantity of unknown contaminants did not interfere with the assay. Possible contaminants that unavoidably disturb adenylation assays are other carboxylate activating enzymes present in every cell, such as aaRSs, precluding measurements of adenylation activity in cellular lysates. In our hands, a single purification step *via* nickel affinity chromatography effectively eliminated background activity arising from these enzymes.

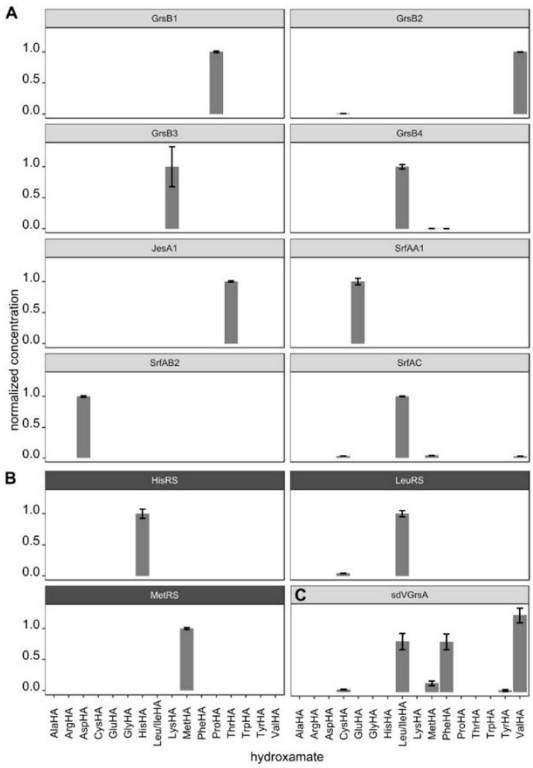


Fig. 3 (A) Specificity profiles of several NRPS modules and (B) aaRSs have been determined using the HAMA assay. (C) Low activity of the engineered NRPS module sdVGrsA[30] is detectable. Error bars indicate the standard deviation from three technical replicates.

In enzyme engineering, low activities are often encountered, which present an analytical challenge. We tested the applicability of HAMA on $sdVGrsA_{ATE}$, a chimeric and poorly active NRPS module with imperfect L-Val specificity (k_{cat}/K_M[Val] = 0.3 mM^{-1} min^{-1}) grafted by subdomain-swapping into L-Phe specific GrsA.[30] Despite the low activity, determination of the sdVGrsA specificity profile succeeded. Reported side-activities for L-Phe and L-Leu next to the designed L-Val activity were confirmed (Fig. 3C).

In order to improve activity of sdVGrsA, we performed a brief directed evolution experiment[37] and characterized selected mutants with HAMA. First, mutants were assessed based on the rate of Val-Pro diketopiperazine formation in a dimodular system together with GrsB1.[30] Three rounds of mutagenesis were targeted to the interface between the grafted subdomain and the surrounding protein. Subdomain residues were reverted to the identity of the corresponding residue in GrsA at 5 positions at the A–T domain interface (1st round) and at 12 positions in the hydrophobic core of the A domain (2nd round). In the 3rd round, beneficial mutations were combined, yielding sdVGrsA-STAP (D306S, N334T, S338A, A356P) which produced 6.2-fold more peptide than sdVGrsA after 3 h at 37 °C. However,

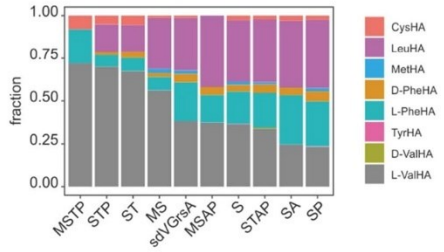


Fig. 4 Specificity profiles of sdVGrsA variants obtained in a directed evolution experiment, sorted by L-Val specificity (MSTP: G243M, D306S, N334T, A356P; STP: D306S, N334T, A356P; ST: D306S, N334T; MS: G243M, D306S; MSAP: G243M, D306S, S338A, A356P; S: D306S; STAP: D306S, N334T, S338A, A356P; SA: D306S, S338A; SP: D306S, A356P) were measured at 33 °C and 3 h.

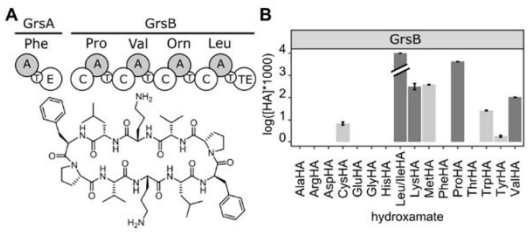


Fig. 5 (A) The tetramodular NRPS GrsB is part of the gramicidin S synthetase. (B) HAMA profile of GrsB. Turnover for Leu was markedly faster than for the other substrates and exceeded 10% conversion. Promiscuous activities are shown in light grey.

a HAMA profile of sdVGrsA-STAP revealed no improvement in substrate specificity compared to the broadly specific starting point sdVGrsA (Fig. 4). Another mutant, sdVGrsA-MSTP (G243M, D306S, N334T, A356P), showed 2-fold higher preference for Val (72%) at only 3-fold higher activity than sdVGrsA. Here, peptide yields in combination with comprehensive HAMA profiles allow to balance activity-specificity trade-offs and to find the most promising pathway for directed evolution.

Since HAMA resolves different products by mass, multimodular NRPSs encoding multiple adenylation activities on one protein can be measured. Heterologously expressed, four-modular GrsB from the gramicidin S synthetase was profiled (Fig. 5). Three out of the expected four hydroxamates (L-Pro, L-Val, L-Orn, L-Leu) were detected together with side activities, in particular for L-Lys which seems to be a good surrogate of L-Orn in GrsB3. L-Orn hydroxamate could not be quantified because synthesis of the standard failed, presumably due to its proclivity to cyclize. Indeed, Lys-containing gramicidin S analogues have been detected in the natural producer by mass spectrometry.[38] It should be noted that a rigorous prediction of intracellular NRPS activities has to take the intracellular amino acid concentrations into account, and these vary from low micromolar to low millimolar.[39]

Conclusions

In summary, we have demonstrated HAMA's utility in substrate profiling of nonribosomal A domains and aaRSs coming from various origins and encoding various specificities. HAMA delivers detailed specificity profiles under competition conditions in a short time and with minimal experimental effort. Being superior to previous assays in several aspects, HAMA has potential to serve as a reliable standard tool in the engineering and discovery of adenylating enzymes. When the goal is to verify the link between genes and natural products in NRPS discovery, HAMA can provide valuable data because it discerns product structures *via* MS fragmentation. Previous MS methods observed acylated natural product synthetases or fragments ejected from acylated prosthetic groups.[40] In contrast to these methods, HAMA calibrated with chemical standards has allowed accurate and highly sensitive quantification of weak, promiscuous activities. Structural information about the products not only alleviates problems caused by common sample contaminants such as amino acids, enzymes, or phosphate but also allows to resolve activities of multimodular NRPSs (Fig. 5). In the age of exploding sequence databases, substrate predictions for NRPSs obtained through HAMA could contribute to the deorphanization of biosynthetic pathways. By resolving multimodular activities, HAMA could also help to better understand the context dependence observed for NRPS substrate preference.[24,41] Efficient substrate profiling will especially benefit screening and directed evolution efforts addressing the specificity of adenylating enzymes. In summary, HAMA offers a new analytical tool to several fields of research where adenylate-forming enzymes play important roles and will potentially expedite the development of tailored proteins and life-saving drugs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We are grateful to Daniel Braga for performing high-resolution MS measurements, to Heike Heinecke and Luka Raguž for help with NMR measurements, and to Oliver Waldmann, Philipp Stephan, and Hsin-Mei Huang for help with protein purification (all HKI Jena). We thank Linda Schlittenbauer (Waters GmbH) for advice on LC-MS/MS method development. This work has been supported by the Daimler und Benz Stiftung (HK). We further acknowledge donation of *Pseudomonas jessenii* genomic DNA from the Stallforth Lab (HKI Jena), and donation of plasmid pTrc99a_GrsBMtoL from David Niquille and Donald Hilvert (ETH Zurich) under an MTA.

Notes and references

1 F. Kudo, A. Miyanaga and T. Eguchi, *J. Ind. Microbiol. Biotechnol.*, 2018, **46**, 515–536.

2 S. Schmelz and J. H. Naismith, *Curr. Opin. Struct. Biol.*, 2009, **19**, 666–671.
3 C. Fan, J. M. L. Ho, N. Chirathivat, D. Söll and Y. S. Wang, *ChemBioChem*, 2014, **15**, 1805–1809.
4 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.
5 A. S. Brown, M. J. Calcott, J. G. Owen and D. F. Ackerley, *Nat. Prod. Rep.*, 2018, **35**, 1210–1228.
6 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chem. Sci.*, 2015, **6**, 50–69.
7 H. Neumann, K. Wang, L. Davis, M. Garcia-Alai and J. W. Chin, *Nature*, 2010, **464**, 441–444.
8 H. Kries, R. Wachtel, A. Pabst, B. Wanner, D. Niquille and D. Hilvert, *Angew. Chem., Int. Ed.*, 2014, **53**, 10105–10108.
9 M. Petchey, A. Cuetos, B. Rowlinson, S. Dannevald, A. Frese, P. W. Sutton, S. Lovelock, R. C. Lloyd, I. J. S. Fairlamb and G. Grogan, *Angew. Chem., Int. Ed.*, 2018, **57**, 11584–11588.
10 R. Hara, K. Hirai, S. Suzuki and K. Kino, *Sci. Rep.*, 2018, **8**, 2950.
11 A. Stanišić and H. Kries, *ChemBioChem*, 2019, **20**, 1347–1356.
12 M. P. Stulberg and G. D. Novelli, *Methods Enzymol.*, 1962, **5**, 703–707.
13 L. G. Otten, M. L. Schaffer, B. R. M. Villiers, T. Stachelhaus and F. Hollfelder, *Biotechnol. J.*, 2007, **2**, 232–240.
14 V. V. Phelan, Y. Du, J. A. McLean and B. O. Bachmann, *Chem. Biol.*, 2009, **16**, 473–478.
15 T. Kittilä, M. Schoppet and M. J. Cryle, *ChemBioChem*, 2016, **17**, 576–584.
16 H. Katano, H. Watanabe, M. Takakuwa, C. Maruyama and Y. Hamano, *Anal. Sci.*, 2013, **29**, 1095–1098.
17 H. Katano, R. Tanaka, C. Maruyama and Y. Hamano, *Anal. Biochem.*, 2012, **421**, 308–312.
18 T. J. McQuade, A. D. Shallop, A. Sheoran, J. E. Delproposto, O. V. Tsodikov and S. Garneau-Tsodikova, *Anal. Biochem.*, 2009, **386**, 244–250.
19 E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. Leigh Allen, G. Skiniotis, C. C. Aldrich and A. M. Gulick, *Nature*, 2016, **529**, 235–238.
20 D. J. Wilson and C. C. Aldrich, *Anal. Biochem.*, 2010, **404**, 56–63.
21 R. Hara, R. Suzuki and K. Kino, *Anal. Biochem.*, 2015, **477**, 89–91.
22 S. D. Copley, *Curr. Opin. Struct. Biol.*, 2017, **47**, 167–175.
23 H. Kaljunen, S. H. H. Schiefelbein, D. Stummer, S. Kozak, R. Meijers, G. Christiansen and A. Rentmeister, *Angew. Chem., Int. Ed.*, 2015, **54**, 8833–8836.
24 S. Meyer, J. C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth and E. Dittmann, *Cell Chem. Biol.*, 2016, **23**, 462–471.
25 J. W. Han, E. Y. Kim, J. M. Lee, Y. S. Kim, E. Bang and B. S. Kim, *Biotechnol. Lett.*, 2012, **34**, 1327–1334.
26 M. Zhu, L. Wang and J. He, *ACS Chem. Biol.*, 2019, **14**, 256–265.
27 F. Ishikawa, A. Miyanaga, H. Kitayama, S. Nakamura, I. Nakanishi, F. Kudo, T. Eguchi and G. Tanabe, *Angew. Chem., Int. Ed.*, 2019, **58**, 6906–6910.
28 K. Throckmorton, V. Vinnik, R. Chawdhury, T. Cook, M. G. Chevrette, C. Maranas, B. Pfleger and M. G. Thomas, *ACS Chem. Biol.*, 2019, **14**(9), 2044–2054.
29 B. Villiers and F. Hollfelder, *Chem. Biol.*, 2011, **18**, 1290–1299.
30 H. Kries, D. L. Niquille and D. Hilvert, *Chem. Biol.*, 2015, **22**, 640–648.
31 B. R. M. Villiers and F. Hollfelder, *ChemBioChem*, 2009, **10**, 671–682.
32 H. Bisswanger, *Enzyme Kinetics*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2008.
33 B. P. Duckworth, D. J. Wilson and C. C. Aldrich, *Measurement of Nonribosomal Peptide Synthetase Adenylation Domain Activity Using a Continuous Hydroxylamine Release Assay*, Springer New York, New York, NY, 2016, vol. 1401.
34 J. Kratzschmar, M. Krause and M. A. Marahiel, *J. Bacteriol.*, 1989, **171**, 5422–5429.
35 B. Kluge, J. Vater, J. Salnikow and K. Eckart, *FEBS Lett.*, 1988, **231**, 107–110.
36 J. Arp, S. Götze, R. Mukherji, D. J. Mattern, M. García-Altares, M. Klapper, D. A. Brock, A. A. Brakhage, J. E. Strassmann, D. C. Queller, B. Bardl, K. Willing, G. Peschel and P. Stallforth, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 3758–3763.
37 F. H. Arnold, *Angew. Chem., Int. Ed.*, 2018, **57**, 4143–4148.
38 K. Hori and T. Kurotsu, *J. Biochem.*, 1997, **122**, 606–615.
39 B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien and J. D. Rabinowitz, *Nat. Chem. Biol.*, 2009, **5**, 593–599.
40 P. C. Dorrestein, J. Blackhall, P. D. Straight, M. A. Fischbach, S. Garneau-Tsodikova, D. J. Edwards, S. McLaughlin, M. Lin, W. H. Gerwick, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 1537–1546.
41 A. Degen, F. Mayerthaler, H. D. Mootz and B. Di Ventura, *Sci. Rep.*, 2019, **9**, 5119.

Supplementary Information

HAMA: A multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes

Aleksa Stanišić, Annika Hüsken and Hajo Kries*

Independent Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans Knöll Institute (HKI Jena), Beutenbergstr. 11a, 07745 Jena, Germany

Contents

Synthesis of amino acid hydroxamates

General procedure

All reagents, amino acid methyl esters and solvents were obtained from commercial suppliers and used without further purification. Amino acid hydroxamates (XaaHAs) were synthesized by treating corresponding amino acid methyl esters with hydroxylamine according to previously published protocols.[1–7] Amino acid methyl ester hydrochloride (0.25-1 g) was dissolved in 10-15 mL MeOH and neutralized by careful dropwise addition of one equivalent of 0.6 M KOH in methanol while stirring on ice. The solution was filtered through a teflon 0.24 µm filter (Labsolute) to remove the precipitated KCl. A solution of hydroxylamine (1 M, 200 mL) was prepared freshly by mixing 140 mL of 1.43 M hydroxylamine hydrochloride solution in methanol with 60 mL of 3.33 M KOH solution in methanol with vigorous stirring on ice. After 30 min, solution was filtered to remove precipitated KCl. Calculated volume of neutralized hydroxylamine solution was added to the neutralized amino acid ester solution up to a final molar ratio of ester and hydroxylamine of 1:6. Reactions were stored at 4°C without stirring to facilitate crystallization. The formation of hydroxamates was detected by formation of a colored Fe^{3+} complex with 3% $FeCl_3$ in 0.1 M perchloric acid in ethanol. In general, hydroxamates of nonpolar amino acids crystallized spontaneously from the reaction mixture after 1-7 days, while polar ones required the addition of organic solvents. Precipitate was filtered, washed with dry methanol, dried under vacuum and stored at -20°C. Yields of hydroxamates were typically low (<20 %) due to the crystallization conditions which were not optimized.

Identity of hydroxamates was confirmed by high resolution mass spectrometry (HRMS; SI Table 3) and NMR (Section 0).

Hydroxamate	Synthetic procedure	NMR shifts of impurities
GlyHA[1,4]	Methyl ester was found to be prone to hydrolysis to free acid in alkaline hydroxylamine solution. Therefore, an incompletely neutralized hydroxylamine solution was used, which was prepared by mixing 13.9 g of hydroxylamine hydrochloride with 8.9 g KOH in 200 mL of methanol.	δ_H 4.02, s (presumably alpha proton of O-glycyl hydroxylamine, 20%)
AlaHA[1]	Reaction mixture concentrated to half the volume to facilitate crystallization. Crystallized after 2 days of storage at 4°C.	δ_H 3.29, s (methanol)
SerHA[4]	Reaction evaporated to dryness, dissolved in methanol and filtered to remove KCl. Diethylether (DET) was added with stirring until the solution turned cloudy. After 30 minutes of stirring, the solution cleared leaving	

	resin on the flask wall. Resin was washed with DET, dried and stored at 4°C.	
ThrHA	Reaction evaporated to dryness and redissolved in a small amount of methanol. Methyl tert-butyl ether (TBME) was added while stirring until the solution turned cloudy. After 15 min of stirring, ThrHA precipitated as a white, hygroscopic solid which was carefully filtered, washed with TBME and dried.	δ_H 3.32, s (methanol) δ_H 1.19, s; 3.20, s (MTBE) δ_C 27.4; 50.4 (MTBE)
Cys_2HA	Cys methyl ester was found to be unstable in alkaline reaction conditions. Therefore, cystine methyl ester was employed for the synthesis of the hydroxamate. Cystine methyl ester dihydrochloride was neutralized with 2 equivalents of KOH and treated with 12 equivalents of hydroxylamine. After two days, reaction was concentrated to half volume and left to precipitate at 4°C overnight. Cysteine hydroxamate is prepared by reducing cystine hydroxamate with 3 equivalents of tris(2-carboxyethyl)phosphine hydrochloride in water at 60°C for 10 min.	δ_H 3.89, s (amino acid methyl ester)
ValHA[1]	Precipitated from the reaction mixture after 3 days of storage at 4°C.	
LeuHA[5]	Precipitated from the reaction mixture after 7 days of storage at 4°C.	δ_H 4.01, dd (alpha proton of free amino acid, 10%)
IleHA[1,5]	Precipitated from the reaction mixture after 7 days of storage at 4°C.	δ_H 3.98, d (alpha proton of free amino acid, 20%) δ_C 172.9 (alpha carbon of free amino acid)
MetHA[5]	Precipitated from the reaction mixture after 2 days of storage at 4°C.	δ_H 4.20, dd (alpha proton of free amino acid, 6%) δ_C 173.0 (alpha carbon of free amino acid)
ProHA*HCl	Reaction evaporated to dryness and redissolved in a small amount of methanol while heating to 60°C. Solution acidified with concentrated HCl under vigorous stirring. DET was added to the solution until it turned cloudy. After 30 min of stirring at room temperature, the hydroxamate precipitated as translucent resin on the flask wall. Solvent was decanted, resin washed with DET, dissolved in methanol, filtered to remove KCl and dried.	δ_H 1.13, d (isopropanol) δ_H 1.89, s (ethyl acetate) δ_H 3.32, s (methanol) δ_H 4.42, dd (alpha proton of free amino acid, 12%)
PheHA[1,2,6,7]	Precipitated from the reaction mixture after 2 days of storage at 4°C.	δ_H 2.50, s (DMSO) δ_C 39.5 (DMSO)
TyrHA[6,7]	Methyl ester was provided as a free base, so the KOH neutralization step was omitted and the ester dissolved directly in hydroxylamine solution. Precipitated from the reaction mixture after 2 days of storage at 4°C.	δ_H 3.30, s (methanol) δ_C 50.3 (methanol)
TrpHA[6]	Reaction evaporated to dryness, redissolved in isopropanol and heated to 60°C until a rose-white precipitate formed which was filtered, washed with isopropanol and dried.	δ_H 1.10, d; 3.96, sept (isopropanol) δ_H 4.35, dd (alpha proton of free amino acid, 9%)

		δ_C 173.2 (alpha carbon of free amino acid)
AspHA	Methyl ester provided as a free base, so the KOH neutralization step was omitted. Ester was dissolved directly in methanolic hydroxylamine solution. Asp methyl ester was found to be prone to hydrolysis to free acid in alkaline hydroxylamine solution. Therefore, an incompletely neutralized hydroxylamine solution was used, which was prepared by mixing 13.9 g of hydroxylamine hydrochloride with 8.9 g KOH in 200 mL of methanol.	δ_H 3.85, s (amino acid methyl ester) δ_H 4.40, dd (alpha proton of free amino acid, 8%) δ_H 4.47, dd (alpha proton of amino acid methyl ester, 3%)
GluHA	Methyl ester was provided as a free base, so the KOH neutralization step was omitted. Ester was dissolved directly in methanolic hydroxylamine solution. Water added dropwise until complete dissolution and the reaction mixture stored at 4°C.	δ_H 3.34, s (methanol) δ_H 4.40, dd (alpha proton of free amino acid, 6%)
HisHA[6]	Reaction evaporated to dryness and redissolved in a small amount of methanol. Isopropanol was added to the solution until a white, extremely hygroscopic precipitate formed which was filtered, washed with isopropanol and dried.	δ_H 1.17, d; 4.02, sept (isopropanol) δ_H 3.35, s (methanol) δ_H 4.40, dd (alpha proton of free amino acid, 7%) δ_C 25.1; 65.6 (isopropanol)
LysHA	Precipitated from the reaction mixture after 2 days of storage at 4°C.	δ_H 4.04, dd (alpha proton of O-lysyl hydroxylamine, 5%) δ_H 4.48, dd (alpha proton of free amino acid, 7%)
ArgHA*2HCl	Reaction evaporated to dryness and redissolved in a small amount of methanol. Solution acidified with concentrated HCl with vigorous stirring. Isopropanol was added to the solution until a white, extremely hygroscopic precipitate formed which was filtered, washed with isopropanol and dried.	δ_H 1.13, d; 3.98, sept (isopropanol) δ_H 3.32, s (methanol)
PipHA	Reaction evaporated to dryness, dissolved in methanol and filtered to remove KCl, DET was added with stirring until the solution turned cloudy. After 30 minutes of stirring, the solution cleared leaving resin on the flask wall. Resin was washed with DET, dried and stored at 4°C.	δ_H 3.33, s (methanol)
Phenylglycine-HA[2]	Precipitated from the reaction mixture after 2 days of storage at 4°C.	δ_H 5.02, dd (alpha proton of O-phenylglycyl hydroxylamine, 3%) δ_H 5.42, dd (alpha proton of free amino acid, 11%)
β-PheHA	Reaction evaporated to dryness and redissolved in a small amount of methanol while heating to 60°C. Solution acidified with concentrated HCl under vigorous stirring. Diethylether (DET) was added to the solution until it turned cloudy. After 30 min of stirring at room temperature, the hydroxamate	δ_H 3.19, s (methanol)

	precipitated as translucent resin on the flask wall. Solvent was decanted, resin washed with DET, dissolved in methanol, filtered to remove KCl and dried.	

Preparation and storage of standard solutions

Individual hydroxamates are stored as 10 mM solutions in 20 mM HCl at -20°C. Very hygroscopic compounds (ArgHA, HisHA, ProHA, SerHA, ThrHA, PipHA, β-PheHA) are stored as 50 mM solutions in water at -20°C. The quantitation standard of amino acid hydroxamates is stored as acidic solution in water: 0.3 mM hydroxamates, 10 mM tris-(2-carboxyethyl) phosphine hydrochloride (TCEP) at -20°C. On the day of the analysis, the standard solution is diluted to 100 µM final concentration in 50 mM TRIS (pH 7.5), 150 mM hydroxylamine (pH 7.5-8), 5 mM ATP. This solution is diluted with the buffer containing assay components (50 mM TRIS [pH 7.5], 150 mM hydroxylamine [pH 7.5-8], 5 mM ATP) to obtain standard solutions (0.032-100 µM) mimicking the assay conditions. All standards are further diluted 10-fold in 95% acetonitrile + 0.1% formic acid before UPLC-MS/MS analysis (0.0032-10 µM). Diluted hydroxamate standards are freshly prepared and used in the course of one day.

UPLC-ESI-HRMS analysis of amino acid hydroxamates

Exact masses of the synthetic hydroxamates were confirmed by high resolution mass spectrometry (**SI Table 3**) on a Dionex Ultimate3000 system combined with a Q-Exactive Plus mass spectrometer (Thermo Scientific) with a heated electrospray ion source (HESI). All masses were detected by ESI as $M+H^+$ adducts in positive mode. The measurement was carried out within a mass range of m/z 50 – 400

Enzymatic product formation under competition is governed by the specificity constant k_{cat}/K_M

The relative product formation rates for two substrates can be derived under steady state conditions in analogy to the Michaelis-Menten equation.[8] The product is formed from the corresponding Michaelis complexes in an irreversible, monomolecular reaction (Eq. 1.1 and 1.2). Under the assumption that the concentrations of both Michaelis complexes remain constant, their concentration can be expressed as a function of the Michaelis constants (e.g. $K_{M1} = [k_{-1} + k_2]/k_1$) and the substrate concentrations (Eq. 2.1 and 2.2). Inserting Eq. (2) into Eq. (1) results in Eq. (3) which describes the ratio of product formation rates which are proportional to the corresponding specificity constants k_2/K_{M1} and k_4/K_{M2} multiplied with the respective substrate concentrations.

$$E + A_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EA_1 \overset{k_2}{\rightarrow} E + P_1$$

$$E + A_2 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} EA_2 \overset{k_4}{\rightarrow} E + P_2$$

$$\frac{d[P_1]}{dt} = k_2[EA_1] = v_1 \tag{1.1}$$

$$\frac{d[P_2]}{dt} = k_4[EA_2] = v_2 \tag{1.2}$$

$$[EA_1] = \frac{K_{M2}[E]_0[A_1]}{K_{M2}[A_1]+K_{M1}[A_2]+K_{M1}K_{M2}} \tag{2.1}$$

$$[EA_2] = \frac{K_{M1}[E]_0[A_2]}{K_{M2}[A_1]+K_{M1}[A_2]+K_{M1}K_{M2}} \tag{2.2}$$

$$\frac{v_1}{v_2} = \frac{\frac{k_2}{K_{M1}}[A_1]}{\frac{k_4}{K_{M2}}[A_2]} \tag{3}$$

Cloning

General cloning

General cloning was carried out in *E. coli* strain NEB 5-alpha (New England Biolabs). Protein expression was carried out in *E. coli* strains NEB BL21 or HM0079.[9] Preparation of plasmid DNA, gel purification of DNA fragments, and purification of PCR products were performed using NucleoSpin Plasmid and Gel and PCR clean-up kits (Macherey Nagel). Purification of the genomic DNA was performed according to a published protocol.[10] PCRs were carried out with Q5 polymerase (New England Biolabs, Massachusetts) or Phusion High-Fidelity DNA Polymerase (New England Biolabs), according to the supplier's instructions. PCR fragments carrying vector-specific overhangs were cloned into vectors linearized by restriction digestions using the InFusion cloning kit (Takara Bio Europe). Oligonucleotide primers (Section 0) were made by custom synthesis and sequence confirmation of PCR amplified inserts was performed using the Mix2Seq service for Sanger sequencing (Eurofins Genomics).

Plasmids

pSU18 and pTrc99a vectors[9] were linearized with NcoI and BamHI while pOPINE[11] was linearized with NcoI and PmeI restriction enzymes. pSU18-TycA, pSU18-sdVGrsA,[12] pMG211-Sfp[13] and pTrc99a-GrsB_MtoL[14] plasmids were kindly provided by Prof. Donald Hilvert (ETH Zurich). The gene encoding the A-T didomain of the Jes-A1 module was amplified from *Pseudomonas aeruginosa* QS1027 genomic DNA[15] and cloned into pTrc99a. Genes encoding the SrfA-C, SrfA-A1 and SrfA-B2 modules were amplified as C-A-T constructs (SrfA-C as C-A-T-Te) by PCR from *Bacillus subtilis* 3610 genomic DNA and cloned into pTrc99a. The genes encoding all four GrsB modules were amplified as C-A-T constructs (GrsB4 as C-A-T-Te) from pTrc99a-GrsB_MtoL. *grsB1* was subcloned into pTrc99a, while *grsB2*, *grsB3* and *grsB4* were subcloned into pSU18. Aminoacyl-tRNA synthetase genes (*hisS*, *leuS*, *metG*) were amplified from *E. coli* NEB 5-alpha genomic DNA and cloned into the pOPINE vector.

To generate mutants of sdVGrsA for the directed evolution experiment, two fragments of *sdVgrsA* were amplified from pSU18-sdVGrsA using mutagenic primers and cloned into pSU18-sdVGrsA linearized with AflII and SacI. The first fragment was amplified with primer sdXGrsA_f and a suitable reverse primer. The second fragment was amplified with a mutagenic forward primer, e.g. D306S_f, and sdXGrsA_r.

Oligonucleotides used as primers

Overhangs for InFusion cloning are underlined

SRFA-A1_F CAA TTT CAC ACA GGA AAC AGA CCA TGT TAA CGG ATG CAC AAA AAC GA
SRFA-A1_R TGG TGA TGG TGA TGA GAT CTG GAT CCT TCC TCT GCA AGA GCC GTA ATC
SRFA-B2_F CAA TTT CAC ACA GGA AAC AGA CCA TGA AGG AGG AGC AGA CGT TTG AA
SRFA-B2_R TGG TGA TGG TGA TGA GAT CTG GAT CCA GCA GAC GCC TCC ATA TAA GC
JESA1_F CAA TTT CAC ACA GGA AAC AGA CCA TGC TCA ATG CCA GCG AAA CCG CG
JESA1_R GGT GAT GGT GAT GAG ATC TGG ATC CAA TCT CGC CGC CCT TGC CAC
GRSB1_F ATT TCA CAC AGG AAA CAG ACC ATG AGT ACA TTT AAA AAA GAA CAT GTT CAG G
GRSB1_R TGG TGA TGA GAT CTG GAT CCC CCG TTT ATA TAA TTA GAG ATT TCC TGA ATG G
HISS_F AGG AGA TAT ACC ATG GCA AAA AAC ATT CAA GCC A
HISS_R GTG ATG GTG ATG TTT ACC CAG TAA CGT GCG CA
METG_F AGG AGA TAT ACC ATG ACT CAA GTC GCG AAG AAA ATT C
METG_R GTG ATG GTG ATG TTT TTT CAC CTG ATG ACC CGG T
LEUS_F AGG AGA TAT ACC ATG CAA GAG CAA TAC CGC C
LEUS_R GTG ATG GTG ATG TTT GCC AAC GAC CAG ATT GAG G
GRSB2_F CAA TTA AGG AGG CAG CAG ATG ATT CAG CCT GTA CCA GAA CAA
GRSB2_R GTG ATG GTG ATG AGA TCT GGA TCC ATC AGC AAT GTA TTG AGC TAA TG
SRFA-C_F ATT TCA CAC AGG AAA CTC GAG ATG AGT CAA TTT AGC AAG GAT CAG G
SRFA-C_R TGG TGA TGA GAT CTG GAT CCT GAA ACC GTT ACG GTT TGT GTA TTA AG
GRSB3_F CAA TTA AGG AGG CAG CAG ATG ATT CAA CCT GTT ACC CCG
GRSB3_R GTG ATG GTG ATG AGA TCT GGA TCC CTC CTC TAT ATA TTT AGC CAG TCC
GRSB4_F CAA TTA AGG AGG CAG CAG ATG GCT ATT CAG CCG GT
GRSB4_R CTT AGT GAT GGT GAT GGT GA

Primers for mutagenesis of sdVGrsA:

SDXGRSA_F GAG CAT AAA GGA ATA AGT AAT CTT AAG G
D306S_F CTT CGC TCC CTA ATT GTA GGT GGA AGC GCC TTG TCT CCG AAA CAC ATC
G243M_F CGT ATA ATA CAG ACC GGA GCA ATT GGA TTC GAT GCA CTG ACA TTT GAA GTT TTT ATG TCA TTG CTG CAT GGA GCT GAA TTG
N334T_F GAA CGG TTA CGG CCC AAC AGA AAC CAC CAC TTT TTC TAC ATG CTT TCT TAT TGA TAA AG
N334T_S338A_F GAA CGG TTA CGG CCC AAC AGA AAC CAC CAC TTT TGC GAC ATG CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TC
S338A_F GAA CGG TTA CGG CCC AAC AGA AAA CAC CAC TTT TGC GAC ATG CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA
TAT TC
A356P_F CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TCC GAT AGG GAA GCC GAT TCA AAA TAC ACA AAT TTA TAT TGT CGA TGA TGA AAA TCT TC
D306_R CCA CCT ACA ATT AGG GAG CGA AGG C
G243_R GCT CCG GTC TGT ATT ATA CGA TCG
N334_R CTG TTG GGC CGT AAC CGT TCC
A356_R GTC ATC ATA TTC TTT ATC AAT AAG AAA GCA TGT AG
SDXGRSA_R GCT AAC CCT TCT CCA CCA ATA CAG

Protein overexpression and purification

Purification protocol

For the overexpression of C-terminally His_6-tagged holo-NRPS proteins, each overexpression plasmid was transformed into *E. coli* HM0079 with genomically integrated 4'-phosphopantheteinyl transferase Sfp.[9] Overexpression of apo-TycA, Sfp and aminoacyl tRNA synthetases was done in *E. coli* BL21 strain. A 2 L flask with 500 mL of 2xYT medium supplemented with antibiotics was inoculated with 0.5 mL of an overnight culture and incubated at 37°C in a rotary shaker at 200 rpm. When the OD_{600} reached 1, cultures were induced with 0.25 mM isopropyl-D-thiogalactoside (IPTG) and grown for another 16-20 hours at 20°C. Cells were harvested by centrifugation and the supernatant was discarded. After resuspending the cell pellet in 30 mL lysis buffer (50 mM TRIS [pH 7.4], 500 mM NaCl, 20 mM imidazole, 2 mM TCEP), 100 µL protease inhibitor mix (Sigma, P8849) were added and cells were lysed by sonication while cooling on ice. The lysate was cleared by centrifugation at 19,000 g for 30 min at 4°C and the supernatant was loaded onto a column packed with 2 mL of Ni-IDA suspension (Rotigarose, Roth) and equilibrated with lysis buffer. After washing the column twice with 20 mL of the lysis buffer, the target protein was eluted with 4 x 0.75 mL elution buffer (50 mM TRIS [pH 7.4], 500 mM NaCl, 300 mM imidazole, 2 mM TCEP). After pooling the protein-containing fractions, they were buffer exchanged with 2-fold concentrated adenylation assay buffer (100 mM TRIS [pH 7.6], 10 mM $MgCl_2$) on 6 mL Vivaspin (Sartorius) filters with 10 kDa cut-off for proteins larger than 30 kDa and 30 kDa cut-off for proteins larger than 90 kDa. Glycerol was added to 10% and protein concentration adjusted to 50 µM. Samples were flash frozen in liquid nitrogen and stored at -20°C. For detailed kinetic analysis, TycA protein samples were further purified by anion exchange chromatography on an NGC Chromatography system (Bio-Rad Laboratories) using a MonoQ 5/50 GL column (GE Healthcare) and eluting with a 20-600 mM NaCl in 20 mM TRIS (pH 8) gradient. Purified protein was washed and prepared for storage as described above. Protein concentrations were determined from the absorbance at 280 nm measured in Take3 plates on an Epoch2 microplate reader (Biotek) using calculated extinction coefficients (www.benchling.com).

SDS-PAGE of overexpressed proteins

Purity of proteins was monitored by SDS-PAGE (SI Fig. 6) using Bolt 4-12% Bis-Tris Plus Gels (ThermoFisher Scientific) with MES-SDS running buffer (Novex). Sample load was 0.3-0.6 µg of protein per lane in Bolt LDS sample buffer and Bolt reducing agent. Triple Color Protein Standard III (Serva) was run alongside the protein samples as

a size standard. The gels were run at 200 V for 22 minutes and stained with Quick Coomassie stain (Serva).

MesG/hydroxylamine spectrophotometric assay

Michaelis-Menten kinetics of TycA

Michaelis-Menten parameters of the adenylation reaction catalyzed by TycA were determined from kinetic data recorded with the MesG/hydroxylamine assay which was performed as described previously with minor modifications.[16] Reactions contained 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 100 µM 7-methylthioguanosine (MesG), 150 mM hydroxylamine (adjusted to pH 7.5-8 with NaOH), 5 mM ATP (A2383, Sigma), 1 mM TCEP, 0.4 U/mL inorganic pyrophosphatase (I1643, Sigma), 1 U/mL of purine nucleoside phosphorylase from microorganisms (N8264, Sigma) and varying amounts of TycA (0.025 – 1 mM) and substrates. In flat-bottom 384-well plates (781620, Brand) 100 µL reactions were started by addition of substrate and the absorbance was followed at 355 nm on a Synergy H1 (BioTek) microplate reader at 30°C. Background activity was recorded in wells containing buffer without substrate and the obtained slopes were subsequently subtracted. Each substrate concentration was measured in duplicate. Initial velocities were divided by the slope of a pyrophosphate calibration curve to obtain the pyrophosphate release rate. Initial velocities $v_0/[E_0]$ were fit to the Michaelis-Menten equation by nonlinear regression using R version 3.4.2 (SI Fig. 2).[17]

Competitive inhibition of TycA with PheHA

For characterizing competitive inhibition of TycA by PheHA, complete L-Phe kinetic profiles were measured at varying PheHA concentrations (0.74 to 540 µM; Fig. 2B) as described for simple Michaelis-Menten kinetics (Section 5.1). Initial velocities v obtained for all combinations of substrate and inhibitor concentrations ([S] and [I]) were fit globally to a competitive inhibition model in R using nonlinear regression and plotted using ggplot2:[17]

```
 dat <- read.table("data.csv", sep=';', header=T)      #data input with
                                                        headers S, I, v
start <- list (kcat=30, Ki=20, Km=0.02)         #starting values
f <- v ~ kcat * S / ((1 + I/Ki)*Km + S)         #kinetic model
m <- nls(f,dat,start=start)                     #nonlinear regression
summary(m)                                      #output of fit param.
```

TycA stability in 150 mM hydroxylamine

To test the stability of TycA in the presence of 150 mM hydroxylamine, a 10 µM enzyme solution containing 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$ and 150 mM hydroxylamine was incubated at room temperature for up to one hour. After the indicated time, initial adenylation velocities were measured with the preincubated enzyme and three different L-Phe concentrations using the MESG/hydroxylamine assay (SI Fig. 1).

Multiplexed hydroxamate assay (HAMA)

Reaction conditions

The hydroxamate formation assay was conducted at room temperature in 100 µL volume containing 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 150 mM hydroxylamine (pH 7.5-8, adjusted with NaOH), 5 mM ATP (A2383, Sigma), 1 mM TCEP and varying concentrations of enzyme. Reactions were started by adding a mix of 5 mM proteinogenic amino acids in 100 mM TRIS (pH 8) to a final concentration of 1 mM or only buffer as a control. For TycA and sdVGrsA assays, L-Phe, L-Val and L-Leu were distinguished from D-Phe, D-Val and L-Ile, respectively by using enantiopure, deuterium labelled standards. Reaction times and temperatures were optimized for each protein. Reactions were quenched at different time points by diluting them 10-fold with 95% acetonitrile in water containing 0.1 % formic acid and submitted to UPLC-MS analysis. Time point t_0 was obtained by quenching the enzyme containing master mix before adding amino acid substrates. To guarantee initial velocity conditions, reactions were quenched before 10% (100 µM) of the most preferred substrate was consumed. We observed a strong impact of sample composition on HILIC separation of hydroxamates. Therefore, care had to be taken that all samples were processed in exactly the same manner without further dilutions, for instance. TycA assays were done in a biological (different enzyme batches) and technical (separate assay reactions) triplicates. Other proteins were assayed from a single protein batch in technical triplicates.

UPLC-MS/MS conditions

Chromatography was performed on a Waters ACQUITY H-class UPLC system (Waters) with an injection volume of 3 µL. Water with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B) were used as strong and weak eluent, respectively. Amino acid hydroxamates were separated on the ACQUITY UPLC BEH Amide column (1.7 µm, 2.1 x 50 mm) with a linear gradient of 10-50% A over 5 min (flow rate 0.4 mL/min) followed by 4 min reequilibration. Water containing 0.1% formic acid was used as a needle wash between the samples. Data acquisition and quantitation were done using the MassLynx and TargetLynx software (version 4.1).

MS/MS analyses were performed on Xevo TQ-S micro (Waters) tandem quadrupole instrument with ESI ionisation source in positive ion mode. Nitrogen was used as a desolvation gas and argon as collision gas. The following source parameters were used: capillary voltage 1.5 kV, cone voltage 65 V, desolvation temperature 500°C, desolvation gas flow 1000 L/h. Compounds were detected via specific mass transitions recorded in multiple reaction monitoring (MRM) mode (SI Table 2).

Standard calibration solutions of hydroxamates were prepared ranging from 0.0032 to 10 µM. In general, detection is very sensitive, limits being in the low nanomolar range. However, at such low concentrations, large loss of the linearity of the response was observed. Therefore, here we are defining limits of quantitation (LOQ) as the lowest concentrations of hydroxamate standards at which the signal response was still linear ($R^2 > 0.95$, deviation < 20%). The upper limit of quantification (10 µM) is given by the requirement not to exceed 10% substrate conversion at 1 mM substrate concentration and 10-fold dilution before injection.

Assay validation with TycA

In order to extend the dynamic range of the assay such that the best six substrates of TycA could be measured across ca. five orders of magnitude in activity (Fig. 2A, SI Table 1), reactions were performed with and without L-Phe. The PheHA and TrpHA concentrations were determined first by incubation of 1 µM enzyme with complete 1 mM substrate mix (L-Phe-d5, D-Phe, L-Ile, L-Leu-d7, L-Val-d8, D-Val, L-Met, L-Tyr, L-Trp) for 3 min. In the second reaction, 1 µM enzyme was incubated with the same substrate mix lacking L-Phe-d5 and D-Phe for 30 min to allow the accumulation of corresponding hydroxamates up to measurable levels. log([XaaHA]/[TrpHA]) ratios were calculated to allow comparison between both reactions.

Progress curve of PheHA formation with TycA

A hydroxamate assay reaction with 200 nM TycA in the presence of 1 mM proteinogenic amino acid mix was allowed to run for up to 20 minutes. Reactions were quenched at seven time points and the concentration of PheHA measured (SI Fig. 5).

Time course of hydroxamate ratios

After a prolonged reaction time, competitive product inhibition will decrease the rate of hydroxamate accumulation,[9] but should not change the ratio of products. Therefore, specificity profiles should remain unaffected. We tested this hypothesis by monitoring hydroxamate ratios over time in the reaction of 1 µM TycA with 1 mM substrates (SI Fig. 4).

DKP formation assay

Reaction conditions

The diketopiperazine (DKP) formation assay was performed in 150 µL volume with 5 mM ATP, 1 mM TCEP, 5 µM GrsB1 and either 5 µM sdVGrsA or a mutant thereof in peptide formation assay buffer (40 mM HEPES, 10 mM $MgCl_2$, 75 mM NaCl, pH 8.0). The reaction was started by addition of L-Val and L-Pro (1 mM each). The resulting solution was incubated at 37 °C and quenched after 3 h by heat denaturation at 95 °C for 3 min. Denatured proteins were precipitated by centrifugation and the supernatant analysed by UPLC-MS/MS.

UPLC-MS/MS conditions

Chromatography was performed on a Waters ACQUITY H-class UPLC system (Waters) with an injection volume of 2 µL. Methanol (A) and water with 0.1 % formic acid (B) were used as strong and weak eluent, respectively. Diketopiperazines were separated on the ACQUITY UPLC BEH C18 column (1.7 µm, 2.1 x 50 mm) with a linear gradient of 20-60% A over 1.5 min (flow rate 0.5 mL/min) followed by 1 min reequilibration. Acetonitrile was used as a needle wash between the samples. Data acquisition and quantitation were done using the MassLynx and TargetLynx software (version 4.1).

MS/MS analyses were performed on a Xevo TQ-S micro (Waters) tandem quadrupole instrument with ESI ionisation source in positive ion mode. Nitrogen was used as desolvation gas and argon as collision gas. The following source parameters were used: capillary voltage 0.5 kV, cone voltage 4 V, desolvation temperature 600°C, desolvation gas flow 1000 L/h. Val-Pro-DKP and was detected via the 197.09>69.95 transition, recorded in multiple reaction monitoring (MRM) mode. Standard calibration solutions of Val-Pro-DKP were prepared ranging from 0.0006 to 10 µM.

Supplementary Tables

SI Table 1. Comparison of kinetic data.

	PP_i exchange	**MesG**	**HAMA (µM)**	
Substrate	k_{cat}/K_M (mM^{-1} min^{-1})[19]	k_{cat}/K_M (mM^{-1} min^{-1})	3 min	30 min
L-Phe	9900 ± 300	1600 ± 85	96 ± 12	
D-Phe	4700 ± 400	2400 ± 120	116 ± 15	
L-Tyr	12.2 ± 1.4	1.7 ± 0.3	0.029 ± 0.002	12.7 ± 1.3
L-Trp	5.4 ± 0.5	3.5 ± 0.3	0.13 ± 0.01	43.0 ± 3.0
L-Met	2.1 ± 0.2	3.6 ± 0.8	0.13 ± 0.009	49.0 ± 4.3
L-Leu	1.26 ± 0.1	1.6 ± 0.1	ND	10.9 ± 0.8
L-Val	0.13 ± 0.008	0.12 ± 0.01	ND	0.045 ± 0.008

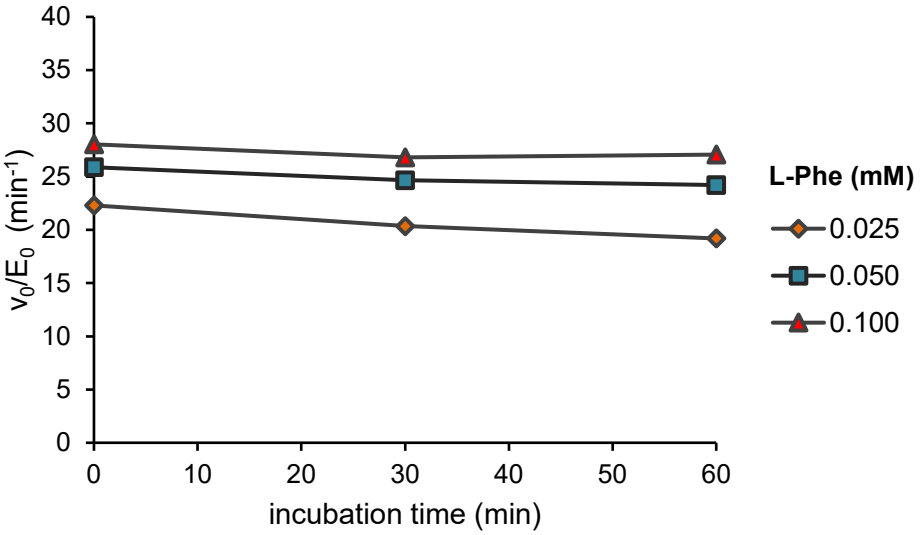
SI Table 2. Acquisition parameters for hydroxamate quantification and limits of quantification (LOQs).

Compound	Parent (m/z)	Cone Voltage (V)	Daughter (m/z)	Collision Energy (V)	LOQ (µM)
AlaHA	104.90	18	43.90	8	0.08
ArgHA	190.02	14	69.94	16	0.016
AspHA	148.95	32	87.92	10	0.08
CysHA	136.87	28	75.87	12	0.0032
GluHA	163.03	24	83.95	18	0.016
GlyHA	90.82	34	29.94	8	0.4
HisHA	171.05	22	109.92	10	0.016
IleHA	147.01	28	85.97	8	0.0032
LysHA	162.02	20	83.94	18	0.08
MetHA	165.03	26	103.88	8	0.0032
D-PheHA	180.99	30	119.94	10	0.0032
ProHA	130.97	24	69.96	12	0.4
ThrHA	134.91	26	73.97	8	0.0032
TrpHA	219.94	30	167.00	16	0.0032
TyrHA	196.98	30	135.95	12	0.0032
D-ValHA	132.87	22	71.91	10	0.016
L-Val-d8-HA	140.92	22	79.96	10	0.016
L-Phe-d5-HA	186.03	30	124.97	10	0.0032
L-Leu-d7-HA	154.05	30	93.01	10	0.0032

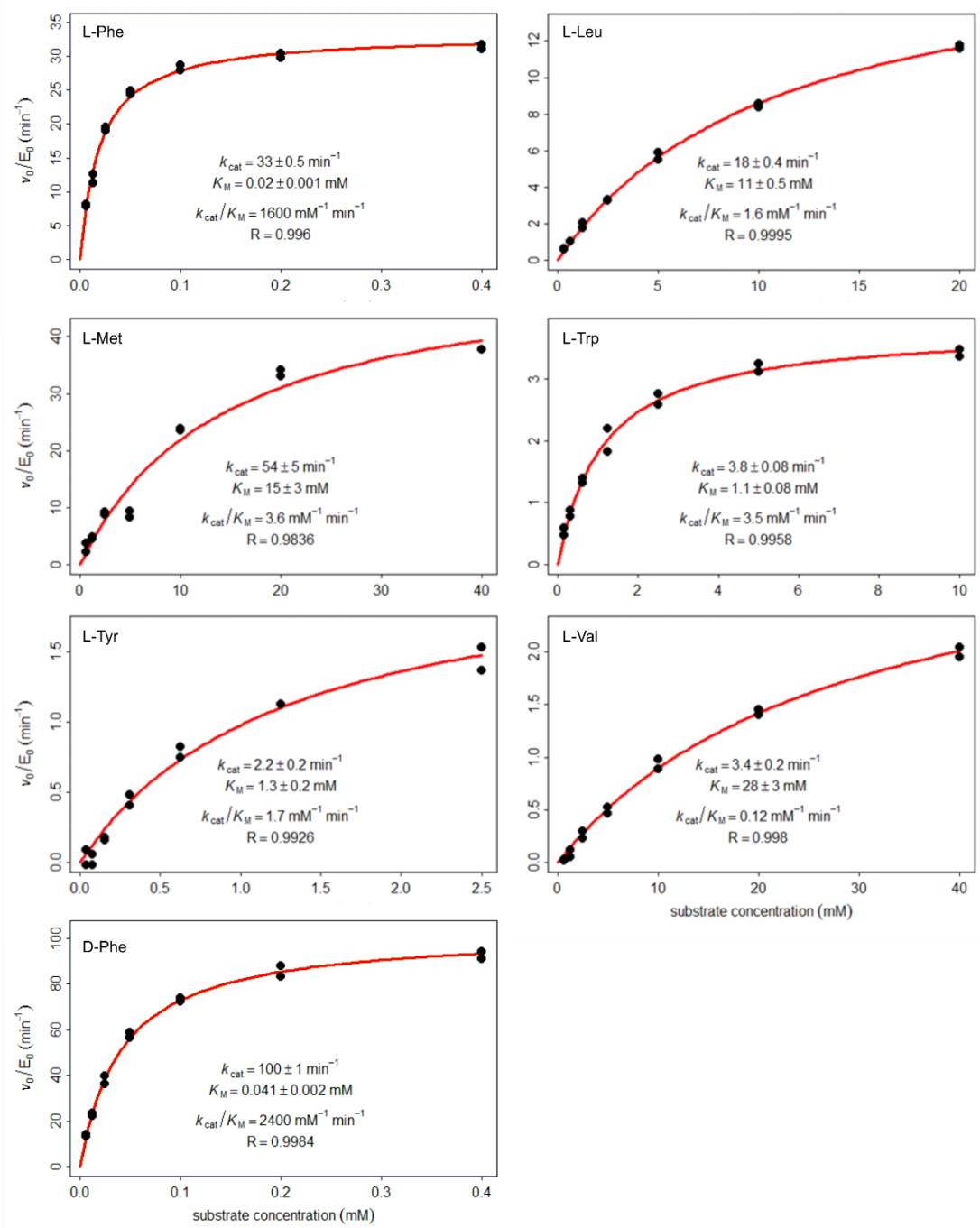
SI Table 3. HRMS of amino acid hydroxamates.

Name	Molecular formula	Expected (m/z)	Found (m/z)	Delta (ppm)
AlaHA	C3H9N2O2+	105.0659	105.0660	1.0
ArgHA	C6H16N5O2+	190.1299	190.1296	1.6
AspHA	C4H9N2O4+	149.0557	149.0555	1.3
CysHA	C3H9N2O2S+	137.0379	137.0378	0.7
GluHA	C5H11N2O4+	163.0713	163.0711	1.2
GlyHA	C2H7N2O2+	91.0502	91.0505	3.3
HisHA	C6H11N4O2+	171.0877	171.0874	1.8
IleHA	C6H15N2O2+	147.1128	147.1126	1.4
LeuHA	C6H15N2O2+	147.1128	147.1126	1.4
LysHA	C6H16N3O2+	162.1237	162.1235	1.2
MetHA	C5H13N2O2S+	165.0692	165.0690	1.2
PheHA	C9H13N2O2+	181.0972	181.0970	1.1
β-PheHA	C9H13N2O2+	181.0972	181.0968	2.2
Phenylglycine HA	C8H11N2O2+	167.0815	167.0813	1.2
Pipecolic acid HA	C6H13N2O2+	145.0972	145.0969	2.1
ProHA	C5H11N2O2+	131.0815	131.0814	0.8
SerHA	C3H9N2O3+	121.0608	121.0608	0.0
ThrHA	C4H11N2O3+	135.0764	135.0763	0.7
TrpHA	C11H14N3O2+	220.1081	220.1079	0.9
TyrHA	C9H13N2O3+	197.0921	197.0919	1.0
ValHA	C5H13N2O2+	133.0972	133.0971	0.8

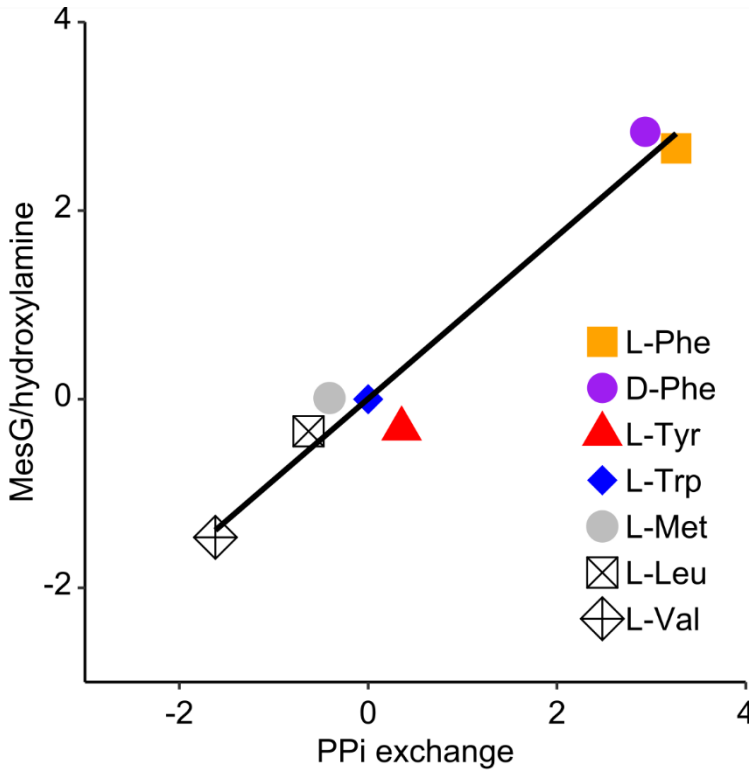
Supplementary Figures



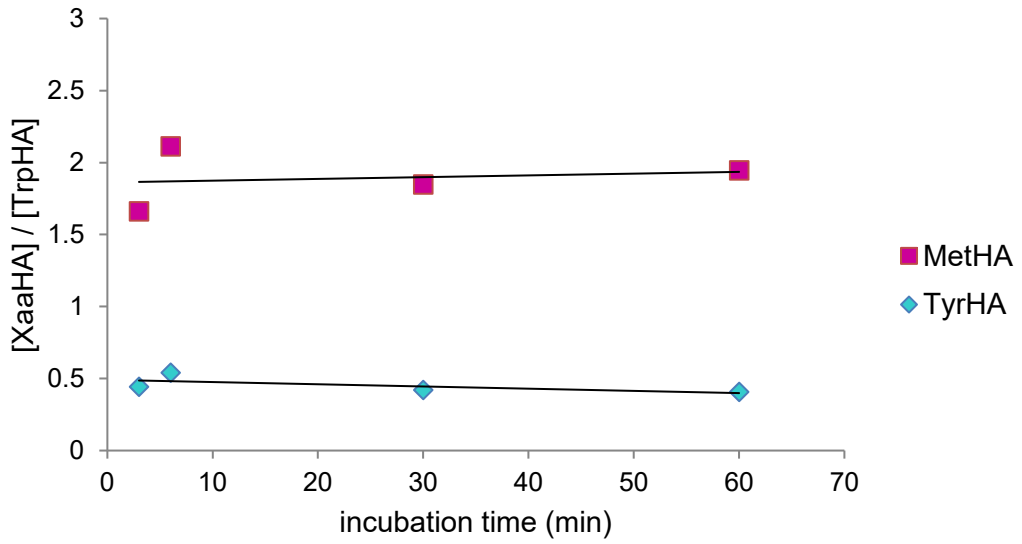
SI Fig. 1. Stability of TycA in hydroxylamine monitored with the MESG/hydroxylamine assay.



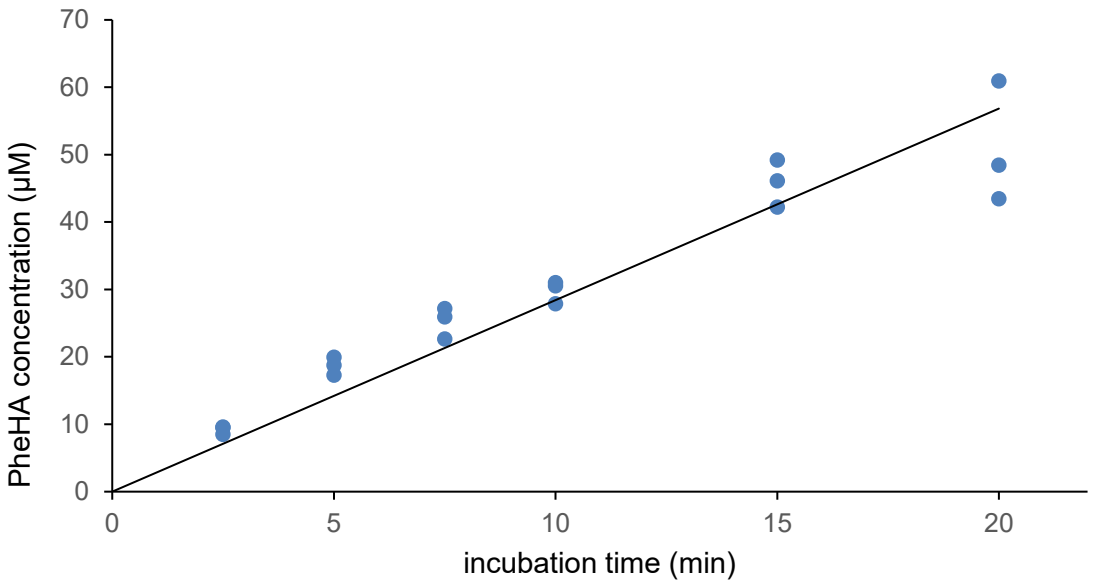
SI Fig. 2. Michaelis-Menten kinetics of TycA.



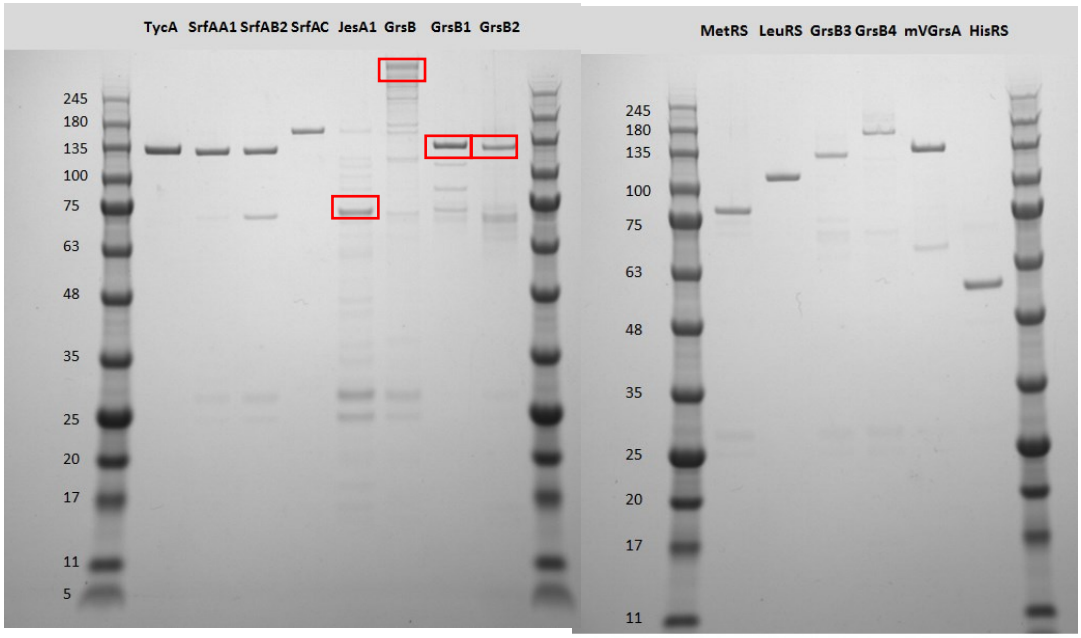
SI Fig. 3. Comparison of TycA parameters measured with MesG/hydroxylamine assay and PPi exchange assay. Data are plotted as log([k_{cat}/K_M]$_{Xaa}$ / [k_{cat}/K_M]$_{Trp}$). Slope: 0.86 ± 0.08; R^2 = 0.957.



SI Fig. 4. Ratios of hydroxamate concentrations during the course of a HAMA assay with TycA. The assay was conducted with the proteinogenic amino acid mix but only MetHA and TyrHA remained in the initial velocity range (<10% conversion) for the entire reaction time.



SI Fig. 5. Progress curve of PheHA formation catalysed by TycA. A linear fit indicates a k_{obs} of 14.2 ± 0.5 min^{-1}. The 3-fold lower turnover rate compared to the k_{cat} determined for pure L-Phe (43 min^{-1}; SI Fig. 2) might be explained by competition with alternative substrates. According to the parameters determined for competitive inhibition (Fig. 2B), the apparent deviation from linearity is not caused by PheHA.



SI Fig. 6. SDS-PAGE of purified proteins. Expected molecular weight of proteins (kDa): TycA (123.6), SrfAA1 (117.1), SrfAB2 (117.0), SrfAC (145.1), JesA1 (67.8), GrsB (510.0), GrsB1 (122.3), GrsB2 (117.9), MetRS (77.2), LeuRS (98.2), GrsB3 (119.3), GrsB4 (152.0), mVGrsA (128.5), HisRS (48.0). Where ambiguous, the protein of interest is labeled with a red rectangle.

Sequences of proteins used in this study

TycA

MVANQANLIDNKRELEQHALVPYAQGKSIHQLFEEQAEAFPDRVAIVFENRRLSYQELNRKANQLARALL
EKGVQTDSIVGVMMEKSIENVIAILAVLKAGGAYVPIDIEYPRDRIQYILQDSQTKIVLTQKSVSQLVHD
VGYSGEVVVLDEEQLDARETANLHQPSKPTDLAYVIYTSGTTGKPKGTMLEHKGIANLQSFFQNSFGVTE
QDRIGLFASMSFDASVWEMFMALLSGASLYILSKQTIHDFAAFEHYLSENELTIITLPPTYLTHLTPERI
TSLRIMITAGSASSAPLVNKWKDKLRYINAYGPTETSICATIWEAPSNQLSVQSVPIGKPIQNTHIYIVN
EDLQLLPTGSEGELCIGGVGLARGYWNRPDLTAEKFVDNPFVPGEKMYRTGDLAKWLTDGTIEFLGRIDH
QVKIRGHRIELGEIESVLLAHEHITEAVVIAREDQHAGQYLCAYYISQQEATPAQLRDYAAQKLPAYMLP
SYFVKLDKMPLTPNDKIDRKALPEPDLTANQSQAAYHPPRTETESILVSIWQNVLGIEKIGIRDNFYSLG
GDSIQAIQVVARLHSYQLKLETKDLLNYPTIEQVALFVKSTTRKSDQGIIAGNVPLTPIQKWFFGKNFTN
TGHWNQSSVLYRPEGFDPKVIQSVMDKIIEHHDALRMVYQHENGNVVQHNRGLGGQLYDFFSYNLTAQPD
VQQAIEAETQRLHSSMNLQEGPLVKVALFQTLHGDHLFLAIHHLVVDGISWRILFEDLATGYAQALAGQA
ISLPEKTDSFQSWSQWLQEYANEADLLSEIPYWESLESQAKNVSLPKDYEVTDCKQKSVRNMRIRLHPEE
TEQLLKHANQAYQTEINDLLLAALGLAFAEWSKLAQIVIHLEGHGREDIIEQANVARTVGWFTSQYPVLL
DLKQTAPLSDYIKLTKENMRKIPRKGIGYDILKHVTLPENRGSLSFRVQPEVTFNYLGQFDADMRTELFT
RSPYSGGNTLGADGKNNLSPESEVYTALNITGLIEGGELVLTFSYSSEQYREESIQQLSQSYQKHLLAII
AHCTEKKEVERTPSDFSVKGLQMEEMDDIFELLANTLRGSRSHHHHHH

SrfAA1

MLTDAQKRIWYTEKFYPHTSISNLAGIGKLVSADAIDYVLVEQAIQEFIRRNDAMRLRLRLDENGEPVQY
ISEYRPVDIKHTDTTEDPNAIEFISQWSREETKKPLPLYDCDLFRFSLFTIKENEVWFYANVHHVISDGI
SMNILGNAIMHIYLELASGSETKEGISHSFIDHVLSEQEYAQSKRFEKDKAFWNKQFESVPELVSLKRNA
SAGGSLDAERFSKDVPEALHQQILSFCEANKVSVLSVFQSLLAAYLYRVSGQNDVVTGTFMGNRTNAKEK
QMLGMFVSTVPLRTNIDGGQAFSEFVKDRMKDLMKTLRHQKYPYNLLINDLRETKSSLTKLFTVSLEYQV
MQWQKEEDLAFLTEPIFSGSGLNDVSIHVKDRWDTGKLTIDFDYRTDLFSREEINMICERMITMLENALT
HPEHTIDELTLISDAEKEKLLARAGGKSVSYRKDMTIPELFQEKAELLSDHPAVVFEDRTLSYRTLHEQS
ARIANVLKQKGVGPDSPVAVLIERSERMITAIMGILKAGGAYVPIDPGFPAERIQYILEDCGADFILTES
KVAAPEADAELIDLDQAIEEGAEESLNADVNARNLAYIIYTSGTTGRPKGVMIEHRQVHHLVESLQQTIY
QSGSQTLRMALLAPFHFDASVKQIFASLLLGQTLYIVPKKTVTNGAALTAYYRKNSIEATDGTPAHLQML
AAAGDFEGLKLKHMLIGGEGLSSVVADKLLKLFKEAGTAPRLTNVYGPTETCVDASVHPVIPENAVQSAY
VPIGKALGNNRLYILDQKGRLQPEGVAGELYIAGDGVGRGYLHLPELTEEKFLQDPFVPGDRMYRTGDVV
RWLPDGTIEYLGREDDQVKVRGYRIELGEIEAVIQQAPDVAKAVVLARPDEQGNLEVCAYVVQKPGSEFA
PAGLREHAARQLPDYMVPAYFTEVTEIPLTPSGKVDRRKLFALEVKAVSGTAYTAPRNETEKAIAAIWQD
VLNVEKAGIFDNFFETGGHSLKAMTLLTKIHKETGIEIPLQFLFEHPTITALAEEGSRSHHHHHH

SrfAB2

MKEEQTFEPIRQASYQQHYPVSPAQRRMYILNQLGQANTSYNVPAVLLLEGEVDKDRLENAIQQLINRHE
ILRTSFDMIDGEVVQTVHKNISFQLEAAKGREEDAEEIIKAFVQPFELNRAPLVRSKLVQLEEKRHLLLI
DMHHIITDGSSTGILIGDLAKIYQGADLELPQIHYKDYAVWHKEQTNYQKDEEYWLDVFKGELPILDLPA
DFERPAERSFAGERVMFGLDKQITAQIKSLMAETDTTMYMFLLAAFNVLLSKYASQDDIIVGSPTAGRTH
PDLQGVPGMFVNTVALRTAPAGDKTFAQFLEEVKTASLQAFEHQSYPLEELIEKLPLTRDTSRSPLFSVM
FNMQNMEIPSLRLGDLKISSYSMLHHVAKFDLSLEAVEREEDIGLSFDYATALFKDETIRRWSRHFVNII
KAAAANPNVRLSDVDLLSSAETAALLEERHMTQITEATFAALFEKQAQQTPDHSAVKAGGNLLTYRELDE
QANQLAHHLRAQGAGNEDIVAIVMDRSAEVMVSILGVMKAGAAFLPIDPDTPEERIRYSLEDSGAKFAVV
NERNMTAIGQYEGIIVSLDDGKWRNESKERPSSISGSRNLAYVIYTSGTTGKPKGVQIEHRNLTNYVSWF
SEEAGLTENDKTVLLSSYAFDLGYTSMFPVLLGGGELHIVQKETYTAPDEIAHYIKEHGITYIKLTPSLF

HTIVNTASFAKDANFESLRLIVLGGEKIIPTDVIAFRKMYGHTEFINHYGPTEATIGAIAGRVDLYEPDA
FAKRPTIGRPIANAGALVLNEALKLVPPGASGQLYITGQGLARGYLNRPQLTAERFVENPYSPGSLMYKT
GDVVRRLSDGTLAFIGRADDQVKIRGYRIEPKEIETVMLSLSGIQEAVVLAVSEGGLQELCAYYTSDQDI
EKAELRYQLSLTLPSHMIPAFFVQVDAIPLTANGKTDRNALPKPNAAQSGGKALAAPETALEESLCRIWQ
KTLGIEAIGIDDNFFDLGGHSLKGMMLIANIQAELEKSVPLKALFEQPTVRQLAAYMEASAGSRSHHHHH
H

SrfAC

MSQFSKDQVQDMYYLSPMQEGMLFHAILNPGQSFYLEQITMKVKGSLNIKCLEESMNVIMDRYDVFRTVF
IHEKVKRPVQVVLKKRQFHIEEIDLTHLTGSEQTAKINEYKEQDKIRGFDLTRDIPMRAAIFKKAEESFE
WVWSYHHIILDGWCFGIVVQDLFKVYNALREQKPYSLPPVKPYKDYIKWLEKQDKQASLRYWREYLEGFE
GQTTFAEQRKKQKDGYEPKELLFSLSEAETKAFTELAKSQHTTLSTALQAVWSVLISRYQQSGDLAFGTV
VSGRPAEIKGVEHMVGLFINVVPRRVKLSEGITFNGLLKRLQEQSLQSEPHQYVPLYDIQSQADQPKLID
HIIVFENYPLQDAKNEESSENGFDMVDVHVFEKSNYDLNLMASPGDEMLIKLAYNENVFDEAFILRLKSQ
LLTAIQQLIQNPDQPVSTINLVDDREREFLLTGLNPPAQAHETKPLTYWFKEAVNANPDAPALTYSGQTL
SYRELDEEANRIARRLQKHGAGKGSVVALYTKRSLELVIGILGVLKAGAAYLPVDPKLPEDRISYMLADS
AAACLLTHQEMKEQAAELPYTGTTLFIDDQTRFEEQASDPATAIDPNDPAYIMYTSGTTGKPKGNITTHA
NIQGLVKHVDYMAFSDQDTFLSVSNYAFDAFTFDFYASMLNAARLIIADEHTLLDTERLTDLILQENVNV
MFATTALFNLLTDAGEDWMKGLRCILFGGERASVPHVRKALRIMGPGKLINCYGPTEGTVFATAHVVHDL
PDSISSLPIGKPISNASVYILNEQSQLQPFGAVGELCISGMGVSKGYVNRADLTKEKFIENPFKPGETLY
RTGDLARWLPDGTIEYAGRIDDQVKIRGHRIELEEIEKQLQEYPGVKDAVVVADRHESGDASINAYLVNR
TQLSAEDVKAHLKKQLPAYMVPQTFTFLDELPLTTNGKVNKRLLPKPDQDQLAEEWIGPRNEMEETIAQI
WSEVLGRKQIGIHDDFFALGGHSLKAMTAASRIKKELGIDLPVKLLFEAPTIAGISAYLKNGGSDGLQDV
TIMNQDQEQIIFAFPPVLGYGLMYQNLSSRLPSYKLCAFDFIEEEDRLDRYADLIQKLQPEGPLTLFGYS
AGCSLAFEAAKKLEEQGRIVQRIIMVDSYKKQGVSDLDGRTVESDVEALMNVNRDNEALNSEAVKHGLKQ
KTHAFYSYYVNLISTGQVKADIDLLTSGADFDMPEWLASWEEATTGVYRVKRGFGTHAEMLQGETLDRNA
EILLEFLNTQTVTVSGSRSHHHHHH

JesA1

MLNASETAQLQAWNAEPQHFAEDRTIHQQFEARAAERPEAVALVYQGESLSYGELNARANQVAHRLLALG
VRPDDRVAICVERGPAMIIGLLGILKSGAGYVPLDPAYPRERLAYTLGDSAPVALLSQHSVQEALPAVKV
PVINLDDADLRDESVRNPQVAVSATHLAYVIYTSGSTGVPKGVMVEHCNVARLFSATDAWFGFNEKDVWA
LFHSFAFDFSVWEIWGALLHGGRLLIVPQLVSRSPEDCYELLCSAGVTVLNQTPSAFRQLIAAQGESGQP
HSLRQVIFGGEALDTAMLKPWYARDLNAATQLVNMYGITETTVHVTYYPLQAEDAQRVGVSPIGRGIPDL
RLYLLDGYGQPLPPGVVGELYVGGAGVARGYLNREELNASRFLDDPFVSTPGARMYRSGDLGRWLADGSL
EYLGRNDEQVKIRGFRIELGEIEAQLAACEGVRDAVVLVREDEPGDKRLVAYVIGKAGVELDAAQLRDQL
RLALAEYMLPSAFVSLESFPLTANGKLDRKALPVPAADAYARREYEAPEGPAETTLAGLWAELLGVEQVG
RHDQFFELGGHSLLAVKLIERMRQVGLSADVRVLFGQPTLAALAAASGKGGEIGSRSHHHHHH

GrsB1

MSTFKKEHVQDMYRLSPMQEGMLFHALLDKDKNAHLVQMSIAIEGIVDVELLSESLNILIDRYDVFRTTF
LHEKIKQPLQVVLKERPVQLQFKDISSLDEEKREQAIEQYKYQDGETVFDLTRDPLMRVAIFQTGKVNYQ
MIWSFHHILMDGWCFNIIFNDLFNIYLSLKEKKPLQLEAVQPYKQFIKWLEKQDKQEALRYWKEHLMNYD
QSVTLPKKKAAINNTTYEPAQFRFAFDKVLTQQLLRIANQSQVTLNIVFQTIWGIVLQKYNSTNDVVYGS
VVSGRPSEISGIEKMVGLFINTLPLRIQTQKDQSFIELVKTVHQNVLFSQQHEYFPLYEIQNHTELKQNL

IDHIMVIENYPLVEELQKNSIMQKVGFTVRDVKMFEPTNYDMTVMVLPRDEISVRLDYNAAVYDIDFIKK
IEGHMKEVALCVANNPHVLVQDVPLLTKQEKQHLLVELHDSITEYPDKTIHQLFTEQVEKTPEHVAVVFE
DEKVTYRELHERSNQLARFLREKGVKKESIIGIMMERSVEMIVGILGILKAGGAFVPIDPEYPKERIGYM
LDSVRLVLTQRHLKDKFAFTKETIVIEDPSISHELTEEIDYINESEDLFYIIYTSGTTGKPKGVMLEHKN
IVNLLHFTFEKTNINFSDKVLQYTTCSFDVCYQEIFSTLLSGGQLYLIRKETQRDVEQLFDLVKRENIEV
LSFPVAFLKFIFNEREFINRFPTCVKHIITAGEQLVVNNEFKRYLHEHNVHLHNHYGPSETHVVTTYTIN
PEAEIPELPPIGKPISNTWIYILDQEQQLQPQGIVGELYISGANVGRGYLNNQELTAEKFFADPFRPNER
MYRTGDLARWLPDGNIEFLGRADHQVKIRGHRIELGEIEAQLLNCKGVKEAVVIDKADDKGGKYLCAYVV
MEVEVNDSELREYLGKALPDYMIPSFFVPLDQLPLTPNGKIDRKSLPNLEGIVNTNAKYVVPTNELEEKL
AKIWEEVLGISQIGIQDNFFSLGGHSLKAITLISRMNKECNVDIPLRLLFEAPTIQEISNYINGGSRSHH
HHHH

GrsB2

MIQPVPEQEYYPVSSVQKRMFILNEFDRSGTAYNLPGVMFLDGKLNYRQLEAAVKKLVERHEALRTSFHS
INGEPVQRVHQNVELQIAYSESTEDQVERIIAEFMQPFALEVAPLLRVGLVKLEAERHLFIMDMHHIISD
GVSMQIMIQEIADLYKEKELPTLGIQYKDFTVWHNRLLQSDVIEKQEAYWLNVFAEEIPVLNLPTDYPRP
TIQSFDGKRFTFSTGKQLMDDLYKVATETGTTLYMVLLAAYNVFLSKYSGQDDIVVGTPIAGRSHADVEN
MLGMFVNTLAIRSRLNNEDTFKDFLANVKQTALHAYENPDYPFDTLVEKLGIQRDLSRNPLFDTMFVLQN
TDRKSFEVEQITITPYVPNSRHSKFDLTLEVSEEQNEILLCLEYCTKLFTDKTVERMAGHFLQILHAIVG
NPTIIISEIEILSEEEKQHILFEFNDTKTTYPHMQTIQGLFEEQVEKTPDHVAVGWKDQTLTYRELNERA
NQVARVLRQKGVQPDNIVGLLVERSPEMLVGIMGILKAGGAYLPLDPEYPADRISYMIQDCGVRIMLTQQ
HLLSLVHDEFDCVILDEDSLYKGDSSNLAPVNQAGDLAYIMYTSGSTGKPKGVMVEHRNVIRLVKNTNYV
QVREDDRIIQTGAIGFDALTFEVFGSLLHGAELYPVTKDVLLDAEKLHKFLQANQITIMWLTSPLFNQLS
QGTEEMFAGLRSLIVGGDALSPKHINNVKRKCPNLTMWNGYGPTENTTFSTCFLIDKEYDDNIPIGKAIS
NSTVYIMDRYGQLQPVGVPGELCVGGDGVARGYMNQPALTEEKFVPNPFAPGERMYRTGDLARWLPDGTI
EYLGRIDQQVKIRGYRIEPGEIETLLVKHKKVKESVIMVVEDNNGQKALCAYYVPEEEVTVSELREYIAK
ELPVYMVPAYFVQIEQMPLTQNGKVNRSALPKPDGEFGTATEYVAPSSDIEMKLAEIWHNVLGVNKIGVL
DNFFELGGHSLRAMTMISQVHKEFDVELPLKVLFETPTISALAQYIADGSRSHHHHHH

GrsB3

MIQPVTPQDYYPVSSAQKRMYILYEFEGAGITYNVPNVMFIEGKLDYQRFEYAIKSLVNRHEALRTSFYS
LNGEPVQRVHQNVELQIAYSEAKEDEIEQIVESFVQPFDLEIAPLLRVGLVKLASDRYLFLMDMHHIISD
GVSMQIITKEIADLYKGKELAELHIQYKDFAVWQNEWFQSDALEKQKTYWLNTFAEDIPVLNLSTDYPRP
TIQSFEGDIVTFSAGKQLAEELKRLAAETGTTLYMLLLAAYNVLLHKYSGQEEIVVGTPIAGRSHADVEN
IVGMFVNTLALKNTPIAVRTFHEFLLEVKQNALEAFENQDYPFENLIEKLQVRRDLSRNPLFDTMFSLSN
IDEQVEIGIEGLNFSPYEMQYWIAKFDISFDILEKQDDIQFYFNYCTNLFKKETIERLATHFMHILQEIV
INPEIKLCEINMLSEEEQQRVLYDFNGTDATYATNKIFHELFEEQVEKTPDHIAVIDEREKLSYQELNAK
ANQLARVLRQKGVQPNSMVGIMVDRSLDMIVGMLGVLKAGGAYVPIDIDYPQERISYMMEDSGAALLLTQ
QKLTQQIAFSGDILYLDQEEWLHEEASNLEPIARPQDIAYIIYTSGTTGKPKGVMIEHQSYVNVAMAWKD
AYRLDTFPVRLLQMASFAFDVSAGDFARALLTGGQLIVCPNEVKMDPASLYAIIKKYDITIFEATPALVI
PLMEYIYEQKLDISQLQILIVGSDSCSMEDFKTLVSRFGSTIRIVNSYGVTEACIDSSYYEQPLSSLHVT
GTVPIGKPYANMKMYIMNQYLQIQPVGVIGELCIGGAGVARGYLNRPDLTAEKFVPNPFVPGEKLYRTGD
LARWMPDGNVEFLGRNDHQVKIRGIRIELGEIEAQLRKHDSIKEATVIAREDHMKEKYLCAYMVTEGEVN
VAELRAYLANDLPAAMIPSYFVSLEAMPLTANGKIDKRSLPEPDGSISIGTEYVAPRTMLEGKLEEIWKD
VLGLQRVGIHDDFFTIGGHSLKAMAVISQVHKECQTEVPLRVLFETPTIQGLAKYIEEGSRSHHHHHH

GrsB4

MAIQPVSGQDYYPVSSAQKRMFIVNQFDGVGISYNMPSIMLIEGKLERTRLESAFKRLIERHESLRTSFE
IINGKPVQKIHEEVDFNMSYQVASNEQVEKMIDEFIQPFDLSVAPLLRVELLKLEEDRHVLIFDMHHIIS

DGISSNILMKELGELYQGNALPELRIQYKDFAVWQNEWFQSEAFKKQEEYWVNVFADERPILDIPTDYPR
PMQQSFDGAQLTFGTGKQLMDGLYRVATETGTTLYMVLLAAYNVLLSKYSGQEDIIVGTPIVGRSHTDLE
NIVGMFVNTLAMRNKPEGEKTFKAFVSEIKQNALAAFENQDYPFEELIEKLEIQRDLSRNPLFDTLFSLQ
NIGEESFELAELTCKPFDLVSKLEHAKFDLSLVAVEKEEEIAFGLQYCTKLYKEKTVEQLAQHFIQIVKA
IVENPDVKLSDIDMLSEEEKKQILLEFNDTKIQYPQNQTIQELFEEQVKKTPEHIAIVWEGQALTYHELN
IKANQLARVLREKGVTPNHPVAIMTERSLEMIVGIFSILKAGGAYVPIDPAYPQERIQYLLEDSGATLLL
TQSHVLNKLPVDIEWLDLTDEQNYVEDGTNLPFMNQSTDLAYIIYTSGTTGKPKGVMIEHQSIINCLQWR
KEEYEFGPGDTALQVFSFAFDGFVASLFAPILAGATSVLPKEEEAKDPVALKKLIASEEITHYYGVPSLF
SAILDVSSSKDLQNLRCVTLGGEKLPAQIVKKIKEKNKEIEVNNEYGPTENSVVTTIMRDIQVEQEITIG
RPLSNVDVYIVNCNHQLQPVGVVGELCIGGQGLARGYLNKPELTADKFVVNPFVPGERMYKTGDLAKWRS
DGMIEYVGRVDEQVKVRGYRIELGEIESAILEYEKIKEAVVMVSEHTASEQMLCAYIVGEEDVLTLDLRS
YLAKLLPSYMIPNYFIQLDSIPLTPNGKVDRKALPEPQTIGLMAREYVAPRNEIEAQLVLIWQEVLGIEL
IGITDNFFELGGHSLKATLLVAKIYEYMQIEMPLNVVFKHSTIMKIAEYITHQESENNVHQPILVNVEAD
REALSLNGEKQRKNIELPILLNEETDRNVFCFAPIGAQGVFYKKLAEQIPTASLYGFDFIEDDDRIQQYI
ESMIQTQSDGQYVLIGYSSGGNLAFEVAKEMERQGYSVSDLVLFDVYWKGKVFEQTKEEEEENIKIIMEE
LRENPGMFNMTREDFELYFANEFVKQSFTRKMRKYMSFYTQLVNYGEVEATIHLIQAEFEEEKIDENEKA
DEEEKTYLEEKWNEKAWNKAAKRFVKYNGYGAHSNMLGGDGLERNSSILKQILQGTFVVKGSRSHHHHHH

GrsB

MSTFKKEHVQDMYRLSPMQEGMLFHALLDKDKNAHLVQMSIAIEGIVDVELLSESLNILIDRYDVFRTTF
LHEKIKQPLQVVLKERPVQLQFKDISSLDEEKREQAIEQYKYQDGETVFDLTRDPLMRVAIFQTGKVNYQ
MIWSFHHILMDGWCFNIIFNDLFNIYLSLKEKKPLQLEAVQPYKQFIKWLEKQDKQEALRYWKEHLMNYD
QSVTLPKKKAAINNTTYEPAQFRFAFDKVLTQQLLRIANQSQVTLNIVFQTIWGIVLQKYNSTNDVVYGS
VVSGRPSEISGIEKMVGLFINTLPLRIQTQKDQSFIELVKTVHQNVLFSQQHEYFPLYEIQNHTELKQNL
IDHIMVIENYPLVEELQKNSIMQKVGFTVRDVKMFEPTNYDMTVMVLPRDEISVRLDYNAAVYDIDFIKK
IEGHMKEVALCVANNPHVLVQDVPLLTKQEKQHLLVELHDSITEYPDKTIHQLFTEQVEKTPEHVAVVFE
DEKVTYRELHERSNQLARFLREKGVKKESIIGIMMERSVEMIVGILGILKAGGAFVPIDPEYPKERIGYM
LDSVRLVLTQRHLKDKFAFTKETIVIEDPSISHELTEEIDYINESEDLFYIIYTSGTTGKPKGVMLEHKN
IVNLLHFTFEKTNINFSDKVLQYTTCSFDVCYQEIFSTLLSGGQLYLIRKETQRDVEQLFDLVKRENIEV
LSFPVAFLKFIFNEREFINRFPTCVKHIITAGEQLVVNNEFKRYLHEHNVHLHNHYGPSETHVVTTYTIN
PEAEIPELPPIGKPISNTWIYILDQEQQLQPQGIVGELYISGANVGRGYLNNQELTAEKFFADPFRPNER
MYRTGDLARWLPDGNIEFLGRADHQVKIRGHRIELGEIEAQLLNCKGVKEAVVIDKADDKGGKYLCAYVV
MEVEVNDSELREYLGKALPDYMIPSFFVPLDQLPLTPNGKIDRKSLPNLEGIVNTNAKYVVPTNELEEKL
AKIWEEVLGISQIGIQDNFFSLGGHSLKAITLISRMNKECNVDIPLRLLFEAPTIQEISNYINGAKKESY
VAIQPVPEQEYYPVSSVQKRMFILNEFDRSGTAYNLPGVMFLDGKLNYRQLEAAVKKLVERHEALRTSFH
SINGEPVQRVHQNVELQIAYSESTEDQVERIIAEFMQPFALEVAPLLRVGLVKLEAERHLFIMDMHHIIS
DGVSMQIMIQEIADLYKEKELPTLGIQYKDFTVWHNRLLQSDVIEKQEAYWLNVFAEEIPVLNLPTDYPR
PTIQSFDGKRFTFSTGKQLMDDLYKVATETGTTLYMVLLAAYNVFLSKYSGQDDIVVGTPIAGRSHADVE
NMLGMFVNTLAIRSRLNNEDTFKDFLANVKQTALHAYENPDYPFDTLVEKLGIQRDLSRNPLFDTMFVLQ
NTDRKSFEVEQITITPYVPNSRHSKFDLTLEVSEEQNEILLCLEYCTKLFTDKTVERMAGHFLQILHAIV
GNPTIIISEIEILSEEEKQHILFEFNDTKTTYPHMQTIQGLFEEQVEKTPDHVAVGWKDQTLTYRELNER
ANQVARVLRQKGVQPDNIVGLLVERSPEMLVGIMGILKAGGAYLPLDPEYPADRISYMIQDCGVRIMLTQ
QHLLSLVHDEFDCVILDEDSLYKGDSSNLAPVNQAGDLAYIMYTSGSTGKPKGVMVEHRNVIRLVKNTNY
VQVREDDRIIQTGAIGFDALTFEVFGSLLHGAELYPVTKDVLLDAEKLHKFLQANQITIMWLTSPLFNQL
SQGTEEMFAGLRSLIVGGDALSPKHINNVKRKCPNLTMWNGYGPTENTTFSTCFLIDKEYDDNIPIGKAI
SNSTVYIMDRYGQLQPVGVPGELCVGGDGVARGYMNQPALTEEKFVPNPFAPGERMYRTGDLARWLPDGT
IEYLGRIDQQVKIRGYRIEPGEIETLLVKHKKVKESVIMVVEDNNGQKALCAYYVPEEEVTVSELREYIA
KELPVYMVPAYFVQIEQMPLTQNGKVNRSALPKPDGEFGTATEYVAPSSDIEMKLAEIWHNVLGVNKIGV

LDNFFELGGHSLRAMTMISQVHKEFDVELPLKVLFETPTISALAQYIADGEKGMYLAIQPVTPQDYYPVS
SAQKRMYILYEFEGAGITYNVPNVMFIEGKLDYQRFEYAIKSLVNRHEALRTSFYSLNGEPVQRVHQNVE
LQIAYSEAKEDEIEQIVESFVQPFDLEIAPLLRVGLVKLASDRYLFLMDMHHIISDGVSMQIITKEIADL
YKGKELAELHIQYKDFAVWQNEWFQSDALEKQKTYWLNTFAEDIPVLNLSTDYPRPTIQSFEGDIVTFSA
GKQLAEELKRLAAETGTTLYMLLLAAYNVLLHKYSGQEEIVVGTPIAGRSHADVENIVGMFVNTLALKNT
PIAVRTFHEFLLEVKQNALEAFENQDYPFENLIEKLQVRRDLSRNPLFDTMFSLSNIDEQVEIGIEGLNF
SPYEMQYWIAKFDISFDILEKQDDIQFYFNYCTNLFKKETIERLATHFMHILQEIVINPEIKLCEINMLS
EEEQQRVLYDFNGTDATYATNKIFHELFEEQVEKTPDHIAVIDEREKLSYQELNAKANQLARVLRQKGVQ
PNSMVGIMVDRSLDMIVGMLGVLKAGGAYVPIDIDYPQERISYMMEDSGAALLLTQQKLTQQIAFSGDIL
YLDQEEWLHEEASNLEPIARPQDIAYIIYTSGTTGKPKGVMIEHQSYVNVAMAWKDAYRLDTFPVRLLQM
ASFAFDVSAGDFARALLTGGQLIVCPNEVKMDPASLYAIIKKYDITIFEATPALVIPLMEYIYEQKLDIS
QLQILIVGSDSCSMEDFKTLVSRFGSTIRIVNSYGVTEACIDSSYYEQPLSSLHVTGTVPIGKPYANMKM
YIMNQYLQIQPVGVIGELCIGGAGVARGYLNRPDLTAEKFVPNPFVPGEKLYRTGDLARWMPDGNVEFLG
RNDHQVKIRGIRIELGEIEAQLRKHDSIKEATVIAREDHMKEKYLCAYMVTEGEVNVAELRAYLANDLPA
AMIPSYFVSLEAMPLTANGKIDKRSLPEPDGSISIGTEYVAPRTMLEGKLEEIWKDVLGLQRVGIHDDFF
TIGGHSLKAMAVISQVHKECQTEVPLRVLFETPTIQGLAKYIEETDTEQYMAIQPVSGQDYYPVSSAQKR
MFIVNQFDGVGISYNMPSIMLIEGKLERTRLESAFKRLIERHESLRTSFEIINGKPVQKIHEEVDFNMSY
QVASNEQVEKMIDEFIQPFDLSVAPLLRVELLKLEEDRHVLIFDMHHIISDGISSNILMKELGELYQGNA
LPELRIQYKDFAVWQNEWFQSEAFKKQEEYWVNVFADERPILDIPTDYPRPMQQSFDGAQLTFGTGKQLM
DGLYRVATETGTTLYMVLLAAYNVLLSKYSGQEDIIVGTPIVGRSHTDLENIVGMFVNTLAMRNKPEGEK
TFKAFVSEIKQNALAAFENQDYPFEELIEKLEIQRDLSRNPLFDTLFSLQNIGEESFELAELTCKPFDLV
SKLEHAKFDLSLVAVEKEEEIAFGLQYCTKLYKEKTVEQLAQHFIQIVKAIVENPDVKLSDIDMLSEEEK
KQILLEFNDTKIQYPQNQTIQELFEEQVKKTPEHIAIVWEGQALTYHELNIKANQLARVLREKGVTPNHP
VAIMTERSLEMIVGIFSILKAGGAYVPIDPAYPQERIQYLLEDSGATLLLTQSHVLNKLPVDIEWLDLTD
EQNYVEDGTNLPFMNQSTDLAYIIYTSGTTGKPKGVMIEHQSIINCLQWRKEEYEFGPGDTALQVFSFAF
DGFVASLFAPILAGATSVLPKEEEAKDPVALKKLIASEEITHYYGVPSLFSAILDVSSSKDLQNLRCVTL
GGEKLPAQIVKKIKEKNKEIEVNNEYGPTENSVVTTIMRDIQVEQEITIGRPLSNVDVYIVNCNHQLQPV
GVVGELCIGGQGLARGYLNKPELTADKFVVNPFVPGERMYKTGDLAKWRSDGMIEYVGRVDEQVKVRGYR
IELGEIESAILEYEKIKEAVVMVSEHTASEQMLCAYIVGEEDVLTLDLRSYLAKLLPSYMIPNYFIQLDS
IPLTPNGKVDRKALPEPQTIGLMAREYVAPRNEIEAQLVLIWQEVLGIELIGITDNFFELGGHSLKATLL
VAKIYEYMQIEMPLNVVFKHSTIMKIAEYITHQESENNVHQPILVNVEADREALSLNGEKQRKNIELPIL
LNEETDRNVFCFAPIGAQGVFYKKLAEQIPTASLYGFDFIEDDDRIQQYIESMIQTQSDGQYVLIGYSSG
GNLAFEVAKEMERQGYSVSDLVLFDVYWKGKVFEQTKEEEEENIKIIMEELRENPGMFNMTREDFELYFA
NEFVKQSFTRKMRKYMSFYTQLVNYGEVEATIHLIQAEFEEEKIDENEKADEEEKTYLEEKWNEKAWNKA
AKRFVKYNGYGAHSNMLGGDGLERNSSILKQILQGTFVVKGSRSHHHHHH

MetRS

MTQVAKKILVTCALPYANGSIHLGHMLEHIQADVWVRYQRMRGHEVNFICADDAHGTPIMLKAQQLGITP
EQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYSRLKENGFIKNRTISQLYDPEKGMFLPDR
FVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEPKSVVSGATPVMRDSEHFFFDLPSFSEMLQAWTRSGA
LQEQVANKMQEWFESGLQQWDISRDAPYFGFEIPNAPGKYFYVWLDAPIGYMGSFKNLCDKRGDSVSFDE
YWKKDSTAELYHFIGKDIVYFHSLFWPAMLEGSNFRKPSNLFVHGYVTVNGAKMSKSRGTFIKASTWLNH
FDADSLRYYYTAKLSSRIDDIDLNLEDFVQRVNADIVNKVVNLASRNAGFINKRFDGVLASELADPQLYK
TFTDAAEVIGEAWESREFGKAVREIMALADLANRYVDEQAPWVVAKQEGRDADLQAICSMGINLFRVLMT
YLKPVLPKLTERAEAFLNTELTWDGIQQPLLGHKVNPFKALYNRIDMRQVEALVEASKEEVKAAAAPVTG
PLADDPIQETITFDDFAKVDLRVALIENAEFVEGSDKLLRLTLDLGGEKRNVFSGIRSAYPDPQALIGRH
TIMVANLAPRKMRFGISEGMVMAAGPGGKDIFLLSPDAGAKPGHQVKKHHHHHH

LeuRS

MQEQYRPEEIESKVQLHWDEKRTFEVTEDESKEKYYCLSMLPYPSGRLHMGHVRNYTIGDVIARYQRMLG
KNVLQPIGWDAFGLPAEGAAVKNNTAPAPWTYDNIAYMKNQLKMLGFGYDWSRELATCTPEYYRWEQKFF
TELYKKGLVYKKTSAVNWCPNDQTVLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDLDKLD
HWPDTVKTMQRNWIGRSEGVEITFNVNDYDNTLTVYTTRPDTFMGCTYLAVAAGHPLAQKAAENNPELAA
FIDECRNTKVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLMEYGTGAVMAVPGHDQRDYEFAS
KYGLNIKPVILAADGSEPDLSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGERKVNYRLRDW
GVSRQRYWGAPIPMVTLEDGTVMPTPDDQLPVILPEDVVMDGITSPIKADPEWAKTTVNGMPALRETDTF
DTFMESSWYYARYTCPQYKEGMLDSEAANYWLPVDIYIGGIEHAIMHLLYFRFFHKLMRDAGMVNSDEPA
KQLLCQGMVLADAFYYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVYTGMSKMSKSKNNGIDP
QVMVERYGADTVRLFMMFASPADMTLEWQESGVEGANRFLKRVWKLVYEHTAKGDVAALNVDALTENQKA
LRRDVHKTIAKVTDDIGRRQTFNTAIAAIMELMNKLAKAPTDGEQDRALMQEALLAVVRMLNPFTPHICF
TLWQELKGEGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRAKITVPVDATEEQVRERAGQEHLVAKYLD
GVTVRKVIYVPGKLLNLVVGKHHHHHH

HisRS

MAKNIQAIRGMNDYLPGETAIWQRIEGTLKNVLGSYGYSEIRLPIVEQTPLFKRAIGEVTDVVEKEMYTF
EDRNGDSLTLRPEGTAGCVRAGIEHGLLYNQEQRLWYIGPMFRHERPQKGRYRQFHQLGCEVFGLQGPDI
DAELIMLTARWWRALGISEHVTLELNSIGSLEARANYRDALVAFLEQHKEKLDEDCKRRMYTNPLRVLDS
KNPEVQALLNDAPALGDYLDEESREHFAGLCKLLESAGIAYTVNQRLVRGLDYYNRTVFEWVTNSLGSQG
TVCAGGRYDGLVEQLGGRATPAVGFAMGLERLVLLVQAVNPEFKADPVVDIYLVASGADTQSAAMALAER
LRDELPGVKLMTNHGGGNFKKQFARADKWGARVAVVLGESEVANGTAVVKDLRSGEQTAVAQDSVAAHLR
TLLGKHHHHHH

sdV-GrsA

MLNSSKSILIHAQNKNGTHEEEQYLFAVNNTKAEYPRDKTIHQLFEEQVSKRPNNVAIVCENEQLTYHEL
NVKANQLARIFIEKGIGKDTLVGIMMEKSIDLFIGILAVLKAGGAYVPIDIEYPKERIQYILDDSQARML
LTQKHLVHLIHNIQFNGQVEIFEEDTIKIREGTNLHVPSKSTDLAYVIYTSGTTGNPKGTMLEHKGISNL
KVFFENSLNVREDDRIIQTGAIGFDALTFEVFGSLLHGAELYPVTKDVLLDAEKLHKFLQANQITIMWLT
SPLFNQLSQGTEEMFAGLRSLIVGGDALSPKHINNVKRKCPNLTMWNGYGPTENTTFSTCFLIDKEYDDN
IPIGKAIQNTQIYIVDENLQLKSVGEAGELCIGGEGLARGYWKRPELTSQKFVDNPFVPGEKLYKTGDQA
RWLSDGNIEYLGRIDNQVKIRGHRVELEEVESILLKHMYISETAVSVHKDHQEQPYLCAYFVSEKHIPLE
QLRQFSSEELPTYMIPSYFIQLDKMPLTSNGKIDRKQLPEPDLTFGMRVDYEAPRNEIEETLVTIWQDVL
GIEKIGIKDNFYALGGDSIKAIQVAARLHSYQLKLETKDLLKYPTIDQLVHYIKDSKRRSEQGIVEGEIG
LTPIQHWFFEQQFTNMHHWNQSYMLYRPNGFDKEILLRVFNKIVEHHDALRMIYKHHNGKIVQINRGLEG
TLFDFYTFDLTANDNEQQVICEESARLQNSINLEVGPLVKIALFHTQNGDHLFMAIHHLVVDGISWRILF
EDLATAYEQAMHQQTIALPEKTDSFKDWSIELEKYANSELFLEEAEYWHHLNYYTENVQIKKDYVTMNNK
QKNIRYVGMELTIEETEKLLKNVNKAYRTEINDILLTALGFALKEWADIDKIVINLEGHGREEILEQMNI
ARTVGWFTSQYPVVLDMQKSDDLSYQIKLMKENLRRIPNKGIGYEIFKYLTTEYLRPVLPFTLKPEINFN
YLGQFDTDVKTELFTRSPYSMGNSLGPDGKNNLSPEGESYFVLNINGFIEEGKLHITFSYNEQQYKEDTI
QQLSRSYKQHLLAIIEHCVQKEDTELTPSDFSFKELELEEMDDIFDLLADSLTGSRSHHHHHH

NMR analysis and spectra

NMR measurements were performed on a Bruker AVANCE II 300 MHz, Bruker AVANCE II 500 MHz and a Bruker AVANCE II 600 MHz spectrometer, equipped with a Bruker Cryoplatform. The chemical shifts are reported in parts per million (ppm) relative to the solvent residual peak of D_2O (^1H: 4.79 ppm, singlet) for ^1H and trifluoroacetic acid (^{13}C: 164.2 ppm, quartet) for ^{13}C spectra. For NMR analysis, hydroxamates and corresponding amino acids were dissolved in 1.8% trifluoroacetic acid (TFA) in D_2O and recorded NMR spectra were compared. The conversion to hydroxamic acid is determined by ~0.2 ppm shift of Cα ^1H and ~5 ppm shift of $^{13}C\alpha$ with respect to the corresponding proton and carbon shifts of free amino acid. The purity of hydroxamates was determined by comparing integral of Cα ^1H of the hydroxamate to the $^{13}C\alpha$ proton of corresponding free amino acid, which was a major impurity. Atoms are labeled according to the atom names, remoteness codes and order indicators for amino acid residues of Protein Data Bank (PDB) nomenclature.

SI Table 4. NMR data.

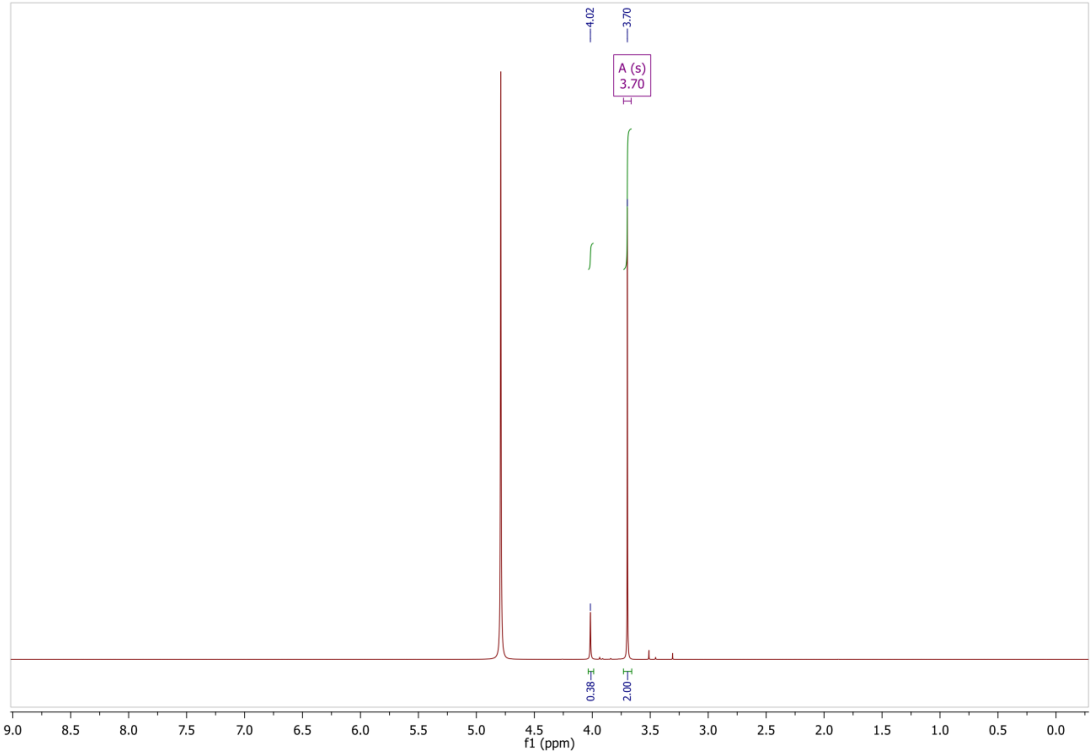
Compound	Position	δ_H, mult. (*J* in Hz)/nH	δ_C
GlyHA	C		165.7
	Cα	3.70, s/1H	39.7
AlaHA	C		169.1
	Cα	3.94, q (7.1)/1H	48.5
	Cβ	1.46, d (7.1)/3H	17.7
SerHA	C		166.3
	Cα	4.05, dd (5.9, 4.7)/1H	54.0
	Cβ	3.96, dd (12.3, 4.7)/1H 3.90, dd (12.3, 5.9)/1H	61.4
ThrHA	C		166.2
	Cα	3.68, d (7.2)/1H	58.5
	Cβ	4.12 – 4.05, m/1H	67.6
	Cγ	1.26, d (6.4)/3H	20.2
CystineHA	C		166.1
	Cα	4.23, dd (6.8)/1H	51.4
	Cβ	3.31, dd (14.8, 6.5)/1H 3.23, dd (14.8, 7.2)/1H	38.5
ValHA	C		167.4
	Cα	3.53, d (7.1)/1H	58.1
	Cβ	2.16 – 2.02, m/1H	31.2
	Cγ1	0.98, d (6.8)/3H	18.8
	Cγ2	0.93, d (6.8)/3H	18.8
LeuHA	C		168.4
	Cα	3.80, dd (7.4)/1H	51.2
	Cβ	1.75 – 1.50, m/3H	40.9
	Cγ		25.3
	Cδ1	0.88, d (2.7)/3H	22.7
	Cδ2	0.86, d (2.6)/3H	22.6
IleHA	C		167.5
	Cα	3.61, d (6.8)/1H	57.0
	Cβ	1.93 – 1.80, m/1H	37.5
	Cγ1	1.52 – 1.39, m/1H 1.26 – 1.09, m/1H	25.9
	Cγ2	0.96 – 0.81, m/6H	15.2
	Cδ		11.6
MetHA	C		167.5
	Cα	3.97, dd (7.1)/1H	51.6
	Cβ	2.63 – 2.50, m/2H	31.0
	Cγ	2.17 – 2.10, m/2H	29.6
	Cε	2.08, s/3H	15.3

SI Table 4. NMR data (continued).

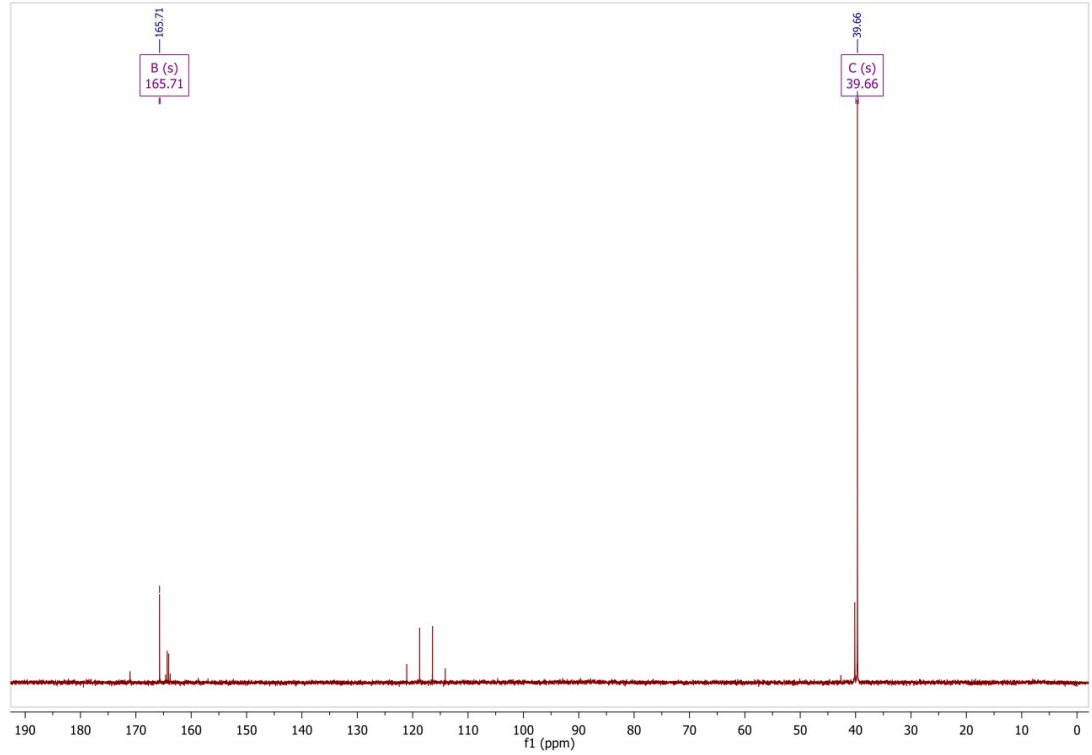
Compound	Position	δ_H, mult. (*J* in Hz)/nH	δ_C
ProHA	C		168.2
	Cα	4.25, dd (7.6)/1H	59.3
	Cβ	2.43 – 2.34, m/1H	31.0
	Cγ	2.10 – 2.01, m/3H	25.3
	Cδ	3.46 – 3.35, m/2H	48.0
PheHA (DMSO-*d*$_6$)	C		164.2
	Cα	3.86 – 3.67, m/1H	51.8
	Cβ	3.03 – 2.96, m/2H	37.1
	Cγ		135.0
	Cδ1	7.39- 7.16, m/5H	128.7
	Cδ2		
	Cε1		129.5
	Cε2		
	Cζ		127.3
TyrHA	C		167.2
	Cα	3.95, dd (6.8, 8.3)/1H	54.1
	Cβ	3.11 – 3.00, m/2H	37.3
	Cγ		126.9
	Cδ1	7.13 – 7.08, m/2H	132.2
	Cδ2		
	Cε1	6.87 – 6.82, m/2H	117.2
	Cε2		
	Cζ		156.5
TrpHA	C		167.8
	Cα	4.07, dd (7.4)/1H	53.2
	Cβ	3.33, d (3.0)/1H 3.31, d (2.2)/1H	28.3
	Cγ		107.8
	Cδ1	7.26, s/1H	126.8
	Cδ2		127.9
	Cε2		137.7
	Cε3	7.61 – 7.57, m/1H	119.6
	Cζ2	7.5 – 7.46, m/1H	113.5
	Cζ3	7.24 – 7.21, m/1H	121.0
	Cη2	7.17 – 7.11, m/1H	123.6
AspHA	C		166.9
	Cα	4.26, dd (6.5)/1H	49.0
	Cβ	3.08 – 2.96, m/2H	36.0
	Cγ		173.7
GluHA	C		167.4
	Cα	3.93, dd (7.0)/1H	51.9
	Cβ	2.16, m/2H	27.1
	Cγ	2.53, dd (12.3, 7.2)/2H	30.4
	Cδ		177.3

SI Table 4. NMR data (continued).

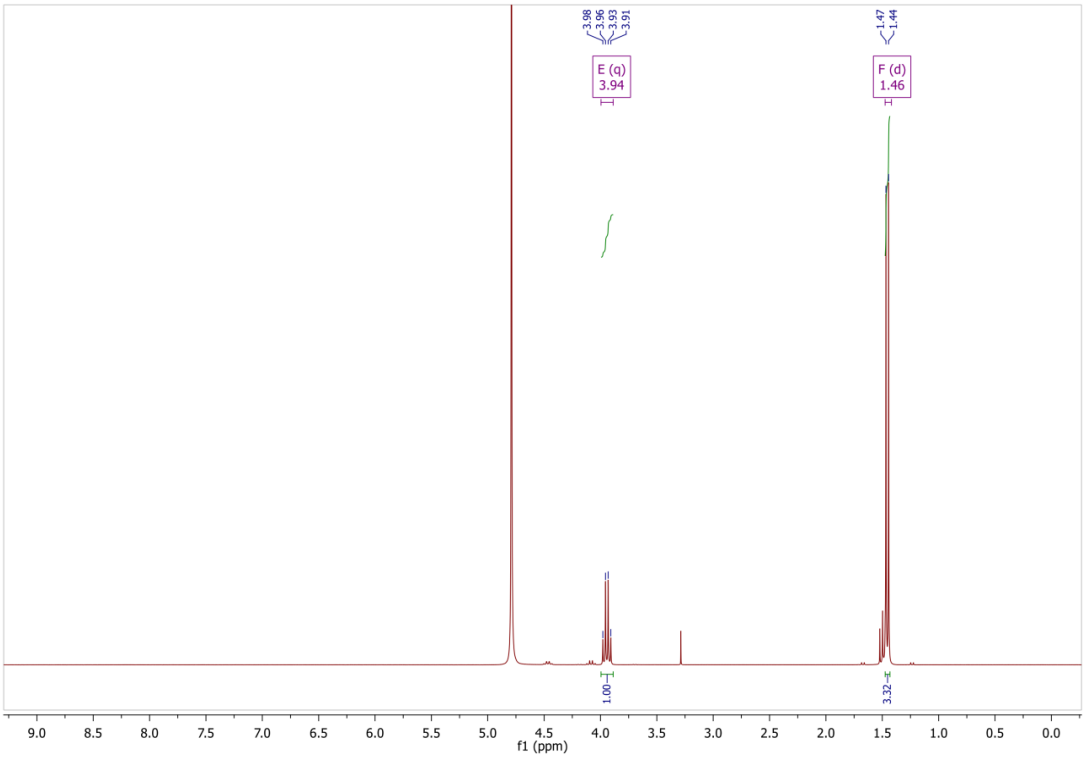
Compound	Position	δ_H, mult. (*J* in Hz)/nH	δ_C
HisHA	C		166.1
	Cα	4.15, dd (7.3)/1H	51.7
	Cβ	3.44 – 3.35, m/2H	27.2
	Cγ		127.1
	Cδ2	7.44, s/1H	119.8
	Cε1	8.72, s/1H	135.8
LysHA	C		167.8
	Cα	3.79, dd (7.1)/2H	52.4
	Cβ	1.89 – 1.79, m/2H	31.5
	Cγ	1.43 – 1.33, m/2H	22.7
	Cδ	1.70 – 1.58, m/2H	27.6
	Cε	2.98 – 2.89, m/2H	40.4
ArgHA	C		167.7
	Cα	3.86, dd (7.0)/1H	52.4
	Cβ	1.93 – 1.86, m/2H	29.3
	Cγ	1.67 – 1.58, m/2H	25.2
	Cδ	3.24 – 3.18, m/2H	41.8
	Cζ		65.8
PipHA	C		168.1
	Cα	3.84, dd (12.0, 3.4)/1H	57.0
	Cβ	2.13 – 2.03, m/1H 1.97 – 1.47, m/5H	28.1
	Cγ		22.3
	Cδ		22.4
	Cε	3.54 – 3.39, m/1H 3.12 – 2.98, m/1H	45.3
Phenylglycine HA	C		167.1
	Cα	4.97, s/1H	55.8
	Cβ		132.8
	Cγ1	7.44 – 7.34, m/5H	131.7
	Cγ2		
	Cδ1		130.9
	Cδ2		
	Cε		129.1
β-PheHA	C		169.0
	Cα	2.82, dd (14.9, 6.7)/1H 2.68, dd (14.9, 8.0)/1H	37.8
	Cβ	4.65 – 4.53, dd (7.3)/1H	53.6
	Cγ		136.0
	Cδ1	7.38 – 7.24, m/5H	130.8
	Cδ2		
	Cε1		131.1
	Cε2		
	Cζ		128.3



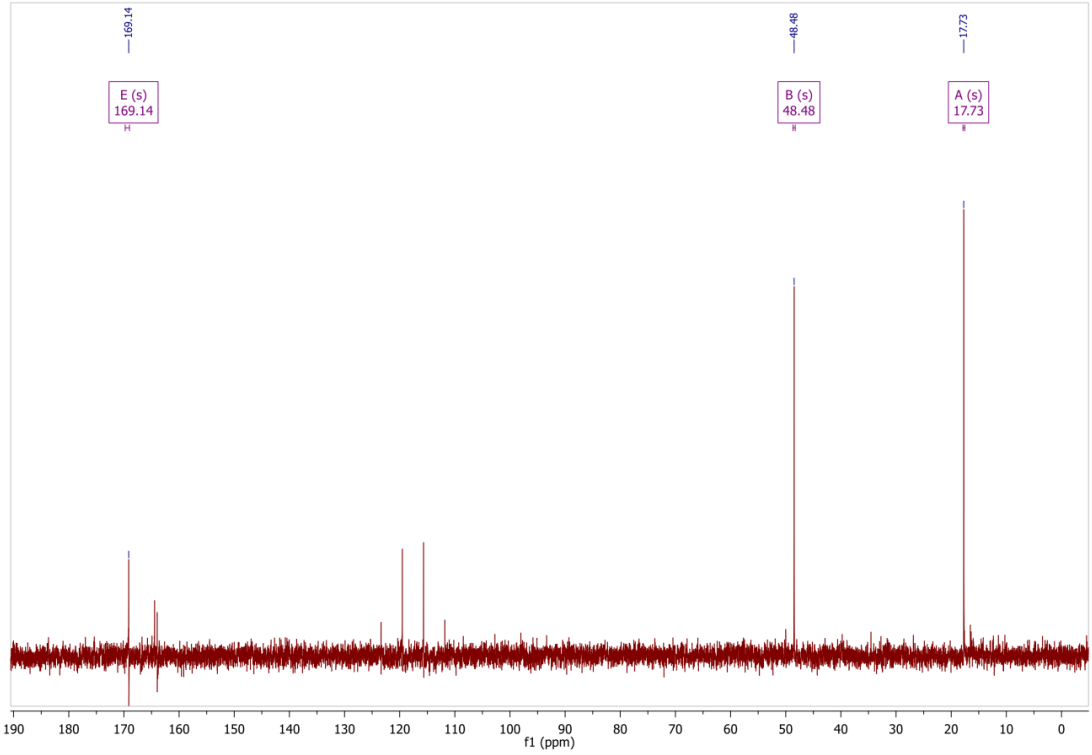
^{1}H NMR spectrum of GlyHA (D_2O + 1.8% TFA, 300 MHz).



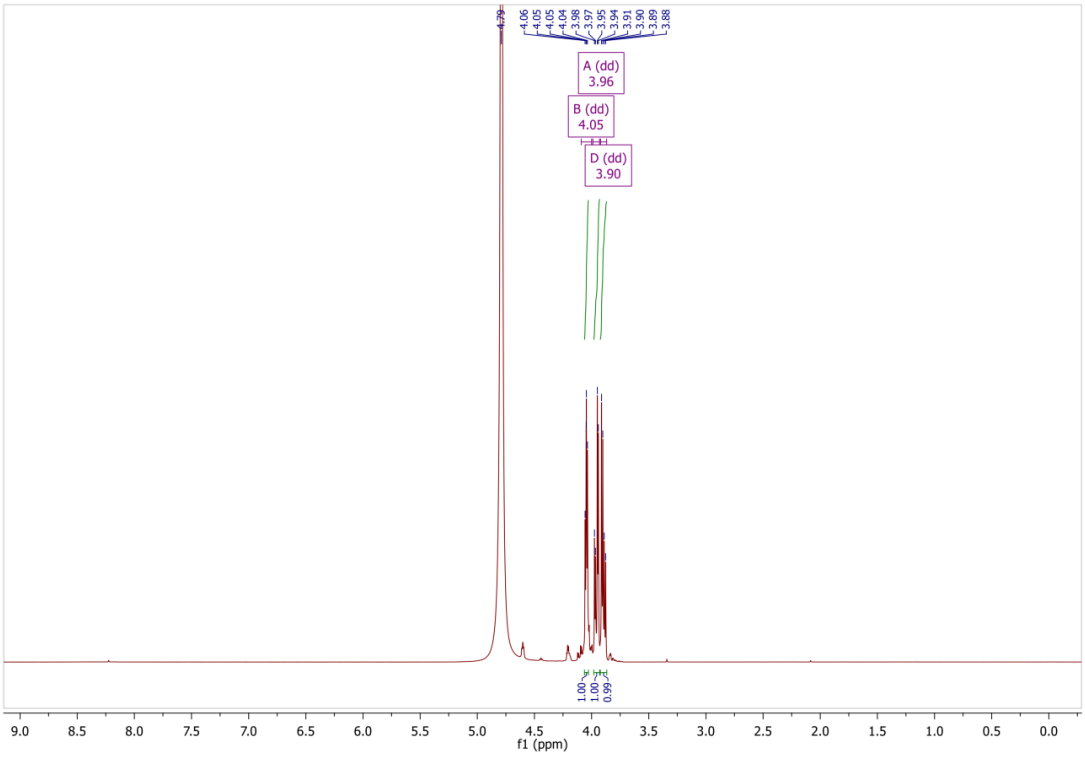
^{13}C NMR spectrum of GlyHA (D_2O + 1.8% TFA, 126 MHz).



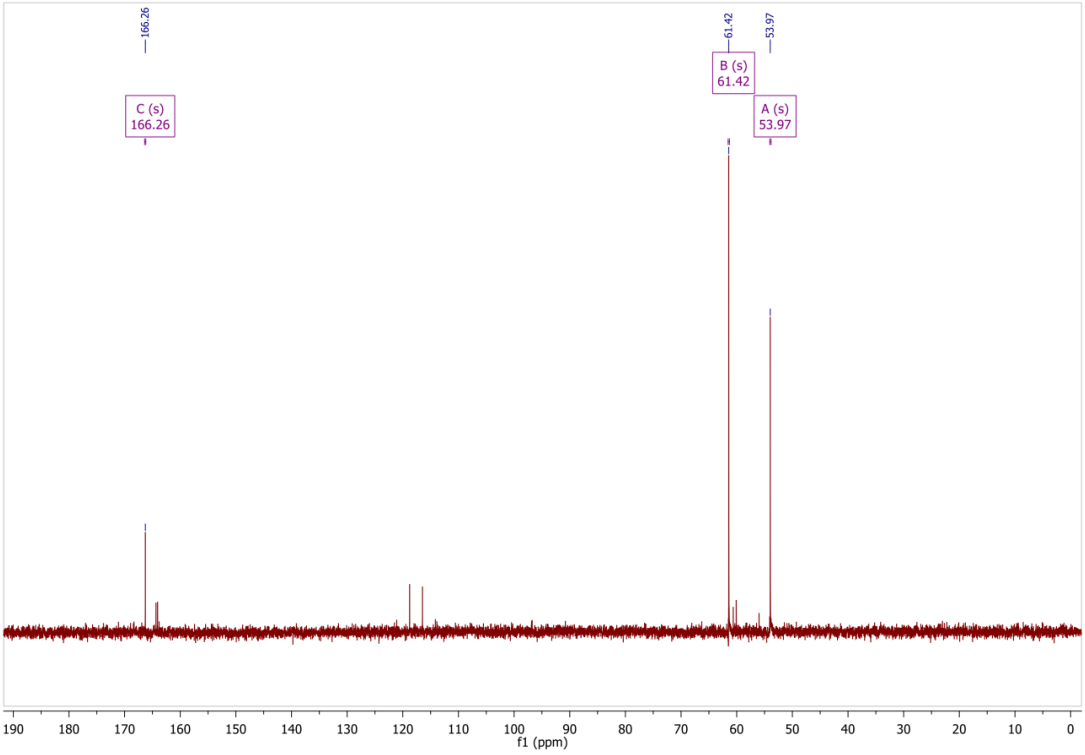
^{1}H NMR spectrum of AlaHA (D_2O + 1.8% TFA, 300 MHz).



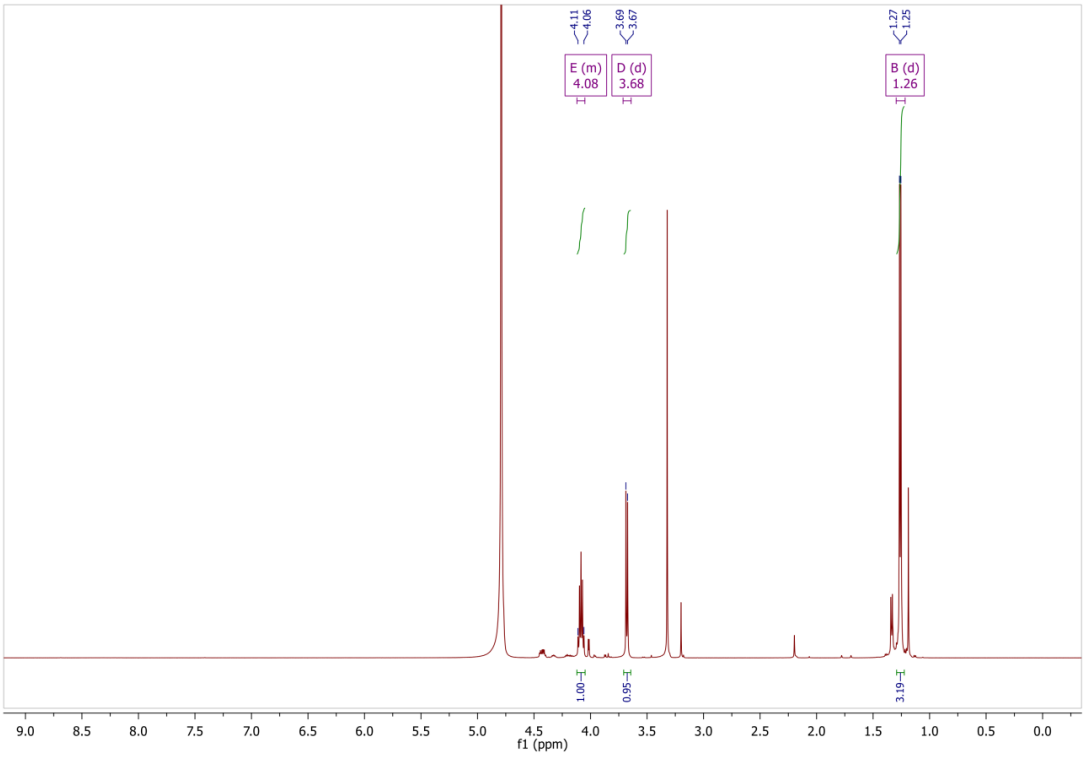
^{13}C NMR spectrum of AlaHA (D_2O + 1.8% TFA, 75 MHz).



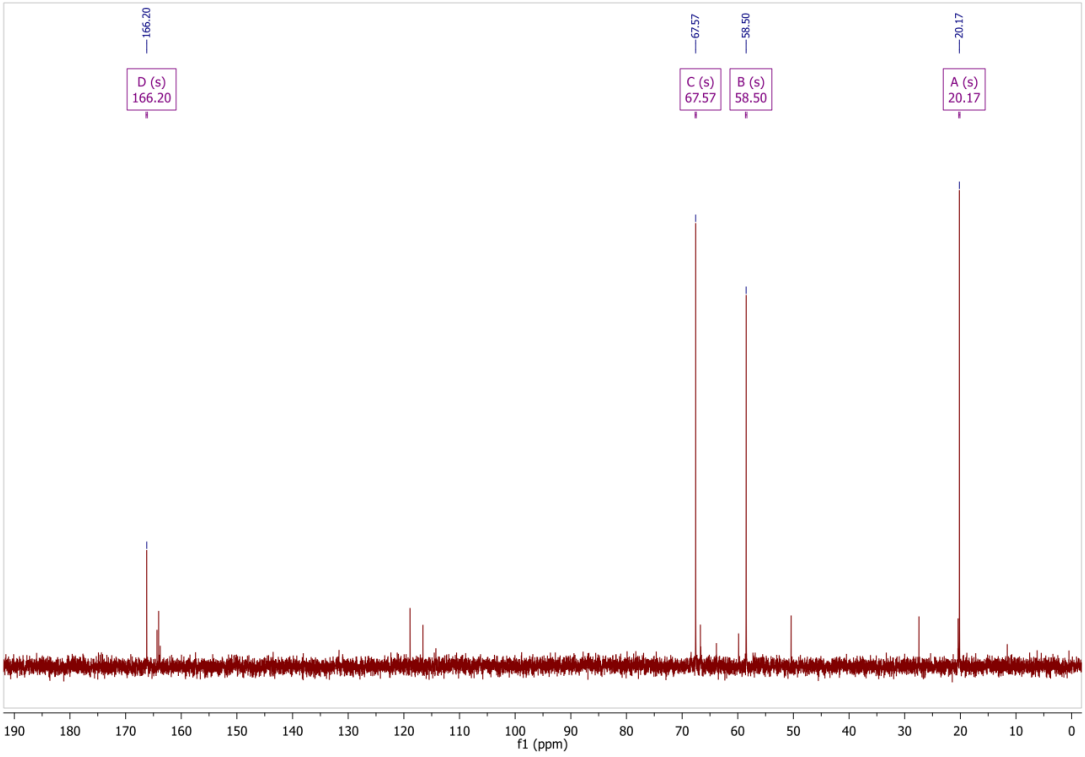
^{1}H NMR spectrum of SerHA (D_2O + 1.8% TFA, 500 MHz).



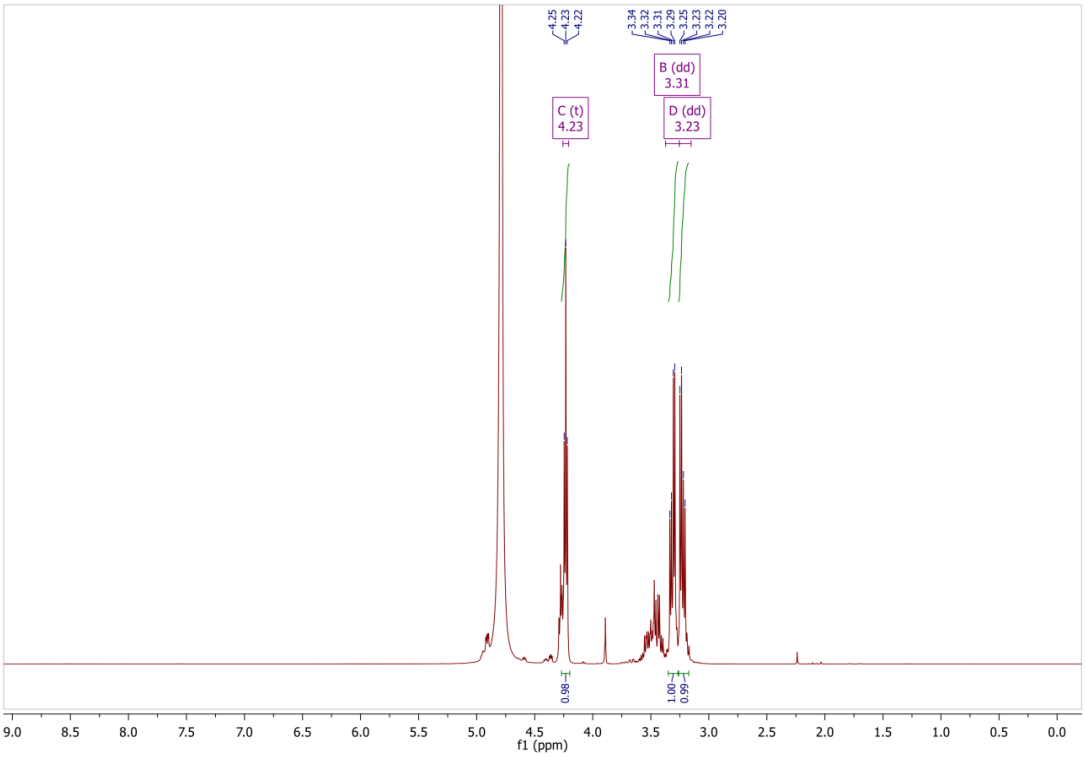
^{13}C NMR spectrum of SerHA (D_2O + 1.8% TFA, 126 MHz).



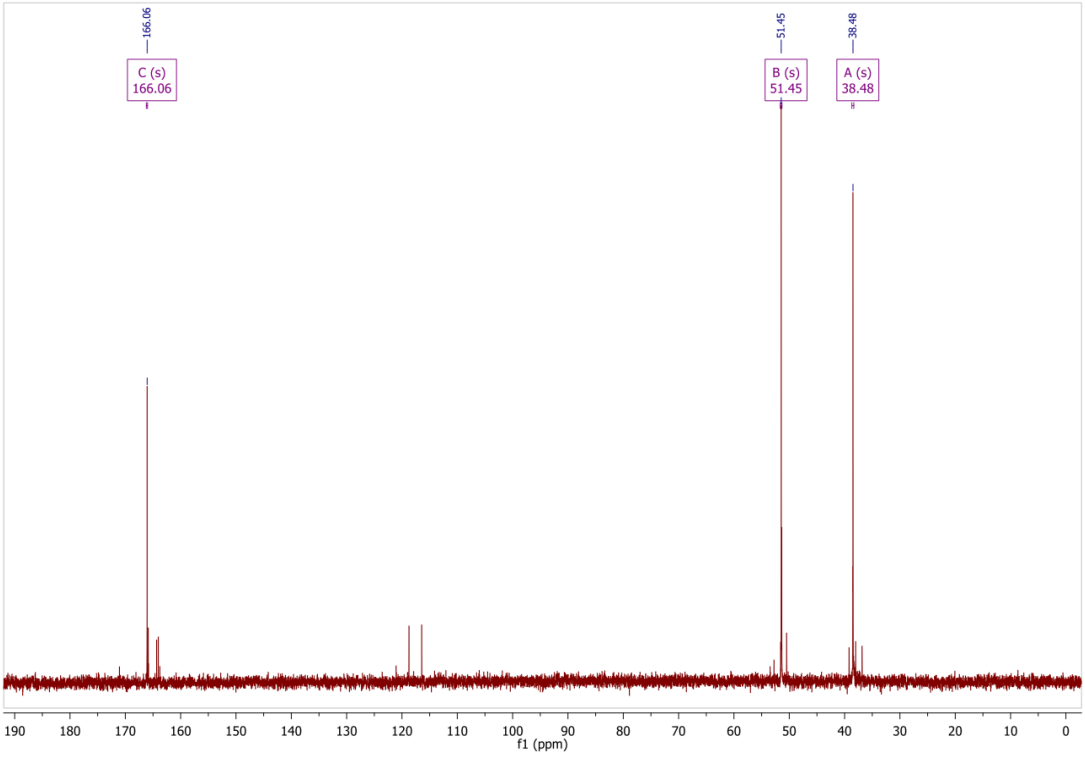
^{1}H NMR spectrum of ThrHA (D_2O + 1.8% TFA, 500 MHz).



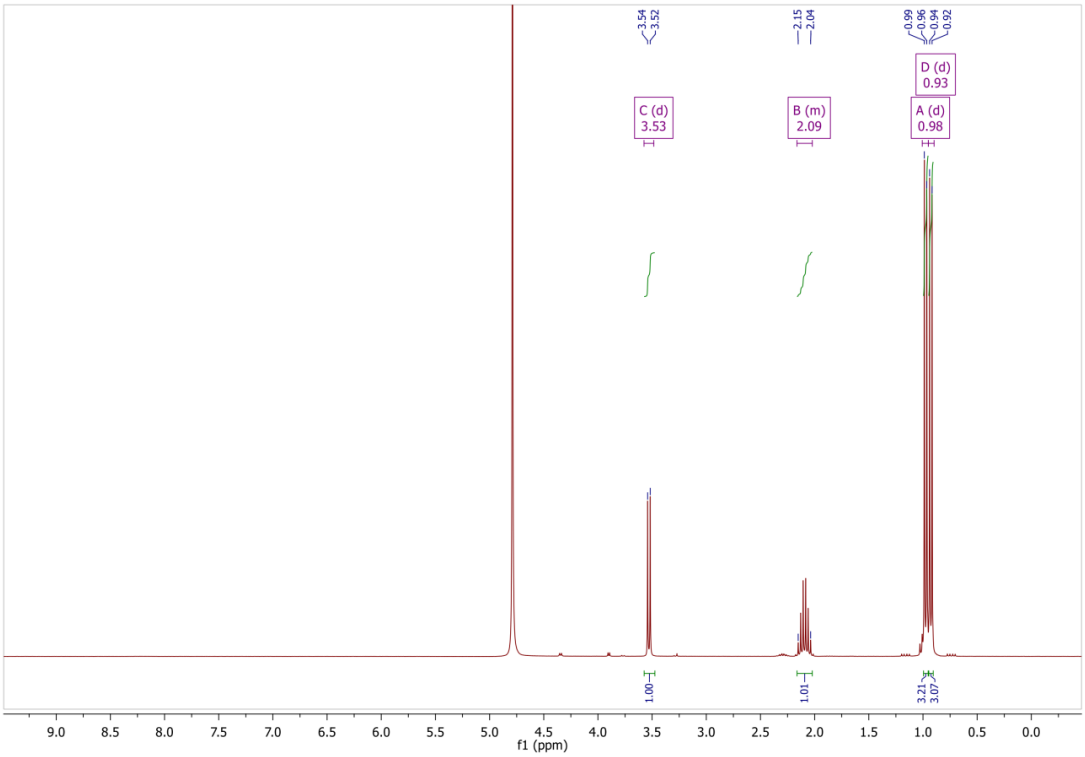
^{13}C NMR spectrum of ThrHA (D_2O + 1.8% TFA, 126 MHz).



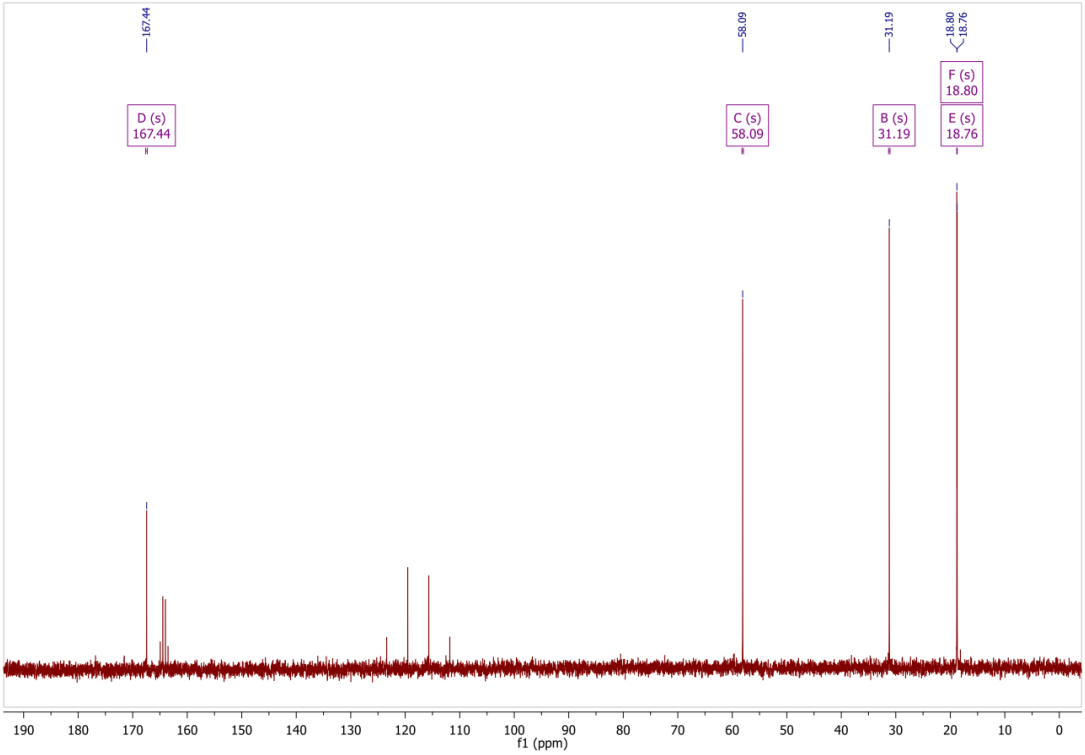
^{1}H NMR spectrum of cystine hydroxamate (D_2O + 1.8% TFA, 500 MHz).



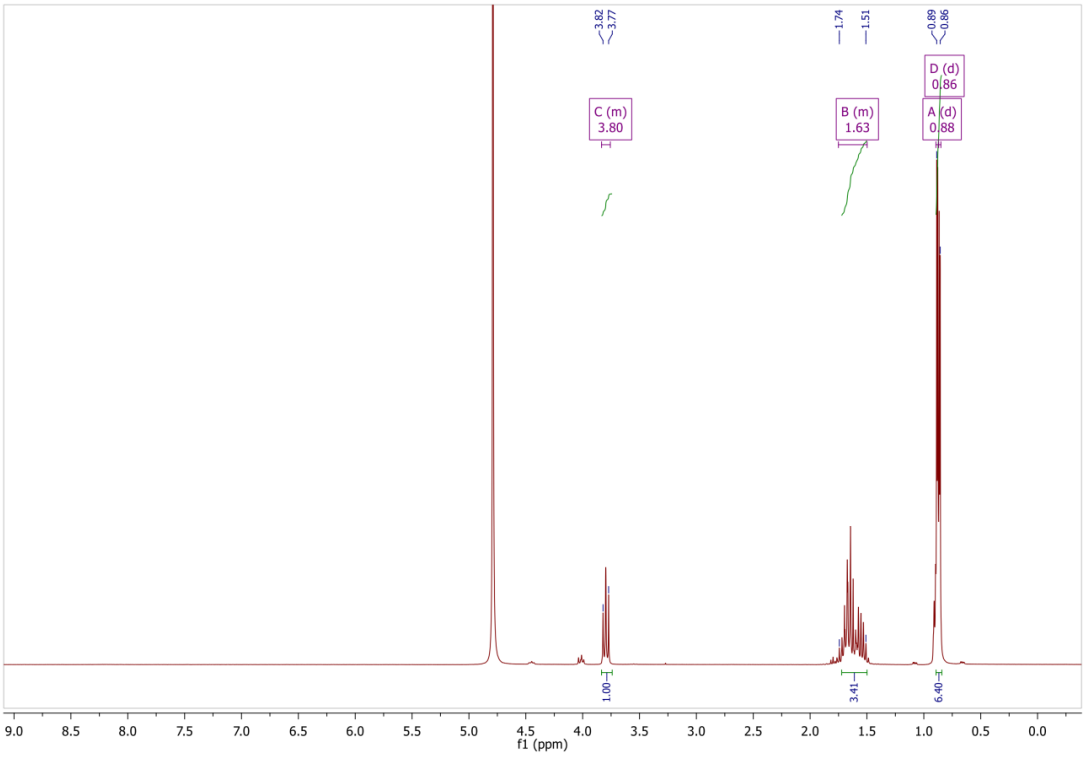
^{13}C NMR spectrum of cystine hydroxamate (D_2O + 1.8% TFA, 126 MHz).



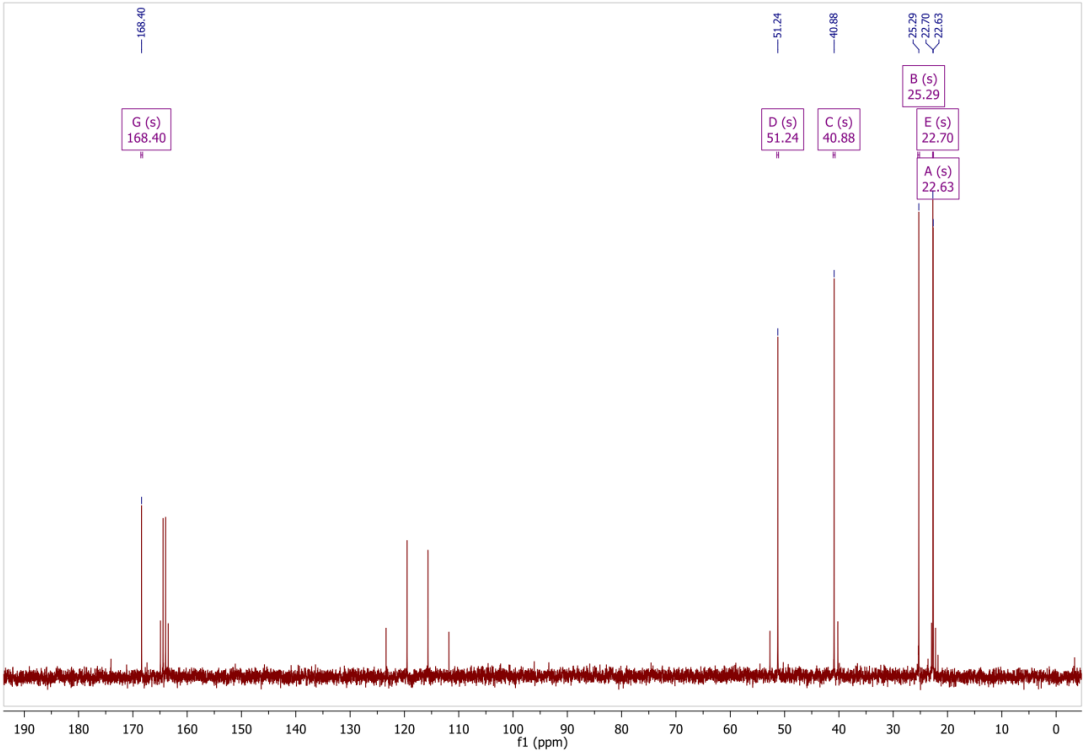
^{1}H NMR spectrum of ValHA (D_2O + 1.8% TFA, 300 MHz).



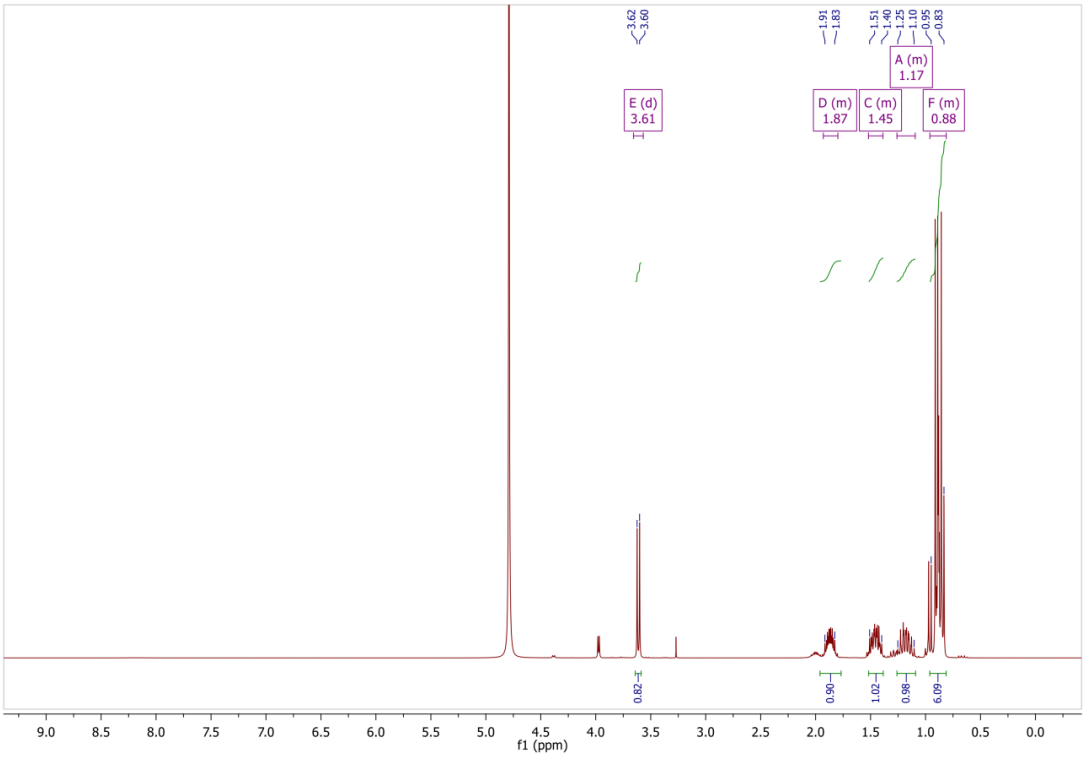
^{13}C NMR spectrum of ValHA (D_2O + 1.8% TFA, 75 MHz).



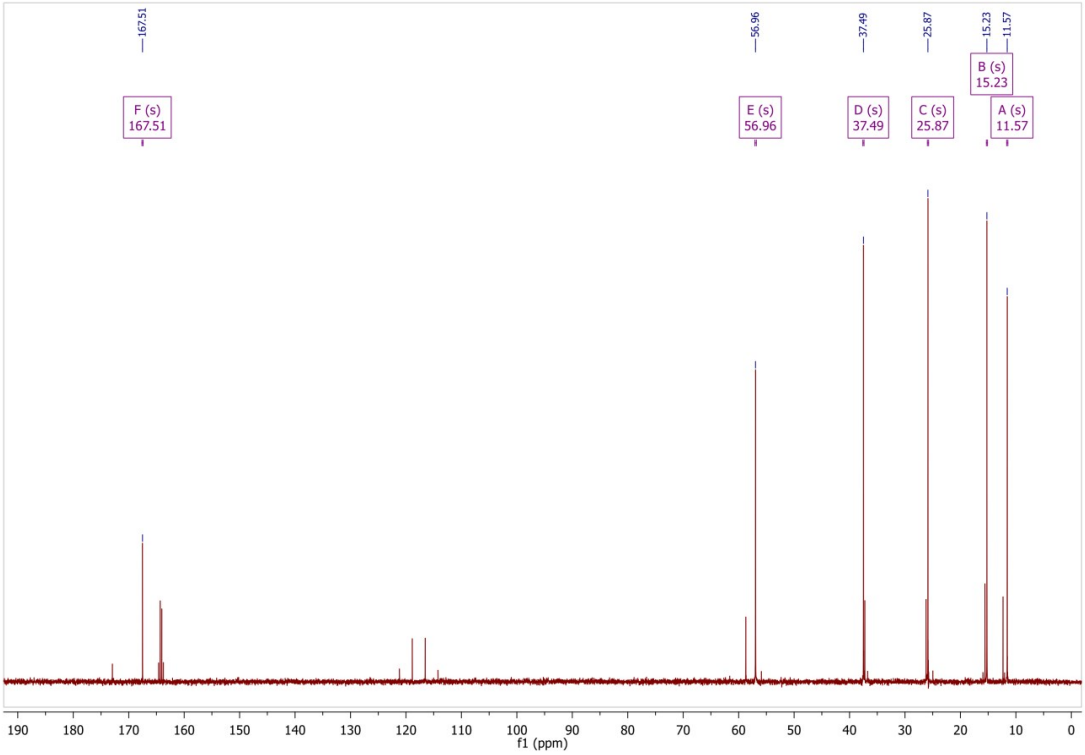
^{1}H NMR spectrum of LeuHA (D_2O + 1.8% TFA, 300 MHz).



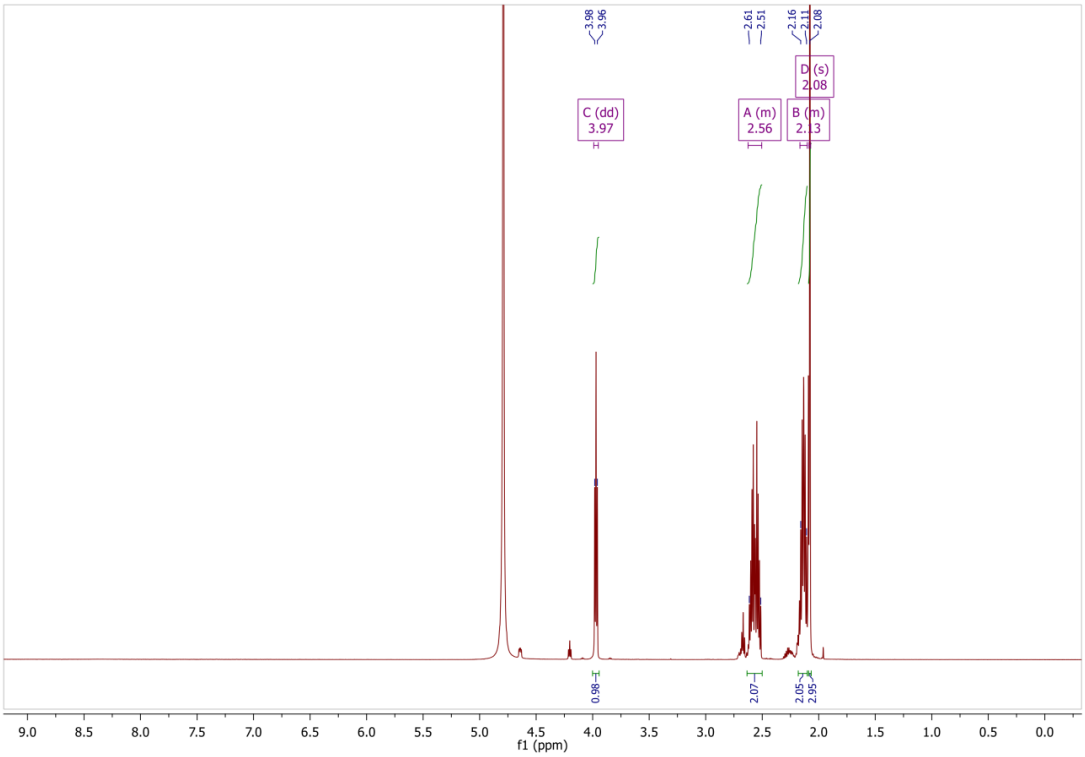
^{13}C NMR spectrum of LeuHA (D_2O + 1.8% TFA, 75 MHz).



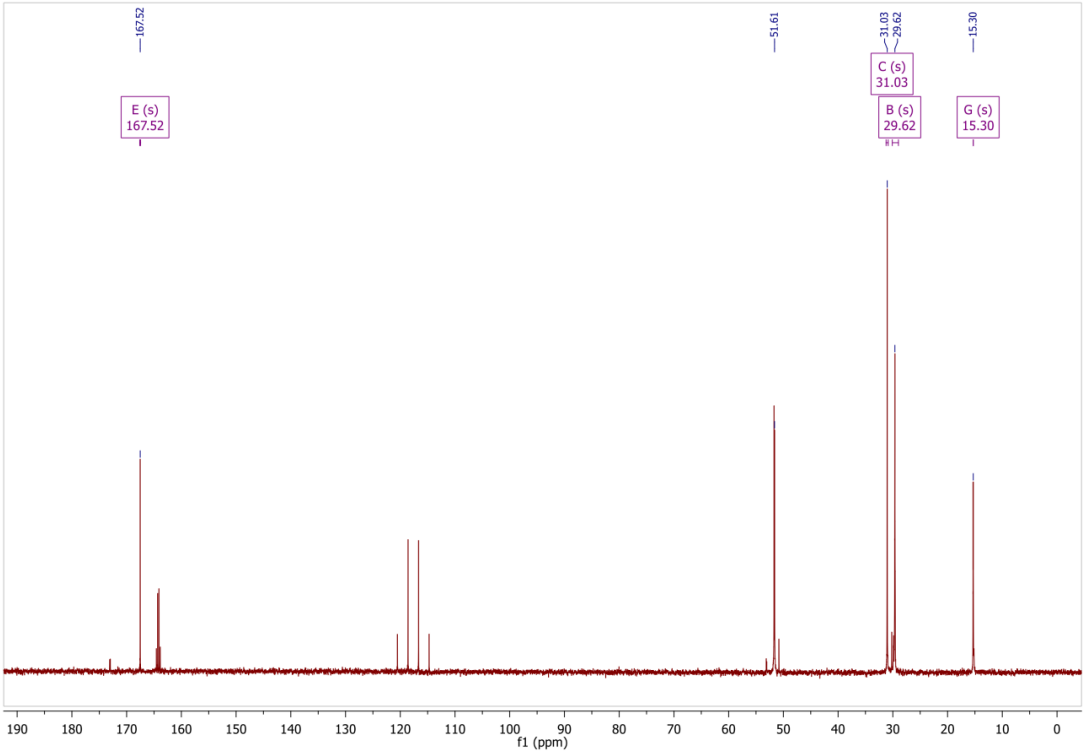
^{1}H NMR spectrum of IleHA (D_2O + 1.8% TFA, 300 MHz).



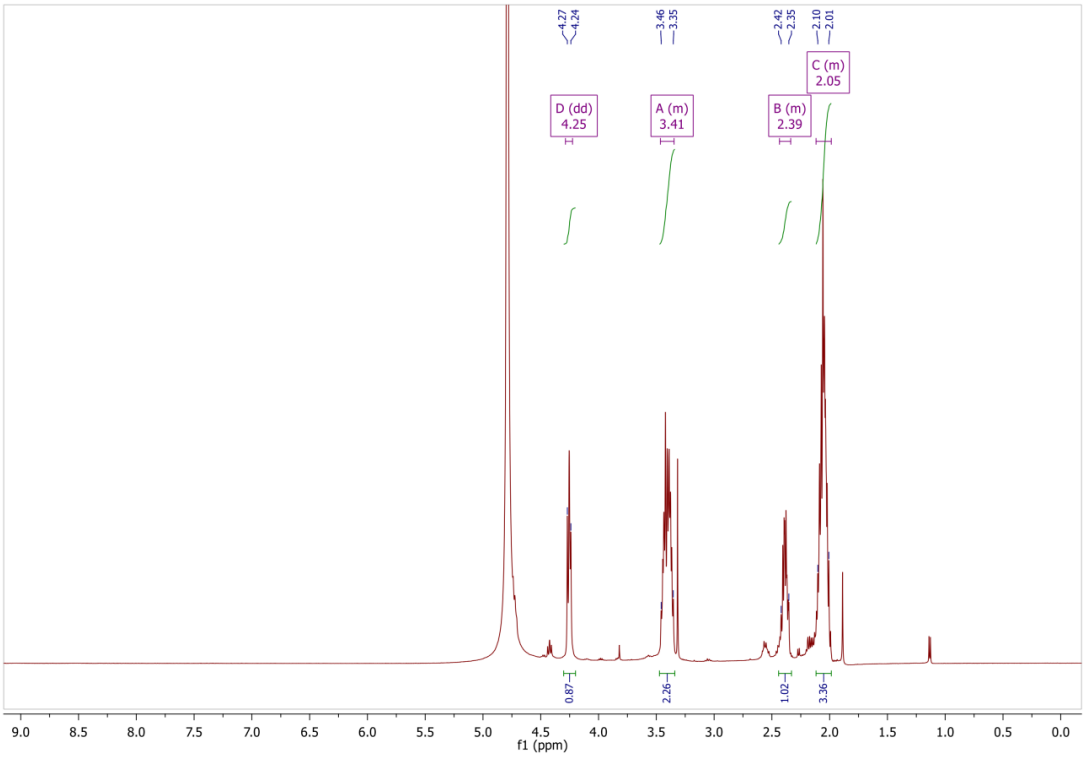
^{13}C NMR spectrum of IleHA (D_2O + 1.8% TFA, 126 MHz).



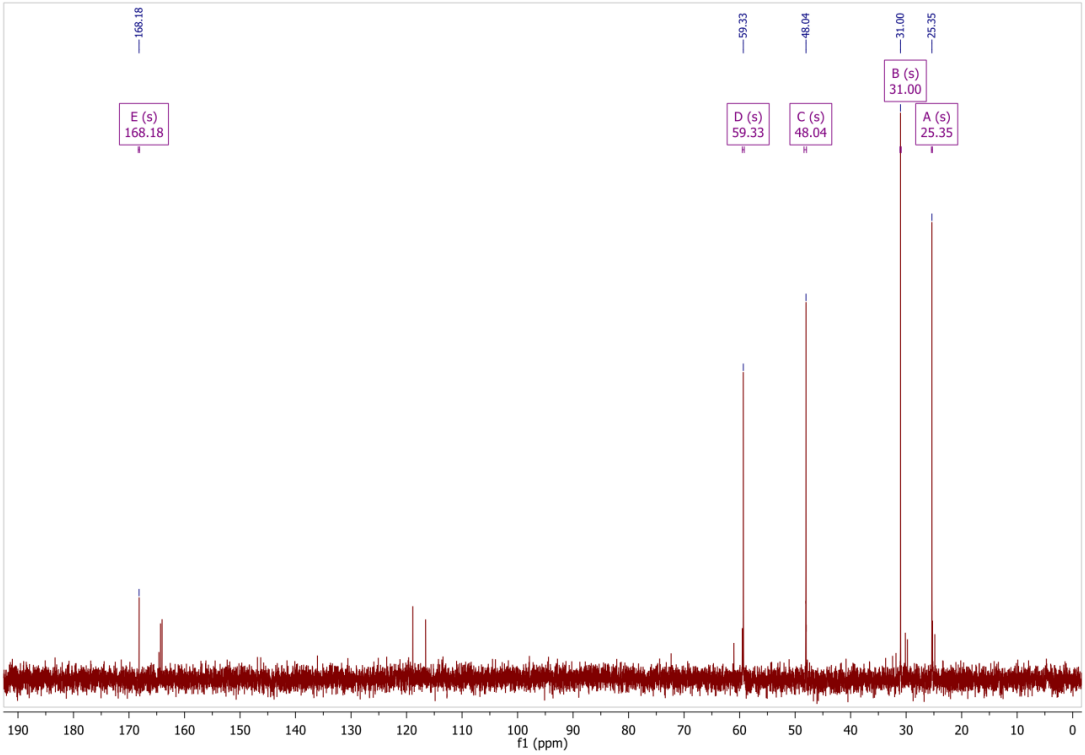
^{1}H NMR spectrum of MetHA (D_2O + 1.8% TFA, 600 MHz).



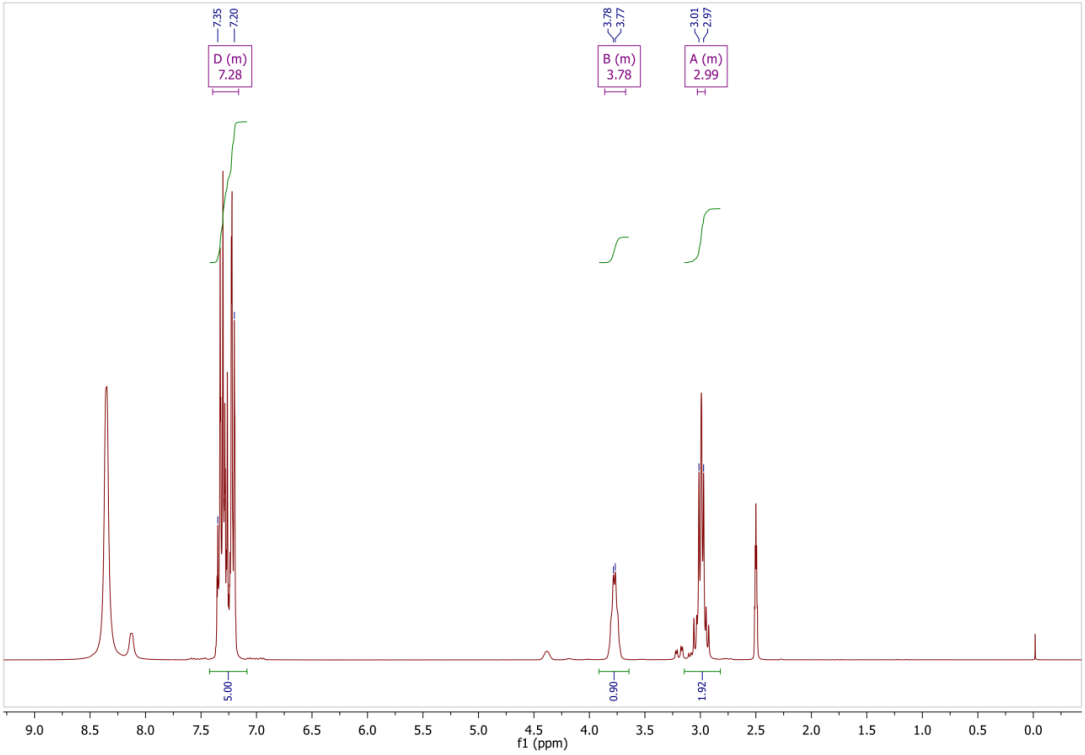
^{13}C NMR spectrum of MetHA (D_2O + 1.8% TFA, 151 MHz).



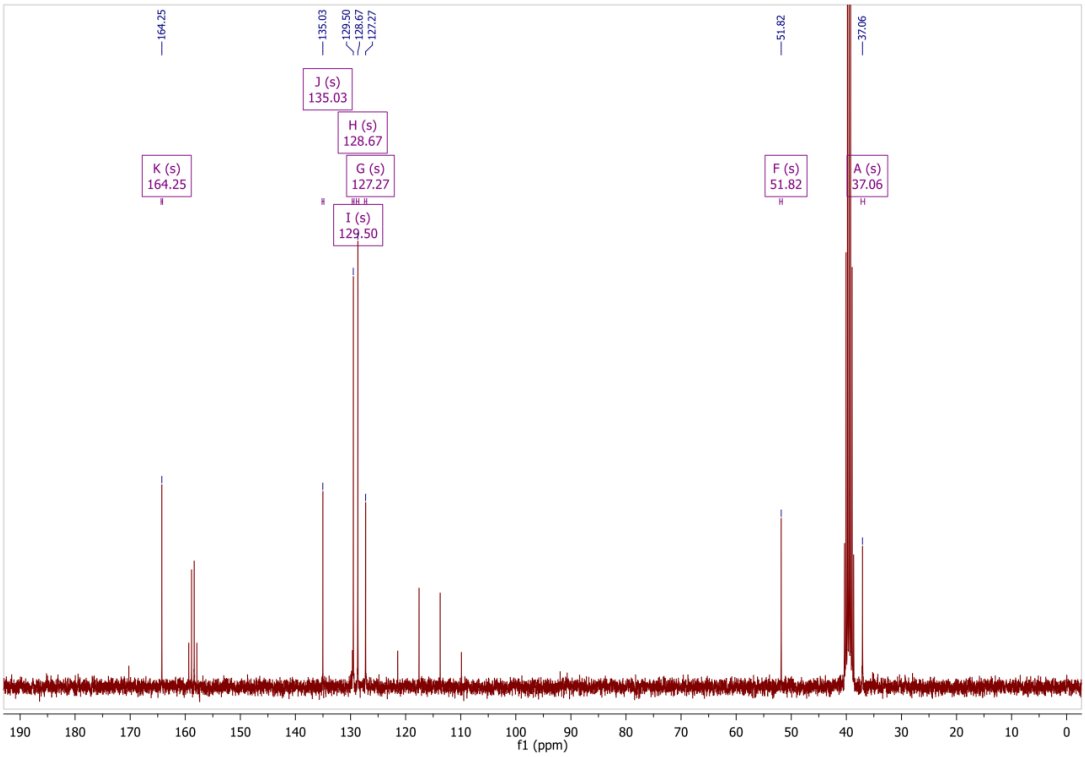
^{1}H NMR spectrum of ProHA (D_2O + 1.8% TFA, 500 MHz).



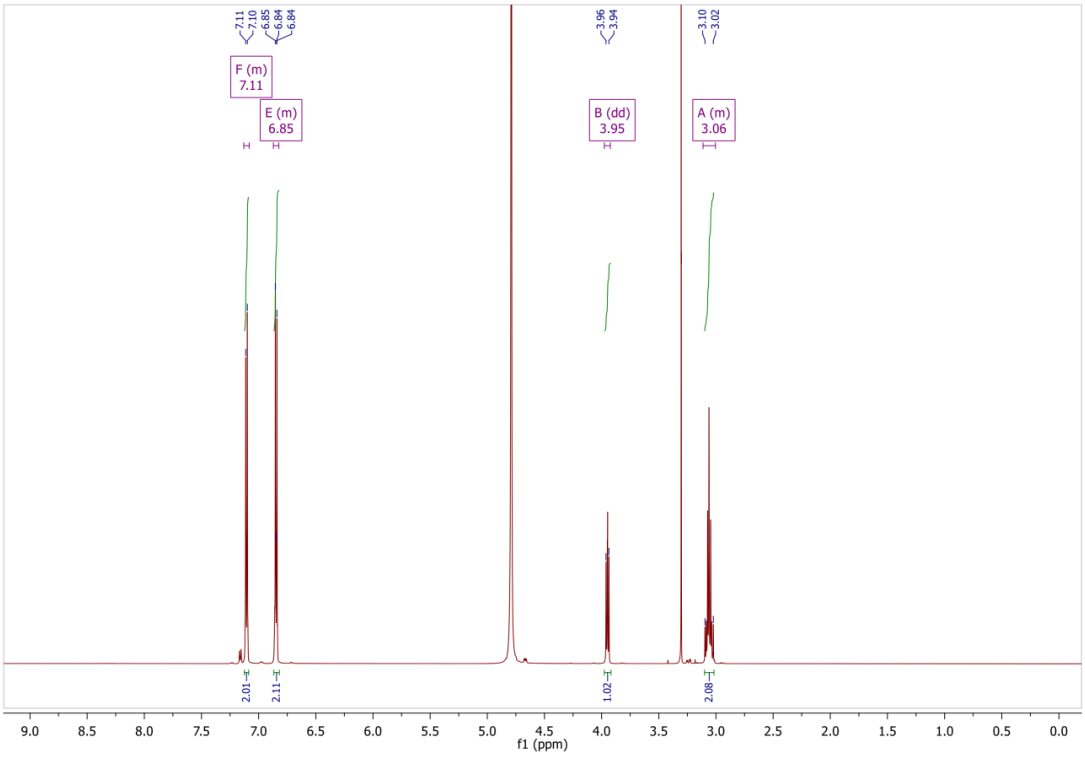
^{13}C NMR spectrum of ProHA (D_2O + 1.8% TFA, 126 MHz).



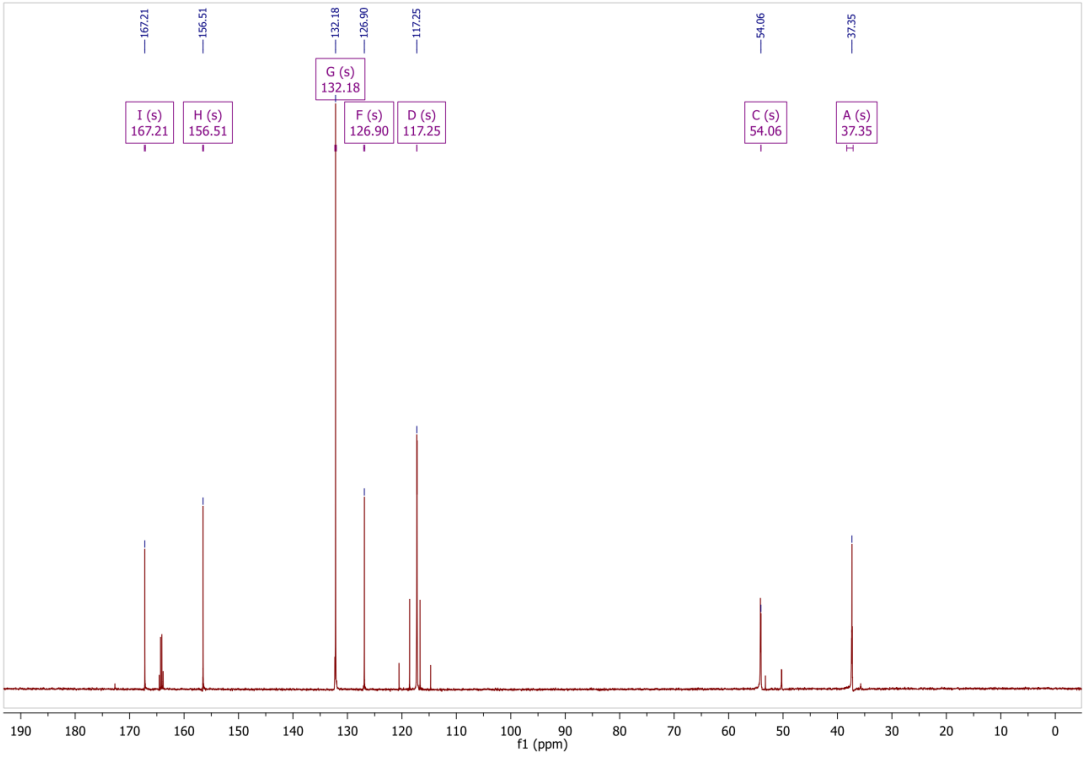
^{1}H NMR spectrum of PheHA (DMSO, 300 MHz).



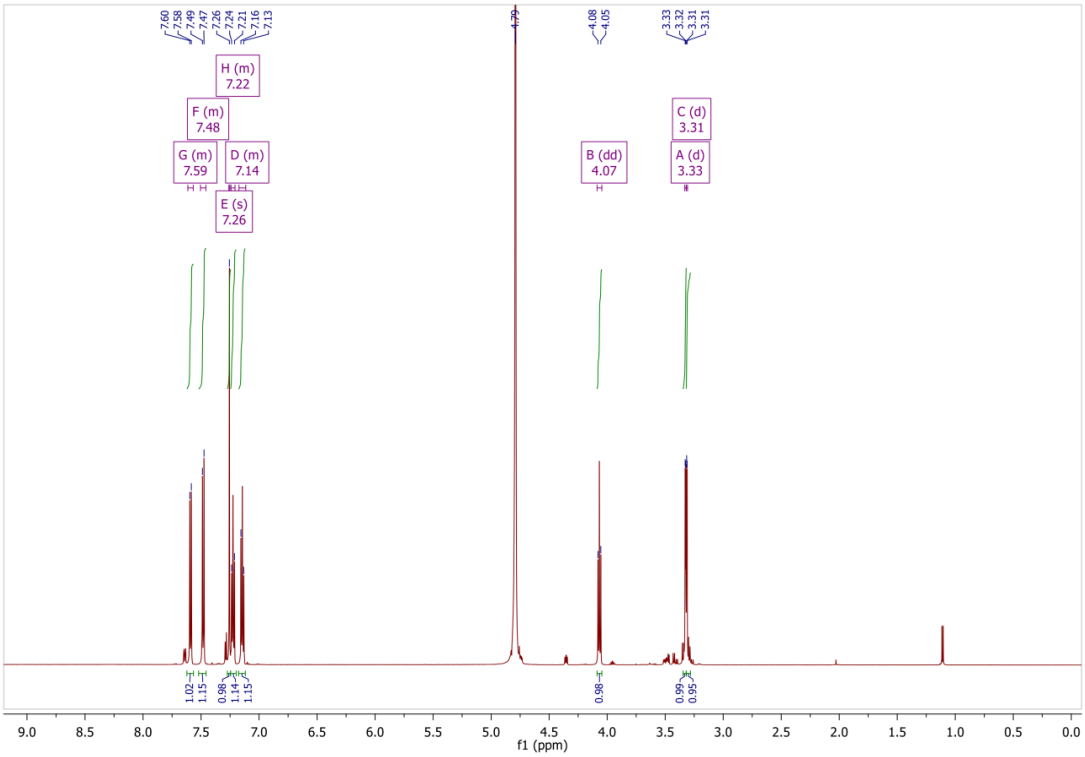
^{13}C NMR spectrum of PheHA (DMSO, 126 MHz).



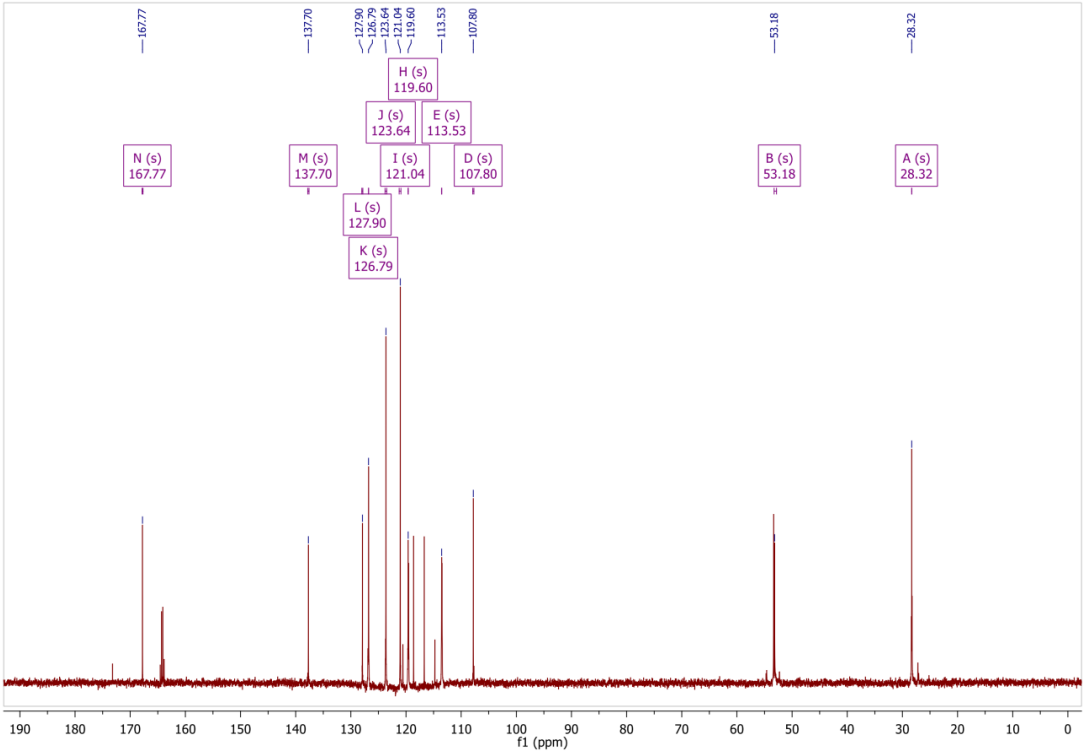
^{1}H NMR spectrum of TyrHA (D_2O + 1.8% TFA, 600 MHz).



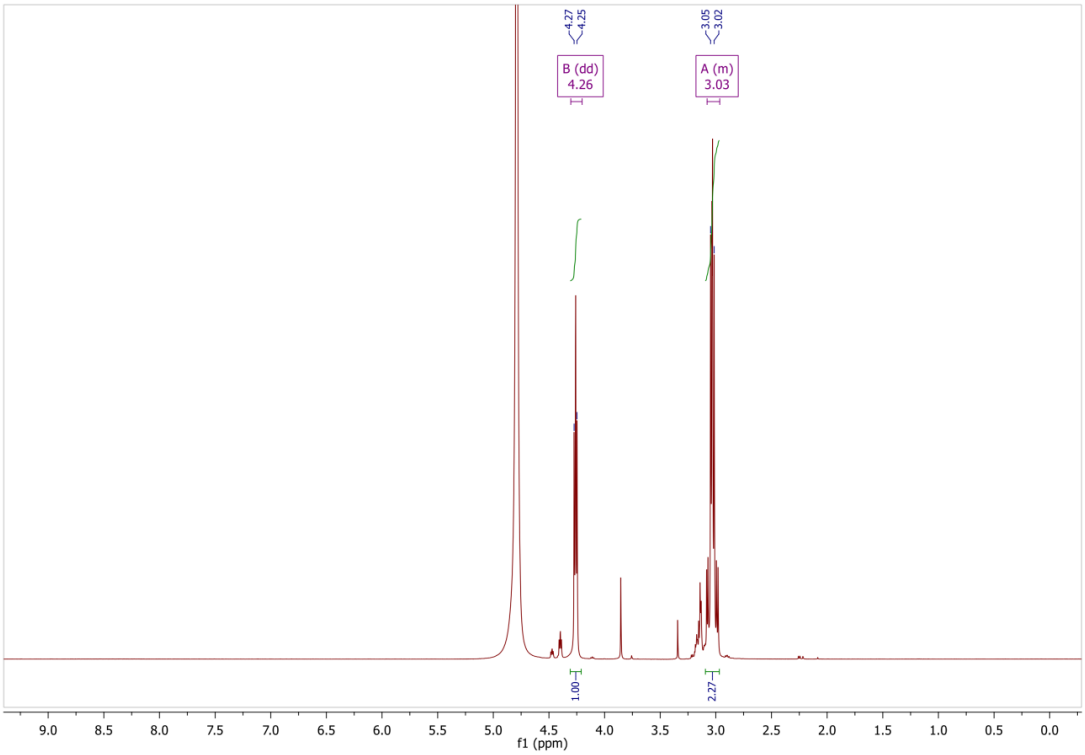
^{13}C NMR spectrum of TyrHA (D_2O + 1.8% TFA, 151 MHz).



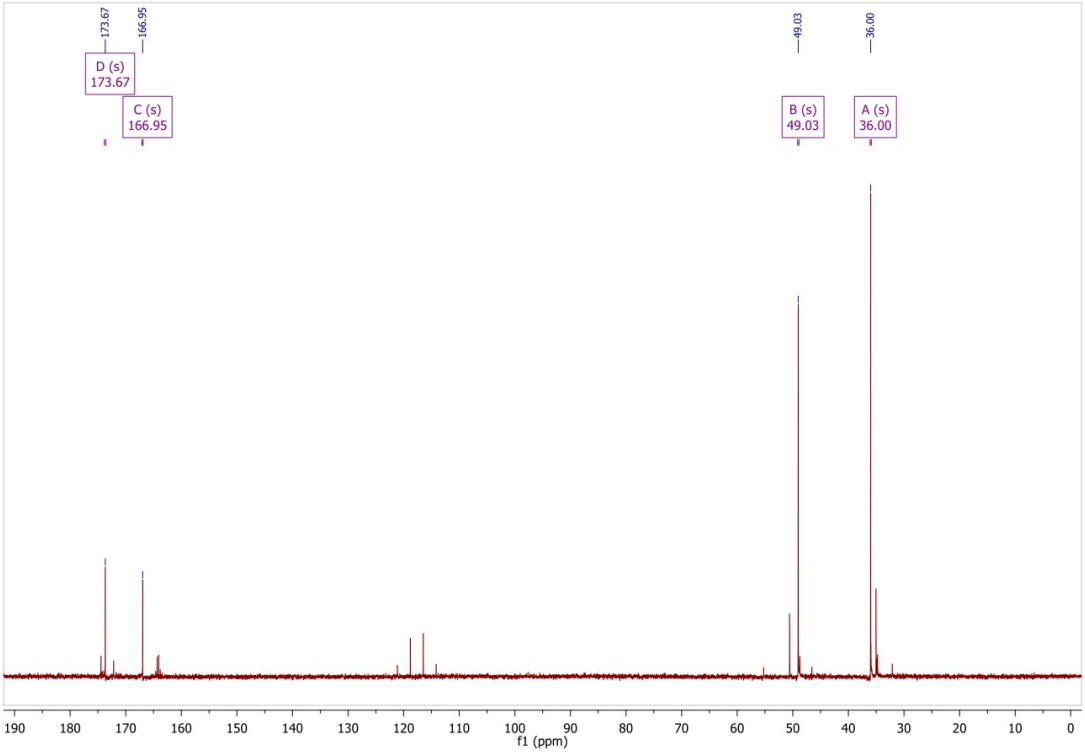
^{1}H NMR spectrum of TrpHA (D_2O + 1.8% TFA, 600 MHz).



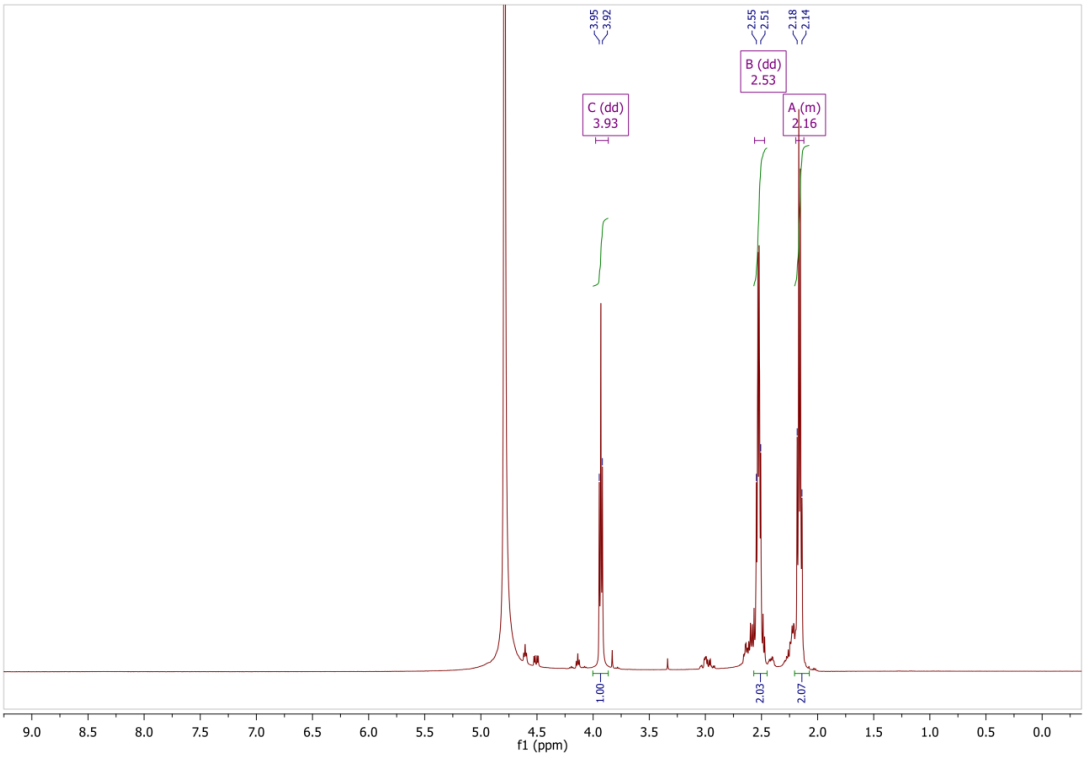
^{13}C NMR spectrum of TrpHA (D_2O + 1.8% TFA, 151 MHz).



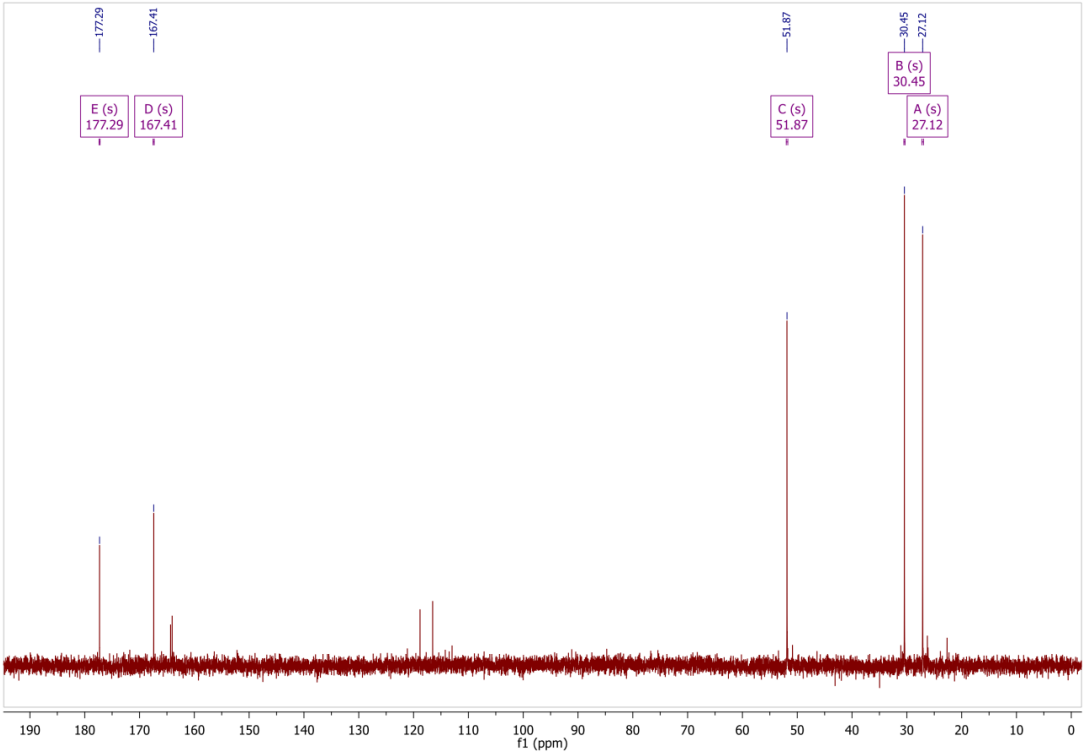
^{1}H NMR spectrum of AspHA (D_2O + 1.8% TFA, 500 MHz).



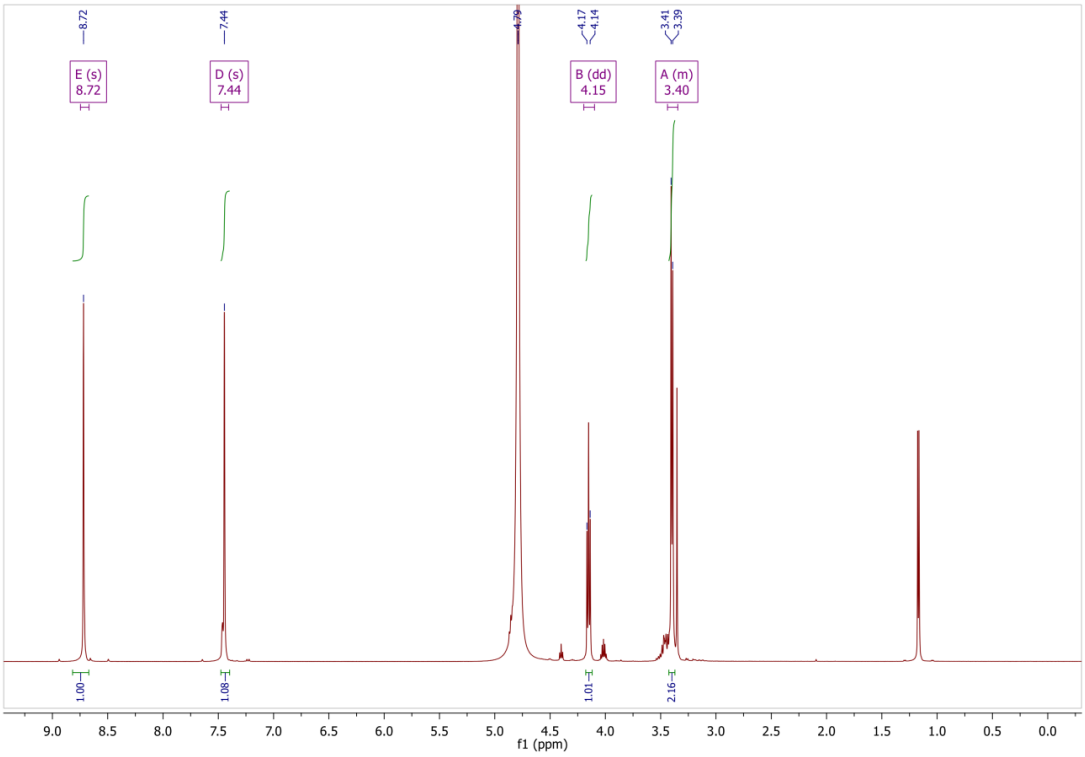
^{13}C NMR spectrum of AspHA (D_2O + 1.8% TFA, 126 MHz).



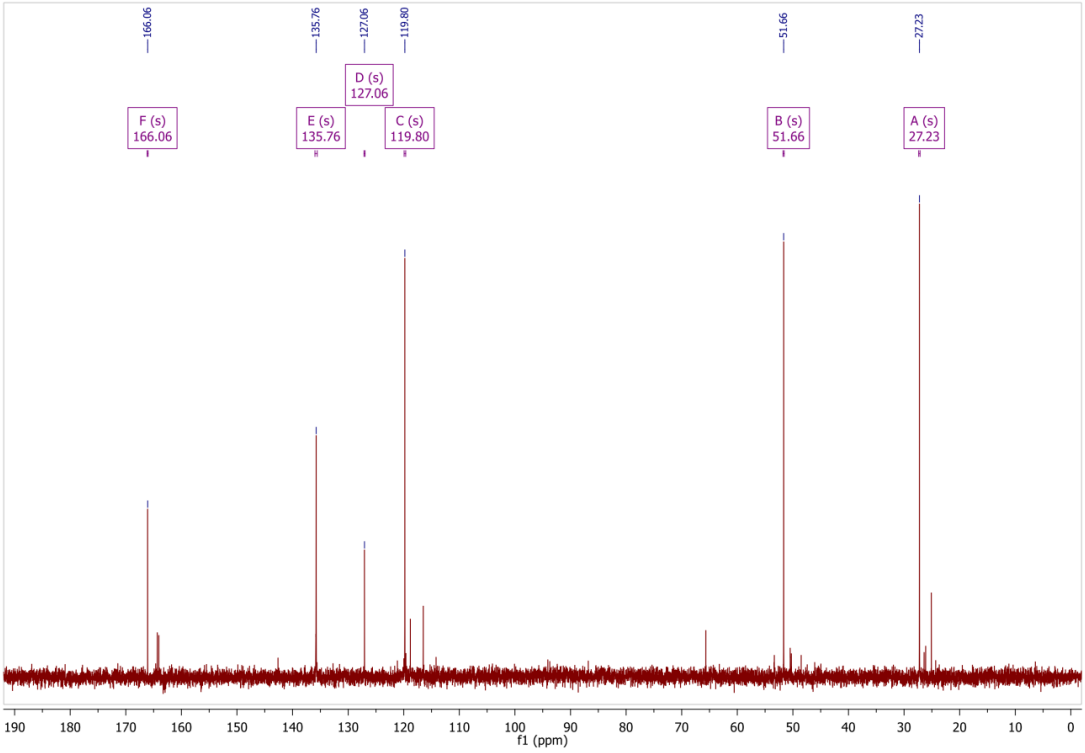
^{1}H NMR spectrum of GluHA (D_2O + 1.8% TFA, 500 MHz).



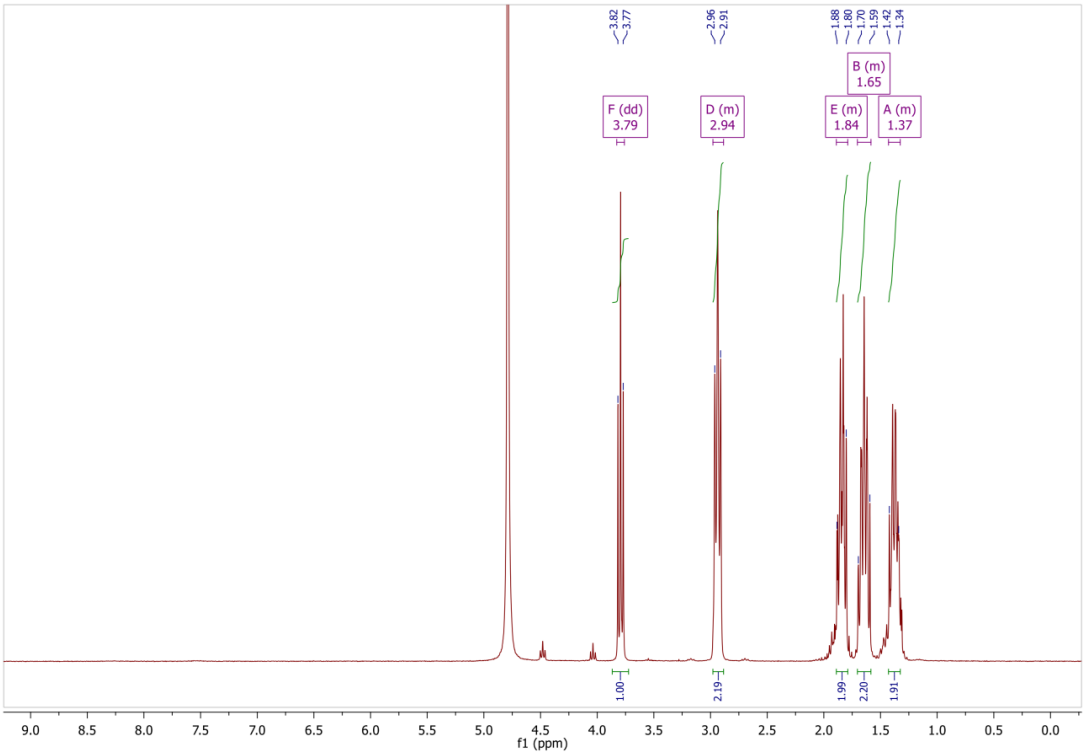
^{13}C NMR spectrum of GluHA (D_2O + 1.8% TFA, 126 MHz).



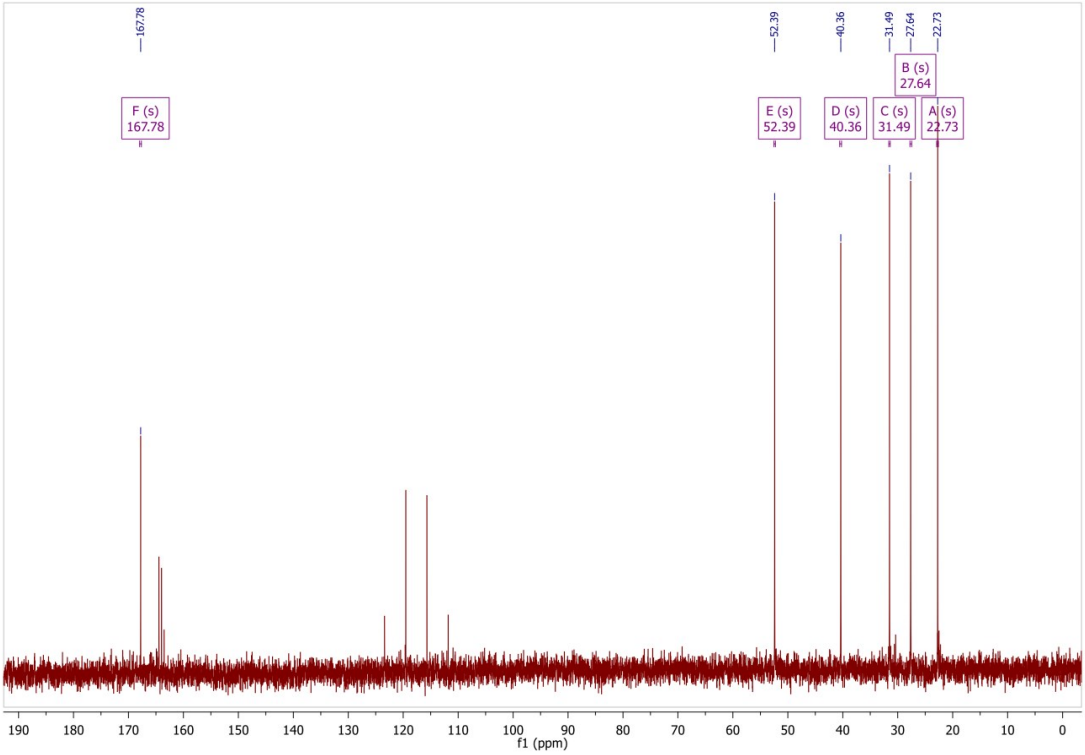
^{1}H NMR spectrum of HisHA (D_2O + 1.8% TFA, 500 MHz).



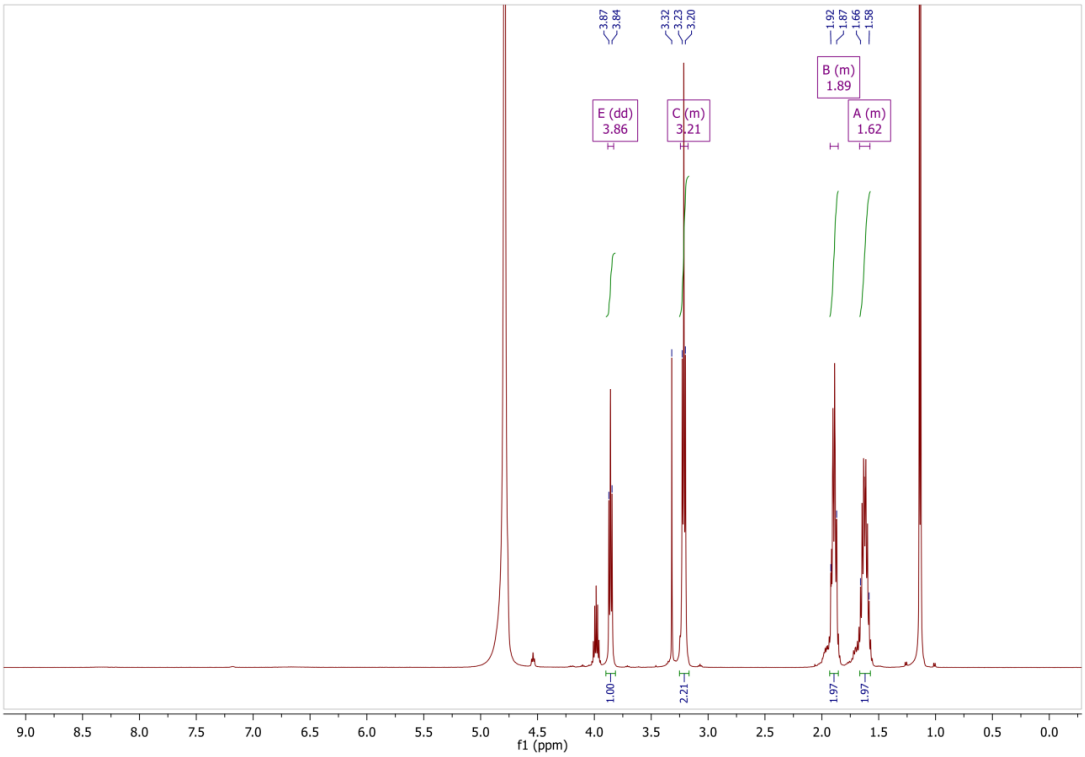
^{13}C NMR spectrum of HisHA (D_2O + 1.8% TFA, 126 MHz).



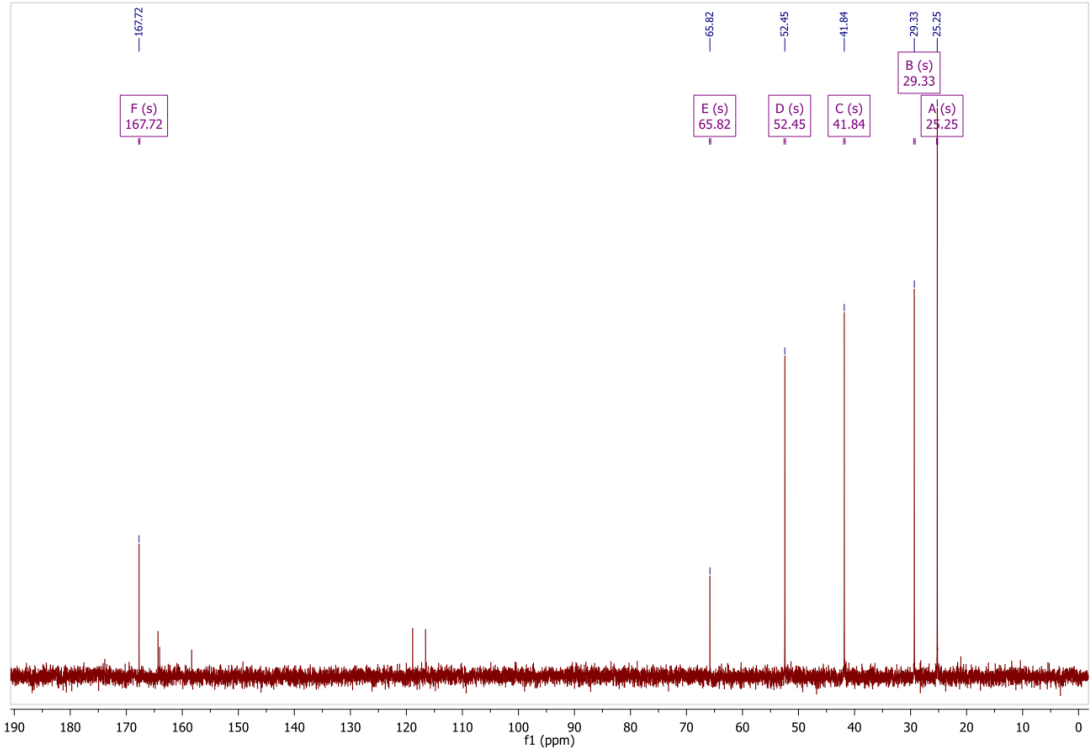
^{1}H NMR spectrum of LysHA (D_2O + 1.8% TFA, 300 MHz).



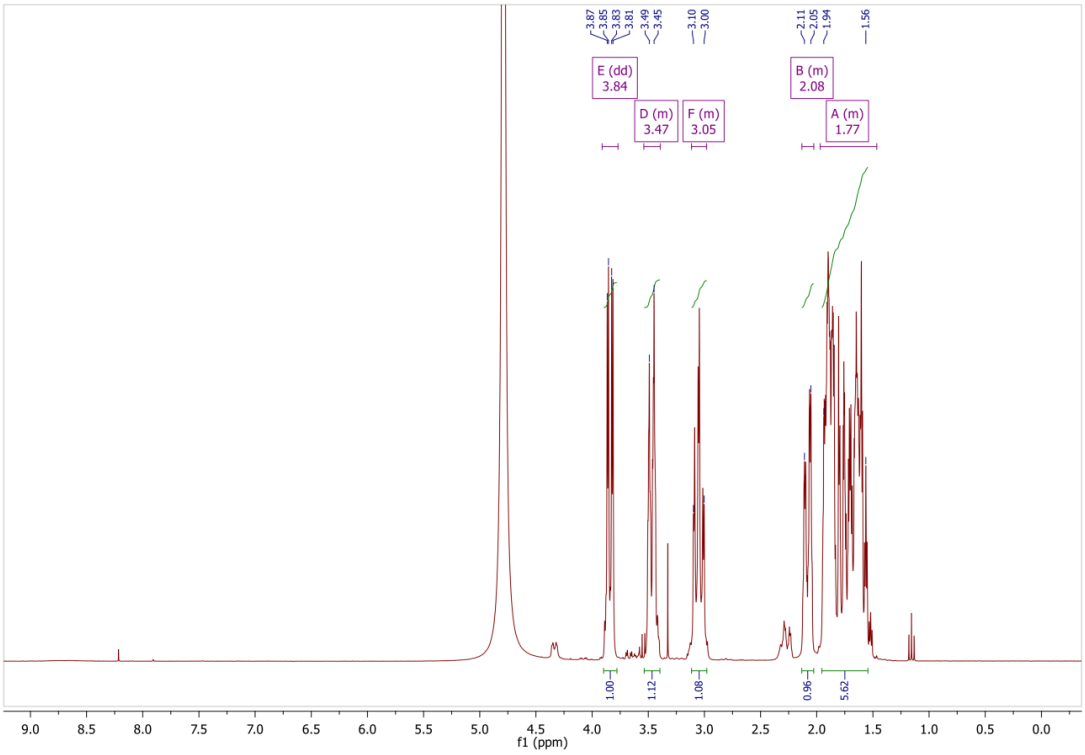
^{13}C NMR spectrum of LysHA (D_2O + 1.8% TFA, 75 MHz).



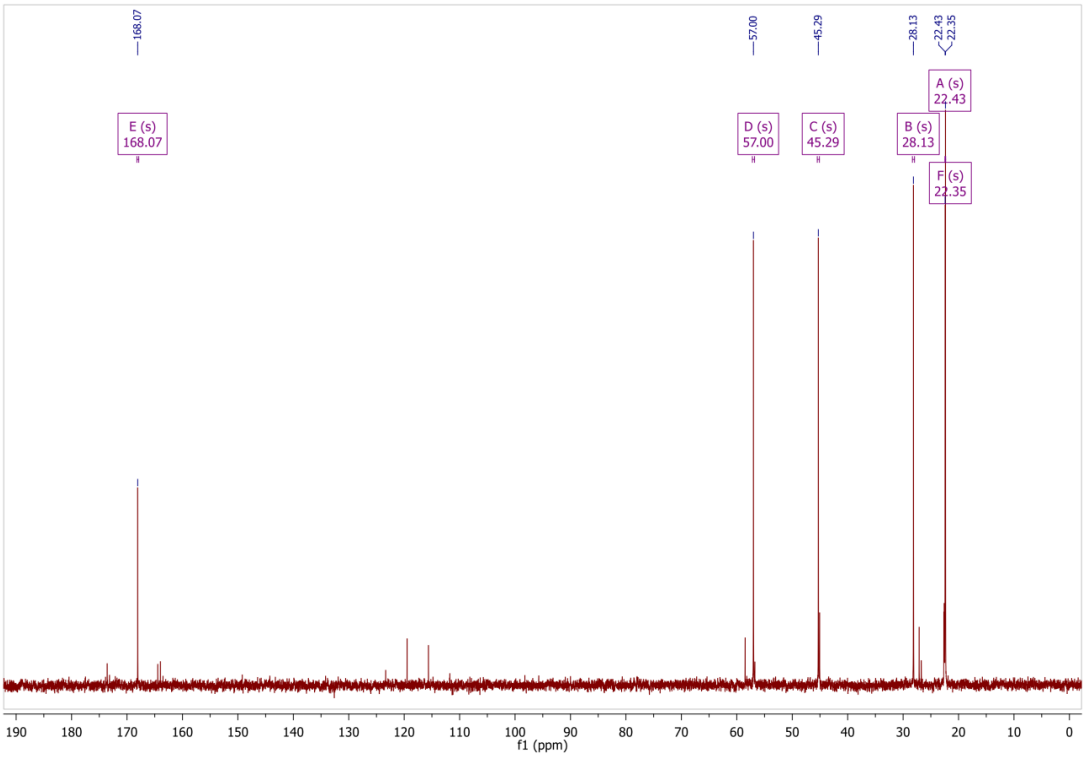
^{1}H NMR spectrum of ArgHA (D_2O + 1.8% TFA, 500 MHz).



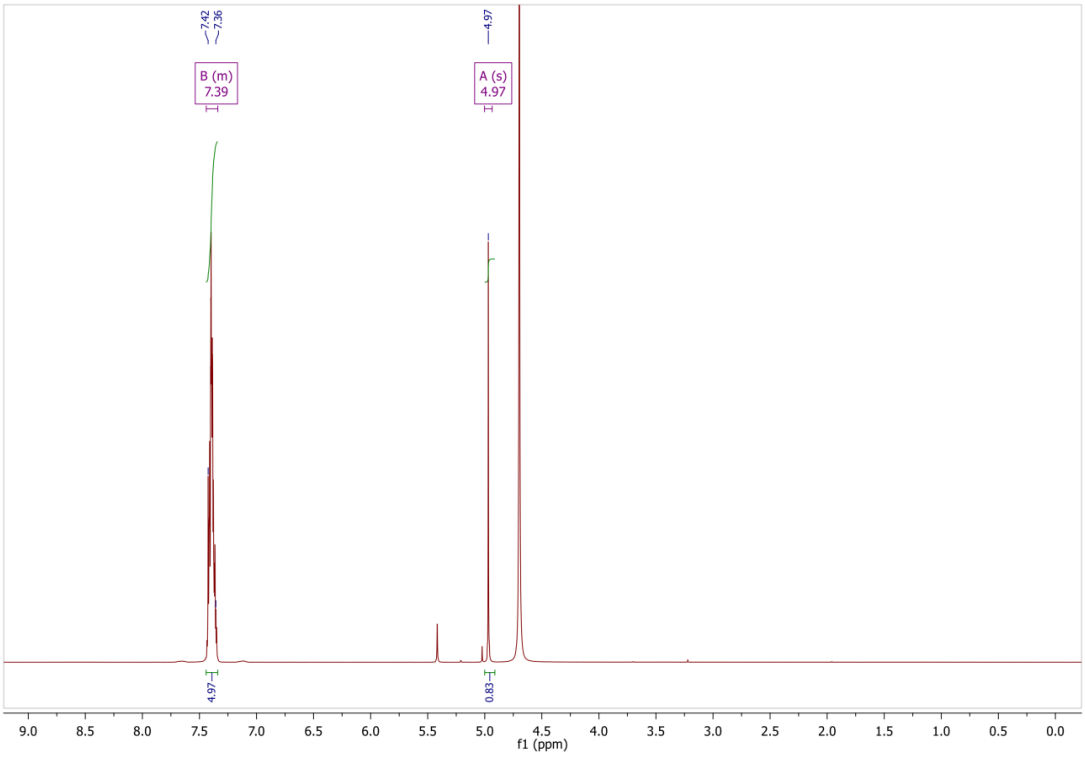
^{13}C NMR spectrum of ArgHA (D_2O + 1.8% TFA, 126 MHz).



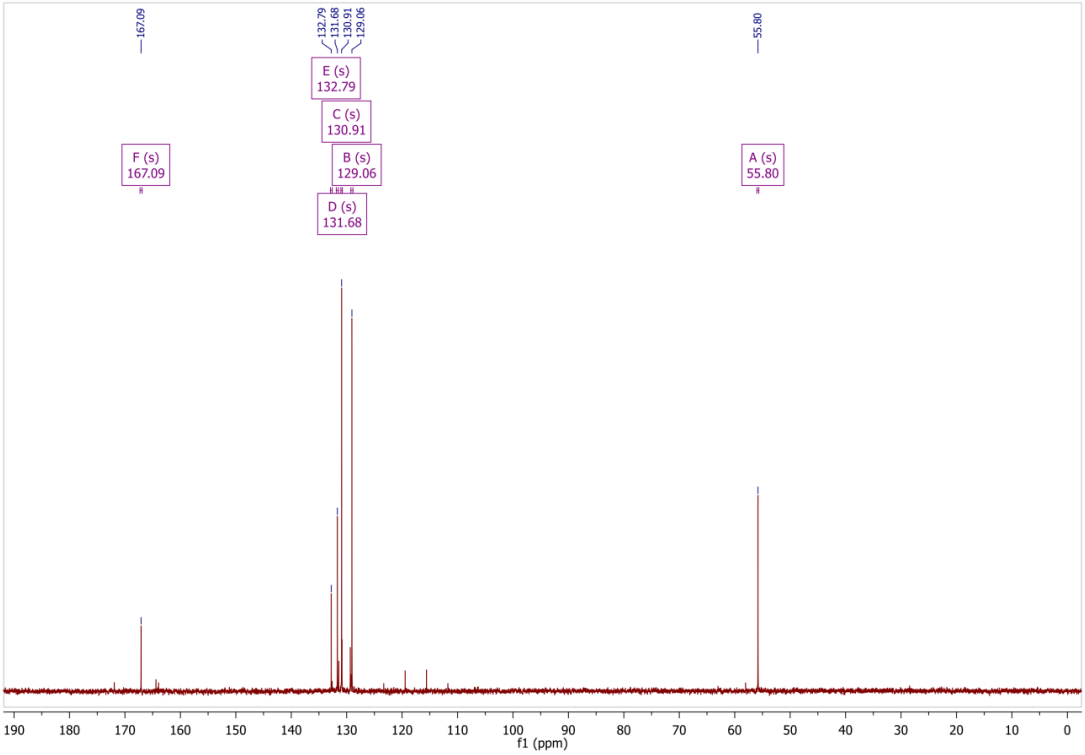
^{1}H NMR spectrum of pipecolic acid HA (D_2O + 1.8% TFA, 300 MHz).



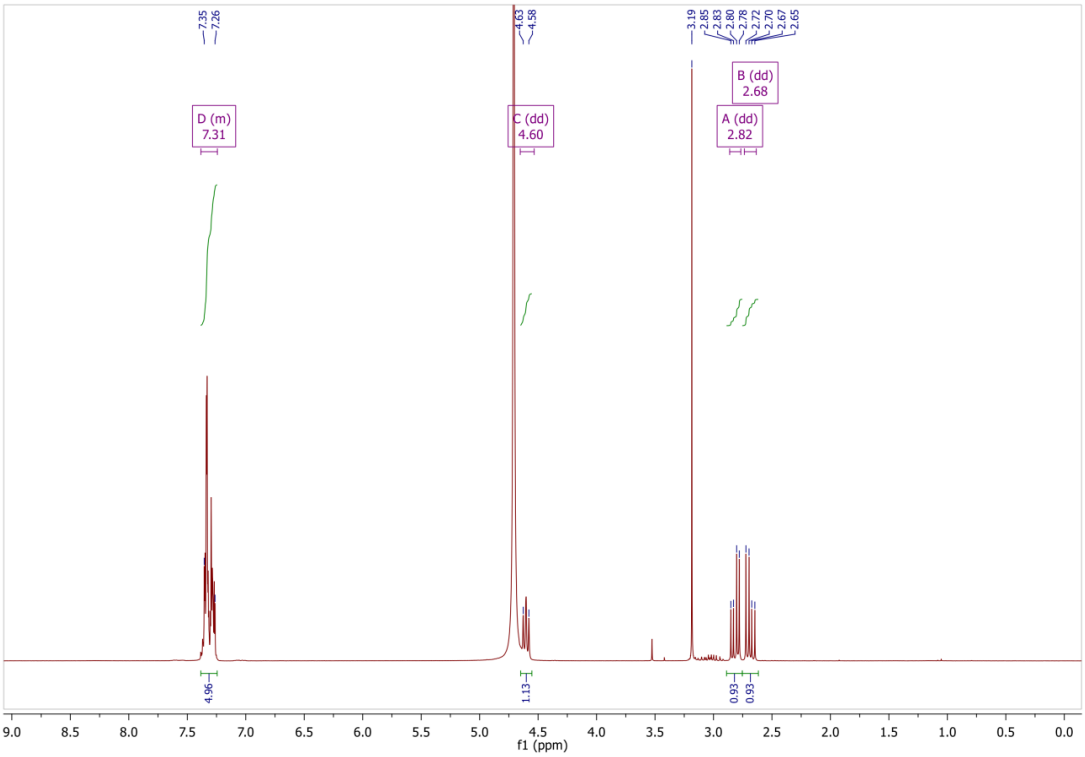
^{13}C NMR spectrum of pipecolic acid HA (D_2O + 1.8% TFA, 75 MHz).



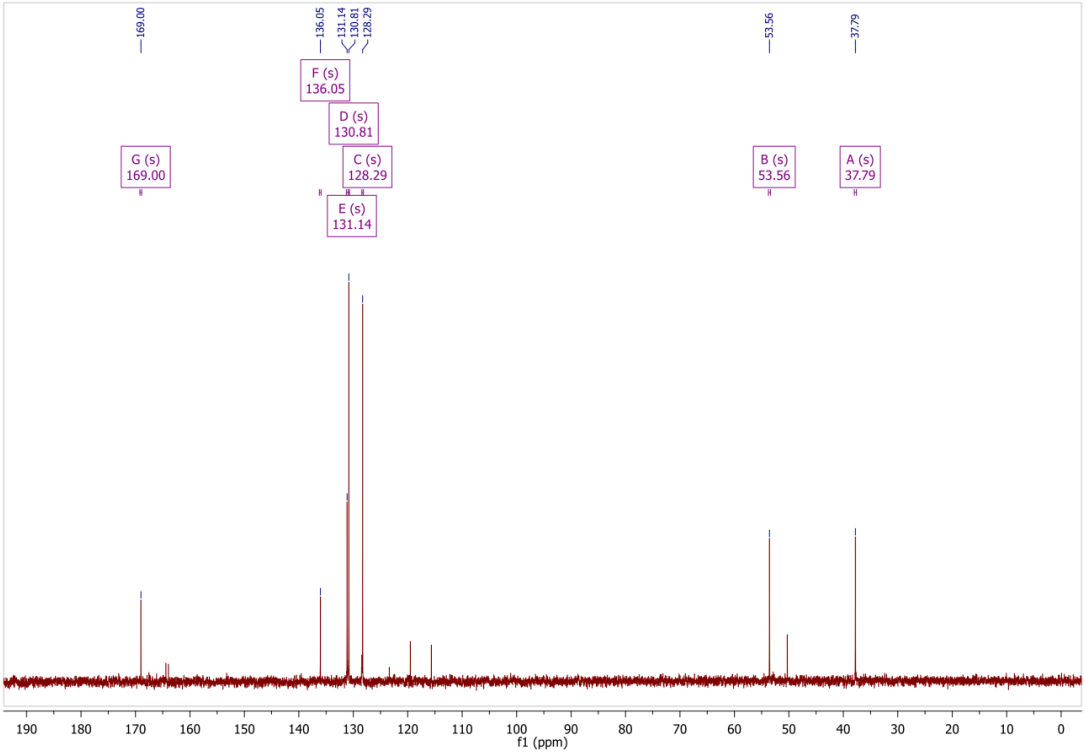
^{1}H NMR spectrum of phenylglycine HA (D_2O + 1.8% TFA, 300 MHz).



^{13}C NMR spectrum of phenylglycine HA (D_2O + 1.8% TFA, 75 MHz).



^{1}H NMR spectrum of β-phenylalanine HA (D_2O + 1.8% TFA, 300 MHz).



^{13}C NMR spectrum of β-phenylalanine HA (D_2O + 1.8% TFA, 75 MHz).

Supplementary References

(1) Cunningham, K. G.; Newbold, G. T.; Spring, F. S.; Stark, J. 449. Hydroxamic Acids. Part II. The Synthesis and Structure of Cyclic Hydroxamic Acids from Pyridine and Quinoline. *J. Chem. Soc.* **1949**, 2091.

(2) Jankolovits, J.; Lim, C. S.; Mezei, G.; Kampf, J. W.; Pecoraro, V. L. Influencing the Size and Anion Selectivity of Dimeric Ln 3+[15-Metallacrown-5] Compartments through Systematic Variation of the Host Side Chains and Central Metal. *Inorg. Chem.* **2012**, *51* (8), 4527–4538.

(3) Tegoni, M.; Furlotti, M.; Tropiano, M.; Lim, C. S.; Pecoraro, V. L. Thermodynamics of Core Metal Replacement and Self-Assembly of Ca 2+ 15-Metallacrown-5. *Inorg. Chem.* **2010**, *49* (11), 5190–5201.

(4) Brown, D. A.; Roche, A. L. Design of Metal Chelates with Biological Activity. 3. Nickel(II) Complexes of Alkyl and Amino Hydroxamic Acids. *Inorg. Chem.* **1983**, *22* (15), 2199–2202.

(5) Zur, A. A.; Chien, H. C.; Augustyn, E.; Flint, A.; Heeren, N.; Finke, K.; Hernandez, C.; Hansen, L.; Miller, S.; Lin, L.; et al. LAT1 Activity of Carboxylic Acid Bioisosteres: Evaluation of Hydroxamic Acids as Substrates. *Bioorganic Med. Chem. Lett.* **2016**, *26* (20), 5000–5006.

(6) Hogg, J. H.; Ollmann, I. R.; Haeggström, J. Z.; Wetterholm, A.; Samuelsson, B.; Wong, C.-H. Amino Hydroxamic Acids as Potent Inhibitors of Leukotriene A4 Hydrolase. *Bioorg. Med. Chem.* **1995**, *3* (10), 1405–1415.

(7) Smissman, E. E.; Warner, V. D. A-Aminohydroxamic Acids as Inhibitors of Histidine Decarboxylase and 3,4-Dihydroxyphenylalanine Decarboxylase. *J. Med. Chem.* **1972**, *15* (6), 681–682.

(8) Bisswanger, H. *Enzyme Kinetics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2008.

(9) 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.* **2004**, *70* (6), 3282–3291.

(10) Pospiech, A.; Neumann, B. A Versatile Quick-Prep of Genomic DNA from Gram-Positive Bacteria. *Trends Genet.* **1995**, *11* (6), 217–218.

(11) Berrow, N. S.; Alderton, D.; Sainsbury, S.; Nettleship, J.; Assenberg, R.; Rahman, N.; Stuart, D. I.; Owens, R. J. A Versatile Ligation-Independent Cloning Method Suitable for High-Throughput Expression Screening Applications. *Nucleic Acids Res.* **2007**, *35* (6).

(12) Kries, H.; Niquille, D. L.; Hilvert, D. A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides. *Chem. Biol.* **2015**, *22* (5), 640–648.

(13) Kries, H.; Wachtel, R.; Pabst, A.; Wanner, B.; Niquille, D.; Hilvert, D. Reprogramming Nonribosomal Peptide Synthetases for "Clickable" Amino Acids. *Angew Chem Int Ed Engl* **2014**, *53* (38), 10105–10108.

(14) Niquille, D. L.; Hansen, D. A.; Mori, T.; Fercher, D.; Kries, H.; Hilvert, D. Nonribosomal Biosynthesis of Backbone-Modified Peptides. *Nat. Chem.* **2018**, *10* (3), 282–287.

(15) Arp, J.; Götze, S.; Mukherji, R.; Mattern, D. J.; García-Altares, M.; Klapper, M.; Brock, D. A.; Brakhage, A. A.; Strassmann, J. E.; Queller, D. C.; et al. Synergistic Activity of Cosecreted Natural Products from Amoebae-Associated Bacteria. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (15), 3758–3763.

(16) Wilson, D. J.; Aldrich, C. C. A Continuous Kinetic Assay for Adenylation Enzyme Activity and Inhibition. *Anal. Biochem.* **2010**, *404* (1), 56–63.

(17) R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria 2017.

.

4 MANUSCRIPT III

Exploring the Functional Sequence Space of Adenylation in SrfA-C by Hydroxamate Profiling

Aleksa Stanišić, Ulrich Ettelt, Carl-Magnus Svensson, Marc Thilo Figge and Hajo Kries

Summary:

Specificity code of the adenylation domain was one of the landmark discoveries in NRPS enzymology enabling the prediction of substrate specificity from the protein sequence. However, attempts to rationally manipulate specificity-conferring residues typically result in conservative specificity switches. Here, we employ HAMA to determine the relative contribution of binding pocket residues to substrate specificity of A-domain of termination module from surfactin synthetase. We develop a promiscuous version of SrfAC and demonstrate the functional flexibility of adenylation reaction by fully randomizing 15 residues in the active site.

The candidate is

☒ First author ☐ Second author ☐ Corresponding author ☐ Coauthor

Status: In preparation for submission to *ACS Catalysis.*

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
AS	80 %	80 %	70 %	80 %	
UE			30 %	5 %	
CS		20 %		5 %	
MF				5 %	
HK	20 %			5 %	

Exploring the Functional Sequence Space of SrfAC Adenylation by Hydroxamate Profiling

Aleksa Stanišić, Ulrich Ettelt, Carl-Magnus Svensson, Marc Thilo Figge, Hajo Kries*

*hajo.kries@leibniz-hki.de
Independent Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans Knöll Institute (HKI Jena), Beutenbergstr. 11a, 07745 Jena, Germany

Abstract

Enzyme promiscuity is a key characteristic required for evolutionary innovation since activation of noncognate substrates can serve as a springboard toward novel activities. We aim to recapitulate this process in the laboratory to generate custom-made catalysts by directed evolution. Ideal candidate enzymes for studying enzyme promiscuity should show broad functional diversity and have a large pool of chemically accessible, potential substrates. Both criteria are fulfilled by nonribosomal adenylation (A-) domains. Nonribosomal peptide synthetases, with their enormous repertoire of building blocks, have great potential to be repurposed as a source of tailored peptides. In nonribosomal peptide synthesis, the A-domain is a main checkpoint for the identity of the incorporated substrate. The discovery of the 'specificity code', a signature of 10 residues in the binding pocket, enabled the prediction of A-domain substrate specificity from sequence data. However, rational modifications of the specificity code achieve mostly conservative changes accompanied by losses of catalytic activity. To improve our understanding of A-domain substrate selection, we take advantage of the hydroxamate specificity assay (HAMA) to determine the relative contribution of specificity code and second shell residues to adenylation promiscuity. First, we use the FuncLib algorithm to develop VSA, a promiscuous, ancestor-like A-domain variant of SrfAC, the termination module of surfactin synthetase. Second, we fully randomize the binding pocket residues of VSA and determine specificity profiles for each single mutant. Thereby, we determine the weight of each position, identify invariable residues, and show a high malleability of adenylation specificity at high catalytic rates. Additionally, we demonstrate that both promiscuity- and specificity-promoting mutations occur at only a few selected specificity code positions. Together, our data reveal the unexplored functional flexibility of A-domains, provide insights for more streamlined A-domain engineering and confirm the evolutionary potential of promiscuous enzymes.

Introduction

A staggering diversity of reactions catalysed by enzymes is a consequence of millions of years of evolution.[1,2] This process typically starts with an initial duplication event lifting one paralog of evolutionary pressure. The newly generated copy is then exposed to neofunctionalization by genetic drift through accumulation of beneficial or neutral mutations.[3–5] Promiscuous activities are an essential component of this process, serving as an evolutionary springboard towards novel functions.[6,7] As the selection progresses, wild type activity is typically retained[8] while the side activities increase, resulting in generalist enzymes. Generalists can become specialized again if the pressure is strong enough to favour the gain of specificity.[9,10]

This process can be hijacked in the laboratory to design customized enzymes with broad applications through directed evolution.[11–14] In directed evolution experiments, libraries of mutants are iteratively screened for improvements towards a desired property. The mutational landscape of an average-sized protein is enormous, far surpassing the capacity of any screening method. Therefore, a key challenge of directed evolution is to design a library of sufficient size which would allow exhaustive sampling of mutants likely to show the desired property. By mimicking natural processes, specialized enzymes can be reversed to a promiscuous, ancestor-like state by amplifying weak activities towards noncognate substrates.[7] These generalist enzymes can serve as a starting point for re-specialization for a new function. Not all enzymes are equally evolvable, however. It has been shown that enzymes showing high natural functional diversity are more amenable to change than those fulfilling identical roles across the homology tree.[15] Secondary metabolism is especially enriched with promiscuous activities[16,17] resulting in diverse mixtures of natural product congeners.[18–20] This makes enzymes from secondary metabolism especially suitable for studying promiscuity and evolutionary pathways between different activities.

Nonribosomal peptides (NRPs) are one of the most widespread natural product class and of great importance for human use as antibiotics, immunosuppressants and anti-cancer drugs.[21] These predominantly cyclic peptides are assembled on large multidomain enzymes termed nonribosomal peptide synthetases (NRPSs). NRPSs consist of domains catalysing individual reactions which are grouped in modules where each incorporates a single substrate into the peptide chain in an assembly line fashion.[22] Substrates are first activated with ATP·Mg by adenylation (A-) domains before being tethered to thiolation (T-) domains and condensed with the substrate from the adjacent module by condensation (C-) domains. The release of the final product is typically achieved by a terminal thioesterase (Te-) domain catalysing hydrolysis or intramolecular cyclization of mature linear peptide. The large variety of NRPS architectures and corresponding NRP products

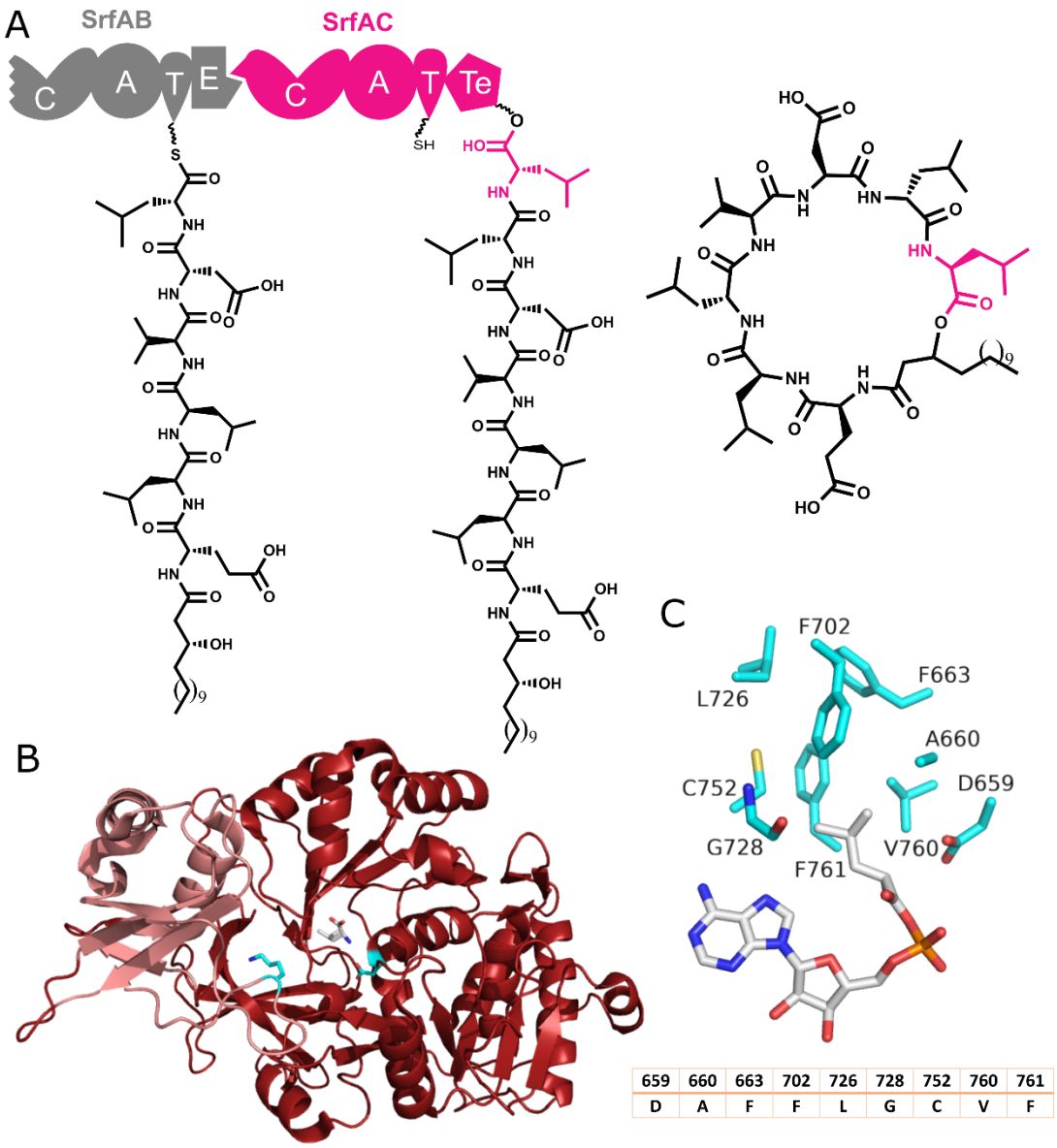


Figure 1. A) Incorporation of a terminal leucine into surfactin. B) Crystal structure of the A-domain of SrfAC (PDB:2vsq) with L-Leu substrate (grey) and invariable D659 and K948 (cyan). N- and C-terminal subunits are coloured in dark and light red, respectively. C) Binding pocket of SrfAC Yasara homology model in complex with L-Leu-AMP with labelled specificity code residues.

must result from fast evolutionary diversification compared to enzymes involved in primary metabolism.[23]

The modular nature of NRPSs makes them an attractive engineering target for sourcing custom-made peptides.[24] A-domains are main specificity gatekeepers controlling the identity of activated and incorporated substrates. They are able to activate more than 500 different monomers[25,26] and exhibit a range of selectivity: from highly specific[27] to bispecific[28,29] and promiscuous[19,30] activities. In depth studies by structural and sequence analysis revealed 'specificity code' residues in the binding pocket that are highly conserved between A-domains activating the same substrate.[31,32] The initial 8-residue code, later amended by 2nd and 3rd shell residues, allowed the development of

algorithms able to predict the identity of the final products from NRPS protein sequence.[33–37] However, the reliability of prediction algorithms depends on the phylogenetic distance of investigated A-domain from A-domains in a reference dataset. Recently, a comprehensive ensemble method integrating existing algorithms and phylogenetic information has been developed to increase prediction accuracy.[38]

First attempts of NRPS reprogramming have used mutagenesis of the A-domain specificity code. However, successful changes were limited to structurally similar substrates indicating that specificity signatures are not readily transferable between A-domains[31,39,40] with attempts at less conservative changes resulting in large losses of catalytic efficiency.[41,42] Nevertheless, the flexibility of A-domain specificity has been demonstrated on Phe-specific GrsA which acquired a $5x10^5$ fold switch in specificity towards "click" amino acid propargyl-Tyr by introducing a single mutation in the binding pocket.[43] Directed evolution has been utilized to bypass limitations imposed by rational A-domain design.[42,44,45] Niquille et al. combined binding pocket randomization with yeast surface display to activate a β-amino acid substrate without losses in catalytic efficiency.[46] While shuffling of whole domains and modules also bears significant promise, more general and reliable strategies for changing A-domain specificity can grant access to non-natural substrates.[47–51]

NRPS engineering is typically focused on developing activity towards one or few products, while the underlying general factors governing substrate selection are unknown. The question remains how natural mutational pathways lead from one substrate to another and how they can be reiterated by NRPS design. Villiers et al. have conducted a thorough study of A-domain promiscuity by measuring complete saturation kinetics of initiation modules of tyrocidine synthetase TycA with a range of natural and synthetic substrates.[27] They revealed a high specificity of TycA for L-Phe with the second preferred natural substrate L-Tyr showing three orders of magnitude lower catalytic efficiency. To explore the functional sequence space of the specificity code, Throckmorton et al. have targeted the binding pocket residues of EntF module from enterobactin synthetase revealing a broad tolerance towards specificity code diversification with a number of unnatural signatures maintaining high functionality.[52] This illustrates a large functional potential of the A-domain and a gap in our understanding of A-domain specificity resulting in poor engineering outcomes.

For A-domain engineering, it is essential to understand the relationship between binding pocket residues and specificity profiles. However, straightforward adenylation specificity assays have been lacking.[53] In previous work, we have developed HAMA which enables us to determine a complete specificity profile of an A-domain in a single reaction, dramatically reducing the workload and facilitating the determination of A-domain specificity.[54] Here, we take advantage of HAMA to investigate the impact of mutations on the specificity landscape of the A-domain from SrfAC, the termination

module from surfactin synthetase.[20] SrfAC is a standalone module with CATTe architecture incorporating the terminal L-Leu into surfactin.[55,20] In nature, surfactins are produced as a mixture of structurally related analogs, a feature utilized for generating modified surfactins by mutasynthesis and rational engineering.[20] This tolerance to substrate selection makes this system ideal for studying enzyme promiscuity. Additionally, the structure of the whole module has been resolved[56] and the protein shows stable expression in *E. coli*. (Figure 1B, Supplementary Figure 1) We reasoned that side activities of an already promiscuous enzyme will be more easily amplified in a subsequent evolutionary step. With the aid of the FuncLib[57] automated algorithm, we have developed VSA - a multispecific, progenitor-like variant of SrfAC with preserved catalytic efficiency and improved stability. In the second step, we selected 15 positions in the binding pocket of VSA for individual, full randomization. Resulting single mutant libraries were screened by determining a complete HAMA specificity profile for each variant. We demonstrate remarkable flexibility of adenylation towards nonpolar substrates in a single evolutionary step and quantify the relative tolerance of each residue to mutation. Additionally, we identify variants with increased promiscuity, tolerance towards D-amino acids, aromatic amino acids, and gain of novel specificities.

Results

Development of a generalist SrfAC

SrfAC incorporates terminal L-leucine with high specificity (Figure 2B). Considering the hypothesis that promiscuous activities are main evolutionary drivers towards novel substrates, our first aim was to develop an ancestor-like version of SrfAC with relaxed specificity. To facilitate this process, we took advantage of Funclib, an automated algorithm using phylogenetic analysis and Rosetta modelling to predict the tolerance of active site residues to mutation.[57] FuncLib draws on homology data and filters out mutations likely to result in inactive variants or introduce clashes in the binding pocket. The output of FuncLib is a selection of active site signatures likely to maintain activity and protein integrity, thus reducing the number of multipoint variants to be screened. First, a model of SrfAC in complex with Leu-AMP was built using the YASARA molecular-graphics and modeling software (Supplementary Figure 2).[58] Second, the Leu-AMP ligand was included in the protein model while the invariable **D659** was fixed by excluding it from FuncLib calculations. We selected the remaining eight specificity code

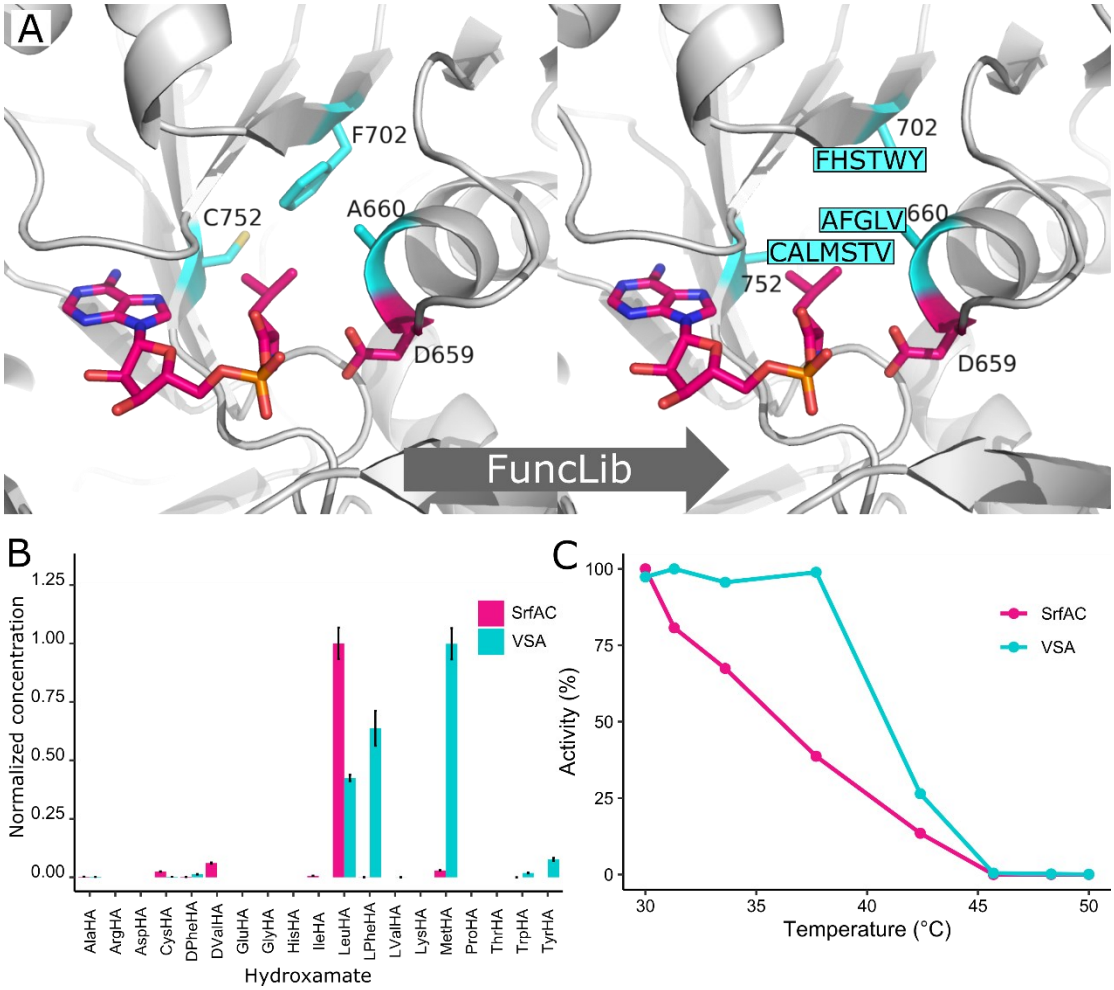


Figure 2. A) Three specificity code residues of SrfAC Yasara model selected for randomization (cyan) and the corresponding list of tolerated residues predicted by FuncLib. Ligand L-Leu-AMP and invariable D659 are fixed (pink). B) HAMA specificity profiles of SrfAC and VSA mutant. C) Thermostability of SrfAC and VSA. Enzymes were incubated for 1 hour at constant temperature and the production of LeuHA is subsequently followed for 1.5 hour at room temperature and 1 µM enzyme. Error bars are standard deviations from three (B) or two (C, too small to be visible) technical replicates.

residues for *in silico* randomization: **A660**, **F663**, **F702**, **L726**, **G728**, **C752**, **V760** and **F761**. From the list of allowed residues at each position (Supplementary Information 3), we selected 3 for experimental randomization and screening. Being located at the entrance (**A660** and **C752**) and the bottom (**F702**) of the binding pocket, we reasoned that they are likely to have a decisive impact on the shape and bulkiness of the side chain of activated amino acid.

SrfAC shows low, but detectable level of promiscuity with side activities 30-fold (Met) or 2000-fold (L-Phe) lower than the wild type substrate L-Leu (Figure 2B). As expected for an A-domain activating a nonpolar amino acid, FuncLib predicted tolerance towards substitution of residues with predominantly aliphatic side chains (Figure 2A). We used the predicted residue tolerance to generate a focused library of triple mutants by partially randomizing positions **660**, **702** and **752** with 5, 6 and 7 residues, respectively (Supplementary Information). This drastically reduces the combinatorial space from 8000 combinations, if three positions were fully randomized, to only 210 variants selected by

FuncLib. The reduced library of mutants was cloned by combining oligonucleotides bearing degenerate codons for each position in appropriate ratios (Supplementary Table 1 and 2). To determine the effect of mutations, the library was expressed in microtiter plates and specificity profiles with 18 proteinogenic and 2 nonproteinogenic substrates were measured with HAMA. To quantify the change in specificity profiles, library members were ranked according to the entropy-based index *I* developed by Nath et al. as a measure of enzyme promiscuity (Supplementary Table 6).[59] The strength of the FuncLib prediction is demonstrated by 46% of library members having detectable activity, remarkably high considering significant losses in activity typically accompanying multisite mutagenesis. Three candidates with highest activity and promiscuity, ASV, ASA and VSA (single letter codes in order of residue sequence: 660-702-752; wild type code: AFC), were selected for further characterization. Remarkably, all three mutants have an expanded substrate scope activating L-Phe and L-Met at wild type-like rates and a range of side activities (Figure 2B, Supplementary Figure 3). They share a Phe to Ser mutation at position **702** at the bottom of the binding pocket coupled with small (Ala) to medium (Val) sized residues at the entrance (**660**, **752**), presumably freeing space for bulkier, hydrophobic amino acids such as Phe and Met.

Mutations can put a significant burden on the structural integrity of an enzyme.[60–62] In addition to broad substrate tolerance, an ancestor-like enzymes must be stable enough to withstand further mutations. To test the influence of mutations on protein stability, the adenylation activity of mutants was followed after incubation at a range of temperatures between 30 and 50 ºC (Supplementary Figure 4). While VSA and ASV maintain activity up to 38 ºC, ASA suffers activity loss at temperatures above 33 ºC. Interestingly, VSA is stabilized even in comparison with the parent SrfAC, which suffers more than 50 % activity loss at 37 ºC (Figure 2C). To characterize the effect of the VSA mutations, saturation kinetics with the three major substrates (L-Leu, L-Phe and L-Met) were measured with the MesG/hydroxylamine assay (Supplementary Figure 5). The adenylation rate k_{cat} for all three substrates is maintained at wild type levels with differences originating in K_M values. $K_{M(Leu)}$ shows a 50-fold increase from 10 µM in SrfAC to 500 µM in VSA while $K_{M(Phe)}$ and $K_{M(Met)}$ of VSA are within 2 and 10-fold of $K_{M(Leu)}$. Consequently, specificity constants k_{cat}/K_M of all three substrates fall within one order of magnitude. Combining high stability and an expanded substrate repertoire at wild-type rates, VSA is ideally suited for further functional diversification.

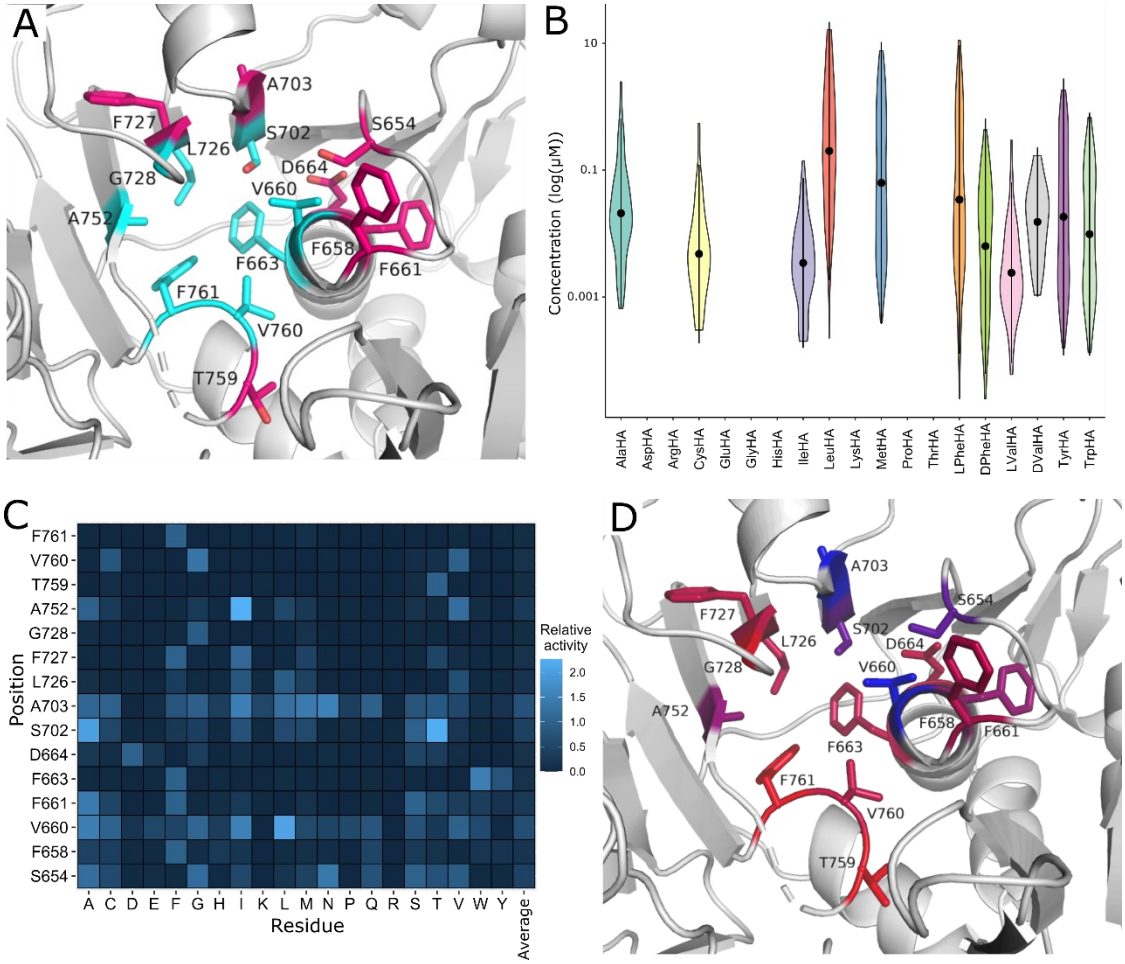


Figure 3. A) Residues in the binding pocket of VSA model selected for saturation mutagenesis. Specificity code resides are cyan and second shell pink. VSA structure is SWISS homology model built against SrfAC (PDB: 2vsq) as a template. B) Logarithmic distribution of concentration of detected hydroxamates pooled from 15 NNK libraries. C) Heatmap of activities of all mutants relative to the progenitor VSA. Activity is calculated as a sum of all formed hydroxamates per mutant. Last column represents the average activity per position. D) Binding pocket of VSA homology model with targeted residues coloured according to the average activity per position, relative to the progenitor VSA. Mutations at blue positions result in highest activities and at red positions, lowest.

Functional sequence space of VSA

We proceeded to thoroughly probe the effects of single point mutations on the specificity landscape of VSA. We aimed to exhaustively cover the binding pocket by generating site-saturation mutagenesis libraries for 15 positions (Figure 3A). In addition to 8 specificity code residues, 7 second shell residues were included to learn more about the weight of each residue for substrate selection which is a crucial information for computational specificity prediction.[33] To ensure 90% coverage of each NNK library, we screened 92 colonies per library with HAMA in microtiter plates. Mutants missing from the random libraries were filled in by cloning them individually.

Activity was detected in 50 % (147/300) of mutants from all libraries. The total scope of detected substrates encompasses a broad range of aliphatic amino acids including the D-amino acids distributed over a wide concentration range (Figure 3B). To test the tolerance of 15 targeted positions to mutation, the activity of each mutant was compared

to the progenitor VSA (Figure 3C). 4 positions in the specificity code (**F663, G728, V760, F761**) and 3 in the second shell (D664, T759, F727) suffer large activity losses upon mutagenesis (specificity code residues in bold font, second shell residues in normal font). This observation is surprising since **F663** and **F761** are considered 'wobble-like' positions with high variability regarding residue usage.[31] In contrast, positions S654, **A660** and A703 are exceptionally permissive to mutations, presumably due to their location on the surface of the enzyme at the entrance to the binding pocket (Figure 3D). The effect of the nature of amino acid substitution is evident in overall low activity of mutants bearing polar (D, E, H, K, R), rigid (P) as well as bulky (W, Y) residues presumably incompatible with the nonpolar VSA pocket.

The analysis of promiscuity index *I* reveals a broad distribution of active mutants showing relaxed as well as constrained substrate tolerance, with 91 mutants (30 %) showing significantly different specificity than the progenitor VSA ($p < 0.05$, FDR 5%) (Figure 4). The majority of mutants with low promiscuity have returned to the L-Leu-specificity of wild type SrfAC but at a lower activity level. Interestingly, these specificity restorations do not rely on reversion mutations. In general, mutations at second shell residues show strong effects on activity but only a marginal effect on promiscuity, consistent with a purely structural role. Small and medium sized residues are well tolerated at positions S654, F658, F661, D664, A703 and F727 and have little effect on substrate selection. The main differences in specificity are observed in four specificity code residues (Figure 5A). **V660** shows exceptionally high influence on specificity profiles, with bulky residues (F,W) increasing L-Leu specificity, small residues (A,S,C,V) increasing the fraction of aromatic substrates and medium sized residues (L, I, M) shifting the specificity towards L-Met (Figures 4, 5B and 5C). Aromatic substrates are also favoured by Ala at positions **V660** and **S702,** while **S702F** mutation results in a remarkably promiscuous mutant activating 7 different substrates (Figure 4 and 5C). Additionally, the activation of D-configured substrates is favoured by Gly substitution at **V760**. **G728** is invariable, except for small to medium size residues (A, M, L) which confer Ala specificity.

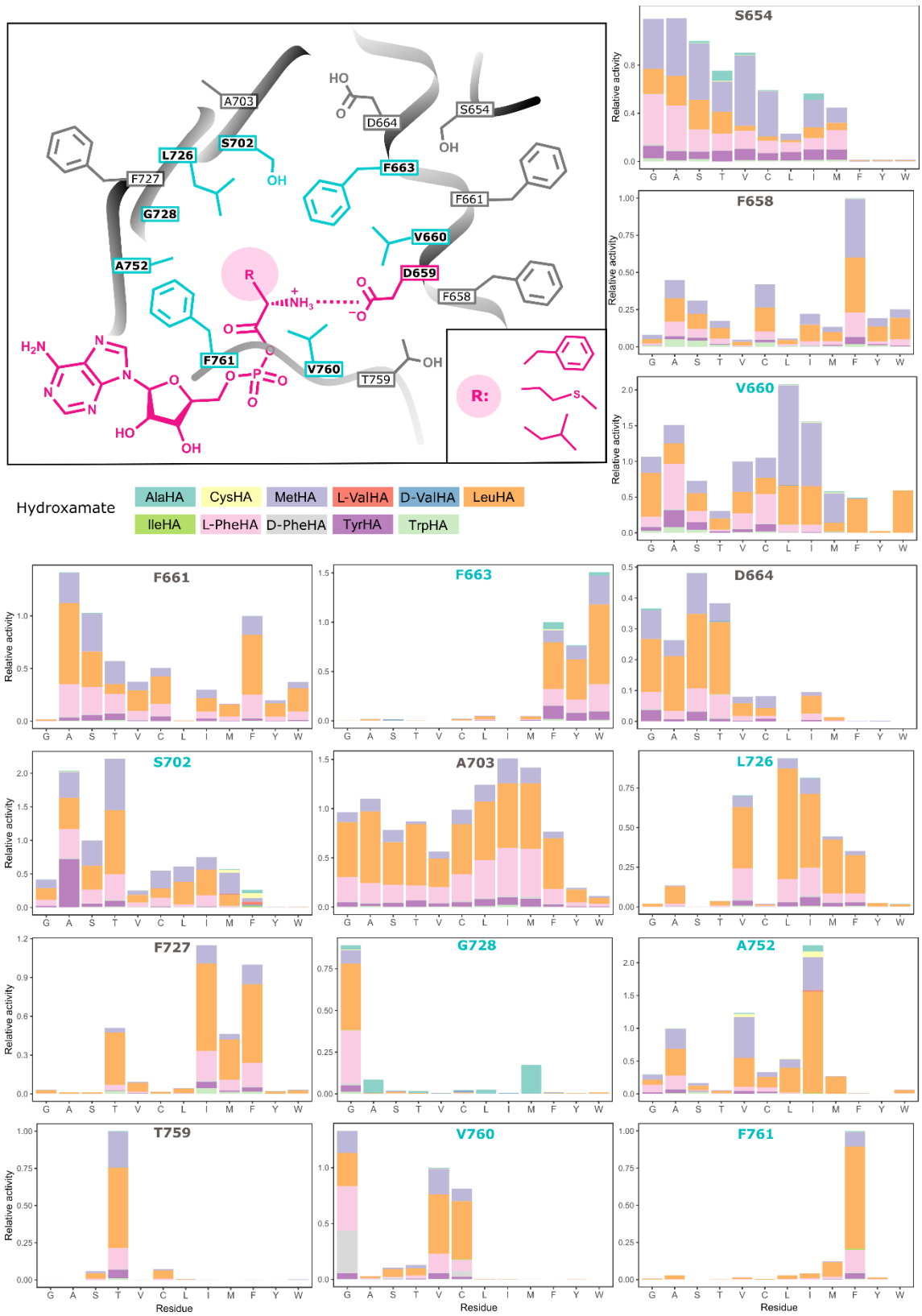


Figure 4. Scheme of the VSA binding pocket and HAMA specificity profiles of VSA variants obtained from screening of NNK libraries. Fractions of individual hydroxamates are plotted against the total activity relative to the progenitor VSA.

Discussion

The discovery of the A-domain specificity code has been one of the most significant breakthroughs in NRPS enzymology.[31,32] However, subsequent attempts to repurpose A-domains to produce tailored peptides were plagued with losses of activity. While specificity signatures for individual substrates are well documented, the rules and mechanisms governing evolution of substrate selection have remained elusive. One main culprit for this lack of understanding has been the lack of an adequate specificity assay. With the development of HAMA profiling, this bottleneck has been cleared and a complete specificity profile under competition conditions is recorded in several minutes.[54] Here, we use HAMA adapted for microtiter plate screening to develop a progenitor-like A-domain and conduct an in depth investigation of its specificity landscape.

First, we have developed a progenitor-like version of SrfAC with the aid of FuncLib.[57] FuncLib does not aim for a specific function, but instead increases the likelihood to obtain enzymes which will retain any function while differing from the wild type as much as possible. Accordingly, 46 % of our FuncLib-inspired library of triple mutants showed activity, an extraordinarily large fraction for multisite mutants. The promiscuous SrfAC variant VSA preserves the turnover rate of the wild type and shows a remarkable 3000- (L-Phe) and 76 (L-Met)-fold increase of side activities according to HAMA. Interestingly, VSA also shows improved thermal stability with respect to wild type SrfAC. Proteins in nature typically attain only marginal stability necessary for functioning in environmental conditions.[60] Khersonsky et al. noted that FuncLib output can result in active-site designs with stabilities surpassing those of the wild type.[57] Additionally, protein stabilization is routinely achieved in directed evolution experiments.[63,64] Stability-conferring mutations are an essential attribute of a generalist enzyme which must be able to withstand the mutational pressure, as exemplified by resurrected ancestors[65,66] and the increased evolvability of stabilized variants[60,67]. Considering that, typically, stabilizing mutations disperse throughout the enzyme structure[68–70] and active site arrangements of enzymes are thermodynamically unfavourable, stabilizing mutations located in this region are expected to cause loss of activity.[71,72] Gain of function mutations typically destabilize the enzyme by increasing the conformational plasticity in the binding pocket.[73–76] In contrast, VSA shows both expanded substrate scope and increased thermostability. We speculate that the space enlarged with **F702S** and **C752A** mutations may allow the accommodation of bulky amino acid side chains while still preserving the hydrophobic attributes of the binding pocket and consequently the integrity of the enzyme.

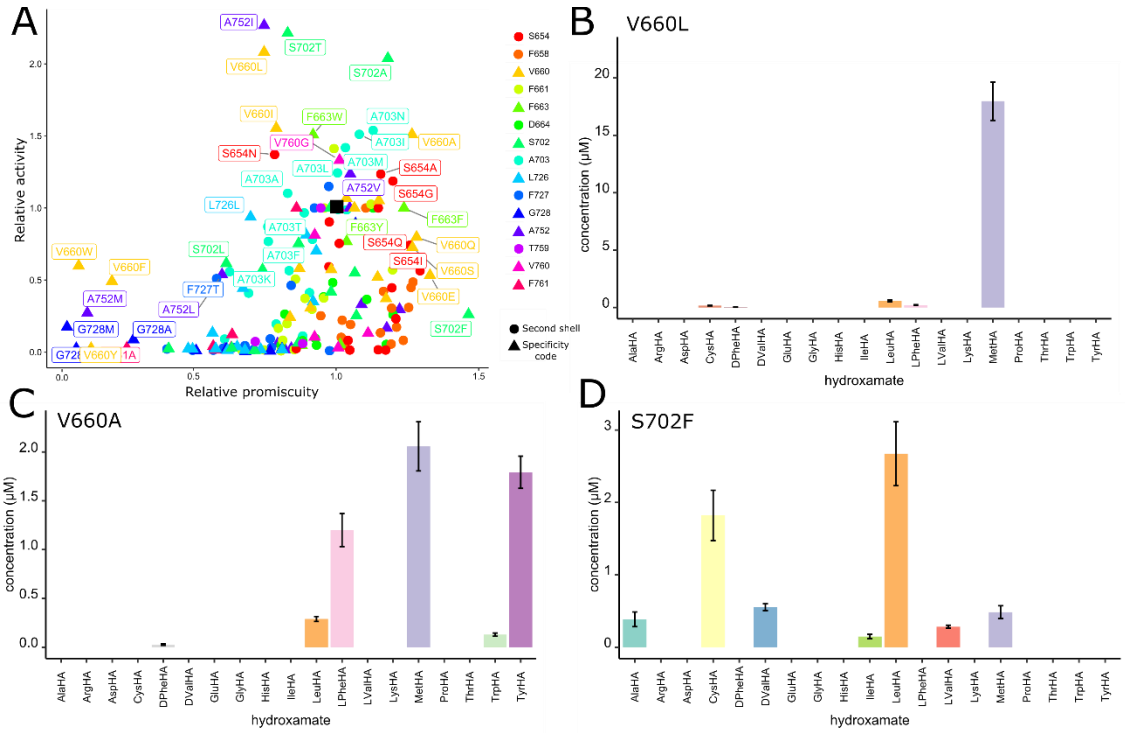


Figure 5. A) Distribution of individual mutants according to the promiscuity and activity, relative to the progenitor VSA (black square). *I* index is used as a measure of enzyme promiscuity. B-D) HAMA specificity profiles of selected VSA mutants. Error bars are standard deviations from three technical replicates at 1 µM enzyme and 1.5 h reaction time.

To completely scan the substrate binding pocket, 7 second shell and 8 specificity code residues were randomized to generate 15 site-saturation mutagenesis libraries and a complete HAMA specificity profile was determined for each. Half of all single mutants were active for at least one of the aliphatic amino acids tested. No hydroxamates of polar amino acids were detected, presumably due to the highly specialized binding pocket of SrfAC for nonpolar residues which is phylogenetically distant from A-domains specific for polar substrates. Considering that A-domain evolution is intimately tied to substrate preference[23] it is likely that more than a single mutation will be required to bridge this evolutionary gap. A combination of polar residues may be required to reduce the hydrophobicity enough for polar substrates to be accommodated even from a promiscuous starting point.

Second shell residues affect predominantly the overall activity, having a marginal effect on specificity suggesting a structural role and a minor influence on substrate selection. Surprisingly, 4 out of 8 specificity code residues (**F663**, **G728**, **V760**, **F761**) turned out to be almost invariable, tolerating only the substitution of structurally similar residues with no effect on the substrate selection. This is surprising considering the variability of specificity code described in phylogenetic analysis of GrsA and a diversity of functional specificity signatures of EntF.[31,52] This functional invariability may point to epistasis which arose early in the evolution of the A-domain and was necessary for the functionality of subsequent mutations.[77] Additionally, this may explain

some previous failures of A-domain engineering by specificity code mutagenesis which were not informed about the susceptibility of individual positions to mutations.

We show a wide distribution of active mutants based on their promiscuity by mutating only 4 specificity code residues. Out of the 5 most specific mutants, 4 regain specificity towards L-Leu without reverting the VSA mutations. However, their activities are reduced compared to the progenitor VSA, illustrating separate evolution of functions of activity and specificity. This is confirmed by a population of most active mutants having a similar specificity profile to VSA. Point mutations at permissive positions can completely change the specificity profile. The most compelling example is **V660** which results in increased selectivity towards L-Leu, aromatic substrates or L-Met depending on the size of the introduced aliphatic residue. Mutant **V660L** shows a complete specificity switch for L-Met. A contrasting case is the most promiscuous mutant **S702F** exhibiting a remarkably expanded specificity profile encompassing 7 different aliphatic substrates with only 4-fold loss of catalytic activity compared to VSA. Interestingly, Ser residue at this position is reverted to Phe, present in wild type SrfAC which leaves only two (**A660V**, **C752A**) promiscuity-enhancing mutations. Selective incorporation of D-amino acids by A-domains is uncommon, however here it is stimulated by the **V760G** mutation, in the loop close to the alpha carbon of the substrate (Figure 4), where impacts on amino acid backbone specificity have been observed before.[46]

Taken together, our results suggest that epistatic effects may not present a serious obstacle for the (directed) evolution of A-domains. Previously reported flexibility of the specificity code[52] and a wide distribution of specificity profiles in single mutants observed here, indicate that there are multiple routes to success. Several in depth studies of evolutionary trajectories of enzymes showed that, while epistasis is a ubiquitous phenomenon of protein fitness landscapes, a significant fraction of pathways consist of simple, incremental improvements.[78–81] Considering the flexibility of A-domain function in a minimal evolutionary step, we predict that extensive sampling of multipoint mutants will usually not be required to reach new activities. However, we cannot exclude that the observed flexibility of adenylation may be an isolated feature of A-domains activating nonpolar substrates. Sequence based prediction of specificity of A-domains activating aromatic and aliphatic substrates is less accurate compared to their counterparts specific for polar substrates.[33] Additional promiscuity studies of A-domains specific for polar substrates will be required to determine whether promiscuity observed here is a universal feature of A-domains.

Here, we utilize HAMA to conduct the most thorough investigation of A-domain specificity to date. We demonstrate the strength of FuncLib-aided screening to provide the ancestor-like SrfAC which is used as a progenitor for detailed investigation of the A-domain binding pocket. We confirm the decisive role of specificity code and show that point mutations at only a few positions can be sufficient to achieve large changes in

specificity, without severe impairment in activity. The obtained quantitative measures of enzyme promiscuity and activity reveal the relative importance of each residue and may be utilized as a training dataset for machine learning algorithms. These can be further employed to deconvolute evolutionary trajectories or generate a workflow containing subsets of preferred mutations for targeted activities. Our results underline the strength of HAMA specificity assay, establish the potential of low-throughput targeted mutagenesis for generating functionally variable A-domains, and confirm the large potential of promiscuous evolutionary intermediates.

References

1. Radzicka, A. & Wolfenden, R. A proficient enzyme. *Science* **267**, 90–93 (1995).
2. Albery, W. J. & Knowles, J. R. Efficiency and evolution of enzyme catalysis. *Angew. Chem. Int. Ed. Engl.* **16**, 285–293 (1977).
3. Copley, S. D. Evolution of new enzymes by gene duplication and divergence. *FEBS J.* **287**, 1262–1283 (2020).
4. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425 (1976).
5. Newton, M. S., Arcus, V. L., Gerth, M. L. & Patrick, W. M. Enzyme evolution: innovation is easy, optimization is complicated. *Curr. Opin. Struct. Biol.* **48**, 110–116 (2018).
6. Glasner, M. E., Truong, D. P. & Morse, B. C. How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. *FEBS J.* **287**, 1323–1342 (2020).
7. Khersonsky, O. & Tawfik, D. S. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471–505 (2010).
8. Guzmán, G. I. *et al.* Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.* **15**, e8462 (2019).
9. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proceedings. Biol. Sci.* **256**, 119–124 (1994).
10. Noda-Garcia, L. & Tawfik, D. S. Enzyme evolution in natural products biosynthesis: target- or diversity-oriented? *Curr. Opin. Chem. Biol.* **59**, 147–154 (2020).
11. Arnold, F. H. Directed evolution: Bringing new chemistry to life. *Angew. Chemie - Int. Ed.* **57**, 4143–4148 (2018).
12. Reetz, M. T. *Directed Evolution of Selective Enzymes*. (Wiley-VCH, Weinheim, 2016).
13. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).
14. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
15. O'Loughlin, T. L., Patrick, W. M. & Matsumura, I. Natural history as a predictor of protein evolvability. *Protein Eng. Des. Sel.* **19**, 439–442 (2006).
16. Fischbach, M. A. & Clardy, J. One pathway, many products. *Nat. Chem. Biol.* **3**, 353–355 (2007).
17. Bar-Even, A. & Salah Tawfik, D. Engineering specialized metabolic pathways--is there a room for enzyme improvements? *Curr. Opin. Biotechnol.* **24**, 310–319 (2013).

18. Lawen, A., Dittmann, J., Schmidt, B., Riesner, D. & Kleinkauf, H. Enzymatic biosynthesis of cyclosporin A and analogues. *Biochimie* **74**, 511–516 (1992).
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. Théatre, A. *et al.* The Surfactin-Like Lipopeptides From Bacillus spp.: Natural Biodiversity and Synthetic Biology for a Broader Application Range. *Front. Bioeng. Biotechnol.* **9**, (2021).
21. 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.
22. 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).
23. Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **10**, (2014).
24. 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).
25. Caboche, S., Leclère, V., Pupin, M., Kucherov, G. & Jacques, P. Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. *J. Bacteriol.* **192**, 5143–5150 (2010).
26. Walsh, C. T., O'Brien, R. V & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed. Engl.* **52**, 7098–124 (2013).
27. Villiers, B. R. M. & Hollfelder, F. Mapping the limits of substrate specificity of the adenylation domain of TycA. *Chembiochem* **10**, 671–682 (2009).
28. 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).
29. Kaljunen, H. *et al.* Structural elucidation of the bispecificity of A domains as a basis for activating non-natural amino acids. *Angew. Chemie Int. Ed.* **54**, 8833–8836 (2015).
30. Zhu, M., Wang, L. & He, J. Chemical Diversification Based on Substrate Promiscuity of a Standalone Adenylation Domain in a Reconstituted NRPS System. *ACS Chem. Biol.* **14**, 256–265 (2019).
31. 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).

32. 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).
33. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).
34. Röttig, M. *et al.* NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, 362–367 (2011).
35. Baranašić, D. *et al.* Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing. *J. Ind. Microbiol. Biotechnol.* **41**, 461–467 (2014).
36. Knudsen, M. *et al.* Computational discovery of specificity-conferring sites in non-ribosomal peptide synthetases. *Bioinformatics* **32**, 325–329 (2016).
37. Khayatt, B. I., Overmars, L., Siezen, R. J. & Francke, C. Classification of the Adenylation and Acyl-Transferase Activity of NRPS and PKS Systems Using Ensembles of Substrate Specific Hidden Markov Models. *PLoS One* **8**, e62136 (2013).
38. Chevrette, M. G., Aicheler, F., Kohlbacher, O., Currie, C. R. & Medema, M. H. SANDPUMA: Ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics* **33**, 3202–3210 (2017).
39. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726 (2002).
40. 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).
41. Chen, C.-Y., Georgiev, I., Anderson, A. C. & Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3764–3769 (2009).
42. Villiers, B. & Hollfelder, F. Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. *Chem. Biol.* **18**, 1290–1299 (2011).
43. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for 'clickable' amino acids. *Angew Chem Int Ed Engl* **53**, 10105–10108 (2014).
44. Fischbach, M. A., Lai, J. R., Roche, E. D., Walsh, C. T. & Liu, D. R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11951–11956 (2007).

45. 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).
46. Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* **10**, 282–287 (2018).
47. Baltz, R. H. Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.* **3**, 748–758 (2014).
48. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
49. 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).
50. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. **11**, (2020).
51. Huang, H.-M., Stephan, P. & Kries, H. Engineering DNA-Templated Nonribosomal Peptide Synthesis. *Cell Chem. Biol.* **28**, 221-227.e7 (2021).
52. Throckmorton, K. *et al.* Directed Evolution Reveals the Functional Sequence Space of an Adenylation Domain Specificity Code. *ACS Chem. Biol.* **14**, 2044–2054 (2019).
53. Stanišić, A. & Kries, H. Adenylation Domains in Nonribosomal Peptide Engineering. *ChemBioChem* **20**, 1347–1356 (2019).
54. 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).
55. Peypoux, F. & Michel, G. Controlled biosynthesis of Val7- and Leu7-surfactins. *Appl. Microbiol. Biotechnol.* **36**, 515–517 (1992).
56. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science (80-.).* **321**, 659–663 (2008).
57. Khersonsky, O. *et al.* Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **72**, 178-186.e5 (2018).
58. Land, H. & Humble, M. S. YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. in *Protein Engineering: Methods and Protocols* (eds. Bornscheuer, U. T. & Höhne, M.) 43–67 (Springer New York, 2018). doi:10.1007/978-1-4939-7366-8_4.
59. Nath, A. & Atkins, W. M. A Quantitative Index of Substrate Promiscuity †. *Biochemistry* **47**, 157–166 (2008).

60. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5869–5874 (2006).
61. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).
62. Bigman, L. S. & Levy, Y. Proteins: molecules defined by their trade-offs. *Curr. Opin. Struct. Biol.* **60**, 50–56 (2020).
63. Xu, Z., Cen, Y.-K., Zou, S.-P., Xue, Y.-P. & Zheng, Y.-G. Recent advances in the improvement of enzyme thermostability by structure modification. *Crit. Rev. Biotechnol.* **40**, 83–98 (2020).
64. Eijsink, V. G. H., Gåseidnes, S., Borchert, T. V & van den Burg, B. Directed e volution of enzyme stability. *Biomol. Eng.* **22**, 21–30 (2005).
65. Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707 (2008).
66. Watanabe, K., Ohkuri, T., Yokobori, S. & Yamagishi, A. Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J. Mol. Biol.* **355**, 664–674 (2006).
67. Bershtein, S., Goldin, K. & Tawfik, D. S. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044 (2008).
68. Rigoldi, F. *et al.* Thermal stabilization of the deglycating enzyme Amadoriase I by rational design. *Sci. Rep.* **8**, 3042 (2018).
69. Bednar, D. *et al.* FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLOS Comput. Biol.* **11**, e1004556 (2015).
70. Jones, B. J., Lim, H. Y., Huang, J. & Kazlauskas, R. J. Comparison of Five Protein Engineering Strategies for Stabilizing an α/β-Hydrolase. *Biochemistry* **56**, 6521–6532 (2017).
71. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, 35–37 (2008).
72. Beadle, B. M. & Shoichet, B. K. Structural Bases of Stability–function Tradeoffs in Enzymes. *J. Mol. Biol.* **321**, 285–296 (2002).
73. Honaker, M. T., Acchione, M., Sumida, J. P. & Atkins, W. M. Ensemble perspective for catalytic promiscuity: calorimetric analysis of the active site conformational landscape of a detoxification enzyme. *J. Biol. Chem.* **286**, 42770–42776 (2011).
74. Hou, L. *et al.* Functional promiscuity correlates with conformational heterogeneity in A-class glutathione S-transferases. *J. Biol. Chem.* **282**, 23264–23274 (2007).

75. Honaker, M. T., Acchione, M., Zhang, W., Mannervik, B. & Atkins, W. M. Enzymatic detoxication, conformational selection, and the role of molten globule active sites. *J. Biol. Chem.* **288**, 18599–18611 (2013).

76. Dellus-Gur, E., Toth-Petroczy, A., Elias, M. & Tawfik, D. S. What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-offs. *J. Mol. Biol.* **425**, 2609–2621 (2013).

77. Ben-David, M. *et al.* Enzyme evolution: An epistatic ratchet versus a smooth reversible transition. *Mol. Biol. Evol.* **37**, 1133–1147 (2020).

78. Bloom, J. D. & Arnold, F. H. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **106 Suppl**, 9995–10000 (2009).

79. Reetz, M. T. & Sanchis, J. Constructing and analyzing the fitness landscape of an experimental evolutionary process. *Chembiochem* **9**, 2260–2267 (2008).

80. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).

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

Supplementary Information

Exploring the Functional Sequence Space of Adenylation in SrfAC by Hydroxamate Profiling

Aleksa Stanišić, Ulrich Ettelt, Carl-Magnus Svensson, Marc Thilo Figge, and Hajo Kries*

Independent Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans Knöll Institute (HKI Jena), Beutenbergstr. 11a, 07745 Jena, Germany

Contents

Protein models

3D-model of SrfAC in complex with L-Leu-AMP ligant was built with YASARA automated software.[1] Model was aligned to the crystal structure of SrfAC (PDB: 2vsq) to confirm that modelling did not change the position of specificity code residues in the binding pocket (Supplementary Figure 2). 3D-model of VSA mutant was created with the help of SWISS-MODEL (https://swissmodel.expasy.org/)[2,3] by modelling on the X-ray crystal structure of SrfAC (PDB: 2vsq)[4] in its thiolation state. Those models were then aligned in PyMOL (https://pymol.org/).

Cloning

General cloning

E. coli HST08 Stellar competent cells (Takara Biotech) were used for the In-Fusion cloning (Takara Bio Europe). For the propagation and storage of plasmids, *E. coli* NEB 5-alpha (New England Biolabs) was used. Holo proteins were expressed in *E. coli* strain HM0079.[5] Plasmid DNA, DNA fragments, and PCR products were purified using NucleoSpin Plasmid and Gel and PCR clean-up kits (Macherey Nagel). DNA amplification was done with with Q5 polymerase (New England Biolabs, Massachusetts) following the supplier's instructions. Two-fragment cloning in linearized vector was done using the InFusion cloning kit (Takara Bio Europe). Oligonucleotide primers were made by custom synthesis and sequence confirmation of assembled constructs was performed using the Mix2Seq service for Sanger sequencing (Eurofins Genomics). pTrc99a-*srfAC* plasmid was kindly provided by Prof. Donald Hilvert (ETH Zurich). For the linearization of pTrc99a-*srfAC*, restriction enzymes BlpI, DraIII and BstBI (New England Biolabs, Massachusetts) were used, depending on the position of the mutation. Libraries were generated by amplification of one or two DNA fragments with primers bearing randomized codons. Before cloning, two-fragment samples were additionally concatenated by assembly PCR to increase the cloning efficiency. Cloned libraries were transformed into E. coli HST08 Stellar Competent Cells. Libraries were purified from overnight liquid cultures grown under ampicillin selection and resulting plasmid mix used for transformation in HM0079 protein expression strain.

Sequencing

The identity of purified plasmids was confirmed by overnight Sanger sequencing service (Eurofins Genomics). The identity of library mutants was determined by *E. coli* plate sequencing service (Microsynth) by withdrawing the aliquot of the saturated preculture shortly before the induction of protein expression.

Library design

FuncLib library of SrfAC

SrfAC randomization with FuncLib

To maintain the essential residues required for adenylation, a SWISS homology model of SrfA-C A-domain was built using crystal structure of EntF (PDB: 5T3D) in complex with serine adenosine vinylsulfonamide inhibitor (Ser-AVS), a nonhydrolyzable analogue of serine-AMP. A structure of L-Leu-AMP was modelled into the SrfAC SWISS model using YASARA molecular graphic software and the resulting model used for FuncLib randomization.[6] Eight specificity code residues of SrfAC were selected for simultaneous randomization by FuncLib: A660, F663, F702, L726, G728, C752, V760 and F761 at default parameters for multiple sequence alignments (Min ID: 35, Max targets: 4000, Coverage: 75, E value: 0.0001). Conformations of AMP and D659 were fixed to maintain the interactions necessary for adenylation. FuncLib generated signatures of residues tolerated at each of the 8 selected positions. Three residues were selected (in bold) for the following construction of library of triple mutants.

A660: AFGLV	F663: FIMWY	**F702: FHSTWY**	L726: LIMVY
G728: GACS	**C752: CALMSTV**	V760: VACFILTY	F761: FHILMWY

Cloning of FuncLib library of SrfAC

For generating FuncLib library of SrfAC triple mutants, a series of oligonucleotides containing degenerate codons coding for predicted residues at three positions were used. A wild type residue is included in each position. Individual oligonucleotides were combined in appropriate ratios and used for PCR amplification of DNA fragments using pTrc99a-SrfAC as a template (Supplementary Table 1). A single DNA fragment for each position is generated and resulting three fragments (A, B, C) are assembled by PCR using two primers with vector-specific overhangs (SrfAC_o_f + SrfAC_o_r, Supplementary Table 2). Assembled fragment is cloned into linearized pTrc99a-SrfAC (DraIII + BstBI) by InFusion cloning and resulting plasmid mix transformed into Stellar competent cells. After the SOC outgrowth phase, 10 µL of transformed culture was inoculated in 3 mL of TB medium with added ampicillin and grown overnight. Plasmid library purified from TB/Stellar culture was transformed into HM0079 for protein expression or NEB 5-alpha for long term storage.

NNK libraries of VSA

Fifteen positions in the binding pocket of VSA were targeted for full randomization to generate 15 NNK libraries of single mutants. DNA fragments containing NNK library were amplified by PCR using pTrc99a-SrfAC-VSA as a template by using NNK oligonucleotides as primers. To prevent the amplification bias from the wild type sequence, each NNK oligo contained a silent mutation adjacent to the NNK codon. Depending on the location of the residue, generated fragments were cloned in one or two steps (Supplementary Tables 3 and 4). Where NNK positions were distant from the restriction site, two fragments were generated and assembled by PCR using two primers with vector-specific overhangs (VSA_Blp_f and SrfAC_o_r). InFusion cloning was done as two fragment assembly with NNK-containing DNA fragment and pTrc99a-SrfAC-VSA linearized with appropriate restriction enzyme pairs. Mutants missing from the libraries were cloned and screened in a separate sample batch (Supplementary Table 5).

Library screening in 96 well plate format

Expression

E. coli HM0079 transformed with pTrc99a library constructs was used for overexpression of C-terminally His_6-tagged holo-NRPS proteins. Precultures were prepared by inoculating the transformants by picking colonies from the agar plate to a round bottom 96-well plate (310 µl, Sarstedt) filled with 150 µl of 2xYT medium supplemented with 100 µg/ml of ampicillin. Each 96-well plate contained four wells with positive control (pTrc99a-SrfaC for FuncLib library, pTrc99a-SrfAC-VSA for NNK libraries) and 4 wells with negative control (pTrc99a-SrfAC with A-domain interrupted with the stuffer fragment). Plates were sealed with a breathable film (Sigma) and incubated for 18 h at 30 °C, 300 rpm in an orbital shaker. For the expression, 20 µl of the preculture was inoculated into a 96 deep-well plate (2 mL, Sarstedt) containing 1 ml 2xYT medium supplemented with 100 µg/ml ampicillin and incubated for 4-6 hours at at 30 °C, 300 rpm until the OD600 reached approximately 1. Prior to induction, 20 µL aliquot was taken from the culture for preparing 25 % glycerol stock for the long-term storage at -80 °C. Additionally, 5 µL aliquot was taken for sequencing. At the induction phase, the temperature was reduced to 18°C for 30 min and induced with 0.25 mM IPTG (Thermo Scientific) and incubated at 18°C, 300 rpm for 18-20 h. Cells were harvested by centrifugation at 3000 g, 15 °C and the supernatant was discarded. The pellet was resuspended in 400 µl lysis buffer (50 mM TRIS [pH 8.0], 100 mM NaCl, 10 mM imidazole, 1.5 mg/mL lysozyme, 50 µL of protease inhibitor mix per plate (Sigma, P8849)) and incubated for 30 min at room temperature. Cells were lysed by a single freeze-thaw cycle at -20 °C.

Purification

After the thawing of the lysate for 1.5 – 2 h at room temperature, 100 µL of DNA removal mix (50 mM TRIS [pH 8.0], 100 mM NaCl, 10 mM imidazole, 10 mM $MgCl_2$, 10 mM TCEP, 15 U/mL Turbonuclease (Jena Bioscience)) was added to reduce the viscosity of the lysate and incubated without shaking at room temperature for 15 min. Cell debris was removed by centrifugation at 3000 g, 6 °C for 30 min. In a separate, 96-well plate (1.8 mL, Sarstedt) compatible with the magnetic separation rack (New England BioLabs), 20 µl of a 25 % Ni-IDA MagBeads (PureCube) suspension was added. The beads were equilibrated with 700 µl lysis buffer and the supernatant was discarded. To purify the released His_6-tagged proteins from the lysate, 400 µl of the lysate supernatant was transferred to the equilibrated beads. The plate was covered with silicon lid, kept at 4 °C in the fridge for 20 min with vigorous shaking every 5 minutes to prevent the aggregation of MagBeads. Beads were subsequently pulled down with the magnetic separator and the supernatant was discarded. To remove the unbound proteins and imidazole, the beads were washed twice with 700 µl of wash buffer (50 mM TRIS [pH 8.0], 100 mM NaCl) with the help of the magnetic separator.

HAMA screening

After the second washing step, 100 µl of freshly prepared HAMA master mix (50 mM TRIS [pH 8.0], 5 mM ATP, 5 mM $MgCl_2$, 100 mM hydroxylamine (adjusted to pH 7.5-8 with NaOH), 1 mM TCEP, 1 mM proteinogenic amino acids) was added directly to the beads containing the adsorbed protein and incubated at room temperature for 1.5 h. We found that enzymes maintain adenylation activity without the elution step and the imidazole from the elution buffer interferes with subsequent hydroxamate detection. After the incubation, 6 µl of the reaction mixture was diluted in 54 µl of analysis solution (95% acetonitrile, 0.1 % formic acid, 1 µM pipecolic acid hydroxamate as an injection control) in a 384 well plate (100 µL, Brandt). After the dilution step, the 384-well plate was immediately placed on ice, covered with aluminium foil to minimize evaporation of the solvent. The plate was analysed immediately by UPLC-MS/MS according to the general HAMA procedure (Supplementary Information 6).

General protein overexpression and purification

For the large-scale expression and purification of individual proteins, saturated *E. coli* HM0079 culture (0.5 mL) with appropriate pTrc99a-SrfAC construct was inoculated in 500 mL of 2xYT medium supplemented with ampicillin in 2 L shaking flask and shaken at 37 °C at 200 rpm. Cultures were grown for 4-6 hours until OD_{600} = 1, induced with 0.25 mM isopropyl-D-thiogalactoside (IPTG) and grown for another 16-20 hours at 20 °C. Cells were pelleted by centrifugation at 8 000 g and the supernatant was discarded. Cell pellets were resuspended in 30 mL lysis buffer (50 mM TRIS [pH 7.4], 500 mM NaCl, 20 mM imidazole, 2 mM TCEP) and 100 µL of protease inhibitor mix (Sigma, P8849) was added before cell lysis by sonication. The lysate was cleared by centrifugation at 19,000 g for 30 min at 4 °C. Proteins were adsorbed on 2 mL of Ni-IDA suspension (Rotigarose, Roth) preequilibrated with lysis buffer by loading the lysate supernatant on the open column. Unbound proteins were washed twice with 20 mL of the lysis buffer before the elution with 4 x 0.75 mL elution buffer (50 mM TRIS [pH 7.4], 500 mM NaCl, 300 mM imidazole, 2 mM TCEP). Fractions containing protein were pooled and the buffer was exchanged with protein storage buffer (50 mM TRIS [pH 7.6], 200 mM NaCl) on 6 mL Vivaspin (Sartorius) filters with 30 kDa cut-off. Glycerol was added to 10% and protein concentration adjusted to 50 µM. Samples were flash frozen in liquid nitrogen and stored at -20 °C. Absorbance at 280 nm measured in Take3 plates on an Epoch2 microplate reader (Biotek) was used for measuring protein concentration, using calculated extinction coefficients (www.benchling.com).

SDS-PAGE of overexpressed proteins

Purity of proteins was determined by SDS-PAGE (Figure 10) using Bolt 4-12% Bis-Tris Plus Gels (ThermoFisher Scientific) with MES-SDS running buffer (Novex). Triple Color Protein Standard III (Serva) was run alongside the protein samples as a size standard. The gels were run at 200 V for 22 minutes and stained with Quick Coomassie stain (Serva).

General hydroxamate specificity assay (HAMA)

Reaction conditions

The hydroxamate formation assay with purified proteins was conducted as described previously.[7] Reactions of 100 µL contained 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 150 mM hydroxylamine (pH 7.5-8, adjusted with NaOH), 5 mM ATP (A2383, Sigma), 1 mM TCEP and 1-5 µM of enzyme. Master mix without the enzyme was prepared and the reaction was initiated by adding enzyme or heat-inactivated enzyme as a control. L-Phe,

L-Val and L-Leu were distinguished from D-Phe, D-Val and L-Ile, respectively by using enantiopure, deuterium labelled standards. Reaction quenching was done after 1 h by 10-fold dilution in acetonitrile containing 0.1 % formic acid and immediately analyzed with UPLC-MS. All assays were done from a single protein batch in technical triplicates.

UPLC-MS/MS conditions

Chromatography was performed on a Waters ACQUITY H-class UPLC system (Waters) with an injection volume of 3 µL. Water with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B) were used as strong and weak eluent, respectively. Separation of amino acid hydroxamates was done on the ACQUITY UPLC BEH Amide column (1.7 µm, 2.1 x 50 mm) with a linear gradient of 10-50% A over 5 min (flow rate 0.4 mL/min) followed by 4 min reequilibration. Data were analyzed with MassLynx and TargetLynx software (version 4.1).

MS/MS detection was performed on Xevo TQ-S micro (Waters) tandem quadrupole instrument with ESI ionisation source in positive ion mode. Nitrogen was used as a desolvation gas and argon as collision gas. The following source parameters were used: capillary voltage 1.5 kV, cone voltage 65 V, desolvation temperature 500 °C, desolvation gas flow 1000 L/h. Specific mass transitions recorded in multiple reaction monitoring (MRM) mode were used to detect and quantify amino acid hydroxamates.[7]

Thermostability assay

Thermal stability of SrfAC mutants from FuncLib library was determined by incubation of enzyme solution at at different temperatures in the range between 30 and 50 °C for 1 hour. Enzymes were subsequently transferred to HAMA master mix and the formation of amino acid hydroxamates is followed over 1.5 hour at room temperature. Assays were done from a single batch of enzyme in two technical replicates.

Saturation kinetics (MesG/hydroxylamine assay)

Michaelis-Menten parameters of the adenylation with L-Leu for SrfAC and additionally with L-Met and L-Phe for VSA were determined using the MesG/hydroxylamine assay.[8] Low activity of SrfAC for L-Phe and L-Met did not allow the determination of kinetic parameters. Reactions contained 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 100 µM 7-methylthioguanosine (MesG), 150 mM hydroxylamine (adjusted to pH 7.5-8 with NaOH), 5 mM ATP (A2383, Sigma), 1 mM TCEP, 0.4 U/mL inorganic pyrophosphatase (I1643, Sigma), 1 U/mL of purine nucleoside phosphorylase from microorganisms (N8264, Sigma) and 5 µM of NRPS. Flat-bottom 384-well plates (100 µL, 781620, Brand) were used for the reactions. Reactions were started by addition of enzyme and the

absorbance was followed at 355 nm on a Synergy H1 (BioTek) microplate reader at 30 °C. Reactions used for background subtraction contained heat-inactivated enzyme. Each substrate concentration was measured in duplicate. Initial velocities ($ODmin^{-1}$) were divided by the slope of a pyrophosphate calibration curve to obtain the pyrophosphate release rate. Initial velocities $v_0/[E_0]$ were fit to the Michaelis-Menten equation by nonlinear regression using RStudio version 1.3.1093 (Supplementary Figure 5).[9]

Data analysis

Random sampling of colonies resulted in variable number of replicates for each mutant. In each batch of samples, hydroxamate concentrations were averaged between the replicated mutants. Total activity of each mutant was calculated as a sum of 19 measured hydroxamates. To minimize the systematic error caused by variable protein expression and purification efficiency in different sample batches, the relative activity is calculated by normalizing the total activity of each mutant by the total activity of the wild type from the same sample batch (Equation 1).

$$A_{rel\,m} = \frac{\sum_{i=19}^{N}[\mathrm{HA}]_m}{\sum_{i=19}^{N}[\mathrm{HA}]_{wt}} \tag{1}$$

Average value of relative activities of 20 mutants for each enzyme position – $A_{rel\,P}$ is used as a measure of tolerance of targeted position to mutations (Equation 2).

$$A_{rel\,P} = \frac{\sum_{i=20}^{N} A_{rel\,m}}{20} \tag{2}$$

The promiscuity of each mutant was calculated based on the model presented by Nath et al.[10] The model uses the Shannon entropy P as a metric for promiscuity (Equation 3) with p_i being the probability that the i'th substrate is converted to a hydroxamate by the enzyme.

$$P = \sum_{i=19}^{N} p_i \times log\, p_i \tag{3}$$

The probability p_i was derived from the proportion of the amino acid hydroxamates (Equation 4).

$$p_i = \frac{[\mathrm{HA}]_i}{\sum_{i=19}^{N}[\mathrm{HA}]} \tag{4}$$

On the basis of P the promiscuity index I of each mutant was calculated for as follows:

$$I = -\frac{1}{log\ 19}P \tag{5}$$

N indicates the number of measured hydroxamates (N = 19). Promiscuity index I can take values between 0 and 1, with 0 corresponding to perfectly specific and 1 to perfectly promiscuous enzyme. To better discern the changes in promiscuity caused by mutations, relative promiscuity index I_{rel} is calculated by normalization by the wild type (Equation 6). This results in I_{rel} values higher than 1 for more promiscuous and lower than 1 for more specific mutants with respect to the wild type.

$$I_{rel} = \frac{I_m}{I_{wt}} \tag{6}$$

To prevent the inclusion of falsely specific mutants, a cut off value for the activity is included to filter out the mutants showing only traces of activity. Namely, due to the different detection limits of hydroxamates, low adenylation activity results in specificity profiles showing traces of individual products which results in low I values. Therefore, before the promiscuity index is calculated, all mutants which accumulate less than 0.2 µM hydroxamates are excluded from the calculation. Mutants were subsequently ranked according to I. P and I were calculated and visualized in R (Supplementary Table 6).

Supplementary Tables

Supplementary Table 1. Oligonucleotide mix for the FuncLib library of SrfAC.

Position	Oligo	Molar ratio	Residues	Oligo mix
A660	SrfAC_660_BTT_f	3	FLV	SrfAC_660_f
	SrfAC_660_GSC_f	2	AG	
F702	SrfAC_702_ASC_f	2	ST	SrfAC_702_f
	SrfAC_702_YAT_f	2	HY	
	SrfAC_702_TTT_f	1	F	
	SrfAC_702_TGG_f	1	W	
C752	SrfAC_752_TGC_f	1	C	SrfAC_752_f
	SrfAC_752_DYG_f	6	ALMSTV	

Supplementary Table 2. PCR amplification and the assembly of fragments for FuncLib library of SrfAC.

Fragment amplification		Fragment assembly
Oligo mix	Fragment	Oligo
SrfAC_660_f	A	SrfAC_o_f
SrfAC_660_r		
SrfAC_702_f	B	
SrfAC_702_r		SrfAC_o_r
SrfAC_752_f	C	
SrfAC_o_r		

Supplementary Table 3. PCR amplification and the assembly of fragments for NNK libraries of VSA.

Library	Fragment	Oligo		Restriction enzyme
VSA-S654NNK	654A	VSA_Blp_f	Assembly PCR	BlpI + DraIII
		VSA_S654NNK_o_r		
	654B	VSA_S654NNK_N655s_f		
		SrfAC_o_r		
VSA-F658NNK		VSA_F658NNK_D659s_f		BstBI + DraIII
		SrfAC_o_r		
VSA-V660NNK		VSA_V660NNK_F661s_f		BstBI + DraIII
		SrfAC_o_r		
VSA-F661NNK		VSA_F661NNK_T662s_f		BstBI + DraIII
		SrfAC_o_r		
VSA-F663NNK		VSA_F663NNK_D664s_f		BstBI + DraIII
		SrfAC_o_r		
VSA-D664NNK		VSA_D664NNK_F665s_f		BstBI + DraIII
		SrfAC_o_r		
VSA-S702NNK	702A	VSA_Blp_f	Assembly PCR	BlpI + DraIII
		VSA_S702NNK_o_r		
	702B	VSA_S702NNK_A703s_f		
		SrfAC_o_r		
VSA-A703NNK	703A	VSA_Blp_f	Assembly PCR	BlpI + DraIII
		VSA_A703NNK_o_r		
	703B	VSA_A703NNK_T704s_f		
		SrfAC_o_r		
VSA-L726NNK	726A	VSA_Blp_f	Assembly PCR	BlpI + DraIII
		VSA_L726NNK_o_r		
	726B	VSA_L726NNK_F727s_f		
		SrfAC_o_r		
VSA-F727NNK	727A	VSA_Blp_f	Assembly PCR	BlpI + DraIII
		VSA_F727NNK_o_r		
	727B	VSA_F727NNK_G728s_f		
		SrfAC_o_r		
VSA-G728NNK	728A	VSA_Blp_f	Assembly PCR	BlpI + DraIII
		VSA_G728NNK_o_r		
	728B	VSA_G728NNK_G729s_f		
		SrfAC_o_r		
VSA-A752NNK		VSA_Blp_f		BlpI + DraIII
		VSA_A752NNK_N751s_r		

VSA-T759NNK	VSA_Blp_f VSA_T759NNK_G758s_r	BlpI + DraIII
VSA-V760NNK	VSA_Blp_f VSA_V760NNK_T759s_r	BlpI + DraIII
VSA-F761NNK	VSA_Blp_f VSA_F761NNK_V760s_r	BlpI + DraIII

Supplementary Table 4. Oligonucleotide sequences for PCR primers. Targeted positions are labelled in bold.

Name	Sequence
SrfAC_o_f	GATCAGGATACGTTCTTGTCTGTTTC
SrfAC_o_r	GAATCCGGCAGATCATGCAC
SrfAC_660_BTT_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCCTTTGAT**BTT**TTTACCTTTGATTTCTATGC
SrfAC_660_GSC_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCCTTTGAT**GSC**TTTACCTTTGATTTCTATGC
SrfAC_660_r	CATGACATTGACATTCTCTTGCAG
SrfAC_702_ASC_f	CAAGAGAATGTCAATGTCATG**ASC**GCGACAACCGCACTATTTAATC
SrfAC_702_YAT_f	CAAGAGAATGTCAATGTCATG**YAT**GCGACAACCGCACTATTTAATC
SrfAC_702_TTT_f	CAAGAGAATGTCAATGTCATG**TTT**GCGACAACCGCACTATTTAATC
SrfAC_702_TGG_f	CAAGAGAATGTCAATGTCATG**TGG**GCGACAACCGCACTATTTAATC
SrfAC_702_r	GTTAATCAGCTTGCCCGGC
SrfAC_752_TGC_f	GCTGCGGATCATGGGGCCGGGCAAGCTGATTAAC**TGC**TACGGGCCGACTGAGGGAAC
SrfAC_752_DYG_f	GCTGCGGATCATGGGGCCGGGCAAGCTGATTAAC**DYG**TACGGGCCGACTGAGGGAAC
VSA_Blp_f	GATGAAAGAACAAGCGGCTGAGCTG
VSA_S654NNK_o_r	ACAGACAAGAACGTATCCTGATCAGAAAATGC
VSA_S654NNK_N655s_f	GATACGTTCTTGTCTGTT**NNKAAC**TACGCCTTTGATGTTTTTACCTTTGATTTC
VSA_F658NNK_D659s_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCC**NNKGACGTT**TTTACCTTTGATTTCTATGCTTCTATGC
VSA_V660NNK_F661s_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCCTTTGAT**NNKTTC**ACCTTTGATTTCTATGCTTCTATGCTG
VSA_F661NNK_T662s_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCCTTTGAT**GTTNNKACG**TTTGATTTCTATGCTTCTATGCTGAATGCG
VSA_F663NNK_D664s_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCCTTTGATGTTTTTACC**NNKGAC**TTCTATGCTTCTATGCTGAATGCG
VSA_D664NNK_F665s_f	GATCAGGATACGTTCTTGTCTGTTTCGAATTACGCCTTTGATGTTTTTACCTTT**NNKTTT**TATGCTTCTATGCTGAATGCGG
VSA_S702NNK_o_r	CATGACATTGACATTCTCTTGCAGG
VSA_S702NNK_A703s_f	CCTGCAAGAGAATGTCAATGTCATG**NNKGCC**ACAACCGCACTATTTAATCTTCTCAC
VSA_A703NNK_o_r	**GCT**CATGACATTGACATTCTCTTGCAGG
VSA_A703NNK_T704s_f	CCTGCAAGAGAATGTCAATGTCATG**AGCNNKACC**ACCGCACTATTTAATCTTCTCACAG
VSA_L726NNK_o_r	TATACAGCGAAGCCCCTTCATC
VSA_L726NNK_F727s_f	GATGAAGGGGCTTCGCTGTATA**NNKTTT**GGCGGAGAGCGCGCGTCAG
VSA_F727NNK_o_r	TAATATACAGCGAAGCCCCTTCATC
VSA_F727NNK_G728s_f	GATGAAGGGGCTTCGCTGTATATTA**NNKGGT**GGAGAGCGCGCGTCAGTG
VSA_G728NNK_o_r	GAATAATATACAGCGAAGCCCCTTC

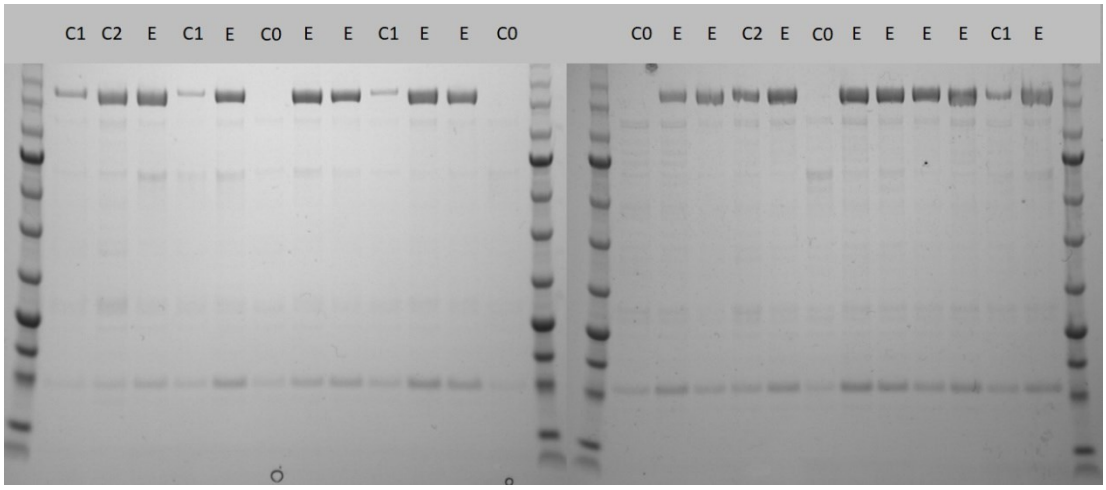
Supplementary Table 5. Overview of NNK libraries of VSA. Numbers in the table denote the frequency of occurrence of the mutant. Missing mutants are labelled in red.

	Mutation																			
Position	**A**	**R**	**N**	**D**	**C**	**Q**	**E**	**G**	**H**	**I**	**L**	**K**	**M**	**F**	**P**	**S**	**T**	**W**	**Y**	**V**
S654	3	7	0	2	5	0	3	5	0	4	13	0	2	9	0	10	0	6	3	8
F658	1	5	3	3	6	3	2	5	0	4	7	3	1	8	2	9	2	5	1	9
V660	1	5	1	3	7	1	1	9	4	6	13	0	1	4	1	4	2	3	4	8
F661	3	3	0	2	0	2	4	10	4	0	5	4	3	7	7	5	1	4	5	11
F663	1	6	5	2	0	7	2	6	0	1	10	4	3	4	5	1	2	5	0	2
D664	0	12	5	6	0	4	1	18	0	1	4	0	3	2	4	2	3	0	0	12
S702	0	4	1	3	1	0	3	8	0	4	8	1	5	10	1	12	0	4	3	7
A703	5	5	0	2	2	0	3	7	4	3	7	1	1	5	1	3	1	7	5	13
L726	2	4	6	1	5	0	4	4	2	7	7	4	2	8	0	5	2	5	3	6
F727	2	1	3	3	6	0	0	5	2	2	14	1	5	11	0	2	3	5	4	7
G728	2	2	3	2	6	2	2	5	0	4	11	0	5	12	0	3	0	3	3	12
A752	10	6	1	2	4	1	3	1	3	0	11	6	1	3	11	4	4	0	0	1
T759	1	7	9	0	1	4	1	0	4	5	5	8	0	5	8	8	8	1	5	3
V760	2	4	2	4	0	2	0	0	4	1	10	10	2	2	6	6	6	1	3	5
F761	4	2	2	7	1	2	0	2	3	4	7	3	6	5	9	7	4	2	5	5

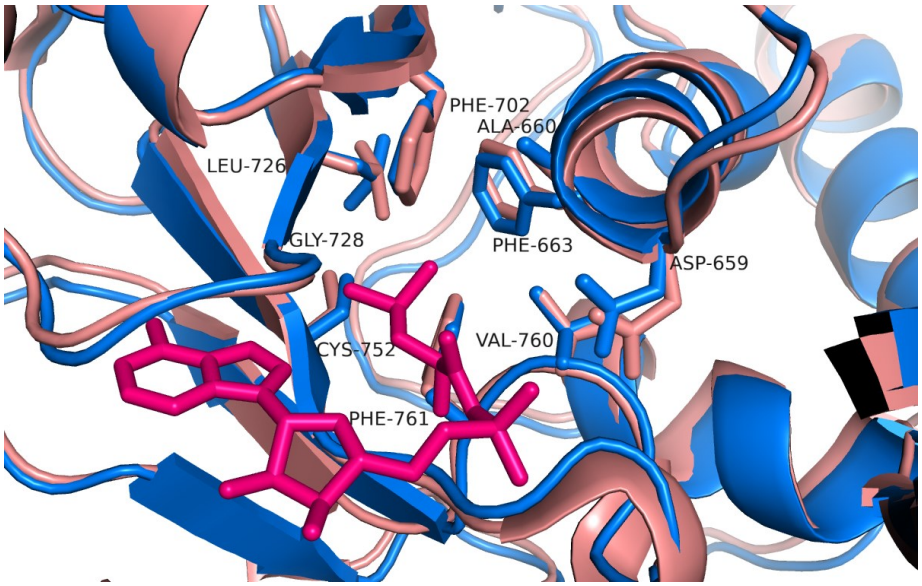
Supplementary Table 6. Top 20 mutants from FuncLib SrfAC library and VSA NNK libraries with highest activity (A_{rel}), promiscuity and selectivity (I_{rel}) relative to the progenitor VSA.

SrfAC FuncLib library				**VSA NNK libraries**					
Activity		**Promiscuity**		**Activity**		**Promiscuity**		**Specificity**	
Mutant	A_{rel}	Mutant	I_{rel}	Mutant	A_{rel}	Mutant	I_{rel}	Mutant	I_{rel}
ASV	3.35	VYS	2.77	A752I	2.27	S702F	1.46	G728M	0.06
ASA	2.96	GWV	2.72	S702T	2.21	V660E	1.33	G728L	0.09
VSA	2.27	ASA	2.66	V660L	2.08	S654I	1.29	V660W	0.10
VSV	2.03	GWS	2.62	S702A	2.04	V660Q	1.28	A752M	0.13
ASL	1.91	ASV	2.45	V660I	1.55	F658Q	1.27	V660Y	0.14
VSL	1.60	ASL	2.34	A703N	1.54	V660S	1.27	V660F	0.21
LSL	1.51	AWS	2.31	V660A	1.51	V660A	1.27	F761A	0.27
GSL	1.46	AWM	2.28	A703I	1.51	F658A	1.26	G728A	0.29
VFA	1.39	VSA	2.24	F663W	1.51	S654Q	1.26	F727Y	0.41
GTV	1.28	VFA	2.19	A703M	1.42	F658S	1.25	S702D	0.41
GSV	1.22	GWC	2.14	F661A	1.41	F663F	1.24	L726D	0.48
GTL	1.15	FWL	2.00	S654N	1.37	F658G	1.24	L726G	0.50
GSC	1.07	FFA	1.98	V760G	1.33	A752G	1.22	F727S	0.50
GST	1.05	VSV	1.91	A703L	1.24	F658T	1.21	G728F	0.51
VFM	1.03	LWA	1.86	A752V	1.24	S654L	1.21	L726A	0.57
AFC	1.02	GYM	1.69	S654A	1.23	D664E	1.20	F761I	0.57
FSV	0.99	FWS	1.66	S654G	1.19	S654G	1.20	F727A	0.57
GYV	0.98	AWL	1.63	F727I	1.15	F661T	1.20	F761V	0.58
FSA	0.96	VYL	1.63	A703A	1.10	S654M	1.20	L726Y	0.58
FSL	0.96	GSL	1.54	V660G	1.07	F658H	1.19	F727T	0.58

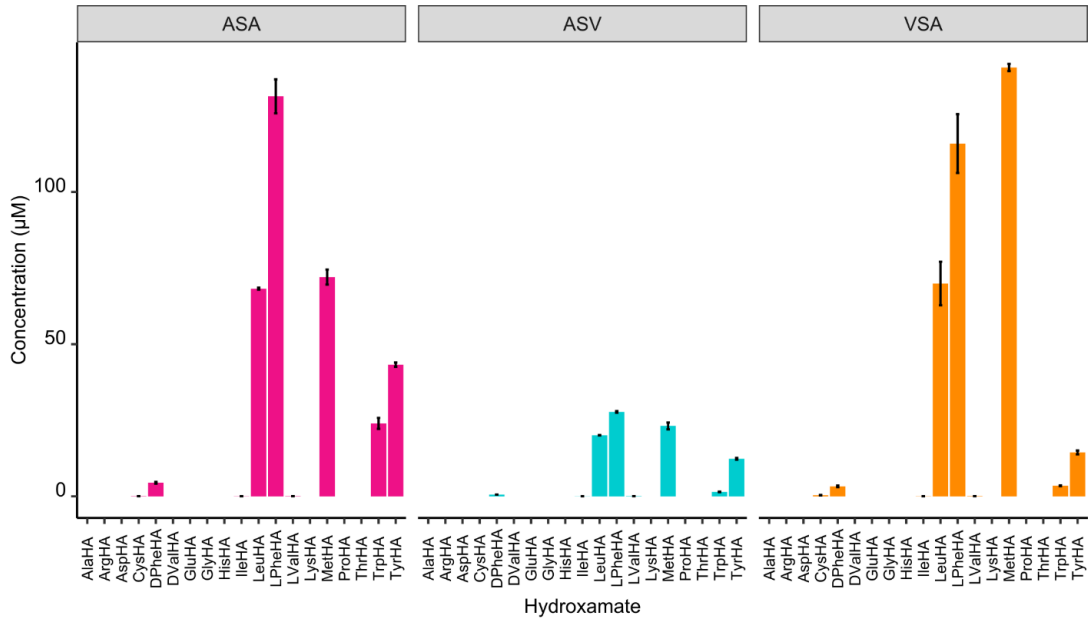
Supplementary Figures



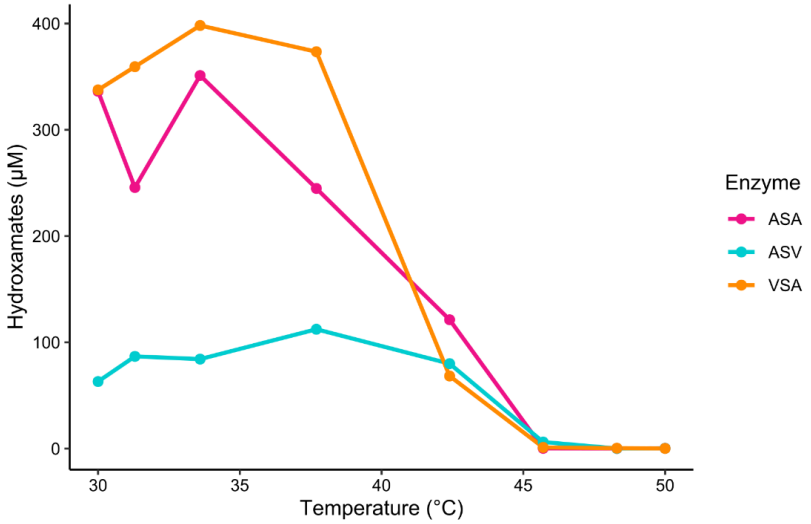
Supplementary Figure 1. SDS PAGE of SrfAC expressed and purified in 96-well plate format. Proteins were eluted from magnetic beads with 200 µL of elution buffer (50 mM TRIS pH 8.0, 200 mM imidazole) and 5 µL was loaded on the gel. E, HM0079 strain with pTrc99a-SrfAC; C0, negative control containing the empty vector; C1, purification control with empty vector and SrfAC added to the cell lysate; C2, purification control with empty vector and SrfAC added to the eluate.



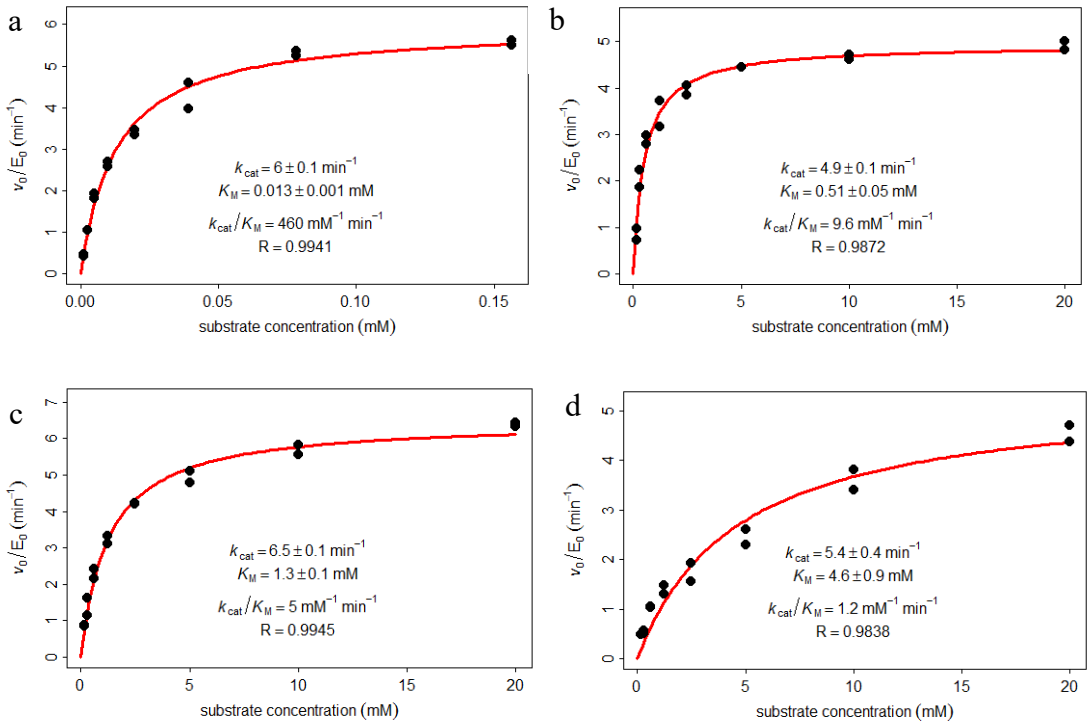
Supplementary Figure 2. Overlay of YASARA model of SrfAC with Leu-AMP (blue) and SrfAC crystal structure (PDB: 2VSQ, pink). Specificity code residues are labeled.



Supplementary Figure 3. HAMA specificity profiles of three mutants with highest activity and promiscuity from FuncLib SrfAC library. Errors are standard deviations from two technical replicates.



Supplementary Figure 4. Thermostability of three mutants with highest activity and promiscuity from SrfAC FuncLib library. Errors are standard deviations from two technical replicates (too small to be visible).



Supplementary Figure 5. Saturation kinetics of SrfAC with L-Leu (a) and VSA with L-Leu (b), L-Phe (c) and L-Met (d) measured with MesG/hydroxylamine spectrophotometric assay. Reactions were measured from a single enzyme batch in technical duplicates.

Sequences of proteins used in this study

A-domain is highlighted in blue. Randomized residues highlighted in red (specificity code) and yellow (second shell).

SrfAC

MSQFSKDQVQDMYYLSPMQEGMLFHAILNPGQSFYLEQITMKVKGSLNIKCLEESMNVIMDRYDVFRTVFIHEKVKRPVQVVLKKRQF
HIEEIDLTHLTGSEQTAKINEYKEQDKIRGFDLTRDIPMRAAIFKKAEESFEWVWSYHHIILDGWCFGIVVQDLFKVYNALREQKPYSLPP
VKPYKDYIKWLEKQDKQASLRYWREYLEGFEGQTTFAEQRKKQKDGYEPKELLFSLSEAETKAFTELAKSQHTTLSTALQAVWSVLISR
YQQSGDLAFGTVVSGRPAEIKGVEHMVGLFINVVPRRVKLSEGITFNGLLKRLQEQSLQSEPHQYVPLYDIQSQADQPKLIDHIIVFENYPL
QDAKNEESSENGFDMVDVHVFEKSNYDLNLMASPGDEMLIKLAYNENVFDEAFILRLKSQLLTAIQQLIQNPDQPVSTINLVDDREREF
LLTGLNPPAQAHETKPLTYWFKEAVNANPDAPALTYSGQTLSYRELDEEANRIARRLQKHGAGKGSVVALYTKRSLELVIGILGVLKAG
AAYLPVDPKLPEDRISYMLADSAAACLLTHQEMKEQAAELPYTGTTLFIDDQTRFEEQASDPATAIDPNDPAYIMYTSGTTGKPKGNITT
HANIQGLVKHVDYMAFSDQDTFLSVSNYAFDAFTFDFYASMLNAARLIIADEHTLLDTERLTDLILQENVNVMFATTALFNLLTDAGED
WMKGLRCILFGGERASVPHVRKALRIMGPGKLINCYGPTEGTVFATAHVVHDLPDSISSLPIGKPISNASVYILNEQSQLQPFGAVGELCIS
GMGVSKGYVNRADLTKEKFIENPFKPGETLYRTGDLARWLPDGTIEYAGRIDDQVKIRGHRIELEEIEKQLQEYPGVKDAVVVADRHES
GDASINAYLVNRTQLSAEDVKAHLKKQLPAYMVPQTFTFLDELPLTTNGKVNKRLLPKPDQDQLAEEWIGPRNEMEETIAQIWSEVLG
RKQIGIHDDFFALGGHSLKAMTAASRIKKELGIDLPVKLLFEAPTIAGISAYLKNGGSDGLQDVTIMNQDQEQIIFAFPPVLGYGLMYQNLS
SRLPSYKLCAFDFIEEEDRLDRYADLIQKLQPEGPLTLFGYSAGCSLAFEAAKKLEEQGRIVQRIIMVDSYKKQGVSDLDGRTVESDVEAL
MNVNRDNEALNSEAVKHGLKQKTHAFYSYYVNLISTGQVKADIDLLTSGADFDMPEWLASWEEATTGVYRVKRGFGTHAEMLQGETL
DRNAEILLEFLNTQTVTVS

SrfAC-VSA

MSQFSKDQVQDMYYLSPMQEGMLFHAILNPGQSFYLEQITMKVKGSLNIKCLEESMNVIMDRYDVFRTVFIHEKVKRPVQVVLKKRQF
HIEEIDLTHLTGSEQTAKINEYKEQDKIRGFDLTRDIPMRAAIFKKAEESFEWVWSYHHIILDGWCFGIVVQDLFKVYNALREQKPYSLPP
VKPYKDYIKWLEKQDKQASLRYWREYLEGFEGQTTFAEQRKKQKDGYEPKELLFSLSEAETKAFTELAKSQHTTLSTALQAVWSVLISR
YQQSGDLAFGTVVSGRPAEIKGVEHMVGLFINVVPRRVKLSEGITFNGLLKRLQEQSLQSEPHQYVPLYDIQSQADQPKLIDHIIVFENYPL
QDAKNEESSENGFDMVDVHVFEKSNYDLNLMASPGDEMLIKLAYNENVFDEAFILRLKSQLLTAIQQLIQNPDQPVSTINLVDDREREF
LLTGLNPPAQAHETKPLTYWFKEAVNANPDAPALTYSGQTLSYRELDEEANRIARRLQKHGAGKGSVVALYTKRSLELVIGILGVLKAG
AAYLPVDPKLPEDRISYMLADSAAACLLTHQEMKEQAAELPYTGTTLFIDDQTRFEEQASDPATAIDPNDPAYIMYTSGTTGKPKGNITT
HANIQGLVKHVDYMAFSDQDTFLSVSNYAFDVFTFDFYASMLNAARLIIADEHTLLDTERLTDLILQENVNVMSATTALFNLLTDAGED
WMKGLRCILFGGERASVPHVRKALRIMGPGKLINAYGPTEGTVFATAHVVHDLPDSISSLPIGKPISNASVYILNEQSQLQPFGAVGELCIS
GMGVSKGYVNRADLTKEKFIENPFKPGETLYRTGDLARWLPDGTIEYAGRIDDQVKIRGHRIELEEIEKQLQEYPGVKDAVVVADRHES
GDASINAYLVNRTQLSAEDVKAHLKKQLPAYMVPQTFTFLDELPLTTNGKVNKRLLPKPDQDQLAEEWIGPRNEMEETIAQIWSEVLG
RKQIGIHDDFFALGGHSLKAMTAASRIKKELGIDLPVKLLFEAPTIAGISAYLKNGGSDGLQDVTIMNQDQEQIIFAFPPVLGYGLMYQNLS
SRLPSYKLCAFDFIEEEDRLDRYADLIQKLQPEGPLTLFGYSAGCSLAFEAAKKLEEQGRIVQRIIMVDSYKKQGVSDLDGRTVESDVEAL
MNVNRDNEALNSEAVKHGLKQKTHAFYSYYVNLISTGQVKADIDLLTSGADFDMPEWLASWEEATTGVYRVKRGFGTHAEMLQGETL
DRNAEILLEFLNTQTVTVS

Supplementary References

1. Land, H. & Humble, M. S. YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. in *Protein Engineering: Methods and Protocols* (eds. Bornscheuer, U. T. & Höhne, M.) 43–67 (Springer New York, 2018). doi:10.1007/978-1-4939-7366-8_4.
2. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, 162–173 (2009).
3. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303 (2018).
4. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science (80-.).* **321**, 659–663 (2008).
5. 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).
6. Khersonsky, O. *et al.* Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **72**, 178-186.e5 (2018).
7. 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).
8. Wilson, D. J. & Aldrich, C. C. A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal. Biochem.* **404**, 56–63 (2010).
9. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (2017).
10. Nath, A. & Atkins, W. M. A Quantitative Index of Substrate Promiscuity †. 157–166 (2008).

5 MANUSCRIPT IV

An Engineered Nonribosomal Peptide Synthetase shows opposite amino acid loading and condensation specificity

Aleksa Stanišić, Annika Hüsken, Philipp Stephan, David L. Niquille, Jochen Reinstein and Hajo Kries

Submitted to: ACS Catalysis and ChemRxiv
doi: https://doi.org/10.26434/chemrxiv.14199695.v1

Summary:
Engineering of NRPSs faces many obstacles, despite being an attractive strategy for generating modified peptides. One of the main issues is the presence of specificity filters at the A- and the C-domain which hinder the incorporation of alternative building blocks. However, the contribution of the C-domain to the peptide formation specificity remains controversial. Here, we take advantage of a dimodular NRPS system with opposite A- and C-domain specificities to determine their relative influence on product formation. We show that A-domain overrules C-domain specificity through dynamic T-domain loading, providing critical insights into A-C interplay during NRPS reaction.

The candidate is

☒ First author ☐ Second author ☐ Corresponding author ☐ Coauthor

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
AS	40 %	65 %	70 %	75 %	
AH	20 %	20 %	25 %	2.5 %	
PS		5 %	5 %	2.5 %	
DN				5 %	
JR		5 %		5 %	
HK	40 %	5 %		10 %	

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

Aleksa Stanišić,[1] Annika Hüsken,[1] Philipp Stephan,[1] David L. Niquille,[2] Jochen Reinstein,[3] Hajo Kries[1]*

[1]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

[2]Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, 500 Technology Square NE47-140, Cambridge, Massachusetts 02139, USA.

[3]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 NRPSs.

Introduction

Compared to the revolutionary advances in the development of antibiotics during the second half of the 20th century, progress in the field has largely stalled for more than 50 years as the infections caused by multidrug resistant bacterial strains increased worldwide.[1] Nonribosomal peptides (NRPs) have long attracted attention due to their impressive structural diversity and antibiotic activities.[2] Nonribosomal peptide synthetases (NRPSs) 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.[3] 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.[4] 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 NRPSs have long inspired engineering attempts aiming at better antibiotics, for instance.[5] 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.[6] As a consequence, several strategies have been tested to edit the specificity of the A-domain.[7–11] Alternatively, domains and modules have been substituted and reshuffled.[12–15] One key emerging issue is the substrate tolerance of follow-up domains after changing the peptide sequence.[16]

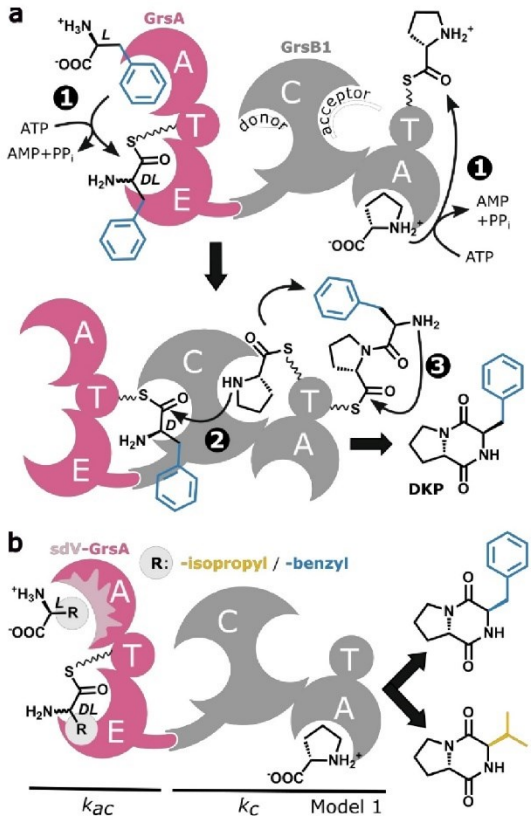


Figure 1. a) Acylation of the T-domain of GrsA (pink) and GrsB1 (grey) with *L*-Phe and *L*-Pro, respectively (step 1). *L*-Phe is racemized at the E-domain of GrsA. Condensation of T-domain loaded *D*-Phe and *L*-Pro occurs in the C-domain of GrsB1 (step 2). Spontaneous cyclization at the T-domain of 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.[17–19] It was also postulated that the C-domain acceptor site shows more stringent proofreading than the donor site.[20–24] 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.[20,23–25] 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.[26] 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.[27] However, a specificity code in the C-domain that would be analogous to the powerful A-domain specificity code describing the substrate binding pocket[28,29] remains elusive. Additionally, it seems that not all C-domains perform stringent proofreading but, in some cases, tolerate alternative substrates well.[30–33] 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.[34] This is in agreement with the outcomes of pyoverdine cluster engineering, which yielded functional chimeras after nonsynonymous A-domain substitutions.[30] 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",[19,35] Kries et al. have generated chimeric initiation module sdV-GrsA$_{ATE}$ (subscript: domain architecture; Figure 1b).[8] 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$_{CAT}$ 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.[25] 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 (*D*F-DKP; indicating chirality and identity of the first amino acid) and *D*-Val-*L*-Pro DKP (*D*V-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 min^{-1} lagging behind adenylation (280 min^{-1}),[36] acylation (500 min^{-1}),[37] and epimerization of *L*-Phe (160 min^{-1}).[37] Consequently, turnover increases with an excess of the second module, as observed in the closely homologous TycA/TycB1 system.[38] 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 ($K_{d,\,app}$) and maximal rate (r_{max}, Table 1). The K_d of 2 to 4 µM measured here is close to that of GrsA and homologous TycB1 (5 µM) obtained through microscale thermophoresis.[39] The enzyme concentration in subsequent peptide formation assays could not always be saturated because prolonged reactions at concentrations above 2.5 µM showed erratic behaviour, presumably due to instability of sdV-GrsA.

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

Product	$K_{d\ app}$ (µM)	r_{max} (min^{-1})
*D*F-DKP	4 ± 1	0.032 ± 0.004
*D*V-DKP	2.0 ± 0.5	0.0088 ± 0.0008

*Apparent K_d and maximal rate (r_{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 µM concentration of one module, we titrated the other. Similar maximum rates for *D*V-DKP formation are reached at a 20-fold excess of sdV-GrsA (0.007 min^{-1}) or GrsB1 (0.014 min^{-1}) relative to the less concentrated module (Figure 2b). Hence, both modules seem to process Val at comparable rates. In contrast, formation of *D*F-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.[40] 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 *D*V-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).[40]

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.

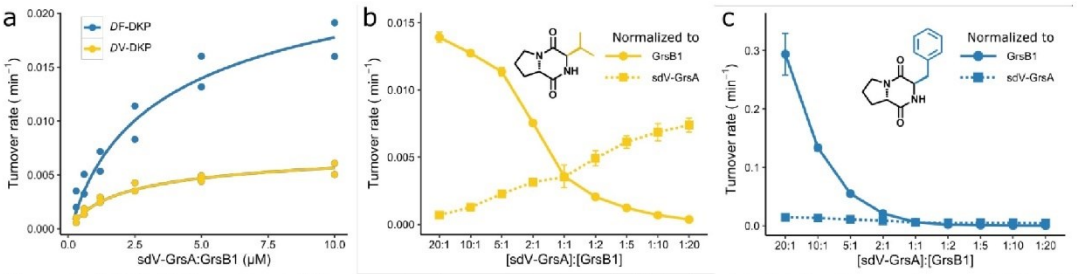


Figure 2. a) *D*V-DKP formation at different equimolar concentrations of sdV-GrsA and GrsB1 (for $K_{d\ app}$ and r_{max} see Table 1). *D*V-DKP (b) and *D*F-DKP (c) formation at different ratios of sdV-GrsA and GrsB1. Measured rates are normalized to GrsB1 (circles, full line) or sdVGrsA (squares, dashed 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)sulfamoyladenosine (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).[36] In the stabilized STAP mutant, the adenylation k_{cat} remains largely unchanged compared to sdV-GrsA (6.0 vs. 8.6 min^{-1}) 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.[6] These affinities were determined with a thermal shift assay[41] 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 *D*V-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 ^{14}C (Figure 3c and Table 2). Compared to sdV-GrsA (0.028 mM^{-1} min^{-1}), acylation with Val has been accelerated to 0.062 mM^{-1} min^{-1} (STAP) and 0.074 mM^{-1} min^{-1} (MS) in the

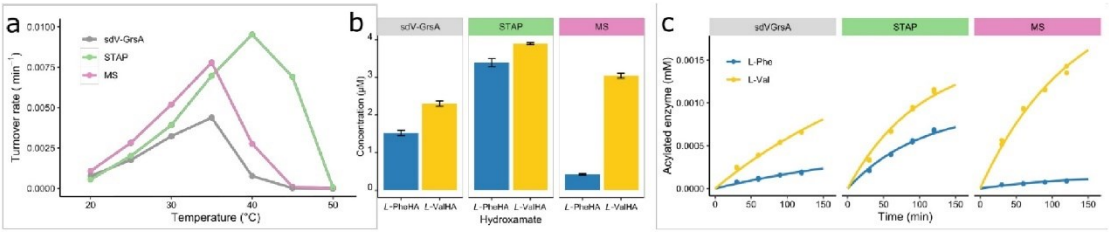


Figure 3. Evolutionary improvement of sdV-GrsA. a) Thermostability of sdV-GrsA and mutants. *D*V-DKP formation rate is measured at 5 µM enzyme and 1 mM amino acid substrates over 30 min at different temperatures. Error bars indicating the standard deviation from two technical replicates are too small to be visible. b) Adenylation specificity of sdV-GrsA and mutants (5 µM) 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 µM) with ^{14}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$		Thermal shift assay§			HAMA*	Acylation#		
	k_{cat} (Val) (min^{-1})	K_M (Val) (mM)	Val-AMS (µM)	Phe-AMS (µM)	T_{m1} (°C)	[ValHA]/ [PheHA]	k_{acV} (mM^{-1} min^{-1})	k_{acF} (mM^{-1} min^{-1})	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

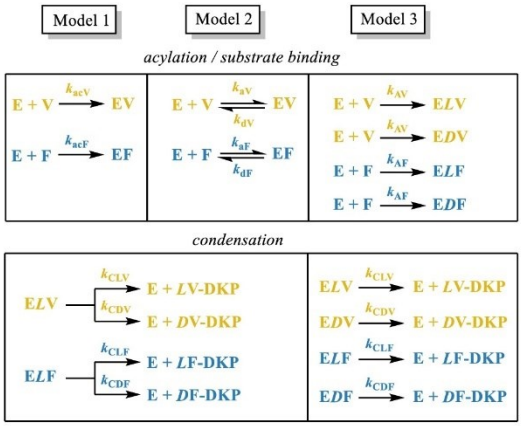
$Kinetic parameters of *L*-Val-adenylation determined with the MESG/NH_2OH assay. Error margins are obtained from a nonlinear fit with technical duplicates to the Michaelis-Menten equation in R. §Dissociation constants of AMS-type inhibitors determined with the thermal shift assay (Supplementary Figures 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). #Experimental acylation rate constants determined with the ^{14}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 *D*F-DKP and acceleration of *D*V-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).[42,43] Progress curves were fit to these models using Dynafit.[44] 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 mM^{-1} 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 *D*F-DKP for-

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

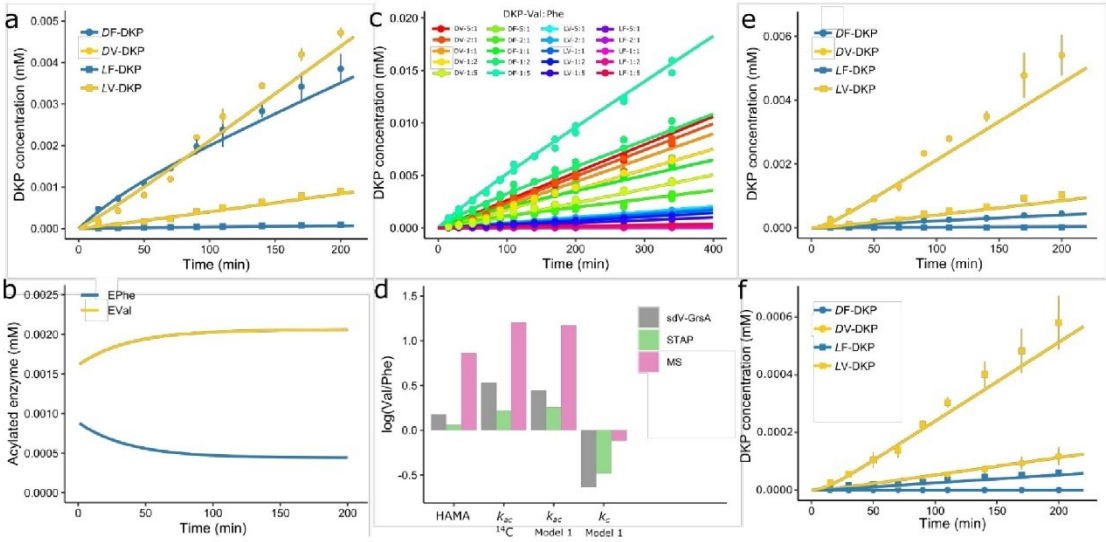


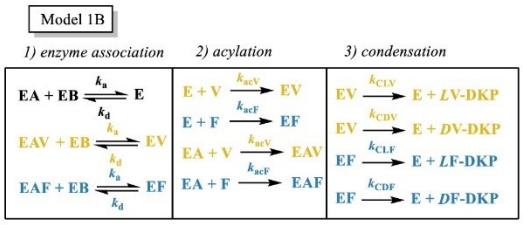
Figure 4. a) Peptide formation reaction with STAP at 1 mM competing *L*-Val and *L*-Phe fit with model 1. b) Concentration change of acylated STAP during the reaction course, as predicted by model 1. c) Global fit of STAP progress kinetic data to model 1 (Table 3). d) Selectivity (Val/Phe) comparison 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 ± 0.01	1.800 ± 0.001	14.49 ± 0.02
	k_{CLV} (min^{-1}) x 10^3	1.3 ±0.1	2.12 ± 0.07	2.01 ± 0.06
	k_{CDV} (min^{-1}) x 10^3	6.8 ± 0.1	11.11 ± 0.08	10.5 ± 0.1
	k_{CLF} (min^{-1}) x 10^3	0.6 ± 0.4	0.6 ± 0.2	0.4 ± 0.4
	k_{CDF} (min^{-1}) x 10^3	29 ±2	33.5 ± 0.5	13.8 ± 0.9
Model 1B	$K_{d\ app}$ (μM)*	2	2	2
	k_{acV}/k_{acF}*	3.4	1.7	14.8
	k_{CLV} (min^{-1}) x 10^3	3.7 ± 0.4	5.4 ± 0.3	5.3 ± 0.2
	k_{CDV} (min^{-1}) x 10^3	19.0 ± 0.5	30.0 ± 0.4	28.1 ± 0.3
	k_{CLF} (min^{-1}) x 10^3	§ND	§ND	5 ± 2
	k_{CDF} (min^{-1}) x 10^3	300 ± 50	89 ± 1	61 ± 7

[#]The values shown are mean and standard deviation of the output of Dynafit's Monte Carlo algorithm. [$]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. §ND: Not determined. Standard errors exceed the value more than 100-fold. Values for k_{CLF} are poorly determined because *L*F-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 *L*F-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,[45] 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 *D*V-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 *L*F-DKP and *L*V-DKP, are now better determined (4.4 ± 0.2 and 2.80 ± 0.02 x 10^{-3} 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.[23] 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.[40] 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 *D*F-DKP at a rate of > 1 min^{-1},[46] while the sdV-GrsA mutants only reach k_{CDF} values between 0.06 and 0.3 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.[47]

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.[48] The only modest differences in condensation rate constants contradict a notion of a previously considered, additional specificity filter.[14,17,23,24] 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 *D*V-DKP becomes the main product late in the reaction. Nevertheless, the product yield

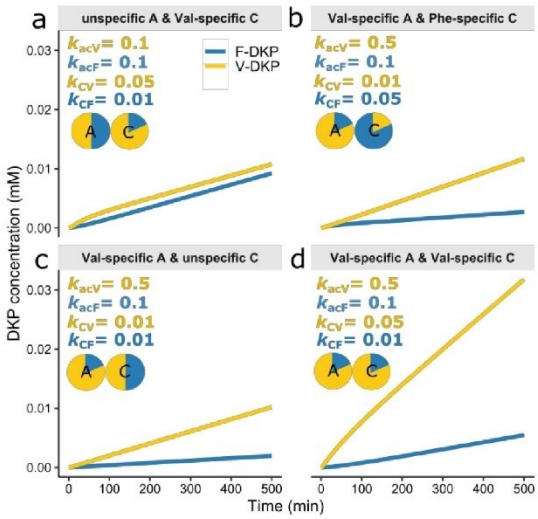


Figure 5. Hypothetical DKP formation by a dimodular NRPS with different combinations of A- and C-domain specificities. The reaction mechanism of model 1 is used to generate simulated progress curves.

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; *D*V-DKP, *D*-Val-*L*-Pro DKP; *D*F-DKP, *D*-Phe-*L*-Pro DKP; *L*V-DKP, *L*-Val-*L*-Pro DKP; *L*F-DKP, *L*-Phe-*L*-Pro DKP; AMS, 5'-O-sulfamoyladenosine; 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).

Supplementary Information

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

Aleksa Stanišić, Annika Hüsken, Philipp Stephan, David L. Niquille, Jochen Reinstein and Hajo Kries*

*hajo.kries@leibniz-hki.de

Independent Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology e.V., Hans Knöll Institute (HKI Jena), Beutenbergstr. 11a, 07745 Jena, Germany

Contents

Protein models

3D-models of sdV-GrsA and mutants were created by first modelling both proteins separately on the X-ray crystal structure of LgrA (PDB: 5ES8)[1] in its thiolation state with the help of SWISS-MODEL (https://swissmodel.expasy.org/).[2,3] Those models were then aligned in PyMOL (https://pymol.org/). Sequence differences between both proteins were highlighted according to the BLOSUM90 matrix using the color_by_mutation script by Christoph Malisi (https://pymolwiki.org/index.php/Color_By_Mutations).

Cloning

General cloning

Cloning was carried out in *E. coli* strain NEB 5-alpha (New England Biolabs). Holo proteins were expressed in *E. coli* strain HM0079.[4] For the purification of plasmid DNA, DNA fragments, and PCR products, NucleoSpin Plasmid and Gel and PCR clean-up kits (Macherey Nagel) were used. DNA fragments were amplified with Q5 polymerase (New England Biolabs, Massachusetts) or Phusion High-Fidelity DNA Polymerase (New England Biolabs), following the supplier's instructions. Assembly of PCR fragments containing vector-specific overhangs and linearized vector was done using the InFusion cloning kit (Takara Bio Europe). Oligonucleotide primers were made by custom synthesis and sequence confirmation of assembled constructs was performed using the Mix2Seq service for Sanger sequencing (Eurofins Genomics).

Plasmids

Cloning of the pSU18-sdV-GrsA constructs

pSU18-*mVGrsA*[5] and pTrc99a-*grsB1*[6] plasmids were kindly provided by Prof. Donald Hilvert (ETH Zurich). Linearization of pSU18 was done with AflII and SacI restriction enzymes. To generate mutants of sdV-GrsA for the directed evolution experiment, two fragments of *mVgrsA* were amplified from pSU18-*mVGrsA* using mutagenic primers and cloned into linearized pSU18-*mVGrsA*. The first fragment was amplified with primer sdXGrsA_f and a suitable reverse primer. The second fragment was amplified with a mutagenic forward primer, e.g. D306S_f, and sdXGrsA_r. The pSU18-*mVgrsA* constructs were created through In-Fusion assembly[7] of two to three fragments with the plasmid backbone. Assembled plasmids were transformed into *E. coli* HST08 Stellar Competent Cells. The identity of the constructs was confirmed by DNA sequencing, before further transforming competent *E. coli* HM0079 for protein expression.

M[a]	Construct	Template	Frag.[b]	Primers[c]	Length (bp)
S1	mVgrsA_D306S	mVgrsA	A	sdXgrsA_f / AH_sdV-GrsA_r	308
			B_a	D306S_f / sdXgrsA_r	272
T	mVgrsA_L308T	mVgrsA	A	sdXgrsA_f / AH_sdV-GrsA_r	308
			B_b	L308T_f / sdXgrsA_r	272
S2	mVgrsA_K311S	mVgrsA	A	sdXgrsA_f / AH_sdV-GrsA_r	308
			B_c	K311S_f / sdXgrsA_r	272
L	mVgrsA_H312L	mVgrsA	A	sdXgrsA_f / AH_sdV-GrsA_r	308
			B_d	H312L_f / sdXgrsA_r	272
K	mVgrsA_N315K	mVgrsA	A	sdXgrsA_f / AH_sdV-GrsA_r	308
			B_e	N315K_f / sdXgrsA_r	272
STSLK	mVgrsA_Ma-e	mVgrsA	A	sdXgrsA_f / AH_sdV-GrsA_r	308
			B_f	STSLK_f / sdXgrsA_r	272
SS1	mVgrsA_G233S _D306S	mVgrsA_D306S	A_1	sdXgrsA_f / AH_sdV-GrsA_2_r	86
			B_g	G233S_f / sdXgrsA_r	491
SS2	mVgrsA_L237S_ D306S	mVgrsA_D306S	A_1	sdXgrsA_f / AH_sdV-GrsA_2_r	86
			B_h	L237S_f / sdXgrsA_r	491
MS	mVgrsA_G243M _D306S	mVgrsA_D306S	A_1	sdXgrsA_f / AH_sdV-GrsA_2_r	86
			B_i	G243M_f / sdXgrsA_r	491
LS	mVgrsA_T255L_ D306S	mVgrsA_D306S	A_2	sdXgrsA_f / AH_sdV-GrsA_3_r	154
			B_j	T255L_f / sdXgrsA_r	425
SI	mVgrsA_D306S_ W326I	mVgrsA_D306S	A_3	sdXgrsA_f / AH_sdV-GrsA_4_r	367
			B_k	W326I_f / sdXgrsA_r	211
ST1	mVgrsA_D306S_ N334T	mVgrsA_D306S	A_4	sdXgrsA_f / AH_sdV-GrsA_5_r	391
			B_l	N334T_f / sdXgrsA_r	186
SA1	mVgrsA_D306S_ S338A	mVgrsA_D306S	A_4	sdXgrsA_f / AH_sdV-GrsA_5_r	391
			B_m	S338A_f / sdXgrsA_r	186
ST2	mVgrsA_D306S_ C340T	mVgrsA_D306S	A_4	sdXgrsA_f / AH_sdV-GrsA_5_r	391
			B_n	C340T_f / sdXgrsA_r	186
SW		mVgrsA_D306S	A_4	sdXgrsA_f / AH_sdV-GrsA_5_r	391

	mVgrsA_D306S_ F341W		B_o	F341W_f / sdXgrsA_r	186
SS3	mVgrsA_D306S_ N350S	mVgrsA_D306S	A_5	sdXgrsA_f / AH_sdV-GrsA_6_r	441
			B_p	N350S_f / sdXgrsA_r	140
SA2	mVgrsA_D306S_ K355A	mVgrsA_D306S	A_5	sdXgrsA_f / AH_sdV-GrsA_6_r	441
			B_q	K355A_f / sdXgrsA_r	144
SP	mVgrsA_D306S_ A356P	mVgrsA_D306S	A_5	sdXgrsA_f / AH_sdV-GrsA_6_r	441
			B_r	A356P_f / sdXgrsA_r	144
MSAP	mVgrsA_G243M _ D306S_S338A_ A356P	mVgrsA_D306S _ A356P	A_1	sdXgrsA_f / AH_sdV-GrsA_2_r	86
			B_2	N334T_f / sdXgrsA_r	186
			C	G243M_f / AH_sdV-GrsA_5_r	325
MSTP	mVgrsA_G243M _ D306S_N334T_ A356P	mVgrsA_D306S _ A356P	A_1	sdXgrsA_f / AH_sdV-GrsA_2_r	86
			B_3	S338A_f / sdXgrsA_r	186
			C	G243M_f / AH_sdV-GrsA_5_r	325
STAP	mVgrsA_D306S_ N334T_S338A_ A356P	mVgrsA_D306S _ A356P	A_4	sdXgrsA_f / AH_sdV-GrsA_5_r	391
			B_4	N334T_S338A_f / sdXgrsA_r	186
SAP	mVgrsA_D306S_ S338A_A356P	mVgrsA_D306S _ A356P	A_4	sdXgrsA_f / AH_sdV-GrsA_5_r	391
			B_2	N334T_f / sdXgrsA_r	186

[a]M: mutant, [b]Frag.: fragment, [c]Sequences can be found in Supplementary Table 3.

Cloning of the pTrc99a-grsB1_corr_CAT construct

The pTrc99a-*grsB1_corr_CAT* construct was cloned based on pTrc99a-*grsB1*.[4] Short sections were removed at the 5' end (30 bp) as well as the 3′ end (153 bp) while retaining the His-tag with a short 12 bp vector derived linker. The removed 3'-section was found to be part of the adjacent module and thus not needed for the expression of GrsB1. The entire corrected *grsB1* frame was PCR amplified as a single fragment using primers GrsB1_pTrc99a_f and GrsB1_pTrc99a_r (Supplementary Table 3) and assembled with a plasmid backbone derived from pTrc99a-*tycB1* by restriction digest with NcoI-HF and BamHI-HF cutting out the complete *tycB1* gene.

Cloning of the pSU18-mVGrsA-MS_Ekn construct

The pSU18-*mVGrsA-MS_Ekn* construct was created by introducing His753Ala mutation, previously shown to abolish epimerization activity.[8] A single DNA fragment was amplified using mutagenic primer GrsA_H753A_f and GrsA_EcoNI_r (Supplementary Table 3) and assembled with a plasmid backbone derived from pSU18-*mVGrsA-MS* by restriction digest with AfeI and EcoNI.

Protein overexpression and purification

Purification protocol

Proteins were overexpressed as C-terminally His_6-tagged holo-NRPS proteins by transforming corresponding pSU18 plasmids into *E. coli* HM0079 with genomically integrated 4'-phosphopantheteinyl transferase Sfp.[4] Saturated *E. coli* culture (0.5 mL) was inoculated in 500 mL of 2xYT medium containing appropriate antibiotic in 2 L shaking flask and incubated at 37 °C in a rotary shaker at 200 rpm. Induction with 0.25 mM isopropyl-D-thiogalactoside (IPTG) was done at OD_{600} = 1 and grown for another 16-20 hours at 20 °C. Cells were harvested by centrifugation and the supernatant was discarded. Cell pellets were resuspended in 30 mL lysis buffer (50 mM TRIS [pH 7.4], 500 mM NaCl, 20 mM imidazole, 2 mM TCEP) and 100 µL of protease inhibitor mix (Sigma, P8849) was added before cell lysis by sonication. The cell debris was cleared by centrifugation at 19,000 g for 30 min at 4 °C and the supernatant was loaded on the open column with 2 mL of Ni-IDA suspension (Rotigarose, Roth) and equilibrated with lysis buffer. Unbound proteins were washed twice with 20 mL of the lysis buffer before the elution with 4 x 0.75 mL elution buffer (50 mM TRIS [pH 7.4], 500 mM NaCl, 300 mM imidazole, 2 mM TCEP). Protein-containing fractions were pooled and the buffer was exchanged with protein storage buffer (50 mM TRIS [pH 7.6], 200 mM NaCl) on 6 mL Vivaspin (Sartorius) filters with 30 kDa cut-off. Glycerol was added to 10% and protein concentration adjusted to 50 µM. Samples were flash frozen in liquid nitrogen and stored at -20 °C. Absorbance at 280 nm measured in Take3 plates on an Epoch2 microplate reader (Biotek) was used for measuring protein concentration, using calculated extinction coefficients (www.benchling.com).

SDS-PAGE of overexpressed proteins

Purity of proteins was determined by SDS-PAGE (Supplementary Figure 10) using Bolt 4-12% Bis-Tris Plus Gels (ThermoFisher Scientific) with MES-SDS running buffer (Novex). Triple Color Protein Standard III (Serva) was run alongside the protein samples as a size standard. The gels were run at 200 V for 22 minutes and stained with Quick Coomassie stain (Serva).

MesG/hydroxylamine spectrophotometric assay

Michaelis-Menten parameters of the adenylation reaction with L-Val were determined using the MesG/hydroxylamine assay.[9] Low activity of the enzymes combined with low solubility of L-Phe did not allow measurement of kinetic parameters for L-Phe. Reactions contained 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 100 µM 7-methylthioguanosine (MesG), 150 mM hydroxylamine (adjusted to pH 7.5-8 with NaOH), 5 mM ATP (A2383, Sigma), 1 mM TCEP, 0.4 U/mL inorganic pyrophosphatase (I1643, Sigma), 1 U/mL of purine nucleoside phosphorylase from microorganisms (N8264, Sigma) and 5 µM of NRPS. Reactions were done on 100 µL scale in flat-bottom 384-well plates (781620, Brand). Reactions were started by addition of substrate and the absorbance was followed at 355 nm on a Synergy H1 (BioTek) microplate reader at 30 °C. Reactions containing buffer without substrate were monitored as a background which was subsequently subtracted. Each substrate concentration was measured in duplicate. Initial velocities (OD min^{-1}) were divided by the slope of a pyrophosphate calibration curve to obtain the pyrophosphate release rate. Initial velocities $v_0/[E_0]$ were fit to the Michaelis-Menten equation by nonlinear regression using RStudio version 1.3.1093 (Supplementary Figure 7).[10]

Thermal shift assay

Thermal shift assays were performed on an Applied Biosystems StepOne Real-Time PCR System using SYPRO Orange (Thermo Fisher Scientific) as fluorescence dye. The assay was carried out using 2 µM enzymes, 0 – 800 µM Phe-AMS, 0 – 400 µM Val-AMS in 50 mM HEPES, 100 mM NaCl and 1mM $MgCl_2$ at pH 8. Val-AMS and Phe-AMS were prepared as 10x concentrated working solutions and SYPRO Orange dye (5000x concentrated) was diluted to a 25x concentrated working solution. The assay was carried out in 20 µl volume using 13 µl of enzyme solution, 2 µl of each inhibitor concentration and 5 µl of fluorescence dye. As negative control, the inhibitor was replaced with buffer.

Temperature was kept at 25 °C for 2 min, increased to 99 °C over 40 min in 1 % increments and maintained at 99 °C for 2 min. All measurements were performed with two biological replicates and three technical replicates. Resulting melting curves were analyzed and the respective melting points calculated using Protein Thermal Shift Software v1.3 (Thermo Fisher Scientific). Shifts in melting points (ΔT_m) for each concentration of inhibitor were calculated by subtracting the melting point of the respective negative control from melting points for each concentration of inhibitor. By plotting ΔT_m against inhibitor concentration using a hyperbolic binding model (Equation 1) K_d values for Val-AMS and Phe-AMS for each enzyme were calculated.

$$\Delta T_m = \frac{[I]}{(K_D + [I])} \qquad (1)$$

Hydroxamate specificity assay (HAMA)

Reaction conditions

The hydroxamate formation assay was conducted at 33 °C as described previously.[11] Reactions of 100 µL contained 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 150 mM hydroxylamine (pH 7.5-8, adjusted with NaOH), 5 mM ATP (A2383, Sigma), 1 mM TCEP and 5 µM of NRPS. Reactions were started by adding a mix of 5 mM proteinogenic amino acids in 100 mM TRIS (pH 8) to a final concentration of 1 mM or only buffer as a control. L-Phe, L-Val and L-Leu were distinguished from D-Phe, D-Val and L-Ile, respectively by using enantiopure, deuterium labelled standards. Reactions were stopped after 3 hours by diluting them 10-fold in acetonitrile containing 0.1 % formic acid and immediately analyzed with UPLC-MS. Time point t_0 was obtained by quenching the enzyme containing master mix before adding amino acid substrates. All assays were done from a single protein batch in technical triplicates.

UPLC-MS/MS conditions

Chromatography was performed on a Waters ACQUITY H-class UPLC system (Waters) with an injection volume of 3 µL. Water with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B) were used as strong and weak eluent, respectively. Separation of amino acid hydroxamates was done on the ACQUITY UPLC BEH Amide column (1.7 µm, 2.1 x 50 mm) with a linear gradient of 10-50% A over 5 min (flow rate 0.4 mL/min) followed by 4 min reequilibration. Data were analyzed with MassLynx and TargetLynx software (version 4.1).

MS/MS detection was performed on Xevo TQ-S micro (Waters) tandem quadrupole instrument with ESI ionisation source in positive ion mode. Nitrogen was used as a desolvation gas and argon as collision gas. The following source parameters were used:

capillary voltage 1.5 kV, cone voltage 65 V, desolvation temperature 500 °C, desolvation gas flow 1000 L/h. Specific mass transitions recorded in multiple reaction monitoring (MRM) mode were used to detect and quantify amino acid hydroxamates (Supplementary Figure 2).[11]

Acylation assay

Acylation of the first module was monitored by measuring the attachment of the ^{14}C labeled substrate to the enzyme as described previously, with minor modifications.[12] Each reaction contained 2.5 µM NRPS, 50 mM TRIS (pH 7.6), 5 mM $MgCl_2$, 5 mM ATP, 1 mM TCEP and 2 mg/mL bovine serum albumin to increase the recovery of precipitated enzyme. Reactions were started by adding substrate mix up to a final concentration of 100 µM. Assay was done in competition conditions, at 100 µM of L-Val and L-Phe in two series, each containing 0.5% of either ^{14}C-L-Phe or ^{14}C-L-Val (0.027 µCi in 500 µL reaction volume). Reactions were quenched after 30-120 min by taking a 100 µL aliquot and transferring to 300 µL of 20% trichloroacetic acid (TCA) to precipitate the protein. After centrifugation at 20 000 g for 15 min, pellets were washed twice with 500 µL of 20% TCA, dissolved in 100 µL of formic acid and assayed by scintillation counting (Perkin Elmer TriCarb 2910 TR). Control reactions were quenched before the addition of substrates. Resulting progress curves were fitted to a bimolecular reaction model with Dynafit. (Supplementary Chapter 10) Single enzyme batch was assayed with two technical replicates.

DKP formation assay

Reaction conditions

The diketopiperazine (DKP) formation assay was performed in 50 µL volume with 5 mM ATP, 1 mM TCEP, and varying concentration of GrsB1 and the first module in peptide formation assay buffer (40 mM HEPES, 10 mM $MgCl_2$, 75 mM NaCl, pH 8.0). Reaction was started by adding the substrate mix containing 1 mM L-Pro and varying L-Val and L-Phe concentrations. Reaction was incubated at 33 °C and quenched by diluting two-fold with water and heat denaturation at 95 °C for 2 min. Denatured proteins were precipitated by centrifugation and the supernatant directly analysed by UPLC-MS/MS. Time point t_0 was obtained by quenching the enzyme containing master mix before adding amino acid substrates. Control reactions contained heat inactivated enzyme.

UPLC-MS/MS conditions

Chromatography was performed on a Waters ACQUITY H-class UPLC system (Waters) with an injection volume of 2 µL. Acetonitrile (A) and water with 0.1 % formic acid (B) were used as strong and weak eluent, respectively. Separation of valine- and phenylalanine-containing diketopiperazines and corresponding diastereomers (L-Val-L-Pro, D-Val-L-Pro, L-Phe-L-Pro and D-Phe-L-Pro) was achieved on the Cortecs UPLC C18 column (1.6 µm, 2.1 x 150 mm) with a linear gradient of 20-95% A over 3 min (flow rate 0.3 mL/min) followed by 0.5 min wash and 2 min reequilibration. Acetonitrile was used as a needle wash solvent between the samples. Data acquisition and quantitation were done using the MassLynx and TargetLynx software (version 4.1).
MS/MS analyses were performed on a Xevo TQ-S micro (Waters) tandem quadrupole instrument with ESI ionisation source in positive ion mode. Nitrogen was used as desolvation gas and argon as collision gas. The following source parameters were used: capillary voltage 0.5 kV, cone voltage 4 V, desolvation temperature 600 °C, desolvation gas flow 1000 L/h. Val-Pro-DKP and Phe-Pro-DKP were detected via the 197.09>69.95 and 245>69.95 transitions, respectively, recorded in multiple reaction monitoring (MRM) mode. Standard calibration solutions of both DKPs were prepared ranging from 0.0006 to 10 µM.

Titration of sdV-GrsA and GrsB1

Ratios

The contribution of the two modules to the overall reaction rate was determined by titrating one module in the presence of fixed concentration of the other and measuring the DKP formation. One enzyme was fixed at the concentration of 0.5 µM while the other was increased from 0.5 (1:1 ratio) up to 10 µM (20:1 ratio). Substrates L-Phe, L-Val and L-Pro were used at 1 mM and both Val-ProDKP and Phe-Pro DKP were measured. Reaction was allowed to run for 60 min at 33 °C. Two different enzyme batches were assayed. Turnover rates were calculated from the DKP concentrations and normalised for both modules individually, according to Equation (2) for the different enzyme ratios.

$$Turnover\ rate = [DKP] * t^{-1} * [E]_0^{-1} \quad (2)$$

Equimolar titration

The concentration at which both modules interact most productively was determined by measuring the DKP formation at different enzyme concentrations, while preserving the 1:1 ratio of the modules. Enzyme concentration was varied from 0.25 to 10 µM. Substrates L-Phe, L-Val and L-Pro were used at 1 mM and both Val-ProDKP and Phe-Pro DKP were measured. Reaction was allowed to run for 60 min at 33 °C. Two different

enzyme batches were assayed. Turnover rates were calculated from the DKP concentrations according to Equation (2). To extract the apparent dissociation constant ($K_{d\ app}$) and maximal velocity (R_{max}), experimental data were fitted to the following bimolecular binding model:

$$R = R_{max} \frac{2C + K_{d\ app} + \sqrt{4K_{d\ app}C + K_{d\ app}{}^2}}{2C} \quad (3)$$

Thermostability of the enzymes

Thermostability of sdV-GrsA, MS and STAP was determined by following the DKP formation at different temperatures. DKP assays were done at 5 µM sdGrsA, GrsB1 and 1 mM L-Val, L-Phe, L-Pro. Reactions were incubated in the range between 20 °C and 50 °C in 5 °C increments and quenched after 30 min. Two different enzyme batches were assayed. Turnover rates were calculated from the DKP concentrations according to Equation 2.

Time courses of DKP formation

ValProDKP and PheProDKP formation over 340 min was followed for the sdV-GrsA, STAP and MS mutant. Reactions were done at 33 °C with 2.5 µM of GrsB1 and the first module. L-Val and L-Phe were added at concentrations of 1-5 mM at five different ratios (5:1, 2:1, 1:1, 1:2, 1:5) to start the reaction. Two different batches of each enzyme were assayed. Reactions were quenched at 13 time points and the DKP concentrations measured.

Stability of enzymes under reaction conditions

The stability of sdV-GrsA, STAP and MS (Figure 1) during the time course was tested by incubating 5 µM of the first module with 5 µM GrsB1, 10 mM ATP, 2 mM TCEP in DKP assay buffer at 33 °C for 6 hours. At seven time points (0-6 h), a 20 µL aliquot was taken and mixed with 20 µL of solution containing 2 mM L-Phe, L-Val and L-Pro. Reaction was allowed to run for 20 min at 33 °C before quenching and UPLC-MS analysis.

Evolution of sdV-GrsA

A common way to look for increased activity in mutant enzymes is microtiter plate screening (MTPS) of large mutant libraries monitored by UPLC-MS/MS. However, having a low activity to begin with, sdV-GrsA/GrsB1 system is not yet sensitive enough to provide reliable and reproducible results. Therefore, we have taken a more focused approach to library design. To keep the library size as small as possible, the smallest amino acid alphabet possible – two per position – can be used and only a small number of residues can be considered. A simpler approach was successfully used by Sun *et al.* to improve the enantioselectivity of limonene epoxide hydrolase.[13] After subdomain swapping, a reasonable amino acid alphabet would only include the wild type and the swapped identity at one position, resulting in binary mutations. Positions to mutate were selected based on structural data. On the one hand, it was considered what is known about interactions of the subdomain with other parts of the NRPS. On the other hand, models of wild type GrsA and sdV-GrsA were overlayed to evaluate differences beside the Stachelhaus code residues.[14] Both enzymes were modelled on the X-ray crystal structure of linear gramicidin synthetase subunit A (LgrA, PDB: 5ES8) in its thiolation state using SWISS-MODEL and subsequently aligned and analysed in PyMOL.[2,3]

Rounds of single mutations were introduced by amplifying gene fragments with mutagenic PCR primers and assembling two to three fragments with a vector backbone via In-Fusion assembly. Protein production was carried out in *Escherichia coli* (*E. coli*) HM0079, a strain containing the PPant-transferase Sfp from *Bacillus subtilis* to release protein in its active *holo*-form.[15,16]

Activity screening was based on measuring VP-DKP formation after 3 h in a non-competitive fashion at 37 °C, the T_{opt} of wild type GrsA. These conditions favor mutants with increased stability and turnover rates for VP-DKP but are insensitive to changes in specificity. Furthermore, the long reaction times make differences more easily detectable. All activities are given relative to wild type sdV-GrsA.

Mutating the interface between A_{NTD} and A_{CTD}

The A domain is divided into two distinct regions, a larger A_{NTD} and a smaller A_{CTD}. Both regions are linked with a short five residue hinge region that enables flexibility necessary for the conformational changes during NRP formation.[1,17] Subdomain swapping changes this interface (Supplementary Figure 3) and thus might especially affect the "closed" state after binding of substrate and ATP, rendering this interface a promising candidate for introducing mutations. Starting close to the binding pocket, five positions were selected to create five sdV-GrsA single mutants (S1, T, S2, L, K) and one mutant combining all five mutations (STSLK, Supplementary Figure 3). According to the model, these

mutations all lie within a loop that might be disordered, because it is poorly predicted and highly diverse between different models.

Testing peptide formation activity revealed that D306S (S1) increased activity by 28% compared to wild type sdV-GrsA, while S2, L and K showed strongly reduced activities (Supplementary Figure 3b). Single mutation L308T (T) and combining all five mutations (STSLK) resulted in complete loss of activity. In this set of positions, reverting back to the identity of GrsA mostly had negative impacts on the activity. However, the D306S mutation seemed to cause a small increase and thus was kept for the next round of mutagenesis.

Second shell mutants

A second round of mutagenesis was focused on interactions between the subdomain and the surrounding GrsA framework in the proximity of the binding pocket. Most of those mutations are located within the second shell, the layer of residues that is in direct contact with residues from the binding pocket, but not with the substrate itself. Mutations at the interface to the GrsA framework were introduced to prevent clashes that could decrease the stability of the enzyme core and thus have a negative effect on enzyme activity. To narrow down the number of positions to mutate in this region, sequence differences between sdV-GrsA and GrsA were weighted according to the Blocks Substitution Matrix 90 (BLOSUM90), a similarity score for local protein alignments of evolutionary closely related proteins, and twelve particularly different positions were chosen (Supplementary Figure 4a). Each mutant also contained D306S, the beneficial mutation from the first round. DKP formation was compared to sdV-GrsA and S1 (Supplementary Figure 4b) which did not show significantly increased activity this time. At equimolar concentrations of both enzyme modules, four mutations were found that improved activity distinctly: G234M, N334T, S338A, A356P. All other mutations showed varying degrees of decreased activity.

Combining mutations

To combine beneficial effects caused by single mutations, four combinations with up to three mutations from the second round and the D306S mutation from the first round were created and analysed (Supplementary Figure 5a). All those mutants contained the A356P mutation that was previously identified to increase activity to 400% (Supplementary Figure 4b). All four mutants exhibited higher activity than sdV-GrsA without mutations, the highest activity being ca. 600% (STAP). In this mutant, all new mutations lie within the same loop of the subdomain (Supplementary Figure 5a). But results also revealed that activities are not directly additive as in that case, STAP would need to show an activity of more than 2600%. Although both mutants that contain G243M and A356P (MSAP and

MSTP) showed higher activities than sdV-GrsA, they were less active than SP with only A356P, indicating negative epistasis. However, this effect was not caused by clashes through direct contact of both amino acids as G243M is located in a helix distant from the other three mutations (Supplementary Figure 5a). Over the course of creating three rounds of mutants, it was found that beneficial mutations also increased the yield of purified protein. Following the same protocol for expression and purification, sdV-GrsA on average yielded 18 mg/L protein. More active mutants all showed higher yields with up to 45 mg/L (MSAP) and with few exceptions, worse mutants all yielded less than 18 mg/L, strongly suggesting that mutations affect the stability of the enzyme.

Dynafit modeling[18–20]

a) Scripts used for the DKP formation progress curves.

Success of the fit between three kinetic models was compared.

```
[task]
       model = Model1 ?
       data = progress discontinuous
       task = fit
      confidence = monte-carlo
[mechanism]
       E + Val ----> EVal    :    kacV
       E + Phe ----> EPhe    :    kacF
       EVal ----> E + LLV    :    kCLLV
       EVal ----> E + DLV    :    kCDLV
       EPhe ----> E + LLF    :    kCLLF
       EPhe ----> E + DLF    :    kCDLF
```

```
[constants] (sdV-GrsA and STAP)    [constants] (MS)
kacV = 0.001 ?                     kacV = 0.01 ?
kacF = 0.001 ?                     kacF = 0.001 ?
kCLLV = 0.001 ??                   kCLLV = 0.001 ??
kCLLF = 0.001 ??                   kCLLF = 0.001 ??
kCDLV = 0.01  ??                   kCDLV = 0.01  ??
kCDLF = 0.01  ??                   kCDLF = 0.01  ??
```

Chemical mechanism and the rate constants. Values for different kinetic constants determine the starting values for the calculation and the question marks indicate that the optimal rate will be identified by the model. Initial acylation constants are set to higher values than condensation. To account for the epimerization step, initial DL-product condensation constants are set to ten-fold higher values. Following set of differential equations is resolved for Model1:

$$\frac{\mathrm{d[E]}}{\mathrm{d}t} = -k_{acV}[\mathrm{E}][\mathrm{Val}] - k_{acF}[\mathrm{E}][\mathrm{Phe}] + kc_{LLV}[\mathrm{EVal}] + kc_{DLV}[\mathrm{EVal}] + kc_{LLF}[\mathrm{EPhe}] + kc_{DLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[Val]}}{\mathrm{d}t} = -k_{acV}[\mathrm{E}][\mathrm{Val}]$$

$$\frac{\mathrm{d[EVal]}}{\mathrm{d}t} = +k_{acV}[\mathrm{E}][\mathrm{Val}] - k_{CLLV}[\mathrm{EVal}] - k_{CDLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d}[Phe]}{\mathrm{d}t} = -k_{acF}[\mathrm{E}][Phe]$$

$$\frac{\mathrm{d[EPhe]}}{\mathrm{d}t} = +k_{acF}[\mathrm{E}][\mathrm{Phe}] - k_{CLLF}[\mathrm{EPhe}] - k_{CDLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[LLVal]}}{\mathrm{d}t} = +k_{CLLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[DLVal]}}{\mathrm{d}t} = +k_{CDLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[LLPhe]}}{\mathrm{d}t} = +k_{CLLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[DLPhe]}}{\mathrm{d}t} = +k_{CDLF}[\mathrm{EPhe}]$$

```
[concentrations]
        E = 0.0025 ; Enzyme concentration in mM
[data]
        directory ./data
monitor EVal, EPhe ; Concentrations of acylated enzyme intermediates over time

mesh from 0 to 500 step 1 ; best-fit model curve
```

The input data is a text file with the first column as a time and additional columns for individual DKP concentrations in mM at different substrate ratios.

```
; P1= LL-VP-DKP, P2= LL-FP-DKP, P3= DL-VP-DKP, P4= DL-FP-DKP ; 11, 21...
are V:F ratios
  column  2 | response P3 = 1 | conc S1 = 1 | conc S2 = 1 | label DLV_11
  column  3 | response P4 = 1 | conc S1 = 1 | conc S2 = 1 | label DLF_11
  column  4 | response P1 = 1 | conc S1 = 1 | conc S2 = 1 | label LLV_11
  column  5 | response P2 = 1 | conc S1 = 1 | conc S2 = 1 | label LLF_11
  column  6 | response P3 = 1 | conc S1 = 5 | conc S2 = 1 | label DLV_51
  column  7 | response P4 = 1 | conc S1 = 5 | conc S2 = 1 | label DLF_51
  column  8 | response P1 = 1 | conc S1 = 5 | conc S2 = 1 | label LLV_51
  column  9 | response P2 = 1 | conc S1 = 5 | conc S2 = 1 | label LLF_51
  column  10 | response P3 = 1 | conc S1 = 2 | conc S2 = 1 | label DLV_21
  column  11 | response P4 = 1 | conc S1 = 2 | conc S2 = 1 | label DLF_21
  column  12 | response P1 = 1 | conc S1 = 2 | conc S2 = 1 | label LLV_21
  column  13 | response P2 = 1 | conc S1 = 2 | conc S2 = 1 | label LLF_21
```

```
  column  14 | response P3 = 1 | conc S1 = 1 | conc S2 = 2 | label DLV_12
  column  15 | response P4 = 1 | conc S1 = 1 | conc S2 = 2 | label DLF_12
  column  16 | response P1 = 1 | conc S1 = 1 | conc S2 = 2 | label LLV_12
  column  17 | response P2 = 1 | conc S1 = 1 | conc S2 = 2 | label LLF_12
  column  18 | response P3 = 1 | conc S1 = 1 | conc S2 = 5 | label DLV_15
  column  19 | response P4 = 1 | conc S1 = 1 | conc S2 = 5 | label DLF_15
  column  20 | response P1 = 1 | conc S1 = 1 | conc S2 = 5 | label LLV_15
  column  21 | response P2 = 1 | conc S1 = 1 | conc S2 = 5 | label LLF_15
 [output]

directory ./Results/
[settings]

{Output}
       XAxisLabel = time (min)
       YAxisLabel = DKP (mM)
```

Second model with the substrate binding equilibrium.

```
[task]
       model = Model2 ?
       data = progress discontinuous
       task = fit
      confidence = monte-carlo


[mechanism]
       E + Val <====> EVal    :    kaV  kdV
       E + Phe <====> EPhe    :    kaF  kdF
       EVal ----> E + LLV    :    kCLLV
       EVal ----> E + DLV    :    kCDLV
       EPhe ----> E + LLF    :    kCLLF
       EPhe ----> E + DLF    :    kCDLF
```

Association of substrate with the enzyme is diffusion controlled, so association constants are fixed to a high value.

```
[constants]
         kaV = 1000000
        kaF = 1000000
        kdV = 1  ??
        kdF = 1  ??
       kCLLV = 0.001 ??
       kCLLF = 0.001 ??
       kCDLV = 0.01  ??
       kCDLF = 0.01  ??
```

Following set of differential equations is resolved for Model 2:

$$\frac{\mathrm{d[E]}}{\mathrm{d}t} = -k_{aV}[\mathrm{E}][\mathrm{Val}] + k_{dV}[\mathrm{EVal}] - k_{aF}[\mathrm{E}][\mathrm{Phe}] + k_{dF}[\mathrm{EPhe}] + k_{CLLV}[\mathrm{EVal}] + k_{CDLV}[\mathrm{EVal}] + k_{CLLF}[\mathrm{EPhe}] + k_{CDLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[Val]}}{\mathrm{d}t} = -k_{aV}[\mathrm{E}][\mathrm{Val}] + k_{dV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[EVal]}}{\mathrm{d}t} = +k_{aV}[\mathrm{E}][\mathrm{Val}] - k_{dV}[\mathrm{EVal}] - k_{CLLV}[\mathrm{EVal}] - k_{CDLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[Phe]}}{\mathrm{d}t} = -k_{aF}[\mathrm{E}][\mathrm{Phe}] + k_{dF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[EPhe]}}{\mathrm{d}t} = +k_{aF}[\mathrm{E}][Phe] - k_{dF}[\mathrm{EPhe}] - k_{CLLF}[\mathrm{EPhe}] - k_{CDLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[LLVal]}}{\mathrm{d}t} = +k_{CLLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[DLVal]}}{\mathrm{d}t} = +k_{CDLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[LLPhe]}}{\mathrm{d}t} = +k_{CLLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[DLPhe]}}{\mathrm{d}t} = +k_{CDLF}[\mathrm{EPhe}]$$

Third model with the integrated acylation and epimerization step.

```
[task]
       model = Model3 ?
       data = progress discontinuous
       task = fit
       confidence = monte-carlo
[mechanism]
E + Val ----> ELVal : kAVal
E + Val ----> EDVal : kAVal
E + Phe ----> ELPhe : kAPhe
E + Phe ----> EDPhe : kAPhe
ELVal ----> E1 + LLV : kCLLV
ELPhe ----> E1 + LLF : kCLLF
EDVal ----> E1 + DLV : kCDLV
EDPhe ----> E1 + DLF : kCDLF
```

```
[constants] (sdV-GrsA and STAP)  [constants] (MS)
kAVal = 0.001 ?                  kAVal = 0.01 ?
kAPhe = 0.001 ?                  kAPhe = 0.001 ?
kCLLV = 0.001 ??                 kCLLV = 0.001 ??
kCLLF = 0.001 ??                 kCLLF = 0.001 ??
kCDLV = 0.01  ??                 kCDLV = 0.01  ??
kCDLF = 0.01  ??                 kCDLF = 0.01  ??
```

Following set of differential equations is resolved for Model 3:

$$\frac{\mathrm{d[E]}}{\mathrm{d}t} = -k_{AVal}[\mathrm{E}][\mathrm{Val}] - k_{AVal}[\mathrm{E}][\mathrm{Val}] - k_{APhe}[\mathrm{E}][\mathrm{Phe}] - k_{APhe}[\mathrm{E}][\mathrm{Phe}]$$

$$\frac{\mathrm{d[Val]}}{\mathrm{d}t} = -k_{AVal}[\mathrm{E}][\mathrm{Val}] - k_{AVal}[\mathrm{E}][\mathrm{Val}]$$

$$\frac{\mathrm{d[ELVal]}}{\mathrm{d}t} = +k_{AVal}[\mathrm{E}][\mathrm{Val}] - k_{CLLV}[\mathrm{ELVal}]$$

$$\frac{\mathrm{d[EDVal]}}{\mathrm{d}t} = +k_{AVal}[\mathrm{E}][\mathrm{Val}] - k_{CDLV}[\mathrm{EDVal}]$$

$$\frac{\mathrm{d[Phe]}}{\mathrm{d}t} = -k_{APhe}[\mathrm{E}][\mathrm{Phe}] - k_{APhe}[\mathrm{E}][\mathrm{Phe}]$$

$$\frac{\mathrm{d[ELPhe]}}{\mathrm{d}t} = +k_{APhe}[\mathrm{E}][\mathrm{Phe}] - k_{CLLF}[\mathrm{ELPhe}]$$

$$\frac{\mathrm{d[EDPhe]}}{\mathrm{d}t} = +k_{APhe}[\mathrm{E}][\mathrm{Phe}] - k_{CDLF}[\mathrm{EDPhe}]$$

$$\frac{\mathrm{d[LLVal]}}{\mathrm{d}t} = +k_{CLLV}[\mathrm{ELVal}]$$

$$\frac{\mathrm{d[DLVal]}}{\mathrm{d}t} = +k_{CDLV}[\mathrm{EDVal}]$$

$$\frac{\mathrm{d[LLPhe]}}{\mathrm{d}t} = +k_{CLLF}[\mathrm{ELPhe}]$$

$$\frac{\mathrm{d[DLPhe]}}{\mathrm{d}t} = +k_{CDLF}[\mathrm{EDPhe}]$$

```
[end]
```

b) Script for model 1B with the integrated module association and acylation step

```
[task]
       model = Model1B
       data = progress discontinuous
       task = fit
      confidence = monte-carlo
[mechanism]
       EA + EB <===> E      :     ka  kd
       EAVal + EB <===> EVal :    ka  kd
       EAPhe + EB <===> EPhe :    ka  kd
       E + Val ----> EVal    :    kacV
       E + Phe ----> EPhe    :    kacF
       EA + Val ----> EAVal  :    kacV
       EA + Phe ----> EAPhe  :    kacF
       EVal ----> E + LLV    :    kCLLV
       EVal ----> E + DLV    :    kCDLV
       EPhe ----> E + LLF    :    kCLLF
       EPhe ----> E + DLF    :    kCDLF
```

```
[constants](sdV-GrsA)
ka = 10000
kd = 20
kacV = 0.028
kacF = 0.0082
kCLLV = 0.01 ?
kCLLF = 0.01 ?
kCDLV = 0.001 ?
kCDLF = 0.001 ?
```

```
[constants](STAP)
ka = 10000
kd = 20
kacV = 0.062
kacF = 0.037
kCLLV = 0.01 ?
kCLLF = 0.01 ?
kCDLV = 0.001 ?
kCDLF = 0.001 ?
```

```
[constants](MS and MS_Eko)
ka = 10000
kd = 20
kacV = 0.074
kacF = 0.005
kCLLV = 0.01 ?
kCLLF = 0.01 ?
kCDLV = 0.001 ?
kCDLF = 0.001 ?
```

$$\frac{\mathrm{d[EA]}}{\mathrm{d}t} = -k_a[\mathrm{EA}][EB] + k_d[\mathrm{E}] - k_{acV}[\mathrm{EA}][\mathrm{Val}] - k_{acF}[\mathrm{EA}][\mathrm{Phe}]$$

$$\frac{\mathrm{d[EB]}}{\mathrm{d}t} = -k_a[\mathrm{EA}][EB] + k_d[\mathrm{E}] - k_a[\mathrm{EB}][EA\mathrm{Val}] + k_d[\mathrm{EVal}] - k_a[\mathrm{EB}][EA\mathrm{Phe}] + k_d[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[E]}}{\mathrm{d}t} = +k_a[\mathrm{EA}][EB] - k_d[\mathrm{E}] - k_{acV}[\mathrm{E}][\mathrm{Val}] - k_{acF}[\mathrm{E}][\mathrm{Phe}] + k_{CLLV}[\mathrm{EVal}] + k_{CDLV}[\mathrm{EVal}] + k_{CLLF}[\mathrm{EPhe}] + k_{CDLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[EAVal]}}{\mathrm{d}t} = -k_a[\mathrm{EB}][\mathrm{EAVal}] + k_d[\mathrm{EVal}] + k_{acV}[\mathrm{EA}][\mathrm{Val}]$$

$$\frac{\mathrm{d[EVal]}}{\mathrm{d}t} = +k_a[\mathrm{EB}][\mathrm{EAVal}] - k_d[\mathrm{EVal}] + k_{acV}[\mathrm{E}][\mathrm{Val}] - k_{CLLV}[\mathrm{EVal}] - k_{CDLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[EAPhe]}}{\mathrm{d}t} = -k_a[\mathrm{EB}][\mathrm{EAPhe}] + k_d[\mathrm{EPhe}] + k_{acF}[\mathrm{EA}][\mathrm{Phe}]$$

$$\frac{\mathrm{d[EPhe]}}{\mathrm{d}t} = +k_a[\mathrm{EB}][\mathrm{EAPhe}] - k_d[\mathrm{EPhe}] + k_{acF}[\mathrm{E}][\mathrm{Phe}] - k_{CLFV}[\mathrm{EPhe}] - k_{CDLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[Val]}}{\mathrm{d}t} = -k_{acV}[\mathrm{E}][Val] - k_{acV}[\mathrm{EA}][Val]$$

$$\frac{\mathrm{d[Phe]}}{\mathrm{d}t} = -k_{acF}[\mathrm{E}][\mathrm{Phe}] - k_{acF}[\mathrm{EA}][\mathrm{Phe}]$$

$$\frac{\mathrm{d[LLVal]}}{\mathrm{d}t} = +k_{CLLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[DLVal]}}{\mathrm{d}t} = +k_{CDLV}[\mathrm{EVal}]$$

$$\frac{\mathrm{d[LLPhe]}}{\mathrm{d}t} = +k_{CLLF}[\mathrm{EPhe}]$$

$$\frac{\mathrm{d[DLPhe]}}{\mathrm{d}t} = +k_{CDLF}[\mathrm{EPhe}]$$

```
[end]
```

c) Script for the experimental acylation progress curves

```
[task]
       model = Acylation
       data = progress discontinuous
       task = fit
      confidence =     monte-carlo
[mechanism]
       E + Val ----> EVal    :    kacV
       E + Phe ----> EPhe    :    kacF
[constants]
       kacV = 1 ??
       kacF = 1 ??
[concentrations]
       E = 0.0025 ; mM
[data]
       directory ./data
mesh from 0 to 150 step 1 ; best-fit model curve
; ES1= E-Val, ES2= E-Phe

column2|response ES1 = 1|conc S1 = 0.1 | conc S2 = 0.1 | label E-Val
column3|response ES2 = 1| conc S1 = 0.1 | conc S2 = 0.1 | label E-Phe
[output]
       directory ./Results/
[settings]
```

```
{Output}
       XAxisLabel = time (min)
       YAxisLabel = Acylated enzyme (mM)
[end]
```

d) Simulation of hypothetical two-module NRPS system progress curves

```
[task]
       task = simulate
       data = progress
[mechanism]
       E + S1 ----> ES1    :    kacV
       E + S2 ----> ES2    :    kacF
       ES1 ----> E + P3    :    kCV
       ES2 ----> E + P4    :    kCF

  [constants]     [constants]     [constants]     [constants]
  (unspecA        (ValspecA       (ValspecA       (ValspecA
  +               +               +               +
  ValspecC)       PhespecC)       unspecC)        ValspecC)

  kacV = 0.1      kacV = 0.5      kacV = 0.5      kacV = 0.5
  kacF = 0.1      kacF = 0.1      kacF = 0.1      kacF = 0.1
  kCV = 0.05      kCV = 0.01      kCV = 0.01      kCV = 0.05
  kCF = 0.01      kCF = 0.05      kCF = 0.01      kCF = 0.01


[concentrations]
       E = 0.0025 ; mM! also substrates are in mM
[data]
       mesh from 0 to 500 step 1 ; best-fit model curve
       ;error constant 1 percent
       directory ./data sheet

; P1= LL-VP-DKP, P2= LL-FP-DKP, P3= VP-DKP, P4= FP-DKP ; 11, 21... are
V:F ratios

column    2 | response P3 = 1 | conc S1 = 1 | conc S2 = 1 | label DLV_11
column    3 | response P4 = 1 | conc S1 = 1 | conc S2 = 1 | label DLF_11
[output]
       directory ./Results/
[settings]
{Output}
       XAxisLabel = time (min)
       YAxisLabel = DKP (mM)
[end]
```

Synthesis of L-Val and L-Phe AMS inhibitors

Analytics

NMR spectra were recorded in deuterated solvents (Carl Roth, Germany) on a Bruker AVANCE II 300 or Bruker AVANCE III 500MHz spectrometer, equipped with a Bruker Cryoplatform. The chemical shifts (δ) are reported in parts per million (ppm) relative to the solvent residual peak of DMSO-d6(1H: 2.50ppm, quintet; 13C: 39.5ppm, heptet). All reagents used were reagent grade and used as supplied (purchased from Sigma-Aldrich, Bachem, Fluorochem or Carl Roth). Reactions were performed at ambient temperature under argon atmosphere in anhydrous solvents (Acros Organics) unless otherwise stated. Analytical thin-layer chromatography was performed on silica 60 F254plates (0.25mm, Merck). Compounds were visualized by dipping the plates in a ninhydrin/acetic acid solution followed by heating.

Synthesis of 2,3-O-Isopropylidenadenosine

Adenosine (**1**, 3 g, 11.2 mmol) and TsOH (2.3 g, 13.6 mmol) were dissolved in acetone (120 ml). 2,2-Dimethoxypropane (5.6 ml, 45.6 mmol) was added and the reaction stirred at room temperature for 3 days. Afterwards, it was neutralized with saturated $NaHCO_3$ solution (100 ml) and extracted with DCM (3x 50 ml). The combined organic phases were washed with brine (50 ml), dried over Na_2SO_4 and volatiles removed under vacuum. Recrystallization of the organic residue from EtOH/EtOAc (9:1) resulted in 2,3-O-isopropylidenadenosine **2** (3.1 g, 10 mmol, 90% yield) as white crystals.
^{1}H NMR (500 MHz, DMSO) δ 8.33 (s, 1H), 8.14 (d, *J* = 5.8 Hz, 1H), 7.37 (s, 2H), 6.11 (d, *J* = 3.1 Hz, 1H), 5.33 (dd, *J* = 6.1, 3.1 Hz, 1H), 5.25 (s, 1H), 4.95 (dd, *J* = 6.1, 2.5 Hz, 1H), 4.20 (td, *J* = 4.8, 2.6 Hz, 1H), 3.60 – 3.46 (m, 2H), 1.53 (s, 3H), 1.31 (s, 3H).

Synthesis of sulfamoyl chloride

Chlorosulfonyl isocyanate (**3**, 6 ml, 67.6 mmol) was placed in an oven dried schlenk tube under N_2 atmosphere and dissolved in dry DCM (6 ml) at 0 °C (ice). Formic acid (3 ml, 81.2 mmol) was added drop wise over 10 min. The reaction became a white suspension with a strong gas evolution. The ice bath was removed and the reaction stirred at room temperature until gas evolution stopped. Then, the reaction mixture was placed at -20 °C for 4 h. Afterwards, the supernatant was decanted, and the residue redissolved in DCM and placed again at – 20 °C for 4 h. The supernatant was discarded again. Drying of the organic residue under vacuum resulted in sulfamoyl chloride **4** (3.5 g, 30 mmol, 45% yield) as white crystals that were used for the following steps without further purification.

Synthesis of Sulfamoyl-Isopropylidenadenosine

2 (1 g, 3.3 mmol) was dissolved in dry DME at 0 °C under N_2 atmosphere. NaH as 60% suspension in mineral oil (117 mg, 4.9 mmol) was added and the reaction stirred at 0 °C for 30 min. Then, **4** (565 mg, 4.9 mmol) dissolved in dry DME (15 ml) was added drop wise. The reaction was allowed to warm to room temperature and stirred for 3 days. Afterwards, the reaction was quenched with saturated $NaHCO_3$ solution (30 ml) and extracted with EtOAc (4x 40 ml). The combined organic phases were washed with brine (100 ml), dried over Na_2SO_4 and concentrated under vacuum. Purification by column chromatography (silica 60, DCM/MeOH 9:1) resulted in sulfamoyl-isopropylidenadenosine **5** (921 mg, 2.4 mmol, 73% yield) as a colourless solid.
^{1}H NMR (500 MHz, DMSO-d_6): δ 8.29 (s, 1H), 8.16 (s, 1H), 7.57 (br s, 2H), 7.33 (br s, 2H), 6.22 (d, J = 2.2 Hz, 1H), 5.42 (dd, J =6.3 Hz, 2.4 Hz, 1H), 5.07 (q, J = 3.1 Hz, 1H), 4.43 - 4.34 (m, 1H), 4.24 (q, J = 5.3 Hz, 1H), 4.11 (dd, J = 10.6 Hz, 6.4 Hz, 1H), 1.54 (s, 3H), 1.33 (s, 3H).
HPLC-MS: m/z = 385.10 $[M-H]^-$

Synthesis of Boc-Phe-OSu

L-Boc-Phe-OH (**6**, 2 g, 7.5 mmol) was dissolved in dry THF (40 ml) at 0 °C under N_2 atmosphere. NHS (874 mg, 7.5 mmol) and DCC (1.6 g, 7.5 mmol) were added and the reaction stirred at 0 °C for 30 min followed by 16 h at room temperature. Afterwards the reaction mixture was filtered. Concentration of the filtrate under vacuum resulted in Boc-Phe-OSu **7** (3 g crude yield) which was used without further purification.
^{1}H NMR (300 MHz, DMSO-d_6): δ 7.64 (d, J = 8.2 Hz, 1H), 7.39 - 7.18 (m, 5H), 4.62 - 4.47 (m, 1H), 3.11 - 2.91 (m, 2H), 2.81 (s, 4H), 1.30 (s, 9H).

Synthesis of Boc-Val-OSu

L-Boc-Val-OH (**8**, 2 g, 9.2 mmol) was dissolved in dry THF (40 ml) at 0 °C under N_2 atmosphere. NHS (1.1 g, 9.2 mmol) and DCC (1.9 g, 9.2 mmol) were added and the reaction stirred at 0 °C for 30 min followed by 16 h at room temperature. Afterwards, the reaction mixture was filtered. Concentration of the filtrate under vacuum resulted in Boc-Phe-OSu **9** (3 g crude yield) which was used without further purification
^{1}H NMR (500 MHz, DMSO): δ 7.58 (d, J = 8.2 Hz, 1H), 4.20 (dd, J = 8.1, 6.7 Hz, 1H), 2.79 (br s, 4H), 2.18 – 2.03 (m, 1H), 1.39 (s, 9H), 1.00 (d, J = 6.7 Hz, 3H), 0.97 (d, J = 6.8 Hz, 3H).

Synthesis of Boc-Phe-AMS

5 (150 mg, 0.39 mmol), **7** (140 mg, 0.39 mmol) and Cs_2CO_3 (140 mg, 0.43 mmol) were dissolved in dry DMF (3.8 ml) under N_2 atmosphere. Reaction was stirred at room temperature for 16 h. Afterwards, volatiles were removed under vacuum and remains redissolved in EtOAc (10 ml). Insoluble parts were filtered off and the filtrate washed with brine (2x 5 ml), dried over Na_2SO_4 and concentrated under vacuum. Purification by column chromatography (silica 60, EtOAc/EtOH 8:2) resulted in Boc-Phe-AMS **10** (166 mg, 0.26 mmol, 67% yield).

^{1}H NMR (300 MHz, DMSO) δ 8.39 (s, 1H), 8.14 (s, 1H), 7.33 (s, 2H), 7.24 – 7.08 (m, 5H), 6.14 (d, *J* = 2.9 Hz, 1H), 6.09 (d, *J* = 8.2 Hz, 1H), 5.34 (dd, *J* = 5.9, 3.0 Hz, 1H), 4.99 (dd, *J* = 6.2, 2.1 Hz, 1H), 4.38 (d, *J* = 2.1 Hz, 1H), 4.08 – 3.77 (m, 3H), 3.03 (dd, *J* = 13.5, 4.3 Hz, 1H), 2.77 (dd, *J* = 13.5, 8.3 Hz, 1H), 1.53 (s, 3H), 1.31 (s, 3H), 1.27 (d, *J* = 5.2 Hz, 9H).

HPLC-MS: *m/z* = 634.35 $[M+H]^+$

Synthesis of Boc-Val-AMS

5 (150 mg, 0.39 mmol), **9** (135 mg, 0.43 mmol) and Cs_2CO_3 (140 mg, 0.43 mmol) were dissolved in dry DMF (3.8 ml) under N_2 atmosphere. Reaction was stirred at room temperature for 18 h. Afterwards, volatiles were removed under vacuum and remains redissolved in EtOAc (10 ml). Insoluble parts were filtered off and the filtrate washed with brine (2x 5 ml), dried over Na_2SO_4 and concentrated under vacuum. Purification by column chromatography (silica 60, DCM/MeOH 85:15) resulted in Boc-Val-AMS **11** (118 mg, 0.23 mmol, 55% yield).

^{1}H NMR (300 MHz, DMSO) δ 8.39 (s, 1H), 8.15 (s, 1H), 7.33 (s, 2H), 6.14 (d, *J* = 3.0 Hz, 1H), 5.76 (d, *J* = 8.6 Hz, 1H), 5.33 (dd, *J* = 6.1, 3.1 Hz, 1H), 5.06 – 4.92 (m, 1H), 4.43 – 4.34 (m, 1H), 4.00 (d, *J* = 6.8 Hz, 2H), 3.61 (dd, *J* = 8.5, 4.7 Hz, 1H), 1.54 (s, 3H), 1.38 (s, 3H), 1.35 (s, 9H), 0.88 – 0.76 (m, 6H).

HPLC-MS: *m/z* = 586.27 $[M+H]^+$

Synthesis of Phe-AMS

To **10** (69 mg, 0.1 mmol) was added 5:1 TFA/H_2O (4 ml) and the reaction stirred at room temperature for 30 min. Afterwards, volatiles were removed under vacuum. Purification by column chromatography (silica 60, EtOAc/MeOH/TEA 65:35:1) resulted in Phe-AMS **12** (42 mg, 0.09 mmol, 90% yield) as triethylammonium salt.

^{1}H NMR (300 MHz, DMSO) δ 8.39 (s, 1H), 8.14 (s, 1H), 7.33 – 7.17 (m, 6H), 5.93 (d, *J* = 5.7 Hz, 1H), 4.61 (t, *J* = 5.3 Hz, 1H), 4.23 – 4.07 (m, 3H), 4.04 (dd, *J* = 9.7, 3.5 Hz,

1H), 3.65 (dd, *J* = 7.4, 5.1 Hz, 1H), 3.14 (dd, *J* = 13.4, 4.1 Hz, 1H), 2.92 (dd, *J* = 14.1, 7.7 Hz, 1H).

^{13}C NMR (75 MHz, DMSO) δ 172.3, 156.0, 152.6, 149.5, 139.4, 136.3, 129.5, 128.3, 126.6, 118.9, 87.1, 82.4, 73.4, 70.7, 67.5, 56.3, 37.4.

HPLC-MS: *m/z* = 494.24 $[M+H]^+$

Synthesis of Val-AMS

To **11** (40 mg, 0.07 mmol) was added 5:1 TFA/H_2O (2 ml) and the reaction stirred at room temperature for 30 min. Afterwards, volatiles were removed under vacuum. Purification by column chromatography (silica 60, EtOAc/MeOH/TEA 65:35:1) resulted in Val-AMS **13** (22 mg, 0.05 mmol, 70% yield) as triethylammonium salt.

^{1}H NMR (300 MHz, DMSO) δ = 8.39 (s, 1H), 8.14 (s, 1H), 7.29 (s, 2H), 5.91 (d, *J* = 5.8 Hz, 1H), 4.61 (t, *J* = 5.3 Hz, 1H), 4.21 – 4.06 (m, 3H), 3.26 (d, *J* = 4.1 Hz, 1H), 2.25 – 2.05 (m, 1H), 0.97 – 0.87 (m, 6H).

^{13}C NMR (75 MHz, DMSO) δ = 172.3, 156.0, 152.7, 149.6, 139.5, 119.0, 87.1, 82.5, 73.5, 70.8, 67.5, 60.3, 29.8, 18.7, 17.4.

HPLC-MS: *m/z* = 446.20 $[M+H]^+$

Supplementary Tables

Supplementary Table 1. Rate constants from Model 1 and 2.

Model 1				Model 2			
sdV-GrsA				sdV-GrsA			
Parameter	Initial	Final	Std. error	Parameter	Initial	Final	Std. error
kacV	0.001	19000	7.50×10^{6}	kacV	1	1.00×10^{-9}	0.00014
kacF	0.001	7000	2.80×10^{6}	kacF	1	0.053	0.0052
kCLLV	0.001	0.0013	0.00011	kCLLV	0.001	0.0013	0.00011
kCDLV	0.01	0.0068	0.00013	kCDLV	0.01	0.0069	0.00013
kCLLF	0.001	0.00060	0.00042	kCLLF	0.001	0.00047	0.00039
kCDLF	0.01	0.029	0.0017	kCDLF	0.01	0.027	0.0014

STAP				STAP			
Parameter	Initial	Final	Std. error	Parameter	Initial	Final	Std. error
kacV	0.001	3100	7.60×10^{5}	kacV	1	0.015	0.0034
kacF	0.001	1700	4.20×10^{5}	kacF	1	0.088	0.014
kCLLV	0.001	0.0021	6.60×10^{-5}	kCLLV	0.001	0.0021	6.6×10^{-5}
kCDLV	0.01	0.011	8.30×10^{-5}	kCDLV	0.01	0.011	8.3×10^{-5}
kCLLF	0.001	0.00063	0.00016	kCLLF	0.001	0.00065	0.00016
kCDLF	0.01	0.033	0.00047	kCDLF	0.01	0.034	0.00049

MS				MS			
Parameter	Initial	Final	Std. error	Parameter	Initial	Final	Std. error
kacV	0.01	1300	1.1×10^{6}	kacV	1	0.0052	0.0047
kacF	0.001	88	72000	kacF	1	0.33	0.072
kCLLV	0.001	0.002	5.7×10^{-5}	kCLLV	0.001	0.002	5.5×10^{-5}
kCDLV	0.01	0.01	9.5×10^{-5}	kCDLV	0.01	0.011	7.7×10^{-5}
kCLLF	0.001	0.00039	0.00042	kCLLF	0.001	0.00022	0.00039
kCDLF	0.01	0.014	0.00096	kCDLF	0.01	0.013	0.00076

*Undefined parameters are shaded in grey. Monomolecular and bimolecular rate constants are given in units of min^{-1} and mM^{-1} min^{-1}, respectively

Supplementary Table 2. Rate constants from Model 1B.

Model 1B			
sdV-GrsA			
Parameter	Initial	Final	Std. error
kCLLV	0.002	0.0037	0.00037
kCDLV	0.0012	0.019	0.00046
kCLLF	0.014	4.1×10^{-11}	3.3×10^{-8}
kCDLF	0.06	0.3	0.052

STAP			
Parameter	Initial	Final	Std. error
kCLLV	0.002	0.0053	0.00028
kCDLV	0.0012	0.03	0.00035
kCLLF	0.014	7.1×10^{-8}	0.00013
kCDLF	0.06	0.089	0.0014

MS			
Parameter	Initial	Final	Std. error
kCLLV	0.002	0.0053	0.00019
kCDLV	0.0012	0.028	0.00025
kCLLF	0.014	0.0048	0.0021
kCDLF	0.06	0.061	0.0066

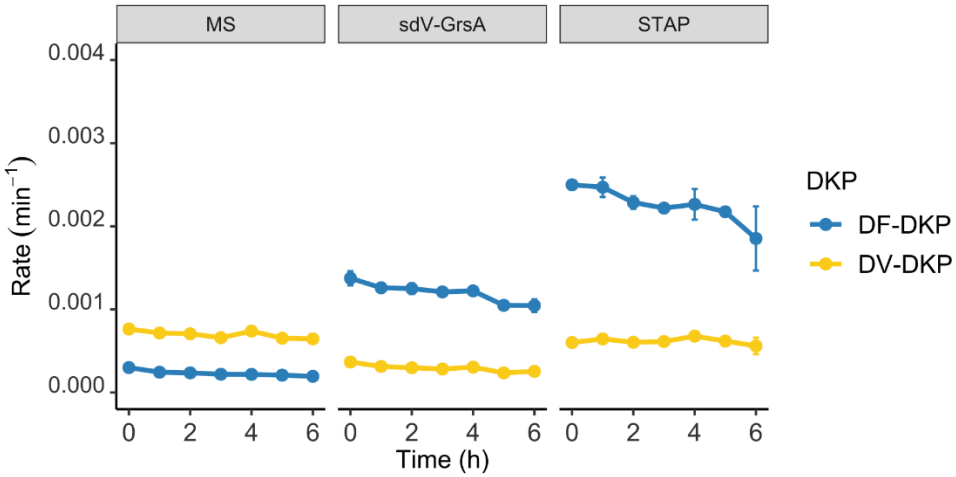
MS_Eko			
Parameter	Initial	Final	Std. error
kCLLV	0.01	0.0028	0.000021
kCDLV	0.001	0.00062	0.00002
kCLLF	0.01	0.0044	0.00019
kCDLF	0.001	1.4×10^{-10}	1.6×10^{-7}

*Undefined parameters are shaded in grey. Monomolecular and bimolecular rate constants are given in units of min^{-1} and mM^{-1} min^{-1}, respectively.

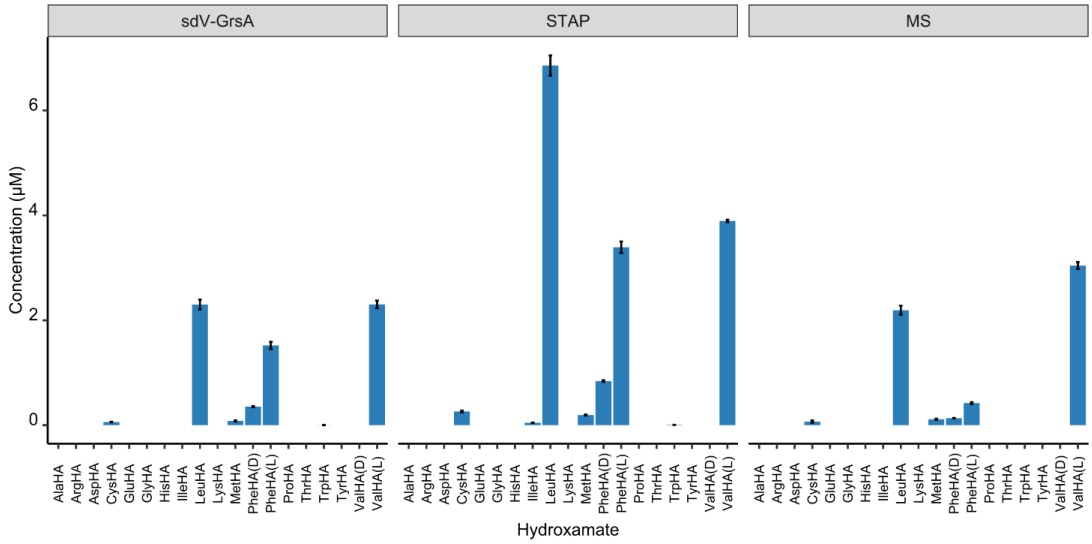
Supplementary Table 3. Oligonucleotide sequences for PCR primers.

Name	Sequence
AH_sdV-GrsA_2_r	GCT CCG GTC TGT ATT ATA CGA TCG
AH_sdV-GrsA_3_r	CAG GAT ACA ATT CAG CTC CAT GC
AH_sdV-GrsA_4_r	TAG TCA GAT TAG GGC ATT TTC GC
AH_sdV-GrsA_5_r	CTG TTG GGC CGT AAC CGT TCC
AH_sdV-GrsA_6_r	GTC ATC ATA TTC TTT ATC AAT AAG AAA GCA TGT AG
AH_sdV-GrsA_r	CCA CCT ACA ATT AGG GAG CGA AGG C
GrsA_EcoNI_r	CAT ATC CGA TTC CTT TGT TAG GTA TTC TG
GrsA_H753A_f	CTC TAG TAA AGA TAG CGC TGT TTC ATA CTC AGA ATG GAG ATC ACC TGT TTA TGG CTA TTC ATG CGT TGG TTG TGG ATG GTA TTT CTT GG
GrsB1_pTrc99a_f	ATT TCA CAC AGG AAA CAG ACC ATG AGT ACA TTT AAA AAA GAA CAT GTT CAG G
GrsB1_pTrc99a_r	GTG ATG GTG ATG GTG ATG AGA TCT GGA TCC CCC GTT TAT ATA ATT AGA GAT TTC CTG AAT GG
D306S_f	CTT CGC TCC CTA ATT GTA GGT GGA AGC GCC TTG TCT CCG AAA CAC ATC
L308T_f	CTT CGC TCC CTA ATT GTA GGT GGA GAT GCC ACC TCT CCG AAA CAC ATC AAT AAT GTA AAG CG
K311S_f	CTT CGC TCC CTA ATT GTA GGT GGA GAT GCC TTG TCT CCG AGC CAC ATC AAT AAT GTA AAG CGA AAA TGC C
H312L_f	CTT CGC TCC CTA ATT GTA GGT GGA GAT GCC TTG TCT CCG AAA CTG ATC AAT AAT GTA AAG CGA AAA TGC CCT AAT CTG
N315K_f	CTT CGC TCC CTA ATT GTA GGT GGA GAT GCC TTG TCT CCG AAA CAC ATC AAT AAA GTA AAG CGA AAA TGC CCT AAT CTG ACT ATG
STSLK_f	CTT CGC TCC CTA ATT GTA GGT GGA AGC GCC ACC TCT CCG AGC CTG ATC AAT AAA GTA AAG CGA AAA TGC CCT AAT CTG ACT ATG
G233S_f	CGT ATA ATA CAG ACC GGA GCA ATT AGC TTC GAT GCA CTG ACA TTT GAA GTT TTT GG
L237S_f	CGT ATA ATA CAG ACC GGA GCA ATT GGA TTC GAT GCA AGC ACA TTT GAA GTT TTT GGC TCA TTG C
G243M_f	CGT ATA ATA CAG ACC GGA GCA ATT GGA TTC GAT GCA CTG ACA TTT GAA GTT TTT ATG TCA TTG CTG CAT GGA GCT GAA TTG
T255L_f	CAT GGA GCT GAA TTG TAT CCT GTT CTG AAA GAC GTG CTA TTA GAT GCA GAG AAA CTA C
W326I_f	GAA AAT GCC CTA ATC TGA CTA TGA TTA ACG GTT ACG GCC CAA CAG AAA AC
N334T_f	GAA CGG TTA CGG CCC AAC AGA AAC CAC CAC TTT TTC TAC ATG CTT TCT TAT TGA TAA AG
N334T_S338A_f	GAA CGG TTA CGG CCC AAC AGA AAC CAC CAC TTT TGC GAC ATG CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TC
S338A_f	GAA CGG TTA CGG CCC AAC AGA AAA CAC CAC TTT TGC GAC ATG CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TC
C340T_f	GAA CGG TTA CGG CCC AAC AGA AAA CAC CAC TTT TTC TAC AAC CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TCC GAT AG
F341W_f	GAA CGG TTA CGG CCC AAC AGA AAA CAC CAC TTT TTC TAC ATG CTG GCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TCC GAT AGG
N350S_f	CTT ATT GAT AAA GAA TAT GAT GAC AGC ATT CCG ATA GGG AAG GCC ATT C
K355A_f	CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TCC GAT AGG GGC GGC CAT TCA AAA TAC ACA AAT TTA TAT TGT CGA TG
A356P_f	CTT TCT TAT TGA TAA AGA ATA TGA TGA CAA TAT TCC GAT AGG GAA GCC GAT TCA AAA TAC ACA AAT TTA TAT TGT CGA TGA TGA AAA TCT TC
sdXgrsA_f	GAG CAT AAA GGA ATA AGT AAT CTT AAG G
sdXgrsA_r	GCT AAC CCT TCT CCA CCA ATA CAG

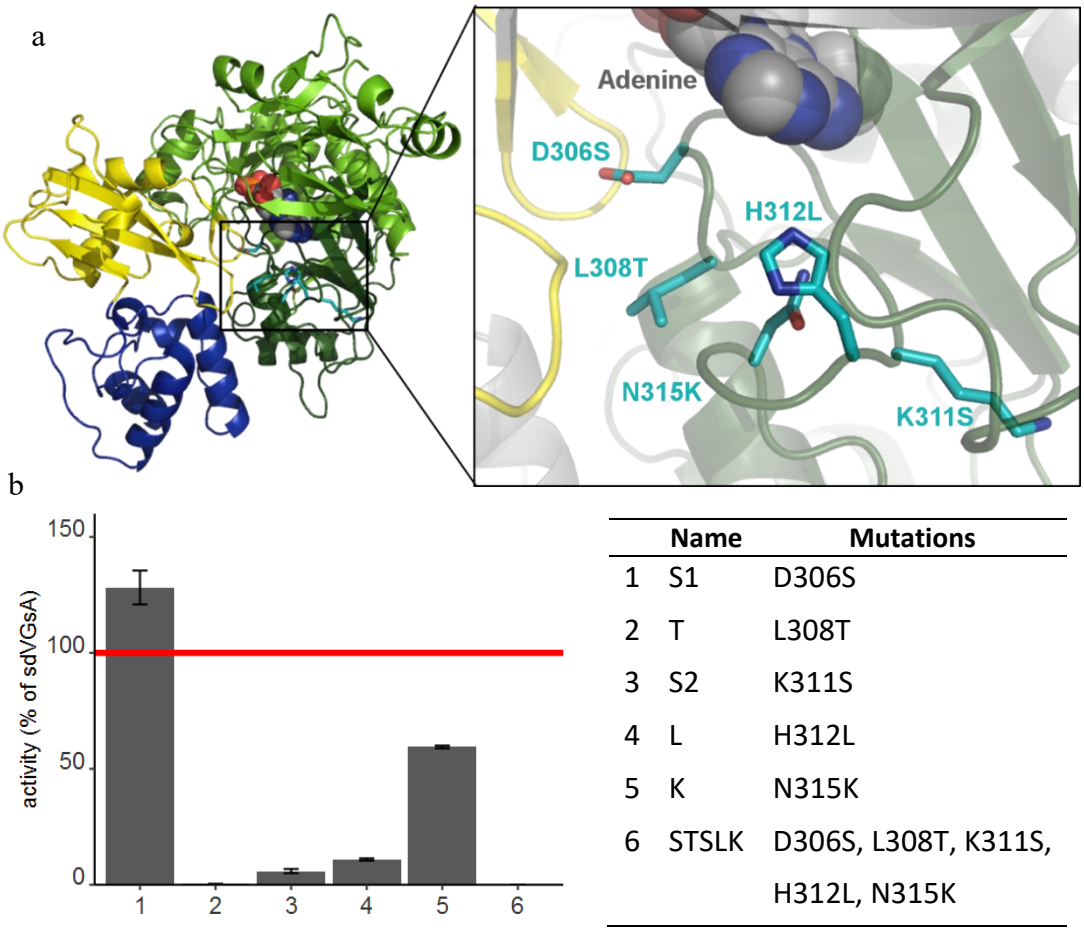
Supplementary Figures



Supplementary Figure 1. Stability of enzymes under reaction conditions monitored with the DKP assay.

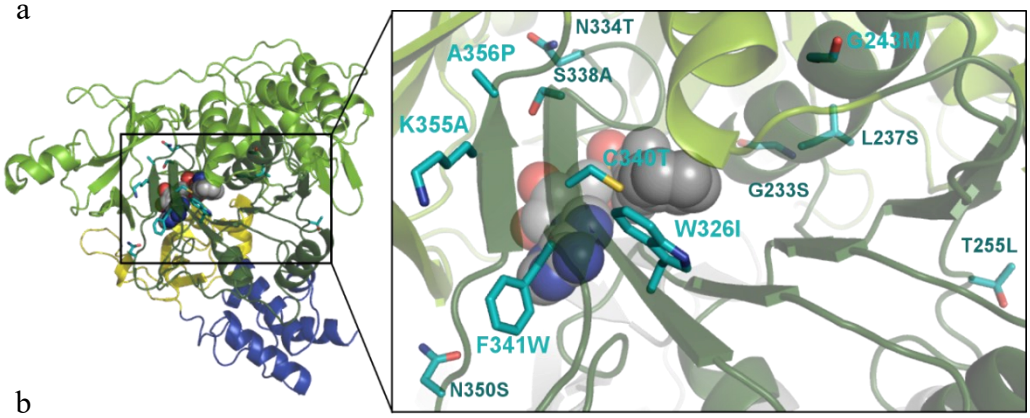


Supplementary Figure 2. HAMA specificity profiles of sdV-GrsA, STAP and MS.

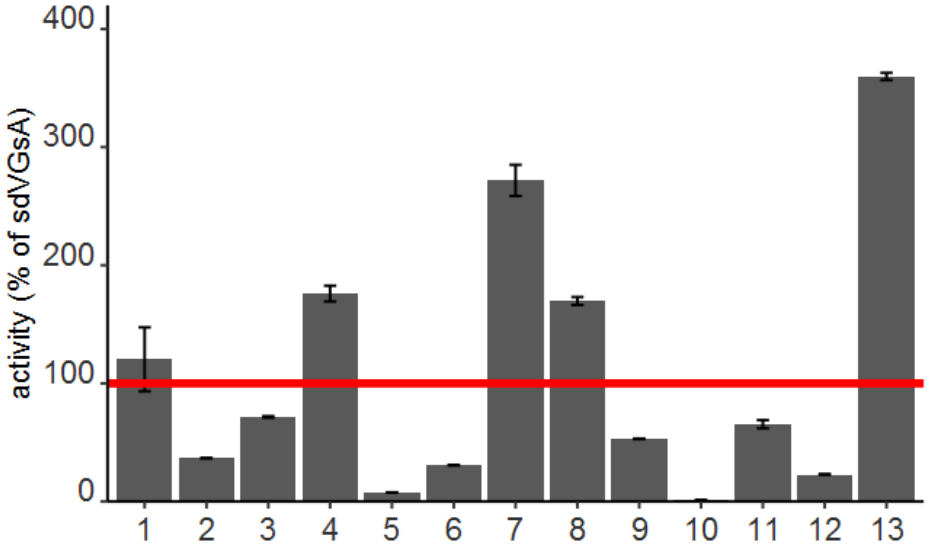


Supplementary Figure 3. First round of sdV-GrsA mutations. (**a**) Model of sdV-GrsA (left) with Phe-AMS as substrate modelled on LgrA (PDB: 5ES8).[1] Positions for mutations are shown as sticks in cyan (right, only side chains). Green: A_{NTD}, dark green: subdomain, yellow: A_{CTD}, dark blue: T domain. (**b**) Results of the DKP formation assay after 3 h with first and second enzyme module at equimolar concentrations (5 µM). TycB1, a close homologue of GrsB1, was used as second module.[21] Activity is given in relation to wild type sdV-GrsA set to 100% (red line). Errors are given as the SD of two technical replicates.

a



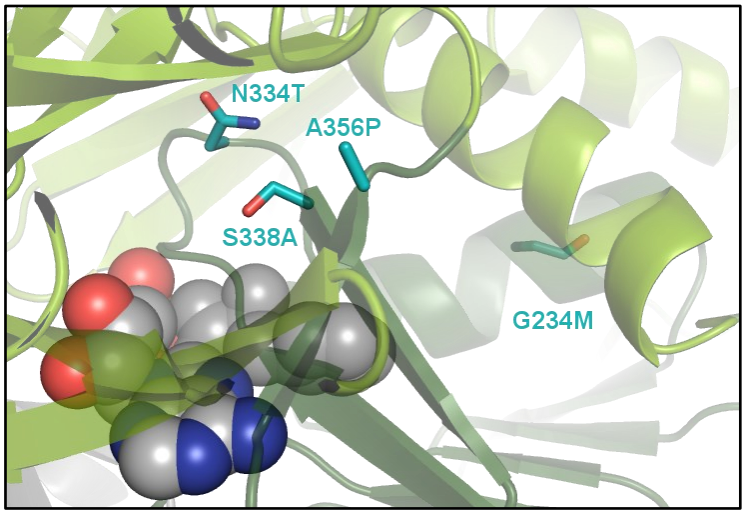
b



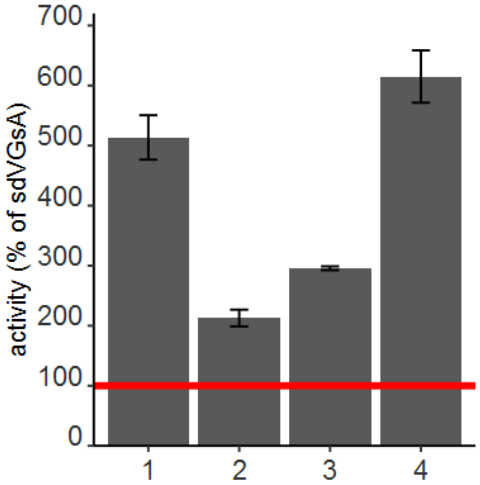
	Name	Mutations		Name	Mutations
1	S1	D306S	8	SA1	D306S, S338A
2	SS1	G233S, D306S	9	ST2	D306S, C340T
3	SS2	L237S, D306S	10	SW	D306S, F341W
4	MS	G243M, D306S	11	SS3	D306S, N350S
5	LS	T255L, D306S	12	SA2	D306S, K355A
6	SI	D306S, W326I	13	SP	D306S, A356P
7	ST1	D306S, N334T			

Supplementary Figure 4. Second round of sdV-GrsA mutations. (**a**) Model of sdV-GrsA (left) with Phe-AMS as substrate modelled on LgrA (PDB: 5ES8).[1] Positions for mutations are shown as light blue sticks (right). Green: A_{NTD}, dark green: subdomain, yellow: A_{CTD}, dark blue: T domain. (**b**) Results of the DKP formation assay after 3 h with first and second enzyme module at equimolar concentration (5 µM, blue bars). Activity is given in relation to wild type sdV-GrsA set to 100% (red line). Errors are given as the SD of two technical replicates.

a

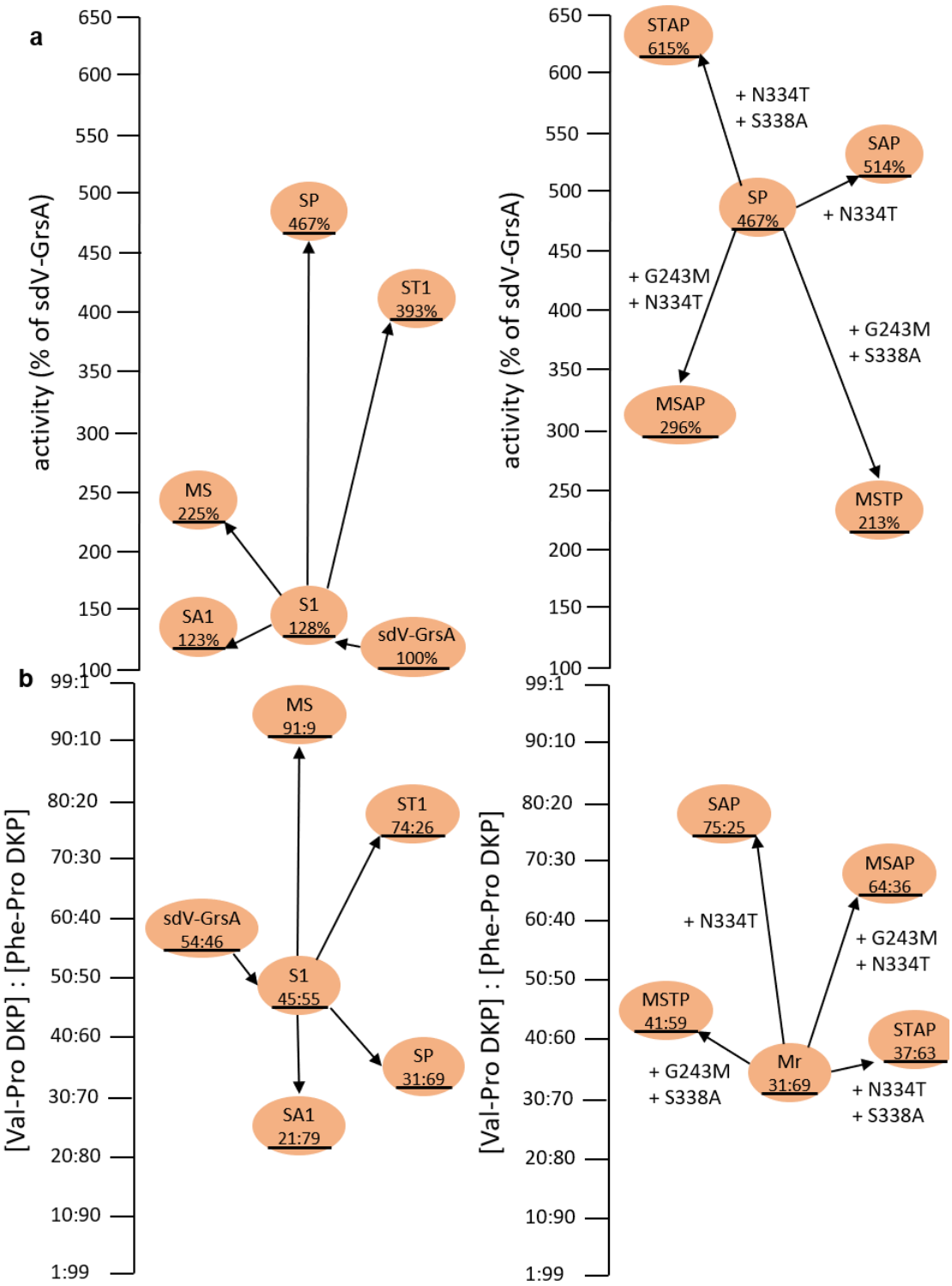


b

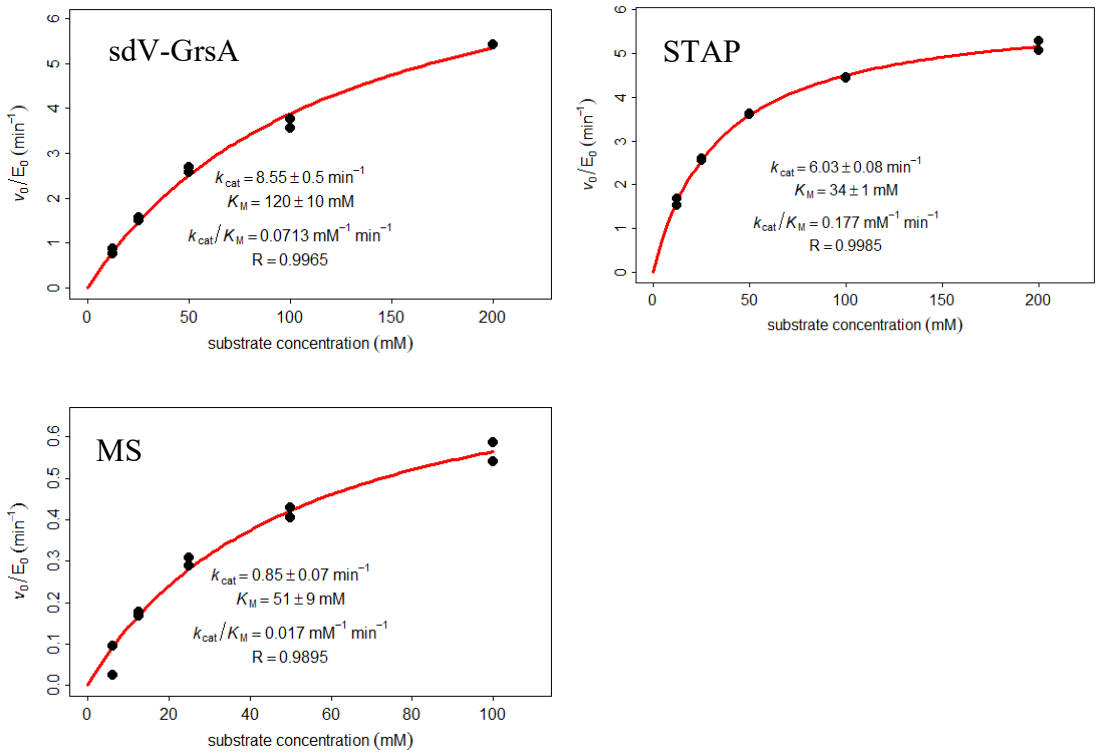


	Name	G243M	D306S	N334T	S338A	A356P
1	SAP		x		x	x
2	MSTP	x	x	x		x
3	MSAP	x	x		x	x
4	STAP		x	x	x	x

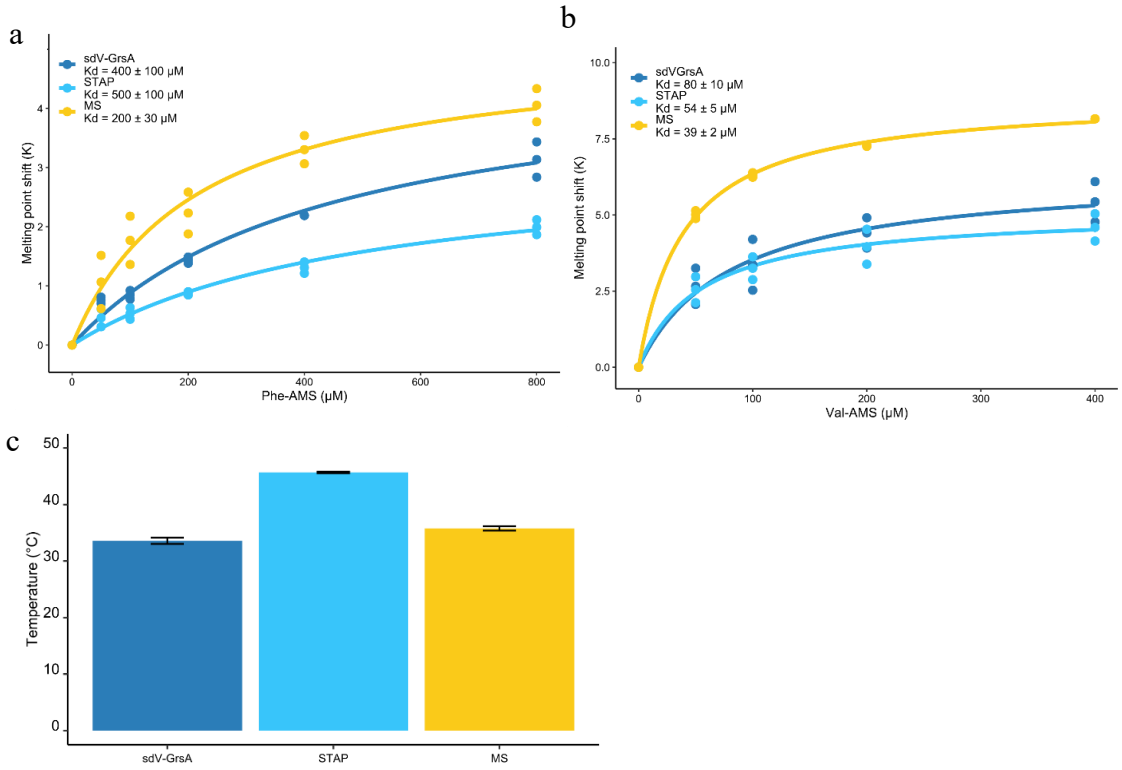
Supplementary Figure 5. Third round of sdV-GrsA mutations. (**a**) Model of positions for beneficial mutations (light blue sticks, side chains only except for glycine) in sdV-GrsA with Phe-AMS as substrate modelled on LgrA (PDB: 5ES8).[1] (**b**) Results of the DKP formation assay after 3 h with first enzyme modules at ten-fold decreased concentrations (0.5 µM) compared to the second enzyme module. Activity is given in relation to wild type sdV-GrsA set at 100% (red line). Errors are given as the SD of two technical replicates.



Supplementary Figure 6. Comparison of beneficial mutants based on activity (**a**) and specificity (**b**). Arrows indicate which mutant emerged from which precursor. Left: first and second round, right: third round (additional mutations listed on the arrows). The earliest precursor is encircled.

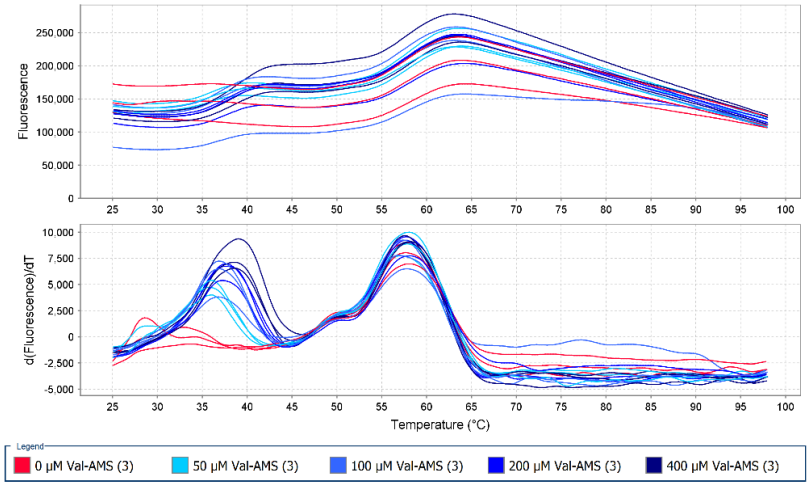


Supplementary Figure 7. Michaelis-Menten kinetics with L-Val.

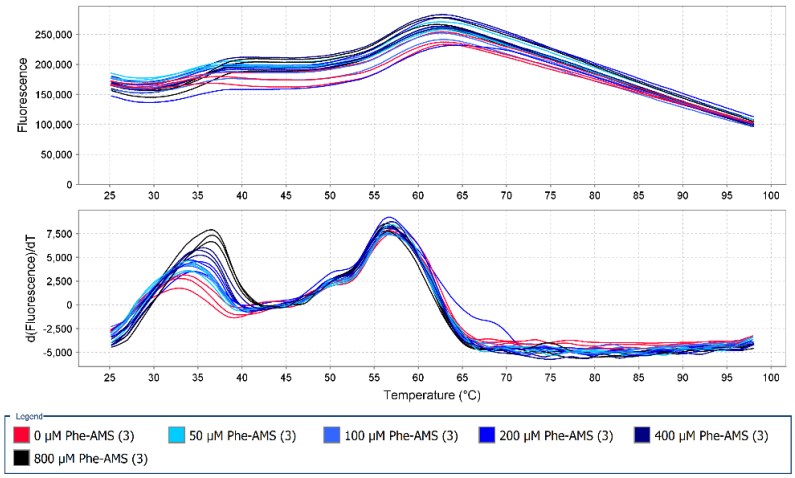


Supplementary Figure 8. Thermal shift assay of sdV-GrsA and mutants with transition state inhibitors L-Phe-AMS (a) and L-Val-AMS (b). K_d values are determined by plotting melting point shifts against inhibitor concentration using a hyperbolic binding model (Equation 1). (c) Melting temperatures of three enzymes in the absence of inhibitor.

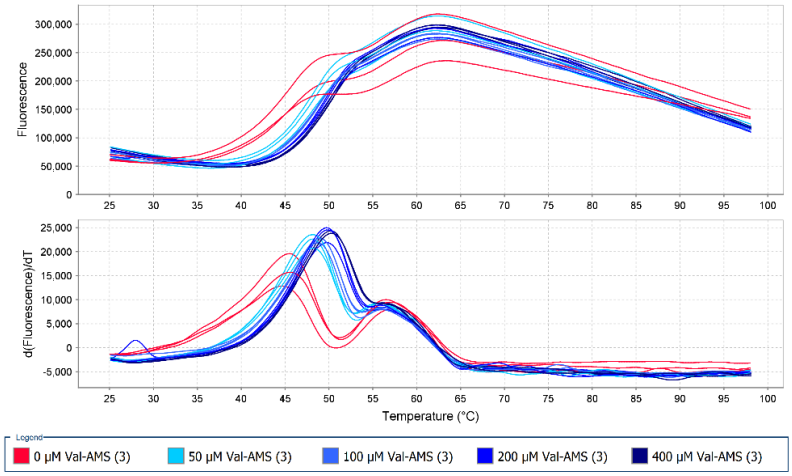
a) sdVGrsA / Val-AMS



b) sdVGrsA / Phe-AMS

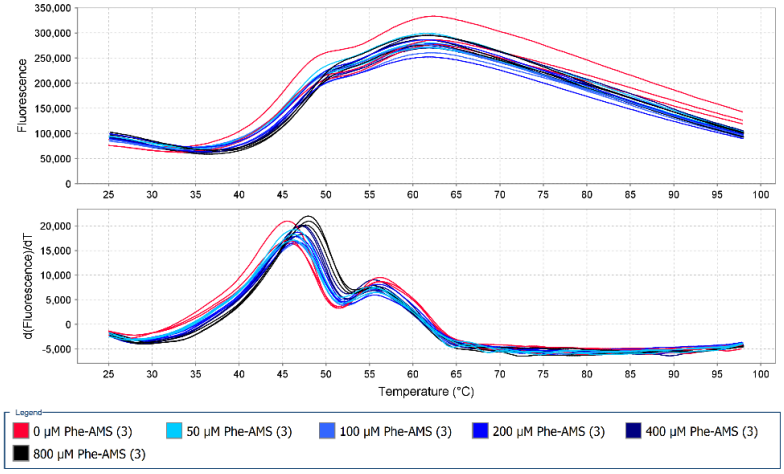


c) STAP/ Val-AMS

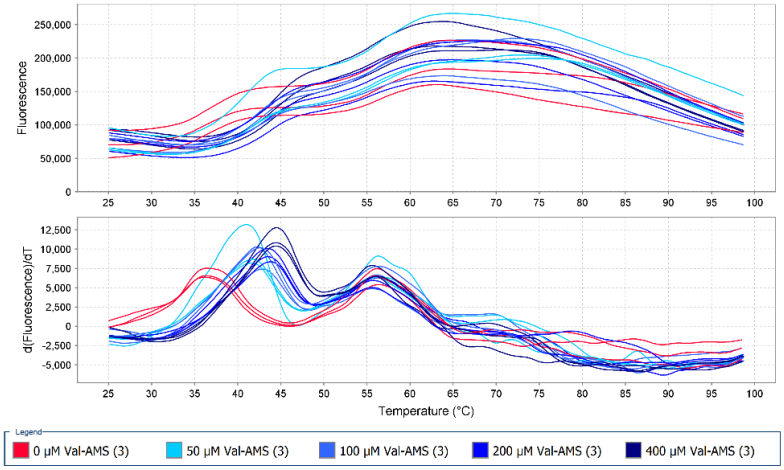


(continued on next page)

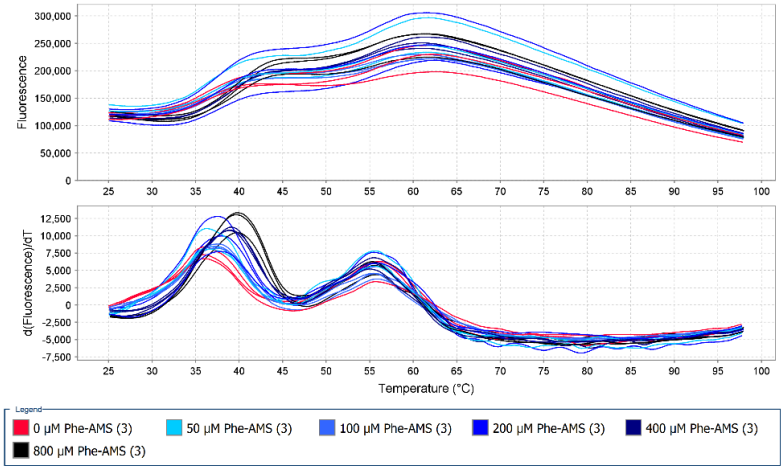
a) STAP / Phe-AMS



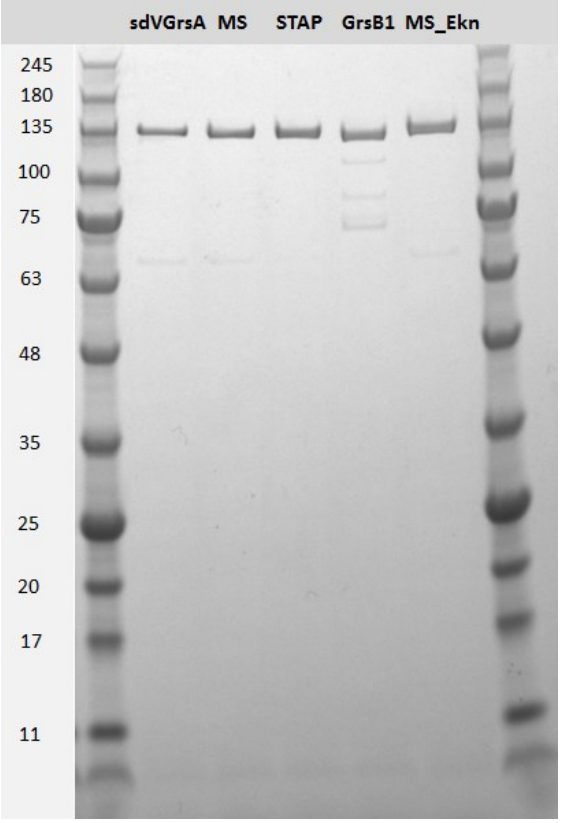
b) MS / Val-AMS



c) MS / Phe-AMS



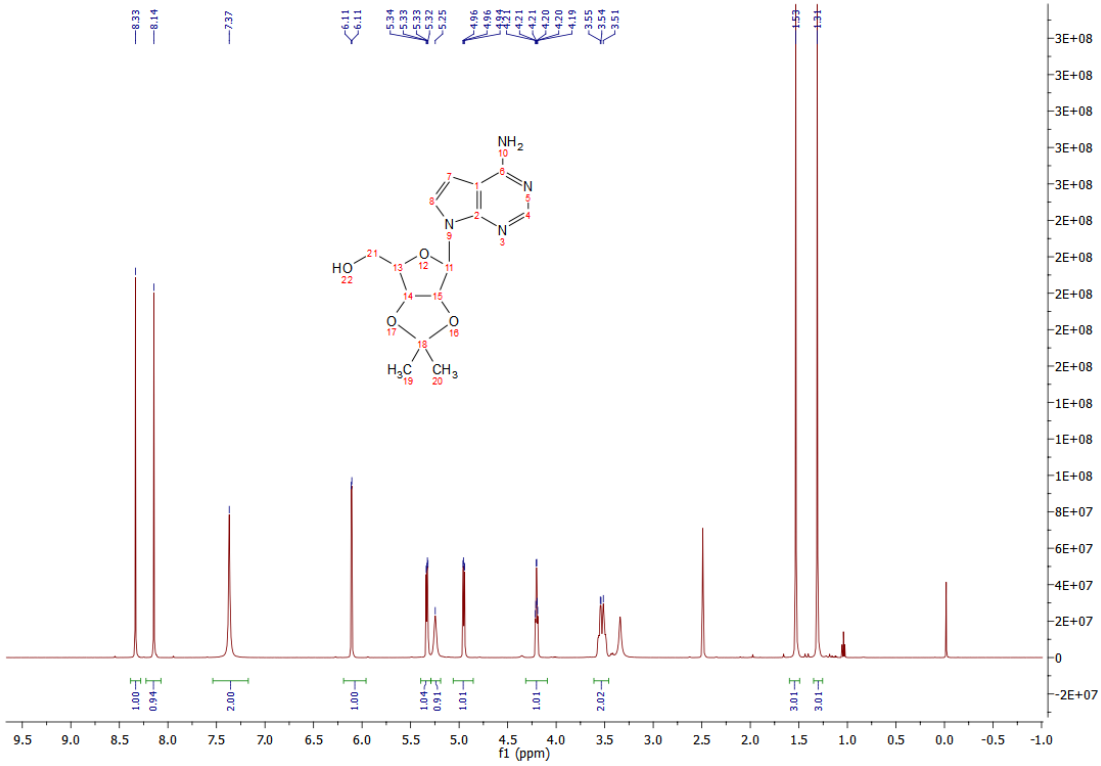
Supplementary Figure 9. Melting curves from the thermal shift assay of sdV-GrsA and mutants with inhibitors L-Phe-AMS and L-Val-AMS.



Supplementary Figure 10. SDS-PAGE of purified proteins.

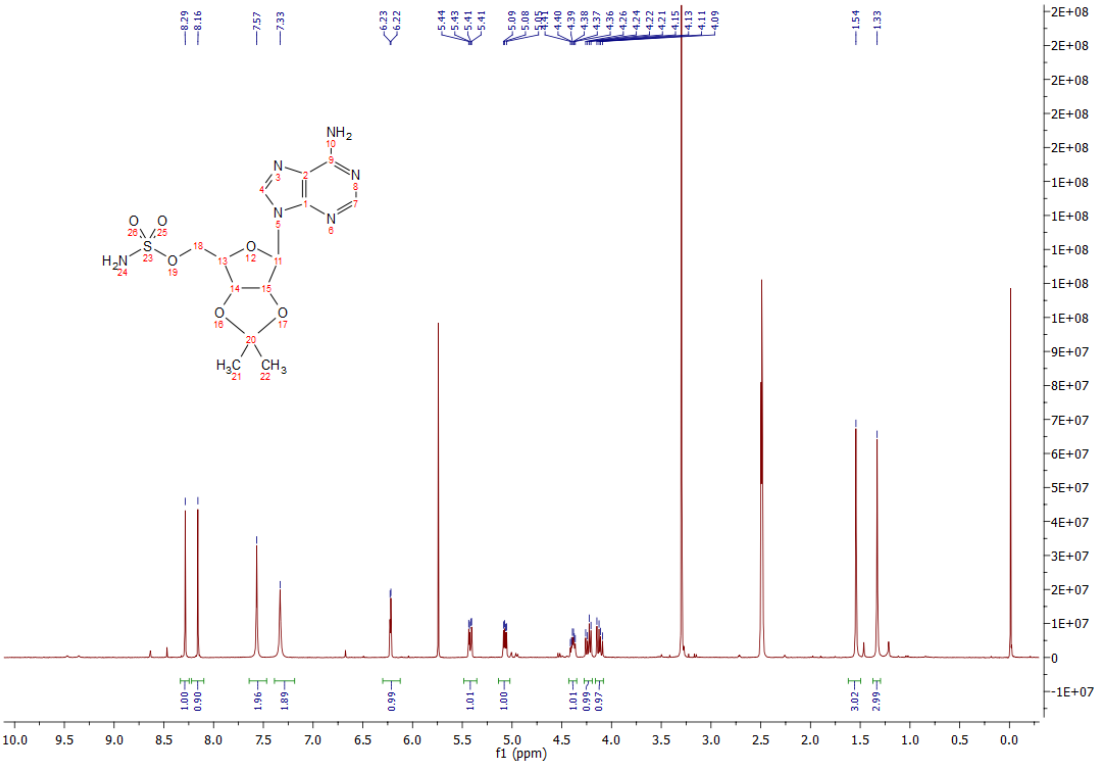
NMR spectra

2,3-O-Isopropylideneadenosine (2), , ^{1}H NMR



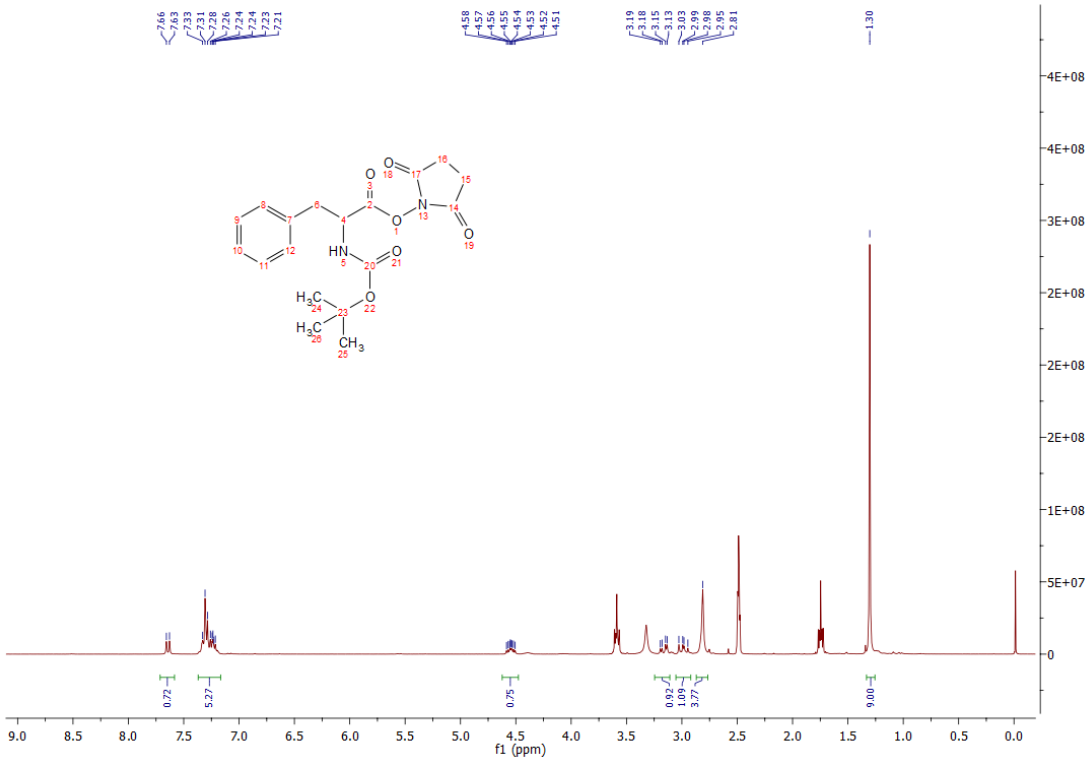
Atom	δ_H, mult. (*J* in Hz)
4	8.33, s
7	8.14, s
10	7.37, br s
11	6.11, d (*J* = 3.1)
13	4.20, td (*J* = 4.8, 2.6)
14	4.95, dd (*J* = 6.1, 2.5)
15	5.33, dd (*J* = 6.1, 3.1)
21	3.60 – 3.46, m
22	5.25, br s
19	1.53, s
20	1.31, s

Sulfamoyl-Isopropylidenadenosine (5), ^{1}H NMR



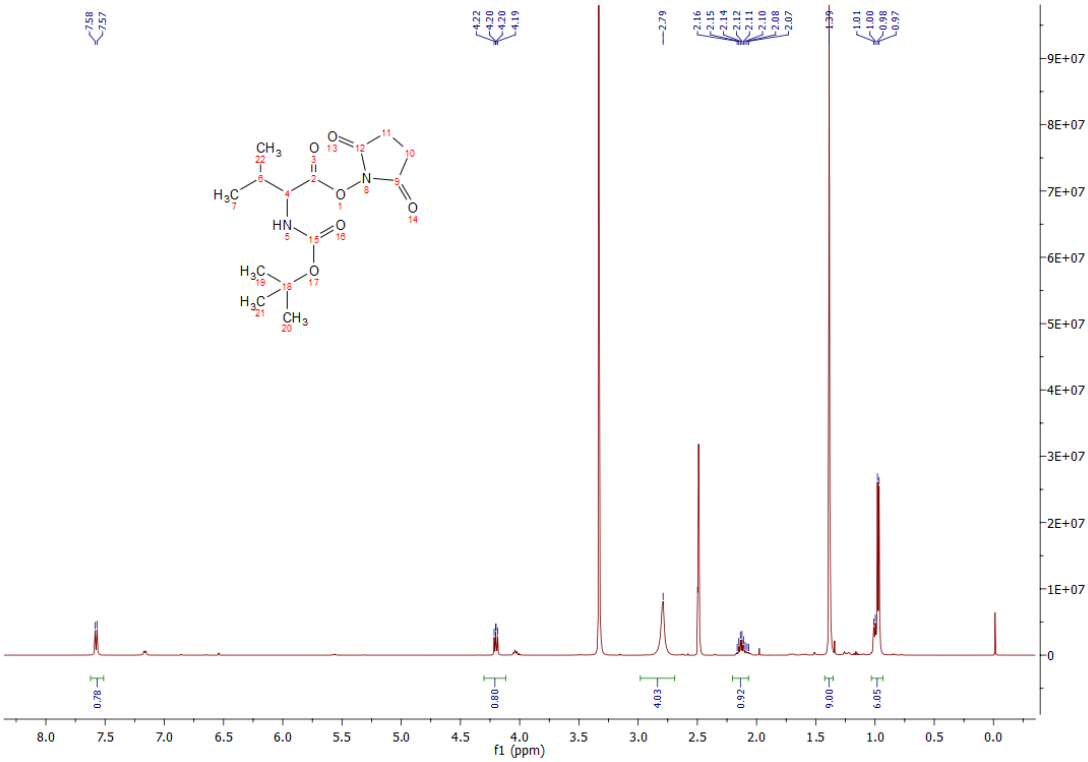
Atom	δ_H, mult. (*J* in Hz)
4	8.16, s
7	8.29, s
10	7.33, br s
11	6.22, d (*J* = 2.4)
13	4.39, ddd (*J* = 3.2, 5.5, 5.8)
14	4.12, dd (*J* = 6.4, 10.6)
15	4.23, dd (*J* = 5.3, 10.6)
18a 18b	5.42, dd (*J* = 2.4, 6.3) 5.07, dd (*J* = 3.2, 6.3)
19	5.25, br s
21	1.54, s
22	1.33, s
24	7.57, br s

Boc-Phe-OSu (7), ^{1}H NMR



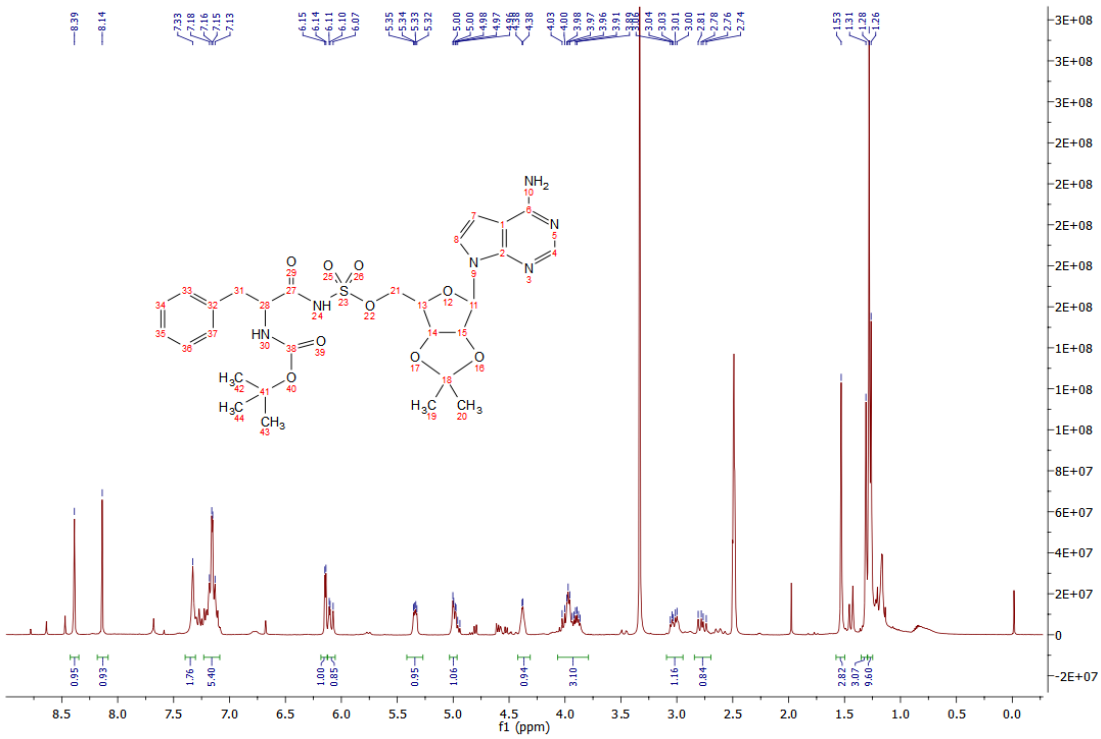
Atom	δ_H, mult. (*J* in Hz)
4	4.55, ddd, (*J* = 10.5, 8.4, 4.6)
5	7.64, d (*J* = 8.4)
6a 6b	3.16, dd (*J* = 13.9, 4.5) 2.99, dd (*J* = 13.8, 10.6)
8	7.37 – 7.16, m
9	7.37 – 7.16, m
10	7.37 – 7.16, m
11	7.37 – 7.16, m
12	7.37 – 7.16, m
11	6.22, d (*J* = 2.4)
15	2.81, br s
16	2.81, br s
24	1.30, s
25	1.30, s
26	1.30, s

Boc-Val-OSu (9), ^{1}H NMR



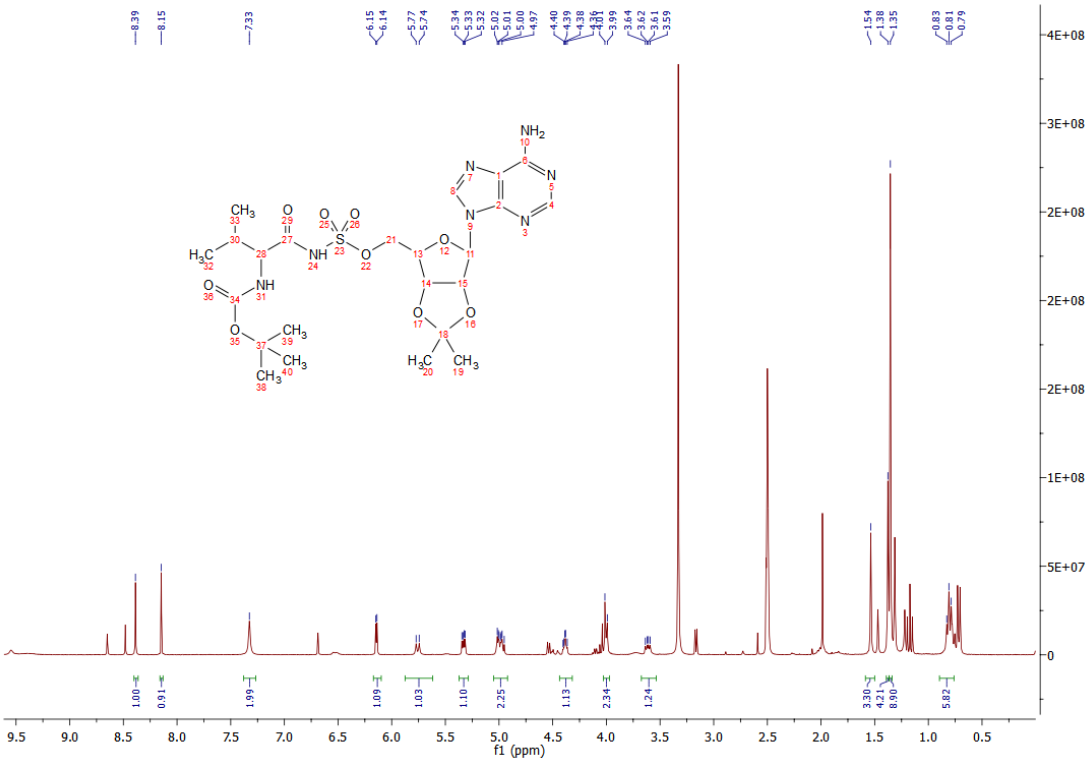
Atom	δ_H, mult. (*J* in Hz)
4	4.20, dd (*J* = 8.1, 6.7)
5	7.58, d (*J* = 8.2)
6	2.20 – 2.04, m
7	0.99, dd (*J* = 15.6, 6.8)
10	2.79, br s
11	2.79, br s
19	1.39, s
20	1.39, s
21	1.39, s s
22	0.99, dd (*J* = 15.6, 6.8)

Boc-Phe-AMS (10), ^{1}H NMR



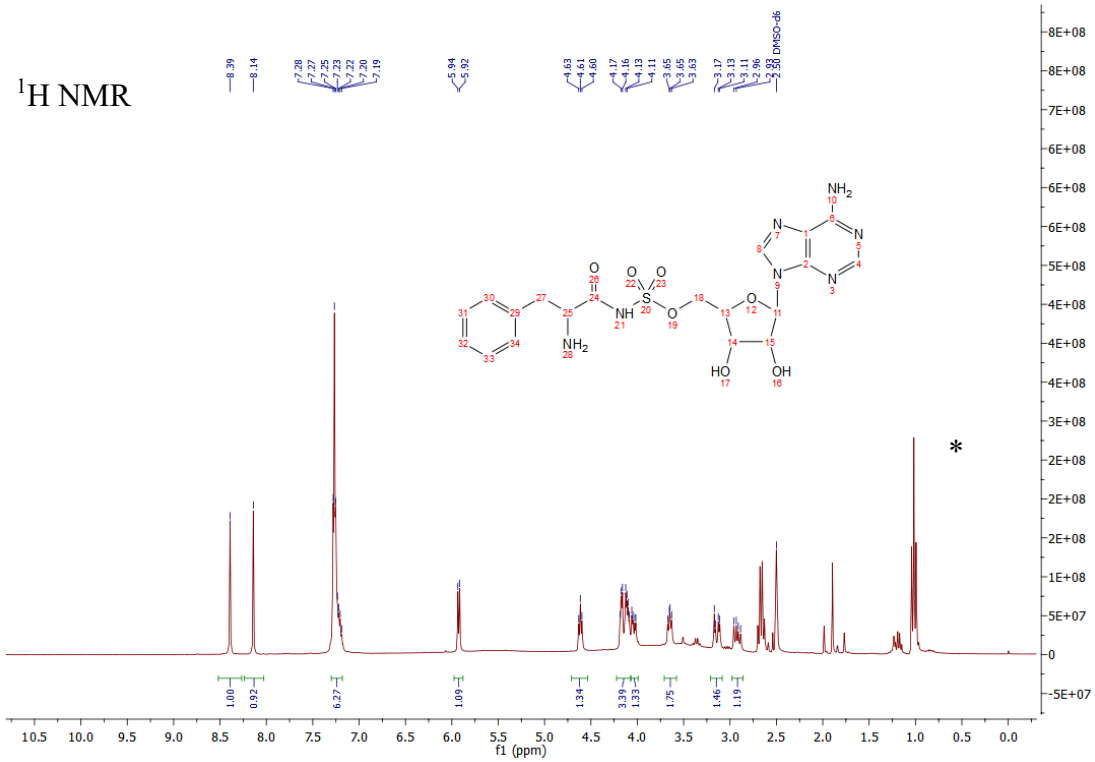
Atom	δ_H, mult. (*J* in Hz)	Atom	δ_H, mult. (*J* in Hz)
4	8.39, s	36	7.24 – 7.08, m
8	8.14, s	37	7.24 – 7.08, m
10	7.33, br s	42	1.27, d (*J* = 5.2)
11	6.14, d (*J* = 2.9)	43	1.27, d (*J* = 5.2)
13	4.08 – 3.77, m	44	1.27, d (*J* = 5.2)
14	4.99, dd (*J* = 6.2, 2.1)		
15	5.34, dd (*J* = 5.9, 3.0)		
19	1.53, s		
20	1.31, s		
21	4.08 – 3.77, m		
28	4.38, d (*J* = 2.1)		
30	6.09, d (*J* = 8.2)		
31a 31b	3.03, dd (*J* = 13.5, 4.3) 2.77, dd (*J* = 13.5, 8.3)		
33	7.24 – 7.08, m		
34	7.24 – 7.08, m		
35	7.24 – 7.08, m		

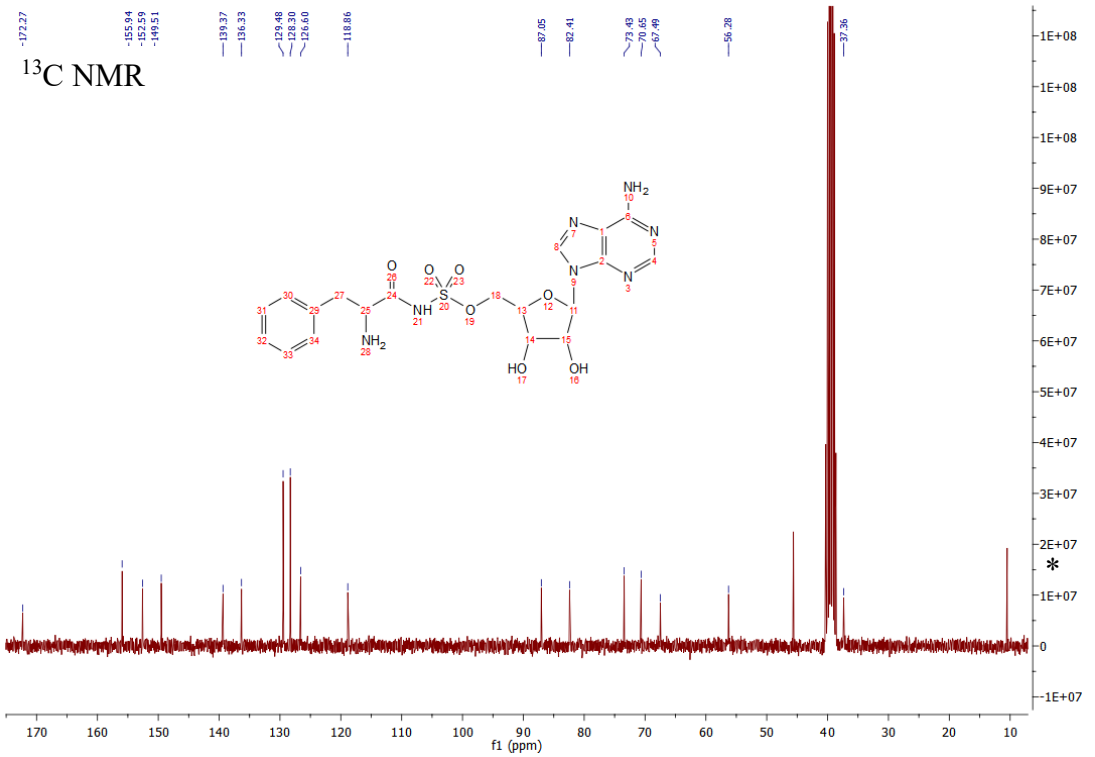
Boc-Val-AMS (11), ^{1}H NMR



Atom	δ_H, mult. (*J* in Hz)	Atom	δ_H, mult. (*J* in Hz)
4	8.39, s	39	1.35, s
8	8.15, s	40	1.35, s
10	7.33, br s		
11	6.14, d (*J* = 3.0)		
13	3.61, dd (*J* = 8.5, 4.7)		
14	5.06 – 4.92, m		
15	5.33, dd (*J* = 6.1, 3.1)		
19	1.38, s		
20	1.54, s		
21	4.00, d (*J* = 6.8)		
28	4.43 – 4.34, m		
31	5.76, d (*J* = 8.6)		
32	0.88 – 0.76, m		
33	0.88 – 0.76, m		
38	1.35, s		

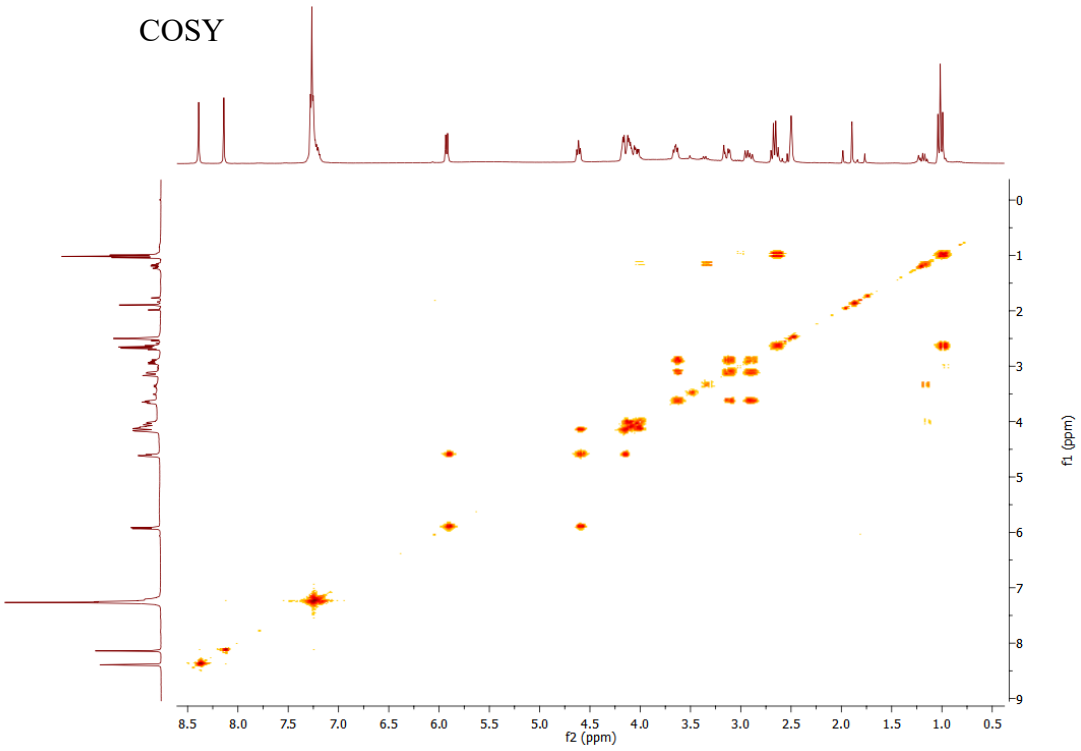
Phe-AMS (12)



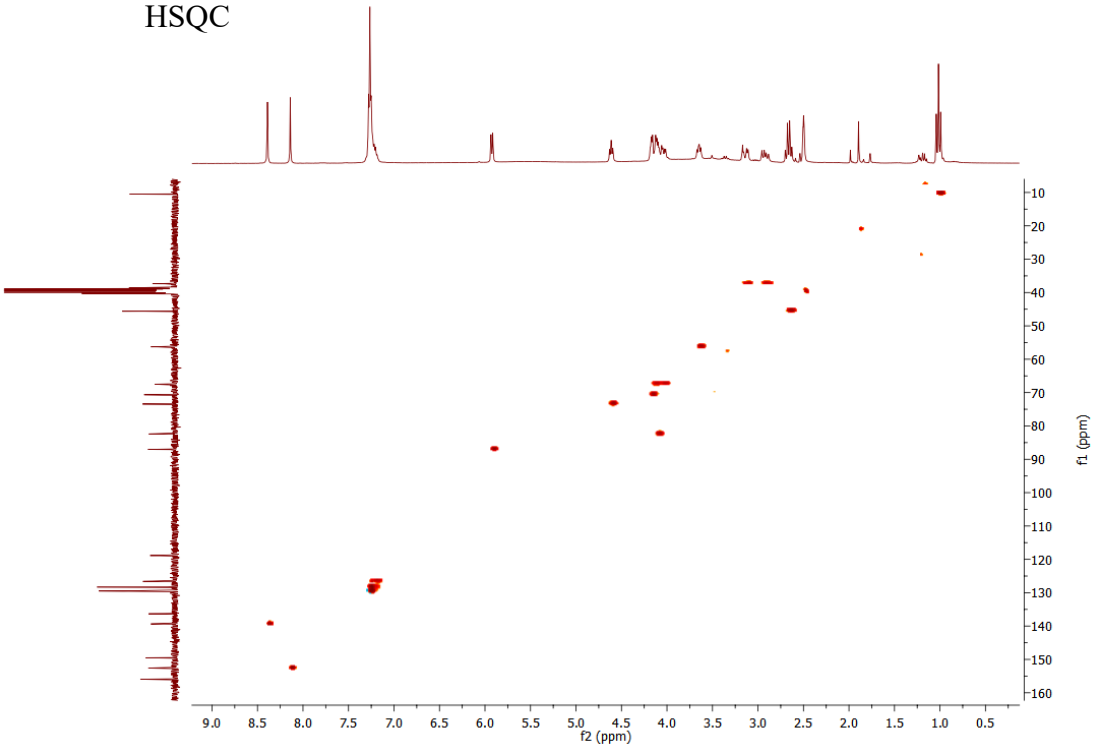


Signals marked with * belong to TEA

COSY



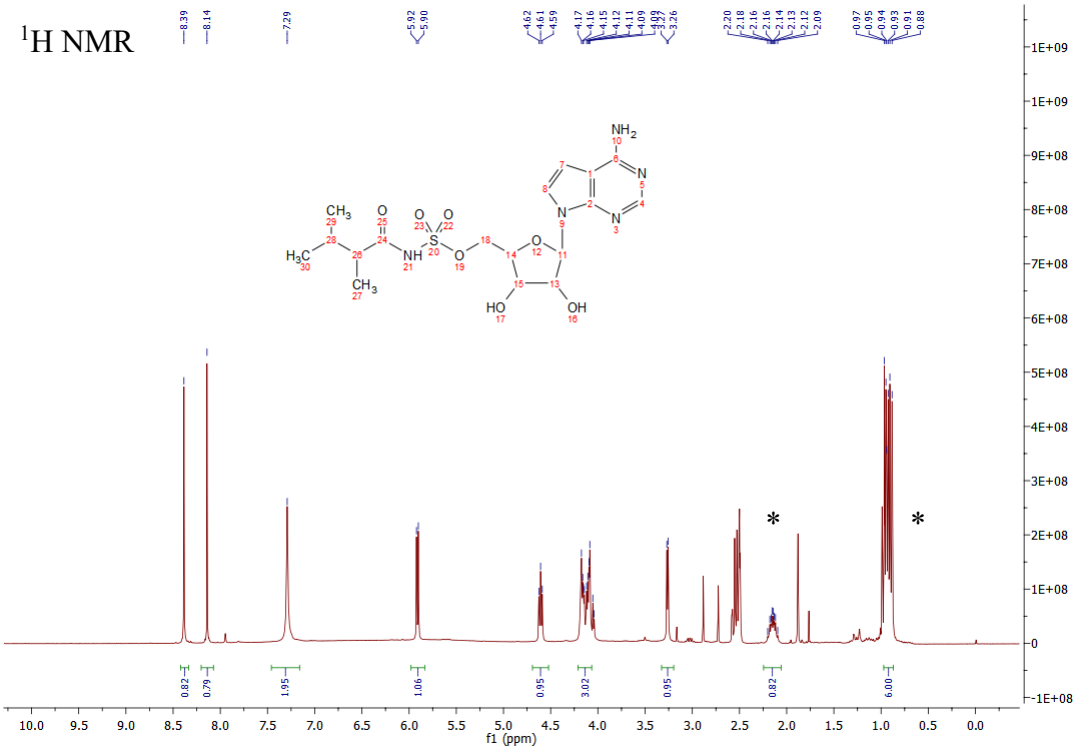
HSQC



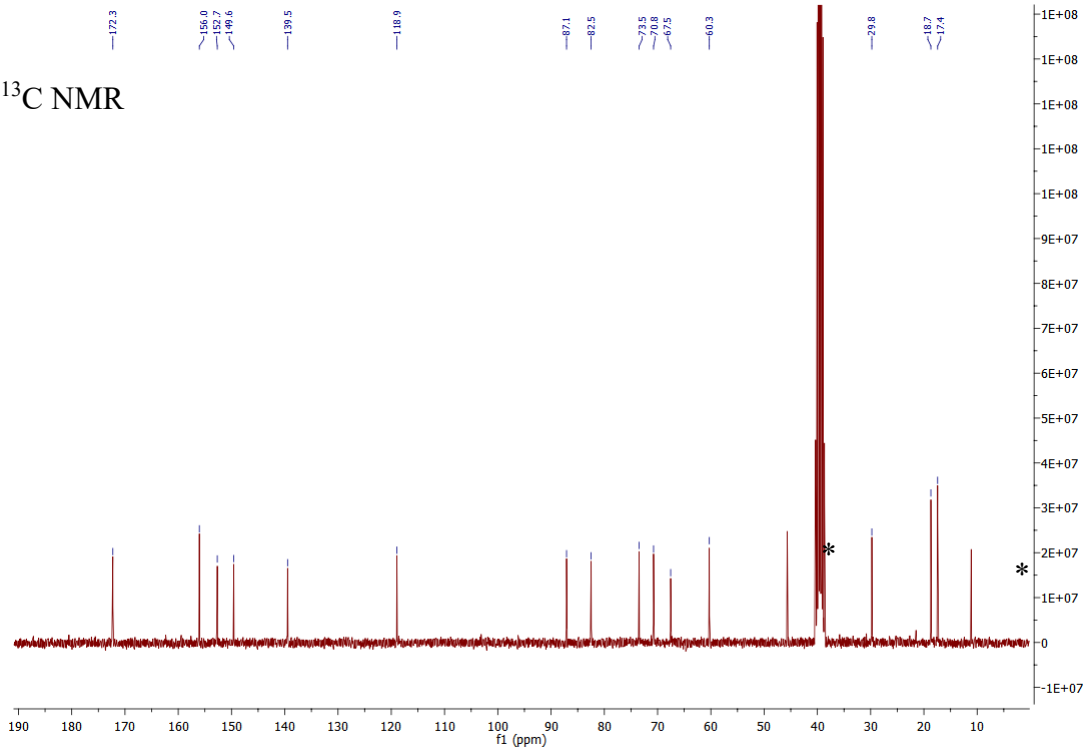
Atom	δH, mult. (J in Hz)	δC
1	-	118.9
2	-	149.5
4	8.39, s	152.6
6	-	156.0
8	8.14, s	139.4
11	5.93, d (J = 5.7)	87.1
13	4.04, dd (J = 9.7, 3.5)	82.4
14	4.23 – 4.07, m	70.7
15	4.61, t (J = 5.3)	73.4
18	4.23 – 4.07, m	67.5
24	-	172.3
25	3.65, dd (J = 7.4, 5.1)	56.3
27a 27b	3.14, dd (J = 13.4, 4.1) 2.92, dd (J = 14.1, 7.7)	37.4
29	-	138.0
30	7.33 – 7.17, m	129.5
31	7.33 – 7.17, m	128.3
32	7.33 – 7.17, m	126.6
33	7.33 – 7.17, m	128.3
34	7.33 – 7.17, m	129.5

Val-AMS (13)

^{1}H NMR

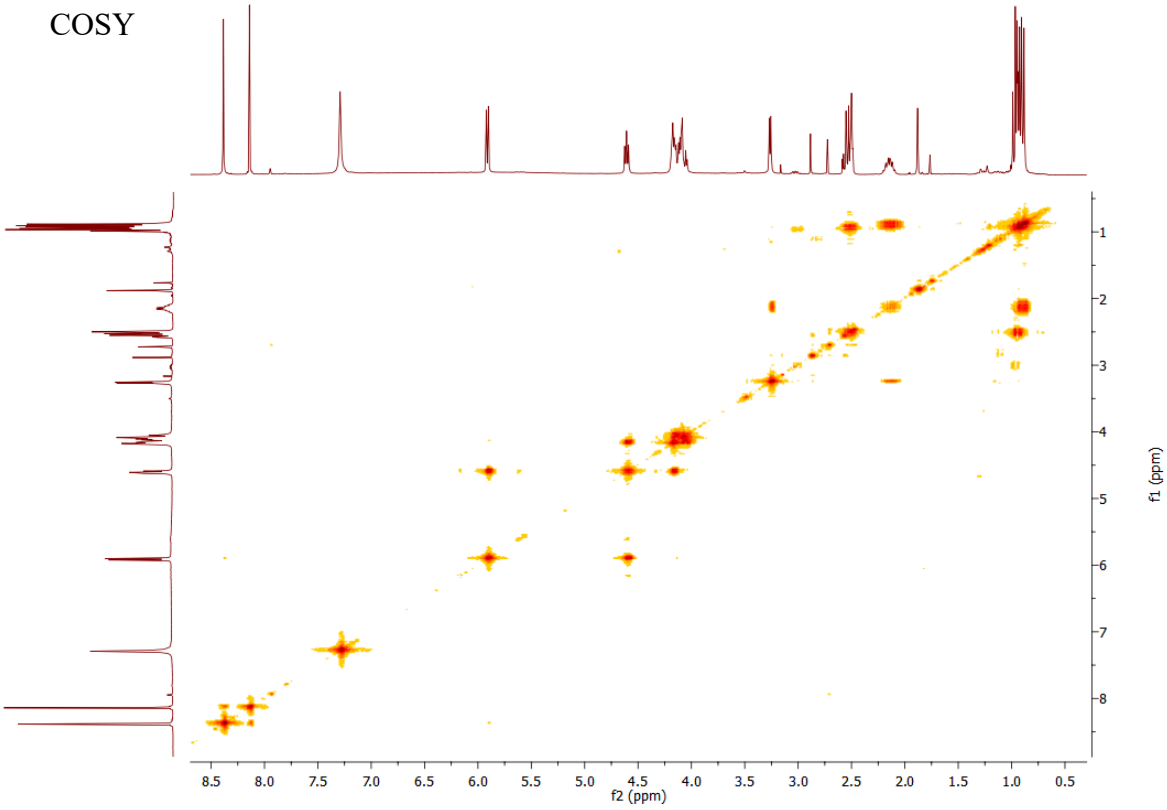


^{13}C NMR

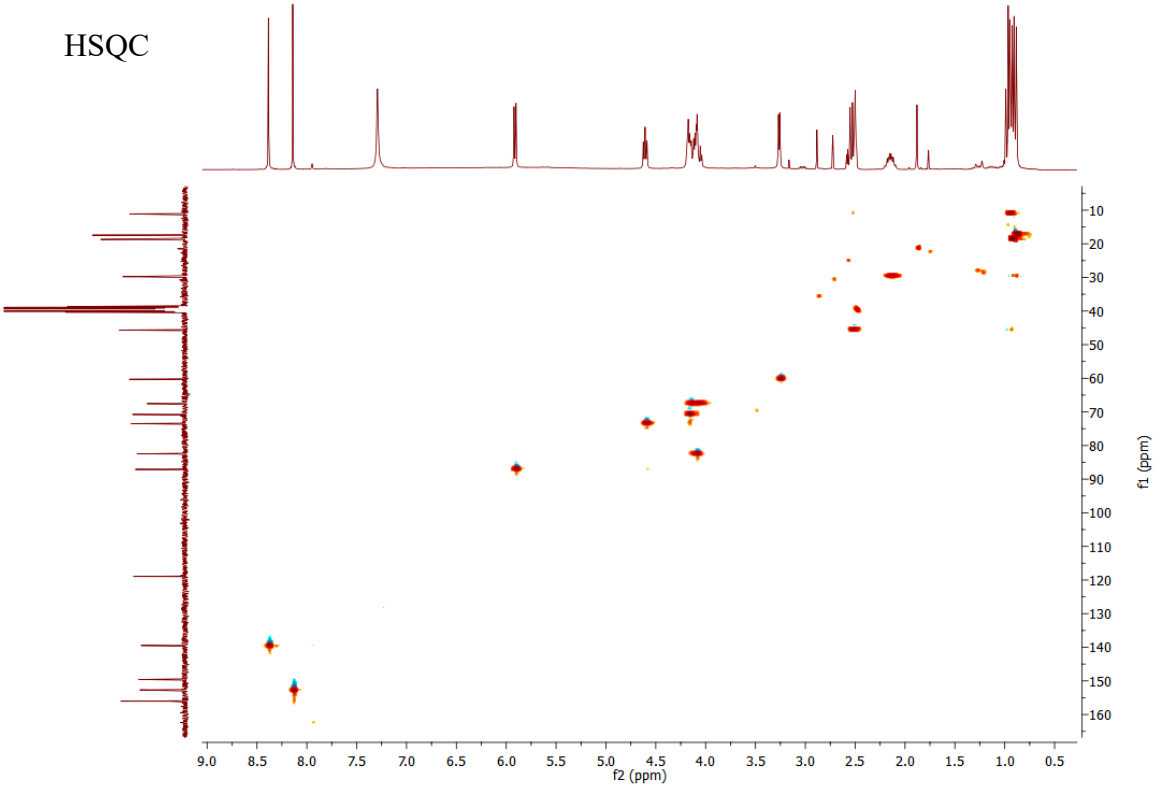


Signals marked with * belong to TEA

COSY



HSQC



Atom	δ_H, mult. (*J* in Hz)	δ_C
1	-	119.0
2	-	149.6
4	8.39, s	152.7
6	-	156.0
8	8.14, s	139.5
11	5.91, d (*J* = 5.8)	87.1
13	4.04, dd (*J* = 9.7, 3.5)	82.5
14	4.21 – 4.06, m	70.8
15	4.61, t (*J* = 5.3)	73.5
18	4.21 – 4.06, m	67.5
24	-	172.3
26	3.26, d (*J* = 4.1)	60.3
27	7.29, br s	-
28	2.25 – 2.05, m	29.8
29	0.97 – 0.87, m	18.7
30	0.97 – 0.87, m	17.4

Sequences of proteins used in this study

Mutations are highlighted in red.

sdV-GrsA

MLNSSKSILIHAQNKNGTHEEEQYLFAVNNTKAEYPRDKTIHQLFEEQVSKRPNNVAIVCENEQLTYHELNVKANQLARIFIEKGIGKDT
LVGIMMEKSIDLFIGILAVLKAGGAYVPIDIEYPKERIQYILDDSQARMLLTQKHLVHLIHNIQFNGQVEIFEEDTIKIREGTNLHVPSKSTD
LAYVIYTSGTTGNPKGTMLEHKGISNLKVFFENSLNVREDDRIIQTGAIGFDALTFEVFGSLLHGAELYPVTKDVLLDAEKLHKFLQANQI
TIMWLTSPLFNQLSQGTEEMFAGLRSLIVGGDALSPKHINNVKRKCPNLTMWNGYGPTENTTFSTCFLIDKEYDDNIPIGKAIQNTQIYI
VDENLQLKSVGEAGELCIGGEGLARGYWKRPELTSQKFVDNPFVPGEKLYKTGDQARWLSDGNIEYLGRIDNQVKIRGHRVELEEVESIL
LKHMYISETAVSVHKDHQEQPYLCAYFVSEKHIPLEQLRQFSSEELPTYMIPSYFIQLDKMPLTSNGKIDRKQLPEPDLTFGMRVDYEAP
RNEIEETLVTIWQDVLGIEKIGIKDNFYALGGDSIKAIQVAARLHSYQLKLETKDLLKYPTIDQLVHYIKDSKRRSEQGIVEGEIGLTPIQH
WFFEQQFTNMHHWNQSYMLYRPNGFDKEILLRVFNKIVEHHDALRMIYKHHNGKIVQINRGLEGTLFDFYTFDLTANDNEQQVICEES
ARLQNSINLEVGPLVKIALFHTQNGDHLFMAIHHLVVDGISWRILFEDLATAYEQAMHQQTIALPEKTDSFKDWSIELEKYANSELFLEE
AEYWHHLNYYTENVQIKKDYVTMNNKQKNIRYVGMELTIEETEKLLKNVNKAYRTEINDILLTALGFALKEWADIDKIVINLEGHGREE
ILEQMNIARTVGWFTSQYPVVLDMQKSDDLSYQIKLMKENLRRIPNKGIGYEIFKYLTTEYLRPVLPFTLKPEINFNYLGQFDTDVKTELF
TRSPYSMGNSLGPDGKNNLSPEGESYFVLNINGFIEEGKLHITFSYNEQQYKEDTIQQLSRSYKQHLLAIIEHCVQKEDTELTPSDFSFKEL
ELEEMDDIFDLLADSLTGSRSHHHHH

MS

MLNSSKSILIHAQNKNGTHEEEQYLFAVNNTKAEYPRDKTIHQLFEEQVSKRPNNVAIVCENEQLTYHELNVKANQLARIFIEKGIGKDT
LVGIMMEKSIDLFIGILAVLKAGGAYVPIDIEYPKERIQYILDDSQARMLLTQKHLVHLIHNIQFNGQVEIFEEDTIKIREGTNLHVPSKSTD
LAYVIYTSGTTGNPKGTMLEHKGISNLKVFFENSLNVREDDRIIQTGAIGFDALTFEVFMSLLHGAELYPVTKDVLLDAEKLHKFLQANQI
TIMWLTSPLFNQLSQGTEEMFAGLRSLIVGGSALSPKHINNVKRKCPNLTMWNGYGPTENTTFSTCFLIDKEYDDNIPIGKAIQNTQIYIV
DENLQLKSVGEAGELCIGGEGLARGYWKRPELTSQKFVDNPFVPGEKLYKTGDQARWLSDGNIEYLGRIDNQVKIRGHRVELEEVESILL
KHMYISETAVSVHKDHQEQPYLCAYFVSEKHIPLEQLRQFSSEELPTYMIPSYFIQLDKMPLTSNGKIDRKQLPEPDLTFGMRVDYEAPR
NEIEETLVTIWQDVLGIEKIGIKDNFYALGGDSIKAIQVAARLHSYQLKLETKDLLKYPTIDQLVHYIKDSKRRSEQGIVEGEIGLTPIQHWF
FEQQFTNMHHWNQSYMLYRPNGFDKEILLRVFNKIVEHHDALRMIYKHHNGKIVQINRGLEGTLFDFYTFDLTANDNEQQVICEESAR
LQNSINLEVGPLVKIALFHTQNGDHLFMAIHHLVVDGISWRILFEDLATAYEQAMHQQTIALPEKTDSFKDWSIELEKYANSELFLEEAE
YWHHLNYYTENVQIKKDYVTMNNKQKNIRYVGMELTIEETEKLLKNVNKAYRTEINDILLTALGFALKEWADIDKIVINLEGHGREEIL
EQMNIARTVGWFTSQYPVVLDMQKSDDLSYQIKLMKENLRRIPNKGIGYEIFKYLTTEYLRPVLPFTLKPEINFNYLGQFDTDVKTELFT
RSPYSMGNSLGPDGKNNLSPEGESYFVLNINGFIEEGKLHITFSYNEQQYKEDTIQQLSRSYKQHLLAIIEHCVQKEDTELTPSDFSFKELE
LEEMDDIFDLLADSLTGSRSHHHHH

MS (E domain knockout)

MLNSSKSILIHAQNKNGTHEEEQYLFAVNNTKAEYPRDKTIHQLFEEQVSKRPNNVAIVCENEQLTYHELNVKANQLARIFIEKGIGKDT
LVGIMMEKSIDLFIGILAVLKAGGAYVPIDIEYPKERIQYILDDSQARMLLTQKHLVHLIHNIQFNGQVEIFEEDTIKIREGTNLHVPSKSTD
LAYVIYTSGTTGNPKGTMLEHKGISNLKVFFENSLNVREDDRIIQTGAIGFDALTFEVFMSLLHGAELYPVTKDVLLDAEKLHKFLQANQI
TIMWLTSPLFNQLSQGTEEMFAGLRSLIVGGSALSPKHINNVKRKCPNLTMWNGYGPTENTTFSTCFLIDKEYDDNIPIGKAIQNTQIYIV
DENLQLKSVGEAGELCIGGEGLARGYWKRPELTSQKFVDNPFVPGEKLYKTGDQARWLSDGNIEYLGRIDNQVKIRGHRVELEEVESILL
KHMYISETAVSVHKDHQEQPYLCAYFVSEKHIPLEQLRQFSSEELPTYMIPSYFIQLDKMPLTSNGKIDRKQLPEPDLTFGMRVDYEAPR
NEIEETLVTIWQDVLGIEKIGIKDNFYALGGDSIKAIQVAARLHSYQLKLETKDLLKYPTIDQLVHYIKDSKRRSEQGIVEGEIGLTPIQHWF
FEQQFTNMHHWNQSYMLYRPNGFDKEILLRVFNKIVEHHDALRMIYKHHNGKIVQINRGLEGTLFDFYTFDLTANDNEQQVICEESAR
LQNSINLEVGPLVKIALFHTQNGDHLFMAIHALVVDGISWRILFEDLATAYEQAMHQQTIALPEKTDSFKDWSIELEKYANSELFLEEAE
YWHHLNYYTENVQIKKDYVTMNNKQKNIRYVGMELTIEETEKLLKNVNKAYRTEINDILLTALGFALKEWADIDKIVINLEGHGREEIL
EQMNIARTVGWFTSQYPVVLDMQKSDDLSYQIKLMKENLRRIPNKGIGYEIFKYLTTEYLRPVLPFTLKPEINFNYLGQFDTDVKTELFT
RSPYSMGNSLGPDGKNNLSPEGESYFVLNINGFIEEGKLHITFSYNEQQYKEDTIQQLSRSYKQHLLAIIEHCVQKEDTELTPSDFSFKELE
LEEMDDIFDLLADSLTGSRSHHHHH

STAP

MLNSSKSILIHAQNKNGTHEEEQYLFAVNNTKAEYPRDKTIHQLFEEQVSKRPNNVAIVCENEQLTYHELNVKANQLARIFIEKGIGKDT
LVGIMMEKSIDLFIGILAVLKAGGAYVPIDIEYPKERIQYILDDSQARMLLTQKHLVHLIHNIQFNGQVEIFEEDTIKIREGTNLHVPSKSTD
LAYVIYTSGTTGNPKGTMLEHKGISNLKVFFENSLNVREDDRIIQTGAIGFDALTFEVFGSLLHGAELYPVTKDVLLDAEKLHKFLQANQI
TIMWLTSPLFNQLSQGTEEMFAGLRSLIVGGSALSPKHINNVKRKCPNLTMWNGYGPTETTTFATCFLIDKEYDDNIPIGKPIQNTQIYIV
DENLQLKSVGEAGELCIGGEGLARGYWKRPELTSQKFVDNPFVPGEKLYKTGDQARWLSDGNIEYLGRIDNQVKIRGHRVELEEVESILL
KHMYISETAVSVHKDHQEQPYLCAYFVSEKHIPLEQLRQFSSEELPTYMIPSYFIQLDKMPLTSNGKIDRKQLPEPDLTFGMRVDYEAPR
NEIEETLVTIWQDVLGIEKIGIKDNFYALGGDSIKAIQVAARLHSYQLKLETKDLLKYPTIDQLVHYIKDSKRRSEQGIVEGEIGLTPIQHWF
FEQQFTNMHHWNQSYMLYRPNGFDKEILLRVFNKIVEHHDALRMIYKHHNGKIVQINRGLEGTLFDFYTFDLTANDNEQQVICEESAR
LQNSINLEVGPLVKIALFHTQNGDHLFMAIHHLVVDGISWRILFEDLATAYEQAMHQQTIALPEKTDSFKDWSIELEKYANSELFLEEAE
YWHHLNYYTENVQIKKDYVTMNNKQKNIRYVGMELTIEETEKLLKNVNKAYRTEINDILLTALGFALKEWADIDKIVINLEGHGREEIL
EQMNIARTVGWFTSQYPVVLDMQKSDDLSYQIKLMKENLRRIPNKGIGYEIFKYLTTEYLRPVLPFTLKPEINFNYLGQFDTDVKTELFT
RSPYSMGNSLGPDGKNNLSPEGESYFVLNINGFIEEGKLHITFSYNEQQYKEDTIQQLSRSYKQHLLAIIEHCVQKEDTELTPSDFSFKELE
LEEMDDIFDLLADSLTGSRSHHHHHH

GrsB1 (adjusted C-A-T frame)

MSTFKKEHVQDMYRLSPMQEGMLFHALLDKDKNAHLVQMSIAIEGIVDVELLSESLNILIDRYDVFRTTFLHEKIKQPLQVVLKERPVQL
QFKDISSLDEEKREQAIEQYKYQDGETVFDLTRDPLMRVAIFQTGKVNYQMIWSFHHILMDGWCFNIIFNDLFNIYLSLKEKKPLQLEAV
QPYKQFIKWLEKQDKQEALRYWKEHLMNYDQSVTLPKKKAAINNTTYEPAQFRFAFDKVLTQQLLRIANQSQVTLNIVFQTIWGIVLQ
KYNSTNDVVYGSVVSGRPSEISGIEKMVGLFINTLPLRIQTQKDQSFIELVKTVHQNVLFSQQHEYFPLYEIQNHTELKQNLIDHIMVIENY
PLVEELQKNSIMQKVGFTVRDVKMFEPTNYDMTVMVLPRDEISVRLDYNAAVYDIDFIKKIEGHMKEVALCVANNPHVLVQDVPLLTK
QEKQHLLVELHDSITEYPDKTIHQLFTEQVEKTPEHVAVVFEDEKVTYRELHERSNQLARFLREKGVKKESIIGIMMERSVEMIVGILGIL
KAGGAFVPIDPEYPKERIGYMLDSVRLVLTQRHLKDKFAFTKETIVIEDPSISHELTEEIDYINESEDLFYIIYTSGTTGKPKGVMLEHKNIV
NLLHFTFEKTNINFSDKVLQYTTCSFDVCYQEIFSTLLSGGQLYLIRKETQRDVEQLFDLVKRENIEVLSFPVAFLKFIFNEREFINRFPTCV
KHIITAGEQLVVNNEFKRYLHEHNVHLHNHYGPSETHVVTTYTINPEAEIPELPPIGKPISNTWIYILDQEQQLQPQGIVGELYISGANVG
RGYLNNQELTAEKFFADPFRPNERMYRTGDLARWLPDGNIEFLGRADHQVKIRGHRIELGEIEAQLLNCKGVKEAVVIDKADDKGGKYL
CAYVVMEVEVNDSELREYLGKALPDYMIPSFFVPLDQLPLTPNGKIDRKSLPNLEGIVNTNAKYVVPTNELEEKLAKIWEEVLGISQIGIQ
DNFFSLGGHSLKAITLISRMNKECNVDIPLRLLFEAPTIQEISNYINGGSRSHHHHHH

Supplementary References

(1) Reimer, J. M., Aloise, M. N., Harrison, P. M. & Martin Schmeing, T. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* 529, 239–242 (2016).

(2) Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30, 162–173 (2009).

(3) Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46, W296–W303 (2018).

(4) 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).

(5) Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* 22, 640–648 (2015).

(6) Niquille, D. L. et al. Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* 10, 282–287 (2018).

(7) Zhu, B., Cai, G., Hall, E. O. & Freeman, G. J. In-FusionTM assembly: Seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *Biotechniques* 43, 354–359 (2007).

(8) Stachelhaus, T.; Walsh, C. T. Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase. *Biochemistry* **2000**, *39* (19), 5775–5787.

(9) Wilson, D. J. & Aldrich, C. C. A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal. Biochem.* 404, 56–63 (2010).

(10) R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria 2017.

(11) 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).

(12) 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).

(13) Sun, Z. et al. Reshaping an Enzyme Binding Pocket for Enhanced and Inverted Stereoselectivity: Use of Smallest Amino Acid Alphabets in Directed Evolution. *Angew. Chemie - Int. Ed.* 54, 12410–12415 (2015).

(14) 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).

(15) Mootz, H. D., Finking, R. & Marahiel, M. a. 4′-Phosphopantetheine Transfer in Primary and Secondary Metabolism of Bacillus subtilis. *Anal. Biochem.* 217, 220–230 (1994).

(16) 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).

(17) Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chemie Int. Ed.* 56, 3770–3821 (2017).

(18) Kuzmič, P. Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. *Anal. Biochem.* **1996**, *237* (2), 260–273. https://doi.org/10.1006/abio.1996.0238.

(19) Kuzmič, P. DynaFit-A Software Package for Enzymology. *Methods Enzymol.* 467, 247–280 (2009).

(20) Kuzmič, P. A generalized numerical approach to rapid-equilibrium enzyme kinetics: Application to 17β-HSD. *Mol. Cell. Endocrinol.* 248, 172–181 (2006).

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

6 MANUSCRIPT V

Bacterial-Like Nonribosomal Peptide Synthetases Produce cyclopeptides in the Zygomycetous Fungus Mortierella alpina

Jacob M. Wurlitzer, **Aleksa Stanišić**, Ina Wasmuth, Sandra Jungmann, Dagmar Fischer, Hajo Kries and Markus Gressler

Published manuscript: *Appl. Environ. Microbiol.* **87**, (2020).
doi: 10.1128/AEM.02051-20

Summary:
In contrast to basidiomycetes and ascomycetes, lower fungi (traditionally zygomycetes) are considered to be poor producers of natural products. Newly discovered malpicyclins (A-E) and known malpibaldins (A-C) are the first described secondary metabolites in zygomycetes that are linked to the corresponding genes *mpcA* and *mpbA* with a close homology to bacterial NRPS genes. These cyclic pentapeptides differ in their amino acid composition in several positions. Here, we elucidate the origins of this variability by characterization of the A-domain specificity with HAMA assay, revealing promiscuous substrate selection. This is a first report on nonribosomal peptide biosynthesis in basal fungi.

The candidate is:

☐ First author ☒ Second author ☐ Corresponding author ☐ Coauthor

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
JW	50 %	80 %	75 %	20 %	
AS		10 %	15 %	5 %	
IW		5 %	5 %	2.5 %	
SJ		5 %	5 %	2.5 %	
DF				5 %	
HK				5 %	
MG	50 %			60 %	

Supplementary datasets can be found under:
https://aem.asm.org/content/87/3/e02051-20 and are included in the enclosed CD-ROM.

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

PHYSIOLOGY

Bacterial-Like Nonribosomal Peptide Synthetases Produce Cyclopeptides in the Zygomycetous Fungus *Mortierella alpina*

Jacob M. Wurlitzer,[a] Aleksa Stanišić,[b] Ina Wasmuth,[a] Sandra Jungmann,[c] Dagmar Fischer,[c] Hajo Kries,[b] Markus Gressler[a]

[a]Department of Pharmaceutical Microbiology, Leibniz Institute for Natural Product Research and Infection Biology (Hans Knöll Institute) and Friedrich Schiller University, Jena, Germany

[b]Junior Research Group Biosynthetic Design of Natural Products, Leibniz Institute for Natural Product Research and Infection Biology (Hans Knöll Institute), Jena, Germany

[c]Pharmaceutical Technology and Biopharmacy, Friedrich Schiller University Jena, Jena, Germany

ABSTRACT Fungi are traditionally considered a reservoir of biologically active natural products. However, an active secondary metabolism has long not been attributed to early-diverging fungi such as *Mortierella*. Here, we report on the biosynthesis of two series of cyclic pentapeptides, the malpicyclins and malpibaldins, as products of *Mortierella alpina* ATCC 32222. The molecular structures of malpicyclins were elucidated by high-resolution tandem mass spectrometry (HR-MS/MS), Marfey's method, and one-dimensional (1D) and 2D nuclear magnetic resonance (NMR) spectroscopy. In addition, malpibaldin biosynthesis was confirmed by HR-MS. Genome mining and comparative quantitative real-time PCR (qRT-PCR) expression analysis pointed at two pentamodular nonribosomal peptide synthetases (NRPSs), malpicyclin synthetase MpcA and malpibaldin synthetase MpbA, as candidate biosynthetic enzymes. Heterologous production of the respective adenylation domains and substrate specificity assays confirmed the existence of promiscuous substrate selection and also confirmed their respective biosynthetic roles. In stark contrast to known fungal NRPSs, MpbA and MpcA contain bacterial-like dual epimerase/condensation domains allowing the racemization of enzyme-tethered L-amino acids and the subsequent incorporation of D-amino acids into the metabolites. Phylogenetic analyses of both NRPS genes indicated a bacterial origin and a horizontal gene transfer into the fungal genome. We report on the as-yet-unexplored nonribosomal peptide biosynthesis in basal fungi which highlights this paraphylum as a novel and underrated resource of natural products.

IMPORTANCE Fungal natural compounds are industrially produced, with applications in antibiotic treatment, cancer medications, and crop plant protection. Traditionally, higher fungi have been intensively investigated for their metabolic potential, but reidentification of already known compounds is frequently observed. Hence, alternative strategies to acquire novel bioactive molecules are required. We present the genus *Mortierella* as representative of the early-diverging fungi as an underestimated resource of natural products. *Mortierella alpina* produces two families of cyclopeptides, designated malpicyclins and malpibaldins, via two pentamodular nonribosomal peptide synthetases (NRPSs). These enzymes are much more closely related to bacterial than to other fungal NRPSs, suggesting a bacterial origin of these NRPS genes in *Mortierella*. Both enzymes were biochemically characterized and are involved in as-yet-unknown biosynthetic pathways of natural products in basal fungi. Hence, this report establishes early-diverging fungi as prolific natural compound producers and sheds light on the origin of their biosynthetic capacity.

KEYWORDS Mortierellales, NRPS, adenylation domain, cyclopeptide, horizontal gene transfer, zygomycetes

Citation Wurlitzer JM, Stanišić A, Wasmuth I, Jungmann S, Fischer D, Kries H, Gressler M. 2021. Bacterial-like nonribosomal peptide synthetases produce cyclopeptides in the zygomycetous fungus *Mortierella alpina*. Appl Environ Microbiol 87:e02051-20. https://doi.org/10.1128/AEM.02051-20.

Editor Ning-Yi Zhou, Shanghai Jiao Tong University

Address correspondence to Markus Gressler, markus.gressler@leibniz-hki.de.

Received 21 August 2020
Accepted 30 October 2020
Accepted manuscript posted online 6 November 2020
Published 15 January 2021

The increasing number of human-pathogenic microbes resistant against well-established antibiotics requires an unbiased search of novel producers of bioactive natural products. Bacteria (e.g., *Actinomyces* and *Streptomyces* spp.) and filamentous higher fungi (e.g., *Aspergillus*, *Penicillium*, and *Fusarium*) are well-investigated resources of natural products such as nonribosomal peptides (NRP), polyketides, and terpenoids (1–4). In contrast, early-diverging fungi—formerly combined in the paraphylum zygomycetes (5, 6)—have long been thought to lack secondary metabolites (7). Still, apart from the C_{18} terpenoid trisporic acid isolated from *Blakeslea trispora* and *Mucor mucedo* as a signaling compound during mating (8–10), secondary metabolites are rarely observed in zygomycetes. *Mortierella alpina*, a species of the subdivision Mucoromycotina, is a strain that is generally regarded as safe (GRAS) and serves as an industrial producer of polyunsaturated fatty acids such as arachidonic acid and linoleic acid (11–13).

Although lipid extracts of *M. alpina* are useful as immunomodulating leukotriene precursor molecules in pharmaceutical applications or as nutritional supplements in baby food (14), the secondary metabolome of the fungus has never been investigated in detail. Case reports on *M. alpina* species indicate the capacity to produce some oligopeptides, such as calpinactam (15) and Ro 09-1679 (16), harboring antimycobacterial and thrombin-inhibiting bioactivities, respectively (Fig. 1A). The metabolic potential of *M. alpina* was expanded by the recent discovery of a family of linear, acetylated hexapeptides, malpinins A to D, that have surface tension-lowering properties (Fig. 1A) (17, 18). Furthermore, the fungus produces the highly hydrophobic cyclopentapeptides known as malpibaldins (compounds 1 to 3) (17). Cyclopentapeptides have been isolated from various biological sources (ascomycetes, algae, and bacteria) and harbor diverse pharmaceutically relevant antiviral, antibiotic, apoptotic, and antiangiogenic properties (Fig. 1B) (19–25). Of interest, malpibaldins from *M. alpina* are structurally similar to luminmide B, a D-amino acid-containing cyclic pentapeptide from the insect-pathogenic gammaproteobacterium *Photorhabdus luminescens* (24). Biosynthetically, luminmide B is produced by a pentamodular NRP synthetase (NRPS), Plu3263, by sequential condensation of five aliphatic L-amino acids activated by specific adenylation (A) domains, and D-amino acids are introduced by the activity of dual epimerization/condensation (E/C) domains (24).

The biosynthetic origin of all above-mentioned peptides from *M. alpina* has never been illuminated. The publicly available 38.38-Mb genome of *M. alpina* (26) harbors 22 genes encoding polymodular NRPS or monomodular NRPS-like proteins, indicating its high biosynthetic potential (7). However, the frequent occurrence of D-amino acids in peptides of Mortierellales is biosynthetically surprising, since mainly bacteria (27), but rarely fungi (28–30), use D-amino acids as building blocks for NRP synthesis. Investigations on the mucoromycete *Rhizopus microsporus* revealed that its D-amino acid-containing peptides rhizonin and heptarhizin are produced not by the fungus but by its in-host endosymbiotic proteobacterium *Paraburkholderia* (syn. *Mycetohabitans*) *rhizoxinica* (31, 32). Similarly, *Burkholderia*-related endosymbionts (BRE) such as *Mycoavidus cysteinexigens* are frequently observed in *Mortierellales* (33, 34). In addition, nonculturable *Mycoplasma*-related obligate endosymbionts (MRE) have been identified by 16S ribosomal DNA (rDNA) sequencing in Mortierellomycotina (35). Therefore, we considered a possible involvement of endosymbionts in the secondary metabolism of *M. alpina*.

Here, we report the biosynthesis of malpibaldins and a novel set of cyclopentapeptides, called malpicyclins A to F (compounds 4 to 9), in *M. alpina*. The latter compounds are composed of D/L-Leu, D/L-Val, D-Phe, D-Trp, D-Tyr, and D-Arg and show moderate activity against Gram-positive bacteria. The production of both peptide families does not rely on fungus-associated bacteria. Instead, two fungal genes encoding pentamodular NRPSs are expressed during metabolite production and their amino acid sequence similarity to bacterial NRPSs suggests a genetic transfer across kingdoms. By combination of differential expression analysis, metabolite profiling, and *in silico* bioinformatic substrate binding studies, the NRPS genes *mpbA* and *mpcA* were specifically

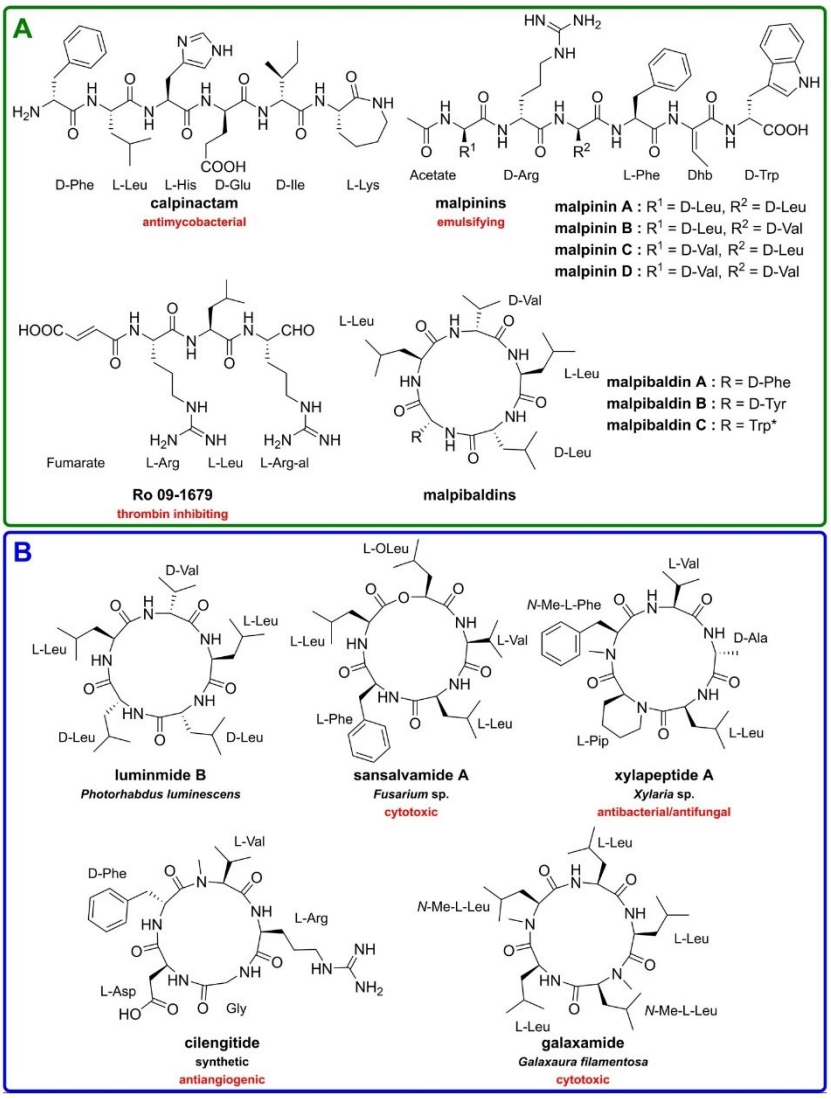
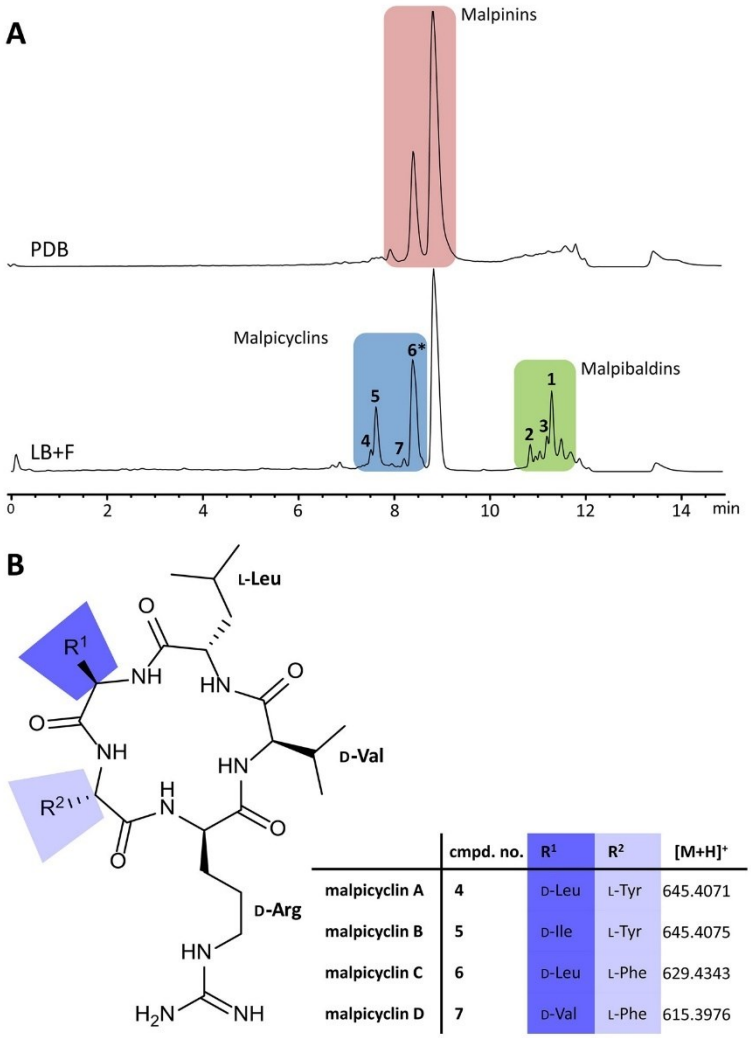


FIG 1 Linear and cyclic oligopeptides from *M. alpina* and other resources. (A) Oligopeptides isolated from *Mortierella alpina* (15–18). (B) Malpibaldin-related cyclopentapeptides from other sources (19, 21–24). *, configuration of the Trp moiety was not determined (17). L-Arg-al, L-arginal; Dhb, dehydrobutyrine; L-Pip, L-pipecolinic acid.

linked to the biosynthesis of malpibaldins and malpicyclins, respectively. Heterologous production of MpbA and MpcA modules in *Escherichia coli* and subsequent substrate turnover assays confirmed their activity.

RESULTS

Metabolic profiling of *M. alpina* revealed the malpicyclin family. The high abundance of the previously reported small peptides (17) prompted us to screen alternative cultivation conditions to induce oligopeptide production in *M. alpina*. Cultivation in potato dextrose broth (PDB) resulted solely in the production of malpinins (17) in the mycelium according to ultrahigh-performance liquid chromatography-mass spectrometry (UHPLC-MS) measurements (Fig. 2A). In contrast, when cultivated in



	cmpd. no.	R^1	R^2	$[M+H]^+$
malpicyclin A	4	D-Leu	L-Tyr	645.4071
malpicyclin B	5	D-Ile	L-Tyr	645.4075
malpicyclin C	6	D-Leu	L-Phe	629.4343
malpicyclin D	7	D-Val	L-Phe	615.3976

FIG 2 Identification of malpicyclins from *M. alpina* extracts. (A) UHPLC-MS profile of mycelial crude extracts from *M. alpina* after cultivation in potato dextrose broth (PDB) or modified lysogeny broth (LB+F, LB medium supplemented with 2% fructose) for 7 days (total ion chromatogram [TIC]). (B) Scheme of the chemical structure of malpicyclins A to D (compounds 4 to 7). *, signal overlap; malpicyclin C (compound 6) coelutes with malpinin B.

lysogeny broth supplemented with 2% fructose (LB+F), the previously reported malpibaldins A to C (compounds 1 to 3) (17) and an additional series of four masses $[M+H]^+$ of unknown nature (compounds 4 to 7) were detected (Fig. 2A). Upscaling of the culture to 4 liters and subsequent isolation of the metabolites by semipreparative HPLC enabled the identification of the four cyclopentapeptides malpicyclins A, B, C, and D (*m/z* 645.4074 $[M+H]^+$, *m/z* 645.4065 $[M+H]^+$, *m/z* 629.4343 $[M+H]^+$, and *m/z* 615.3976 $[M+H]^+$, compounds 4 to 7) by one-dimensional (1D) and 2D nuclear magnetic resonance (NMR) analysis (Fig. 2B, Table 1, Tables S1 to S4, and Fig. S1 to S27). ^{13}C NMR experiments and the calculated molecular formulae revealed by high-resolution mass spectrometry (HR-MS) suggested five carbonyl atoms and eight nitrogen atoms in compounds 4 to 7 and indicate pentapeptides that include a guanidinium group from arginine. MS/MS experiments and $^1H,^1H$ correlation spectroscopy (COSY)

TABLE 1 HR-ESI-MS data of isolated cyclopentapeptides from *M. alpina* ATCC 32222

Compound no.	Name	*m/z* [M + H]$^+$	Planar structure	Reference(s)
1	Malpibaldin A	586.3963	Cyclo(-L-Leu-D-Leu-D-Phe-L-Leu-D-Val-)	17
2	Malpibaldin B	625.4062	Cyclo(-L-Leu-D-Leu-D-Trp-L-Leu-D-Val-)	17
3	Malpibaldin C*	602.3912	Cyclo(-Leu-Leu-Trp-Leu-Val-)	17
4	Malpicyclin A (plactin B)	645.4074	Cyclo(-D-Leu-L-Tyr-D-Arg-D-Val-L-Leu-)	38, 39, 75
5	Malpicyclin B	645.4065	Cyclo(-D-Ile-L-Tyr-D-Arg-D-Val-L-Leu-)	This study
6	Malpicyclin C (plactin D)	629.4343	Cyclo(-D-Leu-L-Phe-D-Arg-D-Val-L-Leu-)	38, 39, 75
7	Malpicyclin D	615.3976	Cyclo(-D-Val-L-Phe-D-Arg-D-Val-L-Leu-)	This study
8	Malpicyclin E[a]	668.4228	Cyclo(-Leu-Trp-Arg-Val-Leu-)	This study
9	Malpicyclin F (MBJ-0174)	631.3959	Cyclo(-D-Val-L-Tyr-D-Arg-D-Val-L-Leu-)	18

[a]Absolute configuration of amino acids was not determined.

analysis suggested a cyclic ring structure which was confirmed by heteronuclear multiple-bond correlation (HMBC) and heteronuclear single quantum coherence (HSQC) 2D analysis experiments. Both COSY and HMBC data allowed the assignment of the amino acid side chains and revealed the planar structure: cyclo-(-Leu/Ile/Val-Tyr/Phe-Arg-Val-Leu-) (Fig. S6). The absolute configuration of the side chains was elucidated by the advanced Marfey's method (36, 37) (Table S5). D-Amino acids are solely incorporated at positions 1, 3, and 4 in compounds 4 to 7. Of note, besides D-Arg at position 3, L-Arg was confirmed as second building unit as evident by dual signals in Marfey's analysis (Table S5) and the presence of two symmetric scalar couplings (δ_H 1.63 and δ_H 1.45 in malpicyclin C) in HSQC spectra (Fig. S21). In addition to compounds 4 to 7, two by-products (compounds 8 and 9) were detected by HR-MS (Table 1 and Tables S1 and S2) but were produced in insufficient amounts for NMR analysis. The MS/MS fragmentation of compound 8 (malpicyclin E, *m/z* 668.4228 [M+H]$^+$) showed a pattern similar to that of compound 4 (Fig. S5) and suggested a tryptophan as the aromatic amino acid at the 3rd position (Table 1). According to HR-MS data, compound 9 (malpicyclin F, *m/z* 631.3959 [M+H]$^+$) is probably identical to the arginine-containing cyclopeptide MBJ-0174, previously isolated from *Mortierella alpina* strain f28740 (18).

Antimicrobial testing of compounds 4 to 7 revealed a moderate antibacterial activity against Gram-positive bacteria, with MIC values ranging from 97.3 to 357.8 μM, while Gram-negative representatives or fungi were not affected (Table 2 and Fig. S28 and S29). Interestingly, compounds 4 and 6 are structurally identical to the cyclopeptides plactin B and plactin D, which have formerly been isolated from an unspecified fungal strain, F165 (38). Plactin D showed blood plasma-dependent fibrinolytic activity by enhancing the prothrombin protease activity (39). Since amphiphilic malpinins have been described as surface-active metabolites, we also checked for tenside properties of malpicyclins by the ring tear-off method (17). Indeed, malpicyclins demonstrated a biphasic profile with a fast decrease of the surface tension up to a concentration of 62.5 μg/ml, followed by a slower decrease (Fig. S30). However, the calculated critical micelle concentration (CMC) of 93.9 μg/ml (147 μM) is 10-fold higher than that of malpinins (17), suggesting only a moderate contribution to *Mortierella*'s biosurfactant activities.

Malpicyclins and malpibaldins are fungal products. In 1984, the macrolactone antibiotic rhizoxin was isolated from the zygomycete *Rhizopus microsporus* and was

TABLE 2 MIC_{50}s of malpicyclins A to D tested against *Bacillus subtilis* and *Escherichia coli*

Compound no. or name	MIC_{50} (μM)	
	B. subtilis	*E. coli*
4	181.1 ± 89.6	>1,000
5	97.3 ± 16.5	>1,000
6	357.8 ± 124.4	>1,000
7	255.9 ± 15.2	>1,000
Ciprofloxacin[a]	0.022 ± 0.002	0.101 ± 0.003

[a]Ciprofloxacin served as a positive control.

seemingly of fungal origin (40). However, 23 years later, Hertweck's group demonstrated in pioneering work that the fungus harbors the endosymbiotic betaproteobacterium *P. rhizoxinica*, which that produces this macrocyclic polyketide (41). Analogously, *Mortierella elongata* can be infected by the endobacterium *Mycoavidus cysteinexigens*, whose genome encodes at least three NRPSs, while the host genome does not (34). Indeed, an approach screening 30 different *Mortierella* isolates revealed that 13% were infected by *Mycoavidus*-related endosymbionts (MRE) (34). To investigate whether cyclopentapeptides in *M. alpina* ATCC 32222 are of fungal or bacterial origin, the strain was repeatedly treated with an antibiotic cocktail known to cure potential infected hyphae (41). However, metabolite production—represented by quantifying the major compound of each metabolite class (malpicyclin C [compound 6] and malpibaldin A [compound 1])—was not affected by the antibiotic treatment, suggesting that the metabolite is produced by the fungus (Fig. S31). Moreover, a subsequent PCR amplification of bacterial 16S rDNA from treated and untreated fungal mycelium of *M. alpina* failed and pointed to the absence of endobacteria (Fig. S32). Both experiments indicate that (i) *M. alpina* isolate ATCC 32222 does not harbor endobacteria and (ii) the isolated metabolites are fungal rather than bacterial products.

Genome mining of *M. alpina* revealed the identification of cyclopeptide synthetases. A whole-genome survey of *M. alpina* (26) using antiSMASH (42) revealed 22 genes encoding NRPS and NRPS-like proteins in *M. alpina*. According to the molecular structure and D-amino acid distribution in malpicyclins and malpibaldins, two five-module NRPS with at least three epimerization domains (E) or dual epimerase/condensation domains (E/C) are required. Indeed, two candidate NRPS genes (*mpcA* and *mpbA*) were identified in the genome (Fig. 3). The potential NRPS gene product MpcA (5,532 amino acids [aa]) comprises a pentamodular NRPS with a scaffold C_s-A-T-E/C-A-T-C-A-T-E/C-A-T-E/C-A-T-TE, which includes an expected pattern of bacterial-like dual E/C domains (in modules 2, 4, and 5) as required for malpicyclin biosynthesis. The high similarity to bacterial domains facilitated the prediction of the *M. alpina* A domain specificities, which can be a difficult task for fungal domains. Analysis of putative substrate preference of the A domains by alignment with the GrsA A domain from *Aneurinibacillus migulanus* (43) suggested acceptance of hydrophobic amino acid substrates in all A domains of MpcA, with exception of domain A3 (Table 3 and Table S6). Here, hydrophilic residues (Ser and Thr) are found at positions 278 and 330 (GrsA numbering) of the NRPS code and, moreover, an Asp at position 331 might ion-pair with cationic amino acids such as Arg (44, 45), which is abundant in all malpicyclins (compounds 4 to 7). Hence, we hypothesized that a positively charged amino acid would be accepted by the MpcA A3 domain.

The second candidate gene encodes an NRPS (MpbA) with an identical domain pattern and a size (5,541 aa) similar to that of MpcA (73.6% aa sequence identity). However, *in silico* substrate specificity analysis revealed exclusively hydrophobic binding pockets in all five A domains, as expected for the malpibaldin building blocks (Table 3). MpbA A3 and MpbA A2 share similar NRPS codes, suggesting that both domains accept the same—probably aromatic—substrates. MpbA is structurally related to the bacterial luminmide B synthetase Plu3263 from *P. luminescens* (50.2% identity; 65.9% similarity), and both NRPSs may produce highly similar cyclopentapeptides NRP (Fig. 1 and 2). However, the specificity codes of the A domains are not identical (Table 3), suggesting that zygomycetous NRPSs use a dissimilar code.

In contrast to Plu3263, both MpcA and MpbA contain a starter condensation (C_S) domain-like N terminus (252 and 247 aa, respectively) known to transfer β-hydroxycarboxylic acid residues from acyl coenzyme A (acyl-CoA) donors to the N termini of bacterial lipopeptides (46). However, the domains are N-terminally truncated and the tandem histidine motifs (HH) responsible for the deprotonation of the substrate prior to the condensation step (47) is missing in the active sites of both domains. Taken together, the proposed domain structure and distribution of dual E/C domains fulfilled the requirement for incorporation of D-amino acids at the expected positions in

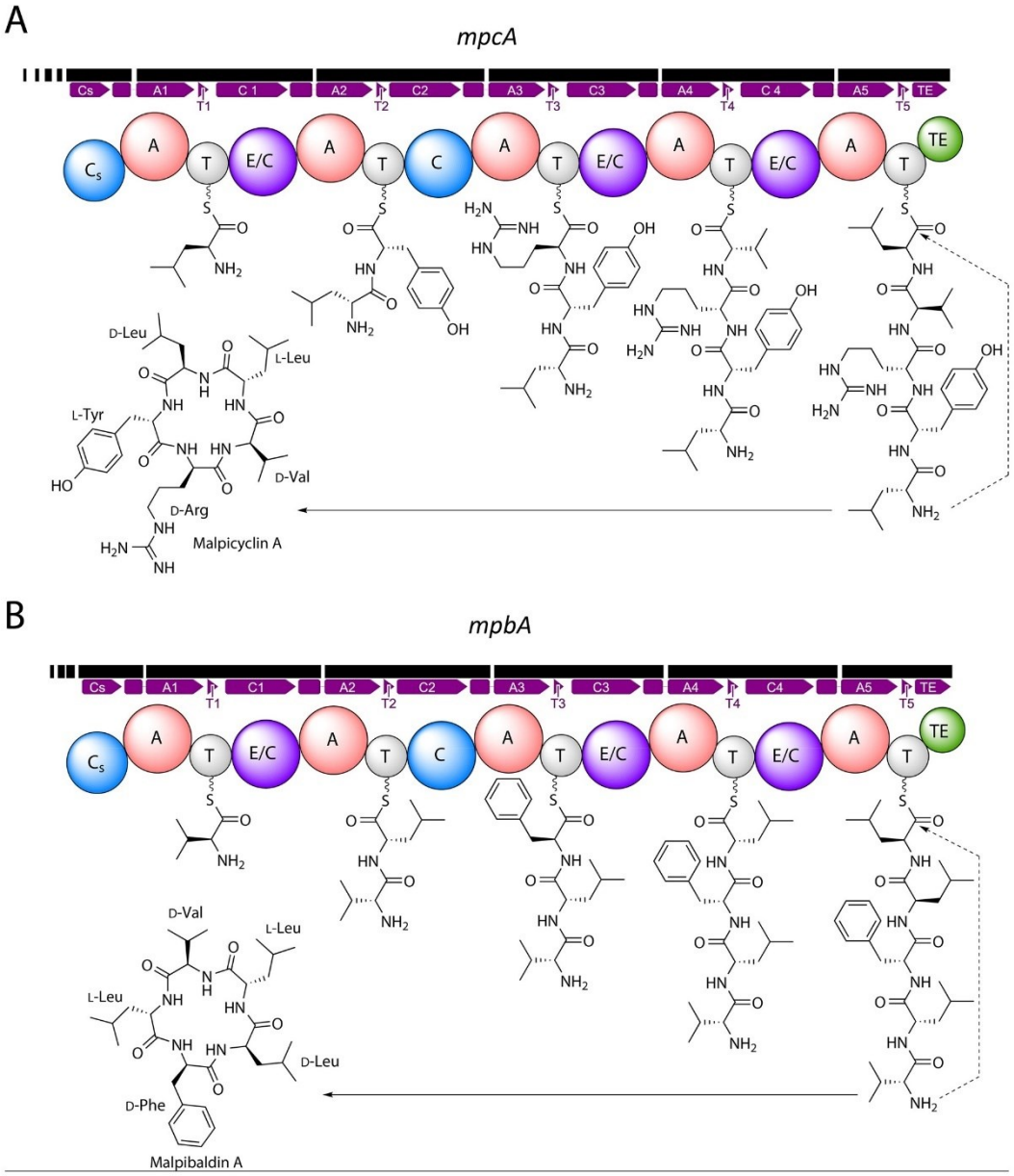


FIG 3 Proposed biosynthesis of malpicyclin A and malpibaldin A. (A) Gene structure of *mpcA* and domain structure of its encoded malpicyclin synthetase, MpcA. (B) Gene structure of *mpbA* and domain structure of its encoded malpibaldin synthetase, MpbA. A, adenylating domain; C condensation domain of the canonical $^{L}C_{L}$ type; E/C, dual epimerization/condensation domain; C_s, truncated starter condensation domain; T, thiolation domain; TE, thioesterase domain. Either gene (bold black lines) is interrupted by nine introns. Vertical white lines represent introns, black rectangles exons.

malpicyclins and malpibaldins and prompted us to study the malpicyclin synthetase gene (*mpcA*) and malpibaldin synthetase gene (*mpbA*) in detail.

Fungal NRPS genes *mpcA* and *mpbA* are coexpressed in the presence of fructose. To confirm active transcription of the NRPS genes, quantitative real-time PCR (qRT-PCR) expression experiments were performed. The fungus was cultivated in media

TABLE 3 NRPS codes for adenylating domains of the *Mortierella* NRPS (MpcA and MpbA) and related biochemically investigated bacterial and fungal NRPS modules[a]

Protein domain	Organism*	substrate	Residue position according to GrsA Phe numbering									
			235	236	239	278	299	301	322	330	331	527
MpcA_A2	*M. alpina*	Tyr/Phe	D	P	F	T	M	G	A	V	V	K
MpbA_A3	*M. alpina*	Trp/Tyr/Phe	D	P	F	V	M	G	G	T	V	K
Plu3263_A3	*P. luminescens*	Phe	D	A	W	C	I	A	A	V	C	K
BacC_A2	*B. subtilis*	Phe	D	A	F	T	V	A	A	V	C	K
CepA_A1	*A. orientalis*	Tyr	D	A	S	T	V	A	A	V	C	K
GliP_A1	*A. fumigatus*	Phe	D	G	S	I	L	G	A	C	A	K
MpbA_A1/ MpcA_A4	*M. alpina*	Val	D	A	F	W	L	G	G	T	F	K
MpcA_A1	*M. alpina*	Leu/Ile/Val	D	A	F	F	I	G	A	M	L	K
MpbA_A2	*M. alpina*	Leu	D	A	F	F	I	G	A	M	V	K
MpbA_A4/A5	*M. alpina*	Leu	D	A	I	F	L	G	A	T	I	K
MpcA_A5	*M. alpina*	Leu	D	A	M	F	I	G	G	T	I	K
Plu3263_A2/ A4/A5	*P. luminescens*	Leu	D	A	W	C	I	G	A	V	C	K
GrsB_A1	*A. migulanus*	Val	D	A	F	W	I	G	G	T	F	K
CssA_A9	*T. inflatum*	Val	D	A	W	M	F	A	A	V	L	K
MpcA_A3	*M. alpina*	Arg	D	A	A	S	V	G	A	T	D	K
SyrE_A5	*P. syringae*	Arg	D	V	A	D	V	C	A	I	D	K
McyC_A1	*M. aeruginosa*	Arg	D	V	W	T	I	G	A	V	D	K
PpzA-1_A2	*E. festucae*	Arg	D	V	S	D	T	G	A	P	T	K

[a]The proposed activated substrates are listed in column 3. The residues are mapped relative to *Aneurinibacillus migulanus* (former *Brevibacillus brevis*) GrsA-A numbering (81). NRPS codes were partially extracted from Bian et al. (82). Amino acid residues in the NRPS code are colored according to their physicochemical properties: acidic (red), small/hydrophobic (gray), aromatic/hydrophobic (amber), hydrophilic (green), and basic (blue). *, organisms are colored according to their phylogenetic origin: basal fungi (blue), higher fungi (brown), and bacteria (green). A detailed alignment is provided as Table S6.

to induce (LB+F) or repress (PDB) cyclopeptide production (Fig. 2A and 4A). Gene expression was determined after 48 h and 4 days of cultivation, and data were normalized against expression profiles derived from freshly germinated *M. alpina* mycelia, which do not produce secondary metabolites (data not shown). Neither NRPS gene *mpcA* nor *mpbA* was expressed in PDB medium, but expression was induced 5.4- or 8.2-fold, respectively, after 48 h of cultivation in LB medium compared to PDB medium. These results matched the observed enrichment of malpibaldin A (compound 7) and malpicyclin C (compound 3) in mycelial metabolite extracts obtained from cultivation in LB medium (Fig. 4B). In contrast, negligible amounts of both metabolites were found in mycelium after cultivation in noninducing PDB medium.

Adenylation domains of MpcA and MpbA activate L-amino acids. Unlike Aspergilli, Mortierellaceae are hardly genetically tractable, and targeted gene deletion occurs with very low frequency and genomic stability (48, 49). Hence, we verified the substrate specificity of the most characteristic modules (C-A-T) in both NRPSs, i.e., module 3 of MpcA (MpcA-m3, 122.45 kDa) and module 3 of MpbA (MpbA-m3, 122.40 kDa), by heterologous production in *E. coli* (Fig. S33). The purified double-His_6-tagged fusion proteins (50) were subjected to specificity testing by the recently established multiplexed hydroxamate-based adenylation domain assay (HAMA) (51), which detects the formation of stable amino acid hydroxamates after enzymatic adenylation by A domains. All proteinogenic L-amino acids and two of its enantiomer counterparts (D-Val and D-Phe) were tested in parallel.

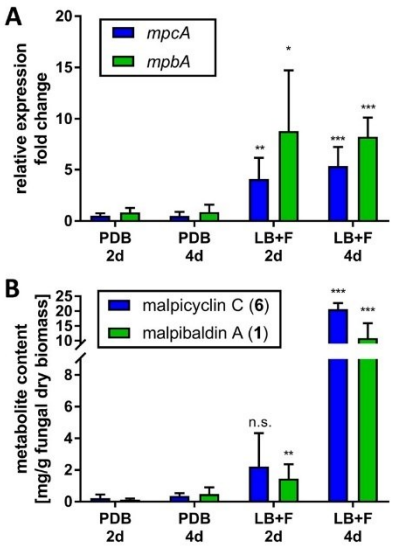


FIG 4 Gene expression of malpicyclin and malpibaldin synthetase genes, *mpcA* and *mpbA*, and metabolite production. *M. alpina* was cultivated under noninducing (PDB) and inducing (LB+F) conditions for up to 4 days. Expression analysis and metabolite quantification by UHPLC-MS were carried out 2 days and 4 days postinoculation (p.i.) (A) Expression analysis of *mpcA* and *mpbA*. Gene expression was normalized against the housekeeping genes *actA*, encoding α-actin, and *gpdA*, encoding the glyceraldehyde-3-phosphate dehydrogenase. cDNA from freshly germinated mitospores (in PDB) served as a reference (set to 1). (B) Metabolite quantification of the malpicyclin C (compound 6) and malpibaldin A (compound 1). Representative of each series, the amount of the most abundant metabolites (malpicyclin C and malpibaldin A) was determined by UHPLC-MS mass chromatograms (EIC) from *M. alpina* mycelium. Statistical significance compared to noninduced control (PDB, 2 days) is indicated as follows: n.s., not significant; *, $P \leq 0.05$; **, $P \leq 0.01$; and ***, $P \leq 0.001$ (paired Student's *t* test). Metabolite production and gene expression correlated with Pearson correlation coefficients of 0.84 ($P = 0.004$) for malpicyclin C/*mpcA* and 0.72 ($P = 0.02$) for malpibaldin A/*mpbA* at day 4.

For MpcA-m3, solely the corresponding L-Arg product formation was observed, confirming the findings of the NMR analysis and *in silico* substrate prediction (Fig. 5A). However, due to a lack of an authentic hydroxamate standard, incorporation of D-Arg could not be excluded. Hence, in a complementary experiment, substrate specificity was confirmed by an ATP-[$^{32}PP_i$] exchange assay (52) (Fig. S34 and Table S7). First, pools

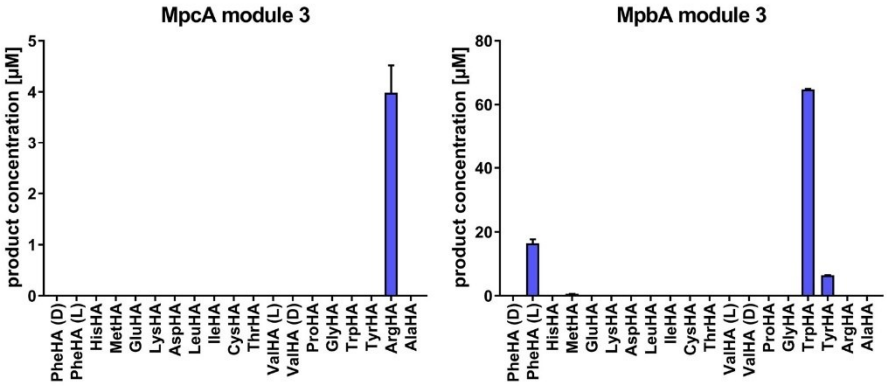


FIG 5 Substrate specificity testing of NRPS modules by the multiplexed hydroxamate assay (HAMA). Modules 3 (C-A-T) of MpcA (A) and of MpbA (B) were separately tested. Substrates were all proteinogenic amino acids (except L-Asn, L-Gln, and L-Ser), D-Phe, and D-Val. Amino acyl hydroxamates were quantified using HAMA ($n = 3$).

of physiochemically similar L-amino acids were tested, followed by a subsequent determination of single amino acids in the active pools. Again, MpcA-m3 showed the highest turnover with L-Arg (2.2×10^5 dpm). However, 11% of ATP-$^{32}PP_i$ conversion was also observed for D-Arg (2×10^4 dpm), suggesting that the A domain is not entirely enantioselective. Poor discrimination against D-amino acids is not surprising when the D-enantiomers are not present as competing substrates. For instance, strong side activities for D-Phe and D-homoserine (D-Hse) have been reported for usually L-amino acid-activating A domains in tyrocidine and ralsolamycin biosynthesis, respectively (52, 53).

The HAMA of MpbA-m3 revealed a promiscuous acceptance of aromatic amino acids, with preference L-Trp > L-Phe > L-Tyr (Fig. 5B). These side chain identities perfectly match residues found in malpibaldins C, A, and B, respectively, but are not in line with the preference for product formation. In *M. alpina*, malpibaldin A (L-Phe derivative) is the major compound (17). Such discrepancies between adenylation preference and relative product formation rate may be caused by side chain specificity of downstream reaction steps or differences in intracellular amino acid availability (51). Promiscuous adenylation leading to the parallel production of multiple products from one NRPS assembly line has been well studied in cyanobacterial enzymes and may be an important springboard for evolutionary diversification (45, 54).

From this knowledge, we predicted the biosynthesis of malpicyclins and malpibaldins by successive activation and, if required, racemization of L-amino acids, which are subsequently condensed to give a nascent, linear pentapeptide tethered to the T domains of the enzymes (Fig. 3). The final cyclization is probably catalyzed in *cis* by the C-terminal type I thioesterase (TE) domain, which has been demonstrated for several cyclopeptides from bacteria (52, 55, 56).

NRPS genes of *M. alpina* are of (endo)bacterial origin. Since no zygomycetous NRPS has been identified before, we were interested in the evolutionary origin of the genes and enzymes. A phylogenetic analysis of the extracted A domains both from *M. alpina* NRPS proteins and from verified fungal and bacterial NRPS, NRPS-like, and PKS-NRPS hybrid proteins was performed. Surprisingly, the zygomycetous A domains clustered in a monophylum with bacterial—but not with fungal—A domains independent of substrate specificity (Fig. 6 and Table S8). Moreover, the A domains share high similarity to A domains from NRPSs of endobacteria such as *Paraburkholderia* or *Mycoavidus*, known to infect zygomycetes, but are more distantly related to other Gram-negative (*Pseudomonas* or *Ralstonia*) or Gram-positive representatives (*Streptomycetes*). Hence, both *Mortierella* NRPSs are most likely of (endo)bacterial but not fungal origin and may have been evolved independently of the fungal counterparts.

A phylogenetic analysis based on the more conserved condensation (C) domains showed a similar outcome (Fig. S35 and S36 and Tables S9 and S10). Most interestingly, C domains from *M. alpina* NRPSs fall into three groups: $^{L}C_L$, common C domains condensing L-amino acids; C_s, N-terminal starter condensation domains; and dual E/C domains, epimerizing enzyme-tethered L-amino acids to their D-enantiomers prior to condensation. The canonical $^{L}C_L$ domains from *Mortierella* cluster with both bacterial and fungal $^{L}C_L$ domains (Fig. S35 and S36). In addition, the truncated C_s domains have most likely evolved from *Mortierella* $^{L}C_L$ domains and show no similarity to acyl-transferring C_s domains found in bacterial lipopeptide NRPSs (Fig. S35). In contrast, the six dual E/C domains from MpcA and MpbA cluster with those of endobacteria such as *Paraburkholderia* (Fig. S35), but they have no close counterpart among fungal C domains (Fig. S36). To date, dual E/C domains have been found solely in bacteria and not in fungi. Indeed, the zygomycetous E/C domains occupy a unique, separate C domain clade among the fungal kingdom. In addition, BLASTP analyses using MpcA and MpbA as queries identified putative NPRS candidates from both bacteria (*Photorhabdus*) and early-diverging fungi (*Basidiobolus meristosporus* and *Mortierella verticillata*) (Tables S11 to S14).

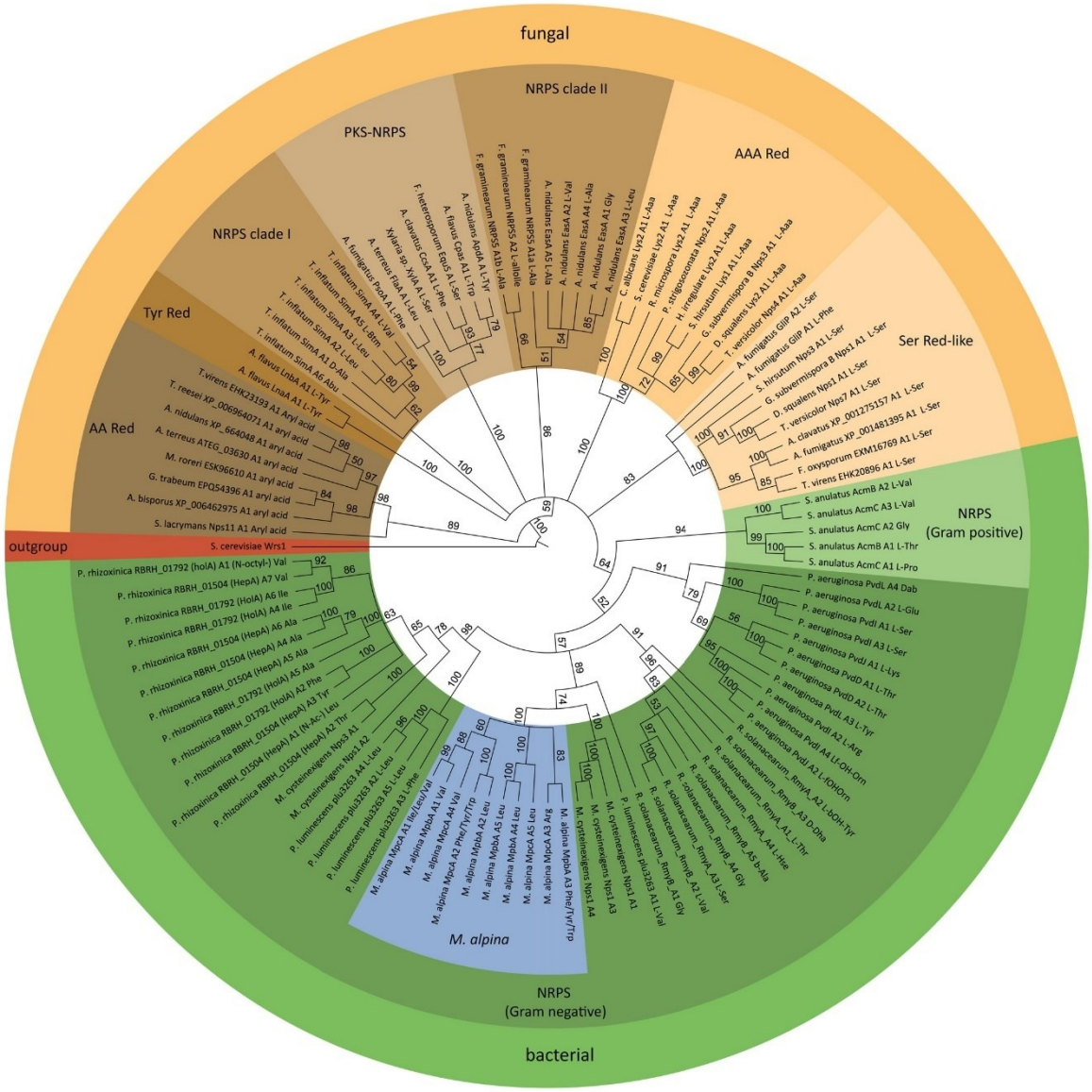


FIG 6 Phylogenetic analysis of A domains from *M. alpina* and other fungal or bacterial representatives. The A domains were extracted from NRPSs, NRPS-like proteins, and PKS/NRPS hybrids from *M. alpina* (blue), (endo)bacteria (green), and higher fungi (taupe/amber) (refer to Table S8). The A domain of the cytoplasmatic tryptophanyl-tRNA synthetase Wrs1 from *Saccharomyces cerevisiae* served as the outgroup (red). Note that the zygomycetous A domains from *M. alpina* cluster together with bacterial A domains. The percentual bootstrap support is labeled next to the branches. AAA red, α-2-aminoadipate reductase; AA Red, aryl acid reductase; NRPS, nonribosomal peptide synthetase; PKS-NRPS, polyketide synthase-nonribosomal peptide synthetase hybrid; Ser Red, serine reductase; Tyr Red, tyrosine reductase.

Taking all these findings together, we postulate a probable horizontal gene transfer (HGT) from an unknown bacterial endosymbiont to the *Mortierella* host. However, the GC contents of *mpcA* (55.0%) and *mpbA* (54.6%) are similar to that of the *M. alpina* genome (51.8%), and the codon usage of their coding sequences is nearly identical to that of the housekeeping genes from *M. alpina*. Both findings suggest an early HGT event in *Mortierella*.

DISCUSSION

Zygomycetes of the order Mortierellales are an established resource for enzymes in industrial detergent manufacturing or for polyunsaturated fatty acids in the food industry. Recent publications revealed highly bioactive compounds such as the antiplasmodial cycloheptapeptide mortiamide A (57) or the surface-active hexapeptide malpinin A (17). Hence, this fungal order seems to be a prolific resource of pharmaceutically relevant natural products, too. However, the biosynthetic basis of the peptide compounds from zygomycetes has never been investigated. We provide here evidence that secondary metabolite genes are actively transcribed in zygomycetes and encode functional enzymes, called zygomycetous NRPSs.

The zygomycetous NRPSs MpcA and MpbA combine properties of both bacterial and fungal NRPSs. A bacterial origin of the NRPS genes is likely since they encode A and C domains, which exclusively cluster with bacterial representatives. Most of the known NRP from *Mortierella* species use D-amino acids as building blocks. To incorporate D-amino acids into the NRP backbone, fungi require a separate, pre-NRPS-acting amino acid racemase as shown for the biosynthesis routes of cyclosporine in *Tolypocladium niveum* or HC toxin in *Cochliobolus carbonum* (28, 58). Alternatively, fungal NRPSs possess a dedicated epimerase domain N-terminally located to a $^{D}C_{L}$ domain (E-C didomain) that solely accepts D-amino acids as the substrate, i.e., in the fusaoctaxin A synthetase NRPS 5 from *Fusarium graminearum* (29). Recently, Tang's group demonstrated that an E-C didomain in the single-module NRPS IvoA catalyzes the ATP-dependent stereoinversion of L- to D-tryptophan, the precursor for the conidiophore pigment in *Aspergillus nidulans* (59). Instead of E-C didomains, zygomycetous NRPSs use bacterial-like dual E/C domains combining both catalytic activities, i.e., epimerization and condensation, in one single domain. Bacterial dual E/C domains are characterized by two consecutive histidine repeats (HH) in condensation domain motifs C1 (HH[I/L]XXXXGD) and C3 (HHXXXGDH) which are required for epimerization and condensation, respectively (60). Similar catalytic HH motifs are present in the *Mortierella* dual E/C domains (C1 ,HH[M/L][M/L]A[T/A]EGD, and C3, HH[I/L][IV][G/I]DH), suggesting a functional racemization.

Another distinctive feature of the zygomycetous NRPSs is the N-terminally truncated starter C-domain (C_s) in both MpcA and MpbA. *Mortierella* C_s domains do not cluster with β-hydroxy acyl-transferring C_s domains from lipopeptide-producing NRPSs from bacteria (46). Instead, our phylogenetic analysis indicated that these domains have evolved from the *Mortierella* canonical $^{L}C_{L}$ domains. Truncated C_s domains with unknown function were also found in fungal NRPSs such as the cyclosporine synthetase SimA from *Tolypocladium inflatum* (61), the tryptoquialanine synthetase TquA from *Penicillium aethiopicum* (62), and the pyrrolopyrazine synthetase PpzA-1 from *Epicholoë festucae* (63) and are thought to be an evolutionary relic required to maintain A domain stability (64). In bacterial macrocyclic-producing NRPSs, the final peptide is cyclized by a terminal *cis*-acting type I TE domain, which is often replaced by a specialized cyclase-like C domain (C_T) in fungi (30). Both domains preferably catalyze a head-to-tail macrolactamization of peptides with D- and L-configured residues at the N and C termini, respectively (65). Hence, the presence of C-terminal TE domains in MpcA and MpbA and the presence of a dual E/C domain at their N termini point at a bacterial-like cyclization mechanism in zygomycetous NRPSs.

Surprisingly, based on the amino acid sequence, MpbA is highly related to bacterial NRPSs, such as the luminmide B synthetase Plu3263 from *P. luminescens*. Both NRPSs produce highly similar NRP with identical amino acid configurations (24). The A domains of Plu3263 are extraordinary flexible and accept a variety of (non)proteinogenic amino acid substrates resulting in the production of novel luminmide variants (luminmides C to I) by simple substrate feeding (66). Similarly, in *M. alpina* quantity and quality of malpibaldins and malpinins are dependent on the amino acids supplied in the medium (17), which was confirmed by the promiscuity of the MpbA A3 domain accepting at least three aromatic substrates. Interestingly, the NRPS code of the

zygomycetous A domains shows higher similarity to bacterial than to fungal A domains. For example, the arginine-activating A3 domain of MpcA comprises a guanidinium-stabilizing aspartate residue which is also present in the arginine-adenylating A domains of the bacterial NRPSs from *Pseudomonas syringae* pv. syringae (67) and *Microcystis aeruginosa* (54) but is absent in fungal NRPSs such as from the ascomycete *E. festucae* (63).

Taking all these finding together and considering the fact that *Mortierella* strains are frequently infected with NRPS-encoding proteobacteria from the genus *Mycoavidus* (35), a horizontal gene transfer from an (endo)bacterial symbiont into the fungal host is highly plausible. A similar phenomenon is postulated for the L-δ-(α-aminoadipoyl)-L-cysteinyl-D-valine synthetase (ACVS) genes responsible for β-lactam antibiotic biosynthesis: the ACVS-like NRPS genes from *A. nidulans* and *Penicillium chrysogenum* have most likely arisen from *Lysobacter*-like or *Streptomyces*-like bacterial ancestors (68–70). In phylogenetic analyses, A domains of fungal ACVS-NRPS clustered with bacterial representatives into one unique monophyletic group (71), similarly to MpcA and MpbA with endobacterial NRPSs in this study. However, neither codon usage nor GC content of the *Mortierella* NRPS genes matches that of the endosymbiotic *Mycoavidus* genes but are highly similar to that of *Mortierella* housekeeping genes. As expected for fungal genes, both *mpcA* and *mpbA* contain introns of an average size of 102 bp, indicating a comparably early gene transfer during evolution of *M. alpina* and a subsequent adaptation of the genes to the requirements of the eukaryotic mRNA processing (72). Recently, HGT events have been postulated as the main driver of secondary metabolism diversity in the zygomycetous genus *Basidiobolus*, as supported by the phylogenetic reconstructions of NRPS gene clusters with bacterial homologs (73). *Basidiobolus* species are common inhabitants of the amphibian gut and, similar to *Mortierella* species, live in close association with proteobacteria (74).

The ecological function of zygomycetous NRPs is still to be deciphered. However, from a pharmaceutical angle, malpicyclins A and C are structurally identical to the cyclopentapeptide plactins B and D, respectively, which were previously isolated from an uncharacterized fungus (38, 75) and exhibit fibrinolytic activities by elevating the activity of cellular urokinase-type plasminogen activator (39). The anticoagulating effects of plactins and derived cyclopentapeptides such as malformin A_1 are under investigation in treatment of thrombic disorders (76). Hence, *M. alpina* is not only of nutritional benefit by production of polyunsaturated fatty acids but also of pharmaceutical interest as a producer of bioactive natural products. Moreover, the abundance of natural products in *M. alpina* under certain growth conditions raises safety concerns regarding the biotechnological use of this strain.

In sum, this report disproves a long-standing dogma of a marginal secondary metabolism in zygomycetes and, instead, establishes Mortierellales as a promising, previously overlooked reservoir for bioactive metabolites.

Conclusion. Zygomycetes have long been industrially used as producers of long-chain unsaturated fatty acids, but their use as a resource for natural products has not been investigated yet. Here, we report on early-diverging fungi as a novel resource of bioactive compounds and demonstrated that their genomes encode functional secondary metabolite genes. The two cyclopentapeptide synthetases, MpcA and MpbA, from *M. alpina* are responsible for malpicyclin and malpibaldin biosynthesis, respectively. The surprising structural and mechanistic similarity to bacterial nonribosomal peptide synthetases (NRPSs) points to an endobacterial origin of *Mortierella* NRPS genes that differ from their asco- and basidiomycete counterparts and may have evolved independently.

MATERIALS AND METHODS

Organisms and culture maintenance. The fungal strain *Mortierella alpina* ATCC 32222 was purchased from the American Type Culture Collection (ATCC). Cultures were maintained on MEP agar plates (30 g/liter of malt extract, 3 g/liter of soy peptone, 18 g/liter of agar) for 7 days at 25°C. For antibiotic treatment of *M. alpina* and subsequent analyses (16S rDNA detection, metabolite quantification), refer to the supplemental experimental procedures. *Bacillus subtilis* subsp. *subtilis* 168 (DSM 23778) and *Esche-*

TABLE 4 Oligonucleotides used in this study

Name	Sequence (5′–3′)	Target	Purpose/restriction site	Primer efficiency (R^2)	Amplicon length cDNA/gDNA
oITS1	TCCGTAGGTGAACCTGCGG	ITS region	Strain identification		
oITS4	TCCTCCGCTTATTGATATGC	ITS region	Strain identification		
oJMW01	CATCGATCTGGCCTACATGG	*gpdA*	qPCR	1.94 (0.999)	80/229
oJMW02	CCACCTTGCCCTTGTAGC	*gpdA*	qPCR	1.94 (0.999)	80/229
oJMW03	GTATGTGCAAGGCCGGTTTCG	*actB*	qPCR	2.10 (0.998)	100/224
oJMW04	CCCATACCGACCATCACACC	*actB*	qPCR	2.10 (0.998)	100/224
oJMW30	GCCACTGCATTGGACTTGC	*mpbA/nps15*	qPCR	1.90 (0.991)	80/140
oJMW31	CCTCTTTGCTTTGCAGTTCGG	*mpbA/nps15*	qPCR	1.90 (0.991)	80/140
oJMW55	GCATAAATTGGTCCACGCTG	*mpcA/nps16*	qPCR	1.92 (0.999)	173/173
oJMW56	CGCTGTCCTCGACGATGAAC	*mpcA/nps16*	qPCR	1.92 (0.999)	173/173
oJMW63	CACGATGATCAAACCCTATCAATCATTC	*mpcA* module 3	Amplification		
oJMW64	CTGTTGAAGGTCGATGAGTGGC	*mpcA* module 3	Amplification		
oJMW67	TATATATATAGGCTAGCATGCACGATGATCAAACCCTATC	*mpcA* module 3	pET28 vector cloning (NheI)		
oJMW68	TATATATACTATGCGGCCGCCTGTTGAAGGTCGATGAGTG	*mpcA* module 3	pET28 vector cloning (NotI)		
oIW009	CAGGATGATCAGTCGCACAAC	*mpbA* module 3	Amplification		
oIW010	CAGGTGTCTCTGGCACTTCAC	*mpbA* module 3	Amplification		
oIW011	TATATAGCTAGCCAGGATGATCAGTCGCACAAC	*mpbA* module 3	pET28 vector cloning (NheI)		
oJMW12	ATATATGCGGCCGCAGGTGTCTCTGGCACTTC	*mpbA* module 3	pET28 vector cloning (NotI)		
oMG342	AGAGTTTGATCCTGGCTCAG	16S-rDNA	Endobacterial rDNA		
oMG343	CGGTTACCTTGTTACGACTT	16S-rDNA	Endobacterial rDNA		

richia coli DSM 498 were maintained on LB agar plates at 37°C. *E. coli* XL1-Blue (Agilent) and *E. coli* KRX (Promega) were used for plasmid propagation and for protein production, respectively, and were maintained in LB medium (5 g/liter of yeast extract, 10 g/liter of tryptone, 10 g/liter of sodium chloride) supplemented with 50 μg/ml of carbenicillin or 100 μg/ml of kanamycin (both Sigma-Aldrich), if applicable.

Chemical analysis and metabolite structure elucidation. (i) General. All chromatographic methods are summarized in Table S15. UHPLC-MS measurements of compounds 1 to 9 were carried out on an Agilent 1290 infinity II UHPLC coupled with an Agilent 6130 single quadrupole mass spectrometer (positive ionization mode) using methods A and B (Table S15) for routine metabolite detection and for metabolite quantification. High-resolution mass spectra and MS/MS fragmentation patterns of compounds 1 to 8 were recorded using a Q Exactive Plus mass spectrometer (Thermo Scientific). Chromatography for determination of amino acyl hydroxamate was performed on a Waters Acquity H-class UPLC system coupled to a Xevo TQ-S micro (Waters) tandem quadrupole instrument with electrospray ionization (ESI) source in positive ion mode (desolvation gas, N_2; collision gas, Ar; capillary voltage, 1.5 kV; cone voltage, 65 V; desolvation temperature, 500°C; desolvation gas flow, 1,000 liters/h). NMR spectra were recorded on a Bruker Avance III 600-MHz spectrometer at 300 K using dimethyl sulfoxide (DMSO) as the solvent and internal standard. Peaks were adjusted to δ_H 2.49 ppm and δ_C 39.5 ppm.

(ii) Metabolite isolation, structure elucidation, and antibiotic testing. Eight flasks with 500 ml of LB medium amended with 2% (wt/vol) fructose were inoculated with six agar blocks (2 by 2 mm) of *M. alpina* grown on MEP agar. After 7 days of cultivation at 160 rpm and 25°C, mycelium was harvested, resuspended in 1 liter of methanol/butanol/DMSO (12:12:1), and homogenized using a blender (Ultra-Turrax; IKA). The extract was filtered, and extraction of the fungal biomass was repeated. Both extracts were pooled and evaporated to dryness. The residue was resuspended in 25 ml of methanol/DMSO (10:1) and subjected to an Agilent Infinity 1260 preparative HPLC equipped with a Luna C_{18} column (250 by 21.2 mm, 10 μm; Phenomenex). The metabolites were separated according to method C (Table S15) (t_R = 8 to 10 min). Subsequently, the compounds were purified using method D (Table S15) on an Agilent 1200 HPLC system. Purified compounds were dissolved in DMSO-d_6 for NMR analyses. Absolute configurations of amino acids in the peptides 4 to 7 were determined using Marfey's method (see supplemental experimental procedures and method E in Table S15). Antimicrobial activities were determined by agar diffusion plates according to a published procedure (17), and MIC analysis was carried out as described in the supplemental experimental procedures.

NRPS identification and expression analysis. The genome of *M. alpina* ATCC 32222 was accessed from NCBI (National Center for Biotechnology Information) under Assembly accession number GCA_000240685.2. Putative NRPS genes were annotated using the fungal antiSMASH 5.0 software (42), and if required, putative intron-exon junctions were curated manually by alignment to fungal/bacterial NRPSs (Tables S8 to S10). Expression primers for *mpcA* and *mpbA* as well as housekeeping genes encoding β-actin (*actB*) and the glyceraldehyde-3-phosphate dehydrogenase (*gpdA*) were designed (cutoff PCR efficiency of at least 0.95) (Table 4). *M. alpina* was grown in LB medium amended with 2% of fructose (LB+F medium) or in potato dextrose broth (PDB; Sigma-Aldrich) at 160 rpm and 25°C for up to 4 days. RNA was extracted using the SV total RNA isolation system (Promega), and residual genomic DNA (gDNA) was digested with Baseline-Zero DNase (Biozym). cDNA was synthesized with RevertAid reverse transcriptase (Thermo Fisher) using oligo$(dT)_{18}$ primers. For quantitative real-time PCR (qRT-PCR), the qPCR Mix EvaGreen (BioSell) was used in a qPCR Cycler qTower³ (Analytik Jena) following the manufacturer's instructions and qPCR protocol: initiation at 95°C for 15 min, followed by 40 amplification

TABLE 5 Plasmids used in this study

Plasmid	Vector backbone	Purpose	Gene product	Source or reference
pJET1.2		Amplification of DNA		Thermo Fisher
pET28a (+)		Expression vector		Agilent
pJMW007	pET28a (+)	Expression of *mpcA*-M3	MpcA module 3	This study
pJMW023	pET28a (+)	expression of *mpbA*-M3	MpcA module 3	This study

cycles (95°C, 15 s; 60°C, 20 s; and 72°C, 20 s) and final recording of a melting curve (60 to 95°C). Expression data were calculated according to the threshold cycle ($\Delta\Delta C_T$) method by Pfaffl (77) using the housekeeping genes as internal, nonregulated reference controls.

Heterologous protein production and determination of enzymatic activity of A domains. For detailed cloning procedures and protein production protocols, refer to the supplemental experimental procedures and Tables 4 and 5. In brief, intron-free coding sequences of MpcA module 3 (*mpcA*-M3) and MpbA module 3 (*mpbA*-M3) were amplified from cDNA and ligated into the blunt pJET1.2 vector system (Thermo) prior to final subcloning into pET28a(+) expression vectors. NRPS modules were produced in *E. coli* KRX (Promega) at 16°C using 0.1% L-rhamnose as the inducer. Proteins were purified from cell-free lysate by metal ion affinity chromatography with Protino nickel-nitrilotriacetic acid (Ni-NTA) agarose (Macherey-Nagel) as the matrix, followed by ultrafiltration (Amicon Ultra-15 centrifugal filter units; Merck).

ATP-[^{32}P]PP$_i$ exchange assay. The assay was carried out as previously described (52, 64) using 5 nM MpcA-m3 in a 100-μl reaction mixture. Initially, pools of L-amino acids were used as substrates (Table S7), followed by testing of single substrates.

Multiplexed hydroxamate based adenylation domain assay (HAMA). The hydroxamate formation assay was conducted as previously described (51). In brief, the assay was carried out at room temperature in a 100-μl volume containing 50 mM Tris (pH 7.6), 5 mM $MgCl_2$, 150 mM hydroxylamine (pH 7.5 to 8, adjusted with NaOH), 5 mM ATP (A2383; Sigma), 1 mM tris(2-carboxyethyl)phosphine (TCEP), and 1 μM enzyme. Reactions were started by adding a mixture of 5 mM amino acids in 100 mM Tris (pH 8) to a final concentration of 1 mM or only buffer as a control. L-Phe, L-Val, and L-Leu were distinguished from D-Phe, D-Val, and L-Ile, respectively, by using enantiopure, deuterium-labeled standards. Reactions were quenched by 10-fold dilution in acetonitrile containing 0.1% formic acid and subjected to UPLC-MS analysis (method F in Table S15). Compounds were detected via specific mass transitions recorded in multiple reaction monitoring (MRM) mode. Data acquisition and quantitation were conducted using the MassLynx and TargetLynx software (version 4.1). Quantitation was done by external calibration with standard solutions of hydroxamates ranging from 0.0032 to 10 μM.

Phylogenetic analysis. The genomes of the zygomycete *M. alpina* ATCC 32222 and the endofungal bacteria *Mycoavidus cysteinexigens* AG77 and *Paraburkholderia* (syn. *Mycetohabitans*) *rhizoxinica* HKI 0454^T were obtained from the JGI fungal genomics resource database or NCBI genome database and were subjected to a screening analysis for fungal and bacterial secondary metabolite gene clusters by the antiSMASH 5.0 software (42). The A and C domains were extracted from putative enzymes and, additionally, from experimentally proven endobacterial, bacterial, and fungal NRPS and NRPS-like enzymes (Tables S8 to S10). For A domain phylogeny, a total set of 108 amino acid sequences (Table S8) were aligned using the ClustalW algorithm implemented in the Geneious 10.2.4 software. For phylogenetic analysis of C domains, altogether 225 bacterial and 90 fungal sequences of C domains of verified NRPS and NRPS-like proteins (Tables S9 and S10) were aligned by MAFFT version 7 using the E-INS-i algorithm and BLOSUM62 scoring matrix (78). The evolutionary history was inferred using the Neighbor-Joining method (79). The evolutionary distances were computed using the Jukes Cantor genetic distance model implemented in the MEGA X software (80). A bootstrap support of ≥50% is given for 1,000 replicates each.

Data availability. The sequence of full-length transcripts of *mpcA* and *mpbA* were deposited in GenBank (accession numbers MT800760 and MT800759).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 7.4 MB.

ACKNOWLEDGMENTS

We are grateful to Heike Heinecke and Andrea Perner (both from Hans Knöll Institute [HKI], Jena, Germany) for their technical assistance in recording NMR and HR-MS/MS spectra, respectively. We thank Kerstin Voigt (Jena Microbial Resource Collection, Jena, Germany) for initial antimicrobial testing and Sarah Niehs (HKI) for donation of genomic DNA from *R. microsporus* and *P. rhizoxinica*.

Hajo Kries gratefully acknowledges a fellowship from the Daimler und Benz Foundation and financial support from the DFG (grant number 441781663) and the Fonds der Chemischen Industrie (FCI).

REFERENCES

1. Smith DJ, Burnham MK, Edwards J, Earl AJ, Turner G. 1990. Cloning and heterologous expression of the penicillin biosynthetic gene cluster from *Penicillium chrysogenum*. Biotechnology (NY) 8:39–41. https://doi.org/10.1038/nbt0190-39.
2. Campbell CD, Vederas JC. 2010. Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. Biopolymers 93:755–763. https://doi.org/10.1002/bip.21428.
3. Gajdacs M. 2019. The continuing threat of methicillin-resistant *Staphylococcus aureus*. Antibiotics (Basel) 8:e52.
4. Weber JM, Wierman CK, Hutchinson CR. 1985. Genetic analysis of erythromycin production in *Streptomyces erythreus*. J Bacteriol 164:425–433. https://doi.org/10.1128/JB.164.1.425-433.1985.
5. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lucking R, Thorsten Lumbsch H, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalpers JA, Vilgalys R, Aime MC, Aptroot A, Bauer R, Begerow D, Benny GL, Castlebury LA, Crous PW, Dai YC, Gams W, Geiser DM, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G, Hosaka K, Humber RA, Hyde KD, Ironside JE, Koljalg U, Kurtzman CP, Larsson KH, Lichtwardt R, Longcore J, Miadlikowska J, Miller A, Moncalvo JM, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, et al. 2007. A higher-level phylogenetic classification of the Fungi. Mycol Res 111:509–547. https://doi.org/10.1016/j.mycres.2007.03.004.
6. Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, James TY, O'Donnell K, Roberson RW, Taylor TN, Uehling J, Vilgalys R, White MM, Stajich JE. 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia 108:1028–1046. https://doi.org/10.3852/16-042.
7. Voigt K, Wolf T, Ochsenreiter K, Nagy G, Kaerger K, Shelest E, Papp T. 2016. Genetic and metabolic aspects of primary and secondary metabolism of the Zygomycetes, p 361–385. *In* Hoffmeister D (ed), The Mycota, vol III. Biochemistry and molecular biology, 3rd ed. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-319-27790-5_15.
8. Sutter RP. 1970. Trisporic acid synthesis in *Blakeslea trispora*. Science 168:1590–1592. https://doi.org/10.1126/science.168.3939.1590.
9. Sutter RP, Capage DA, Harrison TL, Keen WA. 1973. Trisporic acid biosynthesis in separate plus and minus cultures of *Blakeslea trispora*. J Bacteriol 114:1074–1082. https://doi.org/10.1128/JB.114.3.1074-1082.1973.
10. Sahadevan Y, Richter-Fecken M, Kaerger K, Voigt K, Boland W. 2013. Early and late trisporoids differentially regulate beta-carotene production and gene transcript levels in the mucoralean fungi *Blakeslea trispora* and *Mucor mucedo*. Appl Environ Microbiol 79:7466–7475. https://doi.org/10.1128/AEM.02096-13.
11. Bajpai PK, Bajpai P, Ward OP. 1991. Production of arachidonic acid by *Mortierella alpina* ATCC 32222. J Ind Microbiol 8:179–185. https://doi.org/10.1007/BF01575851.
12. Bajpai PK, Bajpai P, Ward OP. 1991. Arachidonic acid production by fungi. Appl Environ Microbiol 57:1255–1258. https://doi.org/10.1128/AEM.57.4.1255-1258.1991.
13. Kawashima H, Akimoto K, Higashiyama K, Fujikawa S, Shimizu S. 2000. Industrial production of dihomo-gamma-linolenic acid by a Delta-5-desaturase-defective mutant of *Mortierella alpina* 1S-4 fungus. J Am Oil Chem Soc 77:1135–1138. https://doi.org/10.1007/s11746-000-0178-2.
14. Tallima H, El Ridi R. 2018. Arachidonic acid: physiological roles and potential health benefits—a review. J Adv Res 11:33–41. https://doi.org/10.1016/j.jare.2017.11.004.
15. Koyama N, Kojima S, Nonaka K, Masuma R, Matsumoto M, Omura S, Tomoda H. 2010. Calpinactam, a new anti-mycobacterial agent, produced by *Mortierella alpina* FKI-4905. J Antibiot (Tokyo) 63:183–186. https://doi.org/10.1038/ja.2010.14.
16. Kamiyama T, Umino T, Nakayama N, Itezono Y, Satoh T, Yamashita Y, Yamaguchi A, Yokose K. 1992. Ro 09-1679, a novel thrombin inhibitor. J Antibiot (Tokyo) 45:424–427. https://doi.org/10.7164/antibiotics.45.424.
17. Baldeweg F, Warncke P, Fischer D, Gressler M. 2019. Fungal biosurfactants from *Mortierella alpina*. Org Lett 21:1444–1448. https://doi.org/10.1021/acs.orglett.9b00193.
18. Kawahara T, Itoh M, Izumikawa M, Sakata N, Tsuchida T, Shin-Ya K. 2017. Novel arginine-containing peptides MBJ-0173 and MBJ-0174 from *Mortierella alpina* f28740. J Antibiot (Tokyo) 70:226–229. https://doi.org/10.1038/ja.2016.116.
19. Burke PA, DeNardo SJ, Miers LA, Lamborn KR, Matzku S, DeNardo GL. 2002. Cilengitide targeting of alpha(v)beta(3) integrin receptor synergizes with radioimmunotherapy to increase efficacy and apoptosis in breast cancer xenografts. Cancer Res 62:4263–4272.
20. Hwang Y, Rowley D, Rhodes D, Gertsch J, Fenical W, Bushman F. 1999. Mechanism of inhibition of a poxvirus topoisomerase by the marine natural product sansalvamide A. Mol Pharmacol 55:1049–1053. https://doi.org/10.1124/mol.55.6.1049.
21. Xu WJ, Liao XJ, Xu SH, Diao JZ, Du B, Zhou XL, Pan SS. 2008. Isolation, structure determination, and synthesis of galaxamide, a rare cytotoxic cyclic pentapeptide from a marine algae *Galaxaura filamentosa*. Org Lett 10:4569–4572. https://doi.org/10.1021/ol801799d.
22. Xu WF, Hou XM, Yao FH, Zheng N, Li J, Wang CY, Yang RY, Shao CL. 2017. Xylapeptide A, an antibacterial cyclopentapeptide with an uncommon L-pipecolinic acid moiety from the associated fungus *Xylaria* sp. (GDG-102). Sci Rep 7:e6937. https://doi.org/10.1038/s41598-017-07331-4.
23. Belofsky GN, Jensen PR, Fenical W. 1999. Sansalvamide: a new cytotoxic cyclic depsipeptide produced by a marine fungus of the genus *Fusarium*. Tetrahedron Lett 40:2919–2926.
24. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Muller R, Stewart AF, Zhang Y. 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol 30:440–446. https://doi.org/10.1038/nbt.2183.
25. Kaweewan I, Hemmi H, Komaki H, Kodani S. 2020. Isolation and structure determination of a new antibacterial peptide pentaminomycin C from *Streptomyces cacaoi* subsp. *cacaoi*. J Antibiot (Tokyo) 73:224–229. https://doi.org/10.1038/s41429-019-0272-y.
26. Wang L, Chen W, Feng Y, Ren Y, Gu Z, Chen H, Wang H, Thomas MJ, Zhang B, Berquin IM, Li Y, Wu J, Zhang H, Song Y, Liu X, Norris JS, Wang S, Du P, Shen J, Wang N, Yang Y, Wang W, Feng L, Ratledge C, Zhang H, Chen YQ. 2011. Genome characterization of the oleaginous fungus *Mortierella alpina*. PLoS One 6:e28319. https://doi.org/10.1371/journal.pone.0028319.
27. Walsh CT, O'Brien RV, Khosla C. 2013. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. Angew Chem Int Ed Engl 52:7098–7124. https://doi.org/10.1002/anie.201208344.
28. Hoffmann K, Schneider-Scherzer E, Kleinkauf H, Zocher R. 1994. Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. J Biol Chem 269:12710–12714.
29. Jia LJ, Tang HY, Wang WQ, Yuan TL, Wei WQ, Pang B, Gong XM, Wang SF, Li YJ, Zhang D, Liu W, Tang WH. 2019. A linear nonribosomal octapeptide from *Fusarium graminearum* facilitates cell-to-cell invasion of wheat. Nat Commun 10:e922. https://doi.org/10.1038/s41467-019-08726-9.
30. Gao X, Haynes SW, Ames BD, Wang P, Vien LP, Walsh CT, Tang Y. 2012. Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. Nat Chem Biol 8:823–830. https://doi.org/10.1038/nchembio.1047.
31. Partida-Martinez LP, de Looss CF, Ishida K, Ishida M, Roth M, Buder K, Hertweck C. 2007. Rhizonin, the first mycotoxin isolated from the zygomycota, is not a fungal metabolite but is produced by bacterial endosymbionts. Appl Environ Microbiol 73:793–797. https://doi.org/10.1128/AEM.01784-06.
32. Niehs SP, Dose B, Scherlach K, Roth M, Hertweck C. 2018. Genomics-driven discovery of a symbiont-specific cyclopeptide from bacteria residing in the rice seedling blight fungus. Chembiochem 19:2167–2172. https://doi.org/10.1002/cbic.201800400.
33. Ohshima S, Sato Y, Fujimura R, Takashima Y, Hamada M, Nishizawa T, Narisawa K, Ohta H. 2016. *Mycoavidus cysteinexigens* gen. nov., sp. nov., an endohyphal bacterium isolated from a soil isolate of the fungus *Mortierella elongata*. Int J Syst Evol Microbiol 66:2052–2057. https://doi.org/10.1099/ijsem.0.000990.
34. Uehling J, Gryganskyi A, Hameed K, Tschaplinski T, Misztal PK, Wu S, Desiro A, Vande Pol N, Du Z, Zienkiewicz A, Zienkiewicz K, Morin E, Tisserant E, Splivallo R, Hainaut M, Henrissat B, Ohm R, Kuo A, Yan J, Lipzen A, Nolan M, LaButti K, Barry K, Goldstein AH, Labbe J, Schadt C, Tuskan G, Grigoriev I, Martin F, Vilgalys R, Bonito G. 2017. Comparative

genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycoavidus cysteinexigens*. Environ Microbiol 19:2964–2983. https://doi.org/10.1111/1462-2920.13669.
35. Desiro A, Hao Z, Liber JA, Benucci GMN, Lowry D, Roberson R, Bonito G. 2018. *Mycoplasma*-related endobacteria within *Mortierellomycotina* fungi: diversity, distribution and functional insights into their lifestyle. ISME J 12:1743–1757. https://doi.org/10.1038/s41396-018-0053-9.
36. Marfey P. 1984. Determination of D-amino acids. 2. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. Carlsberg Res Commun 49: 591–596. https://doi.org/10.1007/BF02908688.
37. Harada K, Fujii K, Mayumi T, Hibino Y, Suzuki M, Ikai Y, Oka H. 1995. A method using LC/MS for determination of absolute configuration of constituent amino acids in peptide—advanced Marfey's method. Tetrahedron Lett 36:1515–1518. https://doi.org/10.1016/0040-4039(95)00078-Q.
38. Inoue T, Hasumi K, Kuniyasu T, Endo A. 1996. Isolation of plactins A, B, C and D, novel cyclic pentapeptides that stimulate cellular fibrinolytic activity. J Antibiot (Tokyo) 49:45–49. https://doi.org/10.7164/antibiotics.49.45.
39. Harada T, Tsuruta T, Yamagata K, Inoue T, Hasumi K. 2009. Dual modulation of prothrombin activation by the cyclopentapeptide plactin. FEBS J 276:2516–2528. https://doi.org/10.1111/j.1742-4658.2009.06976.x.
40. Iwasaki S, Kobayashi H, Furukawa J, Namikoshi M, Okuda S, Sato Z, Matsuda I, Noda T. 1984. Studies on macrocyclic lactone antibiotics. VII. Structure of a phytotoxin "rhizoxin" produced by *Rhizopus chinensis*. J Antibiot (Tokyo) 37:354–362. https://doi.org/10.7164/antibiotics.37.354.
41. Partida-Martinez LP, Hertweck C. 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. Nature 437:884–888. https://doi.org/10.1038/nature03997.
42. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.
43. Stachelhaus T, Mootz HD, Marahiel MA. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol 6:493–505. https://doi.org/10.1016/S1074-5521(99)80082-9.
44. Sohn YS, Nam DH, Ryu DD. 2001. Biosynthetic pathway of cephabacins in *Lysobacter lactamgenus*: molecular and biochemical characterization of the upstream region of the gene clusters for engineering of novel antibiotics. Metab Eng 3:380–392. https://doi.org/10.1006/mben.2001.0200.
45. Kaljunen H, Schiefelbein SH, Stummer D, Kozak S, Meijers R, Christiansen G, Rentmeister A. 2015. Structural elucidation of the bispecificity of A domains as a basis for activating non-natural amino acids. Angew Chem Int Ed Engl 54:8833–8836. https://doi.org/10.1002/anie.201503275.
46. Kraas FI, Helmetag V, Wittmann M, Strieker M, Marahiel MA. 2010. Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. Chem Biol 17:872–880. https://doi.org/10.1016/j.chembiol.2010.06.015.
47. Bergendahl V, Linne U, Marahiel MA. 2002. Mutational analysis of the C-domain in nonribosomal peptide synthesis. Eur J Biochem 269: 620–629. https://doi.org/10.1046/j.0014-2956.2001.02691.x.
48. Kikukawa H, Sakuradani E, Ando A, Shimizu S, Ogawa J. 2018. Arachidonic acid production by the oleaginous fungus *Mortierella alpina* 1S-4: a review. J Adv Res 11:15–22. https://doi.org/10.1016/j.jare.2018.02.003.
49. Kikukawa H, Sakuradani E, Ando A, Okuda T, Ochiai M, Shimizu S, Ogawa J. 2015. Disruption of lig4 improves gene targeting efficiency in the oleaginous fungus *Mortierella alpina* 1S-4. J Biotechnol 208:63–69. https://doi.org/10.1016/j.jbiotec.2015.05.020.
50. Khan F, He M, Taussig MJ. 2006. Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces. Anal Chem 78:3072–3079. https://doi.org/10.1021/ac060184l.
51. Stanišić A, Hüsken A, Kries H. 2019. HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes. Chem Sci 10:10395–10399. https://doi.org/10.1039/c9sc04222a.
52. Baldeweg F, Kage H, Schieferdecker S, Allen C, Hoffmeister D, Nett M. 2017. Structure of ralsolamycin, the interkingdom morphogen from the crop plant pathogen *Ralstonia solanacearum*. Org Lett 19:4868–4871. https://doi.org/10.1021/acs.orglett.7b02329.
53. Villiers BR, Hollfelder F. 2009. Mapping the limits of substrate specificity of the adenylation domain of TycA. Chembiochem 10:671–682. https://doi.org/10.1002/cbic.200800553.
54. Meyer S, Kehr JC, Mainz A, Dehm D, Petras D, Süssmuth RD, Dittmann E. 2016. Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS. Cell Chem Biol 23: 462–471. https://doi.org/10.1016/j.chembiol.2016.03.011.
55. Romans-Fuertes P, Sondergaard TE, Sandmann MI, Wollenberg RD, Nielsen KF, Hansen FT, Giese H, Brodersen DE, Sorensen JL. 2016. Identification of the non-ribosomal peptide synthetase responsible for biosynthesis of the potential anti-cancer drug sansalvamide in *Fusarium solani*. Curr Genet 62:799–807. https://doi.org/10.1007/s00294-016-0584-4.
56. Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT. 2002. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure 10: 301–310. https://doi.org/10.1016/s0969-2126(02)00716-5.
57. Grunwald AL, Berrue F, Robertson AW, Overy DP, Kerr RG. 2017. Mortiamides A–D, cyclic heptapeptides from a novel *Mortierella* sp. obtained from Frobisher Bay. J Nat Prod 80:2677–2683. https://doi.org/10.1021/acs.jnatprod.7b00383.
58. Walton JD. 2006. HC-toxin. Phytochemistry 67:1406–1413. https://doi.org/10.1016/j.phytochem.2006.05.033.
59. Hai Y, Jenner M, Tang Y. 2019. Complete stereoinversion of L-tryptophan by a fungal single-module nonribosomal peptide synthetase. J Am Chem Soc 141:16222–16226. https://doi.org/10.1021/jacs.9b08898.
60. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. 2007. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol 7:78. https://doi.org/10.1186/1471-2148-7-78.
61. Yang X, Feng P, Yin Y, Bushley K, Spatafora JW, Wang C. 2018. Cyclosporine biosynthesis in *Tolypocladium inflatum* benefits fungal adaptation to the environment. mBio 9:e01211-18. https://doi.org/10.1128/mBio.01211-18.
62. Gao X, Chooi YH, Ames BD, Wang P, Walsh CT, Tang Y. 2011. Fungal indole alkaloid biosynthesis: genetic and biochemical investigation of the tryptoquialanine pathway in *Penicillium aethiopicum*. J Am Chem Soc 133:2729–2741. https://doi.org/10.1021/ja1101085.
63. Berry D, Mace W, Grage K, Wesche F, Gore S, Schardl CL, Young CA, Dijkwel PP, Leuchtmann A, Bode HB, Scott B. 2019. Efficient nonenzymatic cyclization and domain shuffling drive pyrrolopyrazine diversity from truncated variants of a fungal NRPS. Proc Natl Acad Sci U S A 116:25614–25623. https://doi.org/10.1073/pnas.1913080116.
64. Kalb D, Lackner G, Hoffmeister D. 2014. Functional and phylogenetic divergence of fungal adenylate-forming reductases. Appl Environ Microbiol 80:6175–6183. https://doi.org/10.1128/AEM.01767-14.
65. Matsuda K, Zhai R, Mori T, Kobayashi M, Sano A, Abe I, Wakimoto T. 2020. Heterochiral coupling in non-ribosomal peptide macrolactamization. Nat Catal 3:507–515. https://doi.org/10.1038/s41929-020-0456-7.
66. Weyler C, Heinzle E. 2017. Synthesis of natural variants and synthetic derivatives of the cyclic nonribosomal peptide luminmide in permeabilized *E. coli* Nissle and product formation kinetics. Appl Microbiol Biotechnol 101:131–138. https://doi.org/10.1007/s00253-016-7770-7.
67. Guenzi E, Galli G, Grgurina I, Gross DC, Grandi G. 1998. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. J Biol Chem 273:32857–32863. https://doi.org/10.1074/jbc.273.49.32857.
68. Landan G, Cohen G, Aharonowitz Y, Shuali Y, Graur D, Shiffman D. 1990. Evolution of isopenicillin N synthase genes may have involved horizontal gene transfer. Mol Biol Evol 7:399–406. https://doi.org/10.1093/oxfordjournals.molbev.a040615.
69. Liras P, Martin JF. 2006. Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate? Int Microbiol 9:9–19.
70. Bushley KE, Turgeon BG. 2010. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. BMC Evol Biol 10:e26. https://doi.org/10.1186/1471-2148-10-26.
71. Lawrence DP, Kroken S, Pryor BM, Arnold AE. 2011. Interkingdom gene transfer of a hybrid NPS/PKS from bacteria to filamentous Ascomycota. PLoS One 6:e28231. https://doi.org/10.1371/journal.pone.0028231.
72. Da Lage JL, Binder M, Hua-Van A, Janecek S, Casane D. 2013. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial alpha-amylase gene to Basidiomycetes. BMC Evol Biol 13:e40. https://doi.org/10.1186/1471-2148-13-40.
73. Tabima JF, Trautman IA, Chang Y, Wang Y, Mondo S, Kuo A, Salamov A, Grigoriev IV, Stajich JE, Spatafora JW. 2020. Phylogenomic analyses of non-Dikarya fungi supports horizontal gene transfer driving diversification of secondary metabolism in the amphibian gastrointestinal symbiont, *Basidiobolus*. G3 (Bethesda) 10:3417–3433. https://doi.org/10.1534/g3.120.401516.

74. Henk DA, Fisher MC. 2012. The gut fungus *Basidiobolus ranarum* has a large genome and different copy numbers of putatively functionally redundant elongation factor genes. PLoS One 7:e31268. https://doi.org/10.1371/journal.pone.0031268.
75. Inoue T, Hasumi K, Sugimoto M, Endo A. 1998. Enhancement of fibrinolysis by plactins: structure-activity relationship and effects in human U937 cells and in mice. Thromb Haemost 79:591–596. https://doi.org/10.1055/s-0037-1614951.
76. Koizumi Y, Nagai K, Gao L, Koyota S, Yamaguchi T, Natsui M, Imai Y, Hasumi K, Sugiyama T, Kuba K. 2018. Involvement of RSK1 activation in malformin-enhanced cellular fibrinolytic activity. Sci Rep 8:e5472. https://doi.org/10.1038/s41598-018-23745-0.
77. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45.
78. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
79. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
80. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.
81. Conti E, Stachelhaus T, Marahiel MA, Brick P. 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. EMBO J 16:4174–4183. https://doi.org/10.1093/emboj/16.14.4174.
82. Bian X, Plaza A, Yan F, Zhang Y, Müller R. 2015. Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives *in vivo*. Biotechnol Bioeng 112:1343–1353. https://doi.org/10.1002/bit.25560.

7 MANUSCRIPT VI

Promiscuity-guided diversification of nonribosomal biosurfactants from Mortierella alpina

Jacob M. Wurlitzer, **Aleksa Stanišić**, Sebastian Ziethe, Sandra Jungmann, Anne Seidlitz, Hajo Kries and Markus Gressler

Summary:

Lower fungi are considered to be poor producers of secondary metabolites. Here, we identify MalA a nonribosomal peptide synthetase (NRPS) responsible for the biosynthesis of surfactants malpinins in *Mortierella alpina*. Characterization of adenylation domains with HAMA indicated relaxed specificity of substrate selection. Precursor-directed biosynthesis yielded 20 new malpinin congeners produced by a single enzyme MalA. The promiscuity of A-domains was further exploited to incorporate click-functionalized amino acids 4-bromo-L-phenylalanine and S-propargyl-L-cysteine. Our results strengthen the substrate specificity screening as an indispensable tool for characterizing flexible NRPSs.

The candidate is:

☐ First author ☒ Second author ☐ Corresponding author ☐ Coauthor

Status: In preparation for submission to *Chemical Science.*

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
JW	50 %	75 %	70 %	20 %	
AS		20 %	20 %	5 %	
SZ			5 %	5 %	
SJ		5%	5 %	5 %	
HK				5 %	
MG	50 %			60 %	

Supplementary datasets are included in the enclosed CD-ROM.

ARTICLE

Promiscuity-guided diversification of nonribosomal biosurfactants from *Mortierella alpina*

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Jacob M. Wurlitzer,[a] Aleksa Stanišić,[b] Sebastian Ziethe,[a] Sandra Jungmann,[c] Hajo Kries,[b] Markus Gressler[a,*]

The biosynthesis of natural products from early diverging fungi such as *Mortierella* species is largely unexplored. Herein, the nonribosomal peptide synthetase (NRPS) MalA responsible for the biosynthesis of the surface-active biotensides, malpinins, has been identified and biochemically characterized. The investigation of substrate specificity of respective A-domains indicated a highly promiscuous enzyme with an unusual, inactive C-terminal NRPS module. Specificity-based precursor-directed biosynthesis led to the identification of 20 new congeners produced by a single enzyme. Moreover, MalA incorporates artificial, click-functionalized amino acids such as 4-bromo-L-phenylalanine or *S*-propagyl-L-cysteine which allowed postbiosynthetic coupling to a fluorophor. The current study demonstrates substrate-specificity testing as a powerful tool to identify flexible NRPS modules and gain chemically tractable natural products.

Introduction

Fungi of the division Mortierellomycotina have traditionally been used as a resource for polyunsaturated fatty acids (such as arachidonic acid) and their fungal oils are widely used as safe additives in food industry.[1] Recent investigation of the secondary metabolism of *Mortierella alpina* revealed an unexpected potential for the production of small oligopeptides of pharmaceutical interest like surface-active compounds, malpinins (compounds **1-5**)[2] (Figure 1), the antimycobacterial agent calpinactam[3, 4] and the thrombin-inhibitor Ro 09-1679[5]. Additionally, cyclic pentapeptides such as malpibaldins and the antibacterial malpicyclins are produced. Their biosynthesis has recently been assigned to two bacterial-like nonribosomal peptide synthetases (NRPSs), malpibaldin synthetase MpbA and malpicyclin synthetase MpcA, respectively.[6] Both enzymes are composed of five consecutive modules (C-A-T) each harboring a condensation (C) domain, an ATP-dependent adenylation (A) domain and a thiolation (T) domain, which act in concert to provide a linear pentapeptide that is subsequently cyclized by a C-terminal thioesterase domain (TE). Whilst L-amino acids are incorporated by the action of canonical C domains, D-amino acid building blocks are introduced by bacterial-like dual epimerization/condensation (E/C) domains. The sequence of C- and dual E/C domains within the peptide chain of both enzymes directly reflects the succession of L- and D-amino acids in the final oligopeptides.

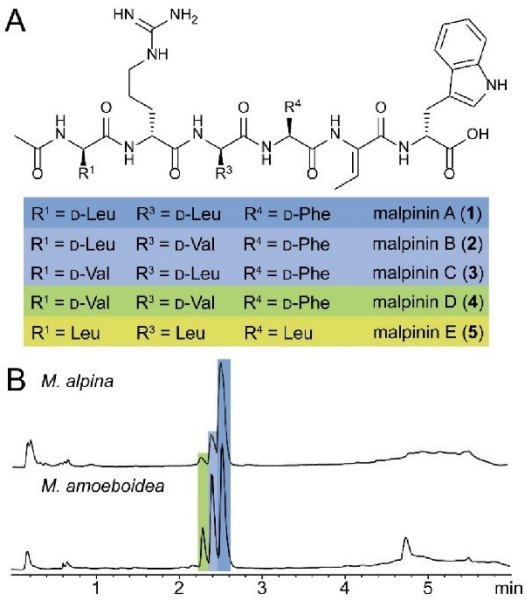


Figure 1. Native malpinins. **A.** Chemical structures of malpinin A-E (**1-5**). The stereoconfiguration of **5** has not been determined. **B.** UV chromatograms (λ = 280 nm) of crude extracts of mycelia from *M. alpina* ATCC32222 and *M. amoeboidea* CBS 889.72 grown on YPD. Compound **1** is the predominant metabolite, whereas compound **5** is detectable only in traces (not depicted).

[a] Department Pharmaceutical Microbiology at the Leibniz Institute for Natural Product Research and Infection Biology (Hans-Knöll-Institute), Friedrich-Schiller-University, Winzerlaer Strasse 2, 07745 Jena (Germany)
[b] Junior Group Biosynthetic Design of Natural Products at the Leibniz Institute for Natural Product Research and Infection Biology (Hans-Knöll-Institute), Beutenbergstrasse 11a, 07745 Jena (Germany)
[c] Pharmaceutical Technology and Biopharmacy, Friedrich-Schiller-University Jena, Lessingstraße 8, 07743 Jena, Germany
* corresponding author: markus.gressler@leibniz-hki.de
Electronic Supplementary Information (ESI) available: The PDF file contains additional experimental procedures, figures S1-S49 and table S1-S14. See DOI: 10.1039/x0xx00000x

In contrast to *Mortierella* cyclopentapeptides, the biosynthetic origin of the biotensides malpinins A-E (**1-5**) is still enigmatic. Malpinins harbor a remarkable surface-tension lowering activity with a critical micelle concentration (CMC) of 14 µM that is 580-fold lower than that of the commercial anionic detergent sodium dodecyl sulfate (SDS).[2] Their low cytotoxicity makes these biotensides appropriate candidates for pharmaceutical or medicinal applications. However, the biosynthesis of this metabolite family is as yet unknown. Malpinins are acetylated hexapeptides (Ac-D-Leu/Val-D-Arg-D-Leu/Val-L-Phe/Leu-Dhb-D-Trp) with two striking structural features (Figure 1): i) a non-canonical amino acid, (*Z*)-dehydrobutyrine (Dhb), at position 5, and ii) a C-terminal D-amino acid, D-tryptophan, that can be oxidized to kynurenine.[2] Moreover, incorporated D-amino acids in position 1, 3 and 4 are variable resulting in the **1**-congeners **2-5**. Consequently, a hexamodular NRPS with promiscuous modules 1, 3 and 4 is expected for the malpinin biosynthesis matching the above-mentioned catalytic features.

Substrate promiscuity is a key feature of many enzymes with large impact both on natural enzyme evolution[7] and enzyme engineering[8] in synthetic biology.[9] In NRPSs, substrate promiscuity has been occasionally encountered at the A domain[10] and harnessed for producing non-natural products[11] but has not been systematically investigated. Despites the widely used ATP/ pyrophosphate (PP_i) exchange assays[12], alternative adenylation assays such as the enzyme-coupled conversion of the chromogenic substrate 7-methyl-6-thioguanosine (MesG)[13] has been established. Recently, the hydroxamate specificity assay (HAMA) has been developed which unravels specificity profiles of A domains under competition conditions in a straightforward fashion. HAMA is based on the quenching of aminoacyl adenylates by hydroxylamine and LC-MS/MS detection of respective hydroxamate products allowing parallel testing of multiple competing substrates.[14] Hence, HAMA is ideally suited for the investigation of putatively promiscuous A domains[6] in the malpinin synthetase.

In this report, we use a combination of state-of-the art genome sequencing techniques and UHPLC-MS/MS-based metabolite screening to identify the malpinin synthase MalA. A thorough substrate specificity analysis of the complete set of its NRPS modules led to the identification of 20 novel malpinin congeners, among them unusual methionine-containing metabolites. Moreover, the enzyme´s relaxed substrate tolerance facilitates incorporations of non-natural and "clickable" amino acid substrates. This work underlines the applicability of the HAMA assay as a tool to determine A-domain flexibility leading to discovery of metabolic diversity. Moreover, it paves the way to chemically tractable compounds to facilitate the identification of their molecular targets.

Results and discussion

Identification of the malpinin synthetase MalA

The publicly available genome of *M. alpina*[15] does not provide appropriate NRPS candidate genes for malpinin biosynthesis. Therefore, a comparative genome analysis of *M. alpina* ATCC32222 and its close relative *Mortierella amoeboidea* CBS889.72[16] was used to identify the malpinin synthetase gene. Both species produce **1-5** as determined by UHPLC-MS and ESI-MS/MS (Figure 1, and Electronic Supplementary Information, Figure S1) and were subjected to genome sequencing using the Oxford nanopore technology (Table S1). Subsequent analysis of both genomes using the ANTISMASH[17] platform lead to the identification of at least 16 potential NRPS and NRPS-like genes in both species (Table S2). As expected from the metabolite screening (Figure S1), both genomes encode the two

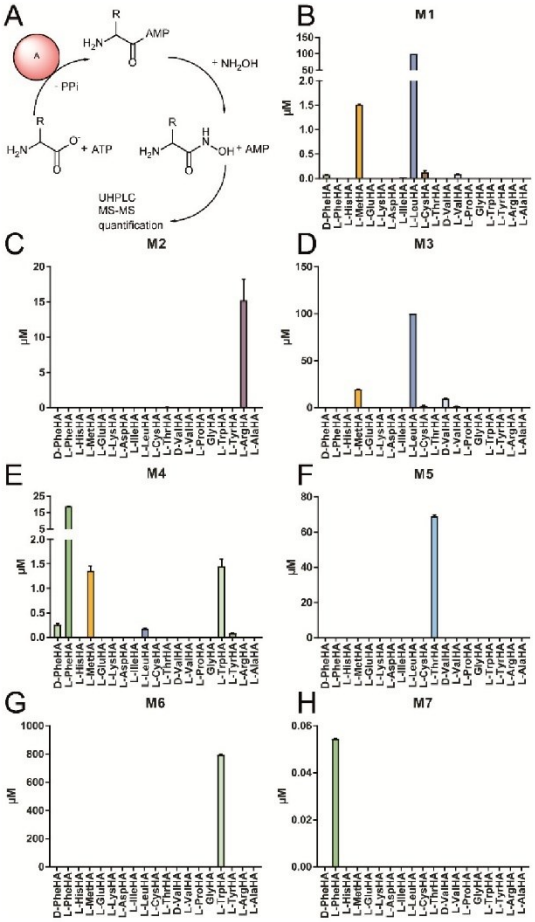


Figure 2. Specificity profiles of the NRPS modules of MalA. **A.** Adenylation reaction and hydroxamate formation during HAMA. Modules and substrates were mixed at a final concentration of 1 µM and 1 mM, respectively, and were incubated for 60 min at 37°C. The resulting aminoacyl hydroxamates (HA) were finally analysed by UHPLC-MS measurements. **B-H.** Specificity profiles of the complete set of the seven NRPS modules of MalA determined by the HAMA assay. Note that module 7 shows lowest activity.

cyclopentapeptide NRPSs MpbA (85 % amino acid identity between both species) and MpcA (90 %), which facilitate the production of malpibaldins and malpicyclins, respectively. In addition, both species share three large genes encoding one hexa-module (Nps5, 73% identity), one hepta-module (Nps3 (MalA), 89% identity) and one octa-module NRPS (Nps2, 90% identity), which might be appropriate candidates for malpinin biosynthesis (Table 1 and Table S2). Whilst the *nps2* gene is hardly expressed under laboratory conditions, transcripts of both *nps3* (*malA*) and *nps5* are highly abundant (Figure S2). However, solely *malA* expression levels correlated with titers of malpinin A, i.e. the major metabolite in *Mortierella* metabolite extracts, according to LC-MS-based metabolite quantification (Figure S2). Indeed, the candidate enzyme MalA shows the required distribution of C and E/C domains (Table 1), but the 7th module seems to be obsolete for the production of a hexapeptide.

Modules 1, 3 and 4 are highly promiscuous

The hepta-module NRPS MalA from *M. alpina* spans over 7,760 aa (853 kDa). Since knock-out strategies are hardly applicable for early diverging fungi[18], we assigned the different A domains of MalA to the specific adenylation steps in the malpinin biosynthesis by specificity profiling of purified proteins (C-A-T modules). Heterologous production of all seven modules (M1-7) as bi-terminal His_6-tagged fusion proteins succeeded in *Escherichia coli* (Figure S3, Table S3). The formation of aminoacyl adenylate during the adenylation reactions was tracked by conversion to stable aminoacyl hydroxamates that were quantified using multiplexed LC-MS/MS measurements (HAMA)[14] (Figure 2).

Table 1. Multi-module NRPSs in *M. alpina*. For domain abbreviations refer to Figure 3.

gene	size incl. introns (bp)	protein size (aa)	domain architecture of the enzyme
nps2	26,963	8,125	A-T-E/C-A-T-C-A-T-E/C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-TE
nps3 (*malA*)	24,471	7,760	C_s-A-T-E/C-A-T-E/C-A-T-E/C-A-T-C-A-T-C*-A-T-E/C-A-T-TE
nps5	20,143	6,489	C_s-A-T-E/C-A-T-C-A-T-E/C-A-T-E/C-A-T-C-A-T-TE

HAMA revealed that both M1 and M3 have a relaxed specificity towards aliphatic amino acids (L-Leu > L-Met > L/D -Val > L-Cys), explaining Val at position 1 and 3 in **1**-congeners **2-4** (Figure 1). The incorporation of L-Leu, but not *N*-acetyl-L-Leu by module 1 was also confirmed (Figure S4) suggesting the *N*-terminal acetylation occurs at later stage of biosynthesis. Similar to M1 and M3, M4 is highly promiscuous and showed the highest activity with L-Phe followed by other hydrophobic amino acids (L-Phe > L-Met = L-Trp). In contrast, M2, M5 and M6 show high

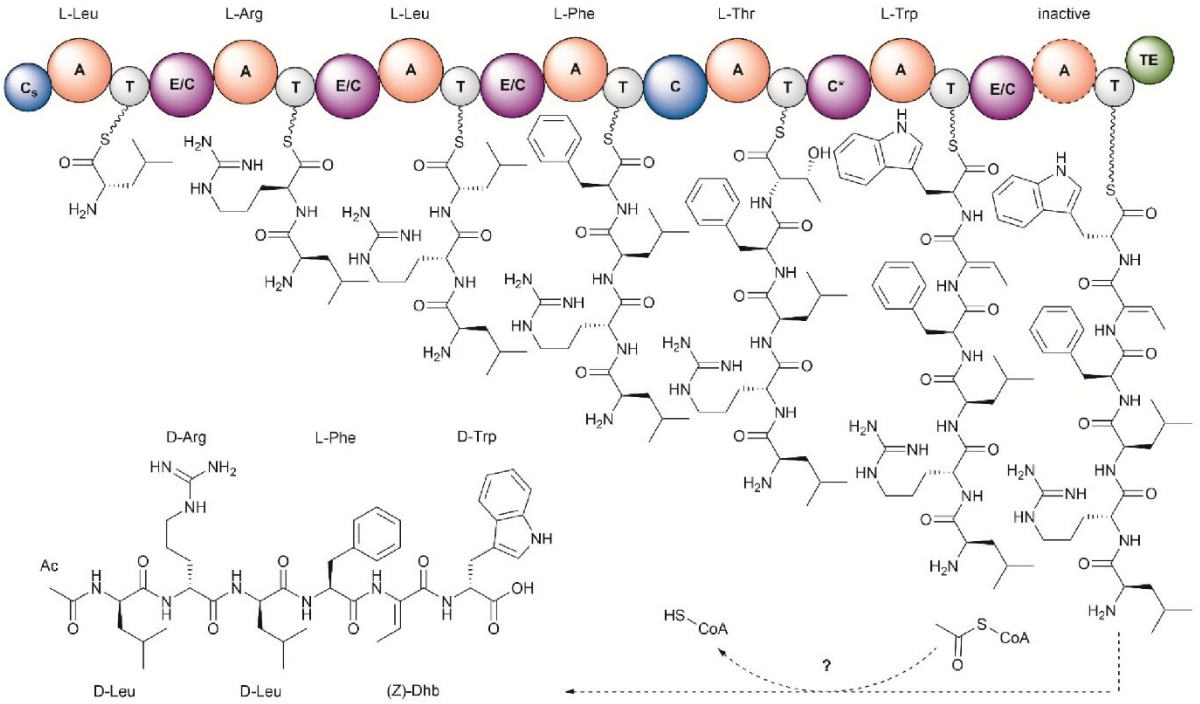


Figure 3. Biosynthesis of malpinins by MalA. The NRPS assembly line is representatively demonstrated for malpinin A (**1**). Involved domains are: A, adenylation domain; C, canonical condensation domain; C_s, starter condensation domain (inactive); C*, dual dehydration/condensation domain; E/C, dual epimerization/condensation domain; T, thiolation domain; TE, thioesterase domain. Note, that the final adenylation domain (in M7) is inactive and offloading occurs either by the final dual E/C domain or the C-terminal T or TE domain.

specificity towards L-Arg, L-Thr, and L-Trp, respectively (Figure 2). Solely, M7 converted its preferred substrate (L-Phe) with a 15 000-fold reduced turnover rate compared to the most active module M6, indicating that its A domain cannot contribute to the malpinin biosynthesis due to low activity. The residual activity for L-Phe may indicate that a 7th residue was present in an evolutionary precursor to the malpinin family. Since the last T domain is apparently not loaded with an amino acid, either the TE domain must act on the preceding T domain or the dual E/C domain of M7 acyl-transfer the final peptide chain to the free acceptor T domain. In any case, this C/E domain must be active because it is required for the stereo-inversion of L-Trp to D-Trp as C-terminal amino acid in **1-5**. A similar mechanism is proposed for the C-terminally located epimerase (E) domain of the β-(L-α-aminoadipyl)-L-cysteinyl-D-valine (LLD-ACV) synthetase from *Penicillium chrysogenum* and other β-lactam synthesizing fungi.[19-21]

Building on our expression analysis and HAMA-based *in-vitro* activity assays, a biosynthetic pathway for malpinins by MalA is proposed, which includes (i) a canonical successive peptide biosynthesis, (ii) epimerizations of L-amino acids by dual E/C domains, if required, and (iii) a peptide offloading at module 7 (Figure 3).

In bacteria, N-terminal acylation of peptides is catalysed by C-starter (C_s) domains that transfer various acyl chains from acyl-CoA, a standalone acyl carrier protein (ACP) or an acylated C-A-T module specifically to the N-terminus of the nascent nonribosomal peptide.[22-24] Recently, a fungal C_s domain has been described to N-terminally acetylate the NRP aspergillicin A from *Aspergillus flavus*.[25] However, MalA lacks a functional tandem His-His motif in its C_s domain (Figure S5) and an acetylation of Leu by M1 could not be determined (Figure S6). Both findings point at a post-synthetic acetylation by an acetyltransferase encoded elsewhere in the genome - a phenomenon that has recently been described for the erinacine biosynthesis in the mushroom *Hericium erinaceus*.[26] Indeed, the precursor deacetyl-**1** (*m/z* 817.4716 $[M+H]^+$) is detectable in *M. alpina* with moderate abundance (Figures S7 and S8).

Remarkably, malpinins contain (*Z*)-Dhb as non-proteinogenic amino acid. Dhb is present in a variety of cyanobacterial NRP, but its enzymatic origin has not been determined.[27-30] To incorporate Dhb in ribosomally and post-translationally modified peptides (RiPPs) such as the lantibiotic precursor prenisin or the lacticin-481 propeptide, Thr residues are post-translationally dehydrated by a downstream processing dehydratase (NisB)[31] or by a bifunctional dehydratase/cyclase (LctM),[32] respectively. However, no NisB- or LctM-homolog is encoded in the genomes of *M. alpina* and *M. amoeboidea*. During malpinin biosynthesis, Thr is incorporated by M5 (Figure 2) and is then dehydrated by the subsequent dual E/C domain of M6 to give the α,β-unsaturated amino acid (*Z*)-Dhb (Figure 3). In dual E/C domains, the double bond is first deprotonated

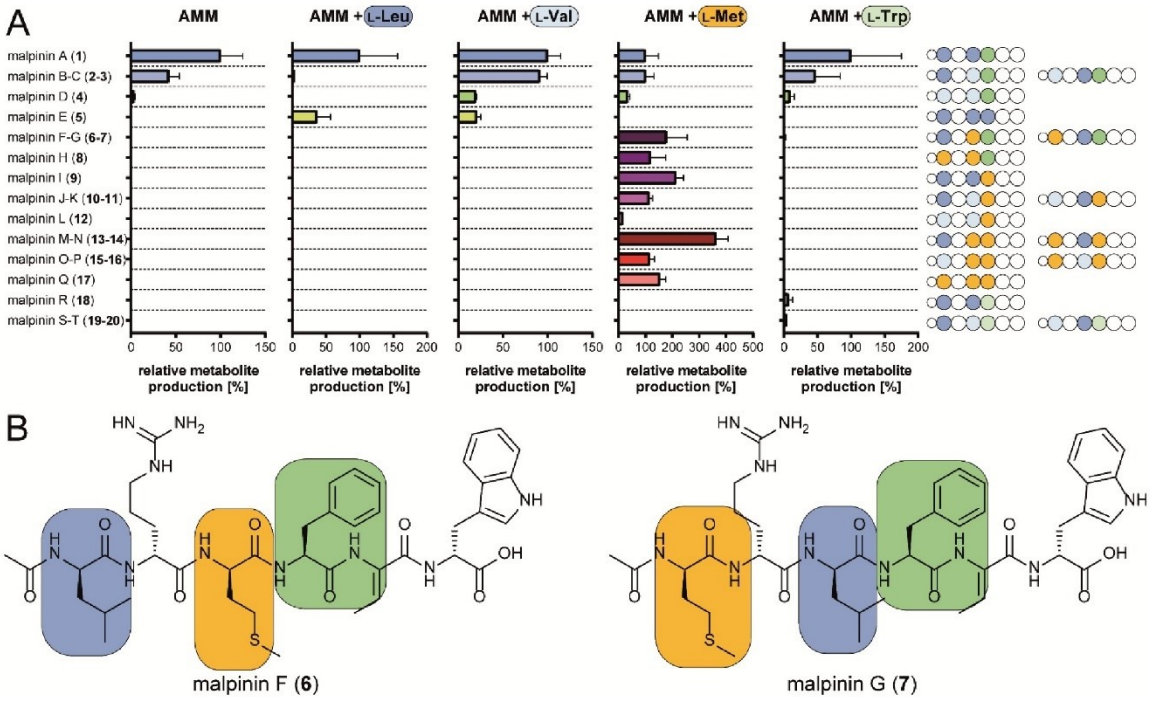


Figure 4. Precursor-directed biosynthesis of novel malpinins. **A.** Distribution of native (**1-5**) and novel malpinins (**6-20**) by feeding amino acids predicted by HAMA. For detailed HR-ESI-MS/MS analysis see Table S6 (properties of metabolites) and Figures S10-18. Experiments were carried out in *Aspergillus* minimal medium without supplementation (AMM) or amended with 5 mM L-Leu, L-Val, L-Met, or L-Trp. Amino acid sequences of **1-20** are shown schematically as stings of beads with variable amino acids highlighted in colour (indigo, Leu; light blue, Val; green, Phe; brown, Met; dark green, Trp). **B.** NMR-verified novel malpinins F **(6)** and G **(7).** For structure elucidation refer to Tables S6 and S8, and Figures S19-S29.

resulting in an enolate intermediate and the epimerization is completed by an electrophilic attack of a proton from the reverse side of the double bond.[33] However, the latter step is avoided by M6 and a hydroxyl group is eliminated instead. Recently, the origin of α,β-unsaturated amino acyl moieties in the nonribosomal peptide albopeptide from *Streptomyces albofaciens* has been assigned to a novel class of dual β-elimination/condensation domains (C*) in the NRPS AlbB.[34] Indeed, the amino acid sequence of the dual E/C domain of M6 shows the same conserved motifs like the C* domains of AlbB, i.e. the ^{136}HHXXXD^{141} motif and a E^{367} residue, which is discussed to be involved in the dehydration reaction.[34] Hence, our findings imply a similar mechanism in incorporation of (*Z*)-Dhb in malpinins. This assumption is further supported by the fact, that both Thr- and Dhb-specific A domains from bacteria and *Mortierella* share the same residues in the specificity determining binding pocket (Table S4).

Promiscuity facilitates the production of diverse malpinin congeners

The substrate specificity data confirmed MalA as malpinin synthetase and disclosed an extraordinary promiscuity of its modules M1, M3, and M4. In-depth analysis of kinetic parameters of the A domain for M3 by MesG activity assays revealed highest specificity for L-Leu (k_{cat}/K_M = 61 mM^{-1} min^{-1}) followed by L-Val (k_{cat}/K_M = 4.8 mM^{-1} min^{-1}) and L-Met (k_{cat}/K_M = 0.47 mM^{-1} min^{-1}) which matches the probability of occurrence of Leu over Val at position 3 in natives malpinins **1-5** (Figure 1, Figure S9). However, Met-containing malpinins have never been detected. Surprisingly, the velocity of adenylation of the three tested substrates is similarly high (k_{cat} ~ 2.2 min^{-1}) and seemingly sufficient to support a typical rate of peptide formation (approx. 1 min^{-1}) with all of them. In other words, the adenylation kinetics suggest that the amino acid composition of malpinins can be simply altered by providing elevated concentrations of alternative substrates such as L-Met and L-Cys (for M1, M3, and M4) or L-Trp (for M4). To test this hypothesis, fungal cultures were supplemented with six different canonical amino acids (L-Leu, L-Val, L-Met, L-Trp, L-Cys, L-Phe) in a precursor-directed biosynthesis approach. While L-Leu feeding resulted in elevated levels of **1** and **5** as sole metabolites, L-Val feeding enlarged the metabolite variety mainly to **2-5**, which is consistent with a previous report.[2] In accordance to the predictions made by the HAMA assay, L-Met supplementation strongly increased the metabolic diversity by at least 12 additional compounds (malpinins F-Q, **6-17**), according to relative quantification by UHPLC-MS (Figure 4, Tables S5-S7). Subsequent ESI-MS/MS-fragmentation confirmed their **1**-derived lead structures and suggested a replacement of Leu and/or Phe by one, two or three Met moieties at the expected flexible positions 1, 2, and/or 4 (Figures S10-S18).

Two metabolites (malpinin F and G, **6** and **7**) show nearly the same molecular weight (*m/z* 877.4368 and 877.4374 [*M*+H]$^+$) and are most likely constitution isomers. They were exemplarily purified from upscaled fungal cultures by preparative HPLC submitted to extensive 1D and 2D NMR analyses (Table S8, Figures S19-S29). ^{13}C NMR spectra of **6** and **7** revealed seven signals above 160 ppm, accounting for seven carbonyl moieties. Fifteen signals in the aromatic range of the spectrum were identified. Five signals ranging from δ_C 50 to 55 ppm hinted to five α-carbon atoms, that could be confirmed by HSQC spectra, linking them to their respective α-C protons (δ_H 4.20-4.70 ppm). As expected, the eight amide protons (δ_H 7.54-9.18 ppm) did not show any scalar couplings in HSQC spectra. The final peptide backbone was constructed using COSY couplings between amide and α-C protons as well as HMBC correlations between α-C protons and the following carbonyl C atoms (Figure S29). Starting from the α-C proton signals, the amino acid side chains were elucidated by COSY, HSQC and HMBC. In **6** and **7**, the ^{13}C signal of the terminal methylene group in the aliphatic side chains of Met (δ_C 29.0 ppm and δ_C 29.5 ppm, respectively) correlated with a highly abundant ^{1}H singlet signal derived from the lone-standing methyl group (δ_H 1.92 ppm and 2.00 ppm). Since the latter H signals are not split into multiple peaks by neighbouring H atoms, the presence of an interrupting heteroatom (such as sulphur) within the side chain is plausible, which was confirmed by the respective HSQC coupling presence of Met in **6** and **7**. With a chemical shift of δ_C 156.6 ppm as part of the guanidinium group of Arg, δ_C 128.0 as double signal of Dhb, and 3 and 4 double bond signals each for Phe and Trp, respectively, a total number of eight C=X double bonds and eight C=C double bonds were identified. The stereo configuration of **6** and **7** was finally determined by advanced Marfey's analysis[35] (Table S9) and revealed incorporation of D-configured Met in both metabolites. In sum, the NMR analysis confirmed the incorporation of D-Met in position 1 and 3 in **6** and **7**, respectively, as suggested by the previous ESI-MS/MS experiments. Among the proteogenic amino acids, Met is underrepresented as building block in NRPs.[36] Met-containing

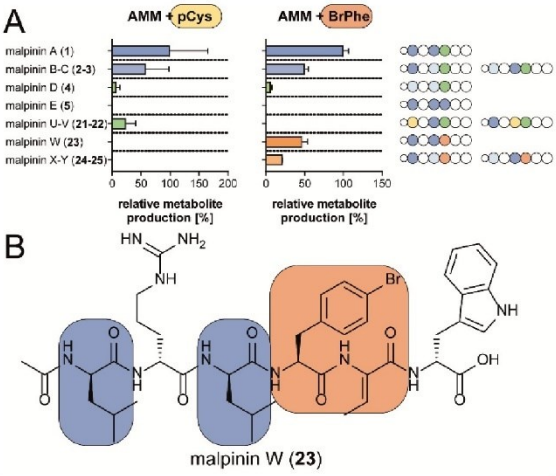


Figure 5. Precursor-directed biosynthesis of novel malpinins with non-natural amino acids. **A.** Distribution of native (**1-5**) and novel malpinins (**21-25**) by feeding *S*-propargyl-L-cysteine (pCys) or 4-bromo-L-phenylalanine (BrPhe). Amino acid sequence of **1-5** and **21-25** are shown schematically as a string of beads with variable amino acids highlighted in colour (indigo, Leu; light blue, Val; green, Phe; golden, pCys; ochre, Br-Phe). For detailed HR-ESI-MS/MS analysis see ESI (Table S6 and Figures S35-37). **B.** NMR-verified, click-enabled malpinin W (**23**). For structure elucidation refer to Table S9-10 and Figures S38-43

peptides have been extracted from cyanobacteria[37] or marine sponges[38-40] and harbour diverse biological activities including anti-cancer and phosphatase-inhibitory properties. However, compounds **6** and **7** do not show antimicrobial activities (Figure S30), but have a similar critical micelle concentration (CMC) as **1** (Figure S31).

Aside from the Met-containing metabolites **6-17**, three Trp-containing **1**-congeners (malpinin R-T, **18-20**) were elicited by L-Trp feeding (Figure 4, Figures S32-S33). In all cases, Trp was incorporated at position 4 in agreement with the relaxed specificity observed in the HAMA profile of M4 (Figure 2). However, no altered metabolite profiles were obtained by feeding L-Cys, or L-Phe (Figure S34).

Promiscuity have been intensively studied for the hexamodular anabaenopeptin synthetase whose first A domain accepts the chemically divergent amino acids Arg and Tyr.[41] Accordingly, substrate flexibility led to the biosynthesis of up to 16 structural variants of mycrosystins in *Phormidium*.[42] Here, promiscuity of A domains 2 and 4 has been assigned to altered residues in the in positions 236, 239 and 278 in the substrate binding pockets. Similar to cyanobacteria, promiscuity is the major driver of biodiversity in *Mortierellaceae* resulting in 20 natural malpinins (**1-20**). To achieve this, modules M1, M3, and M4 of MalA are promiscuous and incorporate hydrophobic amino acids such as L-Met or L-Trp - additionally to their native substrates L-Leu, L-Val or L-Phe *in vitro* and *in vivo*. In contrast, M2, M5 and M6 are highly specific.

Clickable amino acids enable synthesis of malpinin conjugates

Malpinins possess surface-active properties, but marginal cytotoxicity, and are hence promising for medicine and material science. Despites the Dhb moiety, the bottleneck in chemical tractability is the availability and accessibility of coupling moieties for "click" chemistry within the molecules. To investigate a potential incorporation of non-proteinogenic, but click-functionalized amino acids, fungal cultures were fed with the Met-congener *S*-propargyl-L-cysteine (pCys) or Phe-homolog 4-bromo-L-phenylalanine (BrPhe) and malpinin derivative production was quantified by UHPLC-MS (Figure 5).

Feeding with pCys led to poor growth of the fungus and moderate incorporation (24%) into malpinins: The Leu moieties in positions 1 and 3 were replaced by pCys in malpinin U (**21**), *m/z* 887.4224 $[M+H]^+$ and malpinin V (**22**), *m/z* 887.4224 $[M+H]^+$, respectively (Figure S35, Table S5-6). In contrast, BrPhe was incorporated in suitable amounts (47%) in place of Phe in position 4: the **1**-congener malpinin W (**23**, *m/z* 937.3920 $[M+H]^+$), and the two minor **2-** and **3**-congeners malpinin X (**24**, *m/z* 923.3766 $[M+H]^+$) and malpinin Y (**25**, *m/z* 923.3767 $[M+H]^+$) were detected (Figures S36-S37, Table S5). This relaxed substrate specificity is a remarkable feature of MalA, since acceptance of non-natural amino acids of A domains usually requires a time-consuming, systematic mutagenesis of residues in the enzyme´s substrate binding pockets.[8] A preferred incorporation of halogenated and other Phe analogs has been previously achieved by a Trp-to-Ser point mutation in module 1 of the gramicidin S synthetase GrsA and module 4 of the tyrocidine synthetase TycA.[43, 44] However, MalA M4 contains the respective Trp (W^{3852}) and the intrinsic flexibility might be due to an enlarged cavity by other small residues in the active-site, e.g. G^{3915} (Table S4).

We isolated 12.8 mg of **23** from 3 L of culture and confirmed its structure by NMR (Table S10, Figures S38-S43) and Marfey's (Table S9) analysis. The inspection of the ^{1}H-NMR spectrum revealed the absence of the δ_H 7.17 signal corresponding to the replacement of the C4 proton by a bromine atom in the Phe moiety at position 4 when compared to the ^{1}H-NMR spectrum of **1** (Figure S38, Table S10). The high field chemical shift of C4

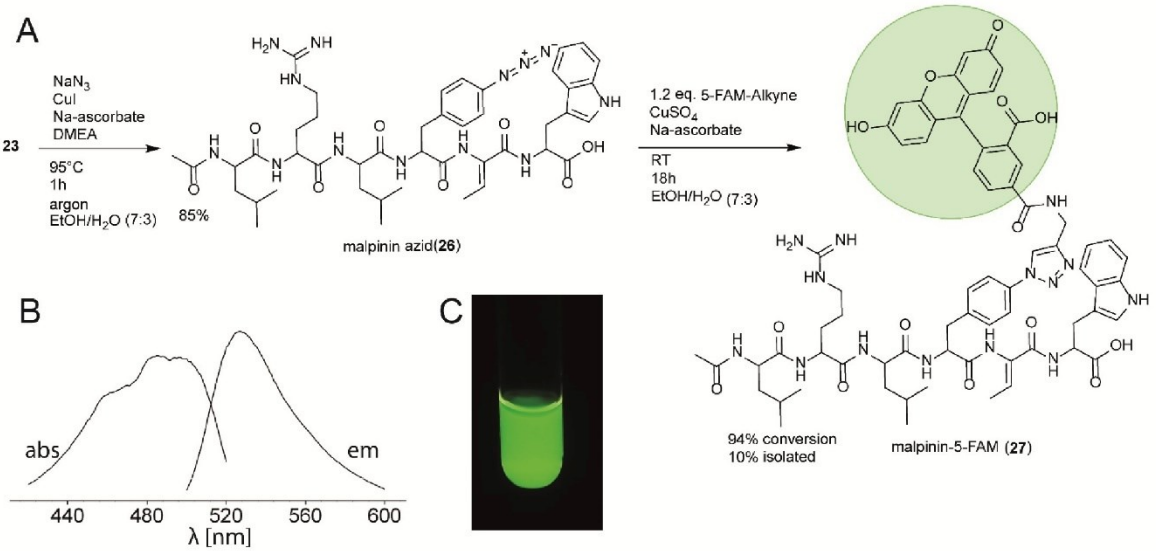


Figure 6. Synthesis of the malpinin-conjugated 5-FAM dye (**27**). **A.** Scheme of the azidation of **23** and CuAAC reaction resulting in triazole-linked conjugate **27**. **B.** Absorption and emission spectrum of **27**. **C.** Fluorescent solution of **27** under exposure to UV light (λ =384 nm).

from δ_C 126.3 ppm (in **1**) to δ_C 119.5 ppm (in **23**) in the ^{13}C-NMR spectrum confirmed the successful 4-bromo-L-Phe integration (Figure S39).
To test the suitability of the compound for further click chemistry, the aryl-halide (**23**) was substituted by an aryl-azide (**26**, *m/z* 900.4836 [*M*+H]$^+$, Figure S44) via an Ullmann-type copper catalysed nucleophilic aromatic substitution[45] (Figure 6). As a proof of principle, the click-enabled **26** was successfully coupled to the fluorescent dye 5-FAM-alkyne by a Cu(I)-catalysed azide-alkyne click reaction (CuAAC) to finally yield 2 mg **27** (*m/z* 1313.5734 [*M*+H]$^+$, Figure S45). The structure of the product was confirmed by ^{1}H- and ^{13}C-NMR (Figure S46-S47). Similar to 5-FAM, the malpinin-conjugate **27** showed fluorescent properties (Figure 6, Figure S48), i.e. an emission maximum at λ = 526 nm.

The current study demonstrates the potential of promiscuity profiling of A domains for biosynthetic diversification of nonribosomal peptides. Feeding alternative substrates of a promiscuous NRPS from early diverging fungi strongly shifted product profiles and enabled the production of click-functionalised compounds. Bioorthogonal-labelling may help to investigate the ecological and biological role of malpinins which are the predominant compounds in these fungi. Moreover, promiscuous A domains from early diverging fungi strongly enlarges the molecular toolbox in synthetic biology and eventually will bring up more chimeric natural products.

Experimental section

Organisms and culture maintenance.

The fungal strains *Mortierella alpina* ATCC 32222 and *Mortierella* (syn. *Linnemannia*) *amoeboidea* CBS 889.72 were purchased from the American Type Culture Collection (ATCC) and the Westerdijk Fungal Biodiversity Institute (CBS), respectively (Table S11). Cultures were maintained on MEP agar plates (30 g L^{-1} malt extract, 3 g L^{-1} soy peptone, 18 g L^{-1} agar) for 7 days at 25°C.
Liquid cultures (100 mL) for metabolite quantification and expression analyses were inoculated with five agar blocks (2 x 2 mm) and incubated at 25°C and 140 RPM for 3, 7 or 14 days (depending on the experiment). Media were MEP, potato dextrose broth (PDB, Sigma Aldrich), yeast extract peptone dextrose medium (YPD; 20 g L^{-1} peptone, 10 g L^{-1} yeast extract, 20 g L^{-1} glucose), lysogeny broth (LB; 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride), or hay medium (HM, 25 g L^{-1} hay, extracted with hot water, 100 mM phosphate buffer, pH 5.6). For the precursor-directed biosynthesis, 100 mL *Aspergillus* minimal medium[46] supplemented with 100 mM glucose and 20 mM ammonium nitrate was amended with 5 to 10 mM respective amino acids.
Escherichia coli strains (Table S11) used for plasmid propagation or heterologous protein production were maintained in LB medium amended with carbenicillin (50 µg L^{-1}) or kanamycin (100 µg L^{-1}), if required.

Molecular biological techniques.

Details on isolation of nucleic acids, genome sequencing, genome annotation, cloning of DNA into expression vectors and gene expression analysis are provided in the ESI. Constructed plasmids and oligonucleotides are listed in Tables S12- S13.

Protein purification.

Heterologous protein production and purification was carried out as previously described.[6] For details on the purification procedure, protein yields and SDS-PAGE gels see ESI (Figure S3, Table S3).

Adenylation enzyme activity assays.

Multiplexed hydroxamate assay (HAMA). The HAMA was carried out as previously described.[14] The hydroxamate samples were quantified on a Waters ACQUITY H-class UPLC system coupled to a Xevo TQ-S micro (Waters) tandem quadrupole instrument with ESI ionisation source in positive mode (method A, Table S14), by external calibration using a serial dilution of synthetic authentic hydroxamate standards.
MesG assay. The determination of kinetic parameters for MalA module 3 was conducted as previously described[14] using a continuous kinetic adenylation assay (MesG assay).[13] Enzymatic reaction was started by addition of 4 µM enzyme to a final reaction volume of 100 µl. Absorbance of released 7-methyl-6-thioguanin was monitored at λ_{max} = 355 nm on a Synergy H1 (BioTek) microplate reader at 30°C.
Acetylation assay. The determination of the acetylation activity of the C_s domain of MalA-M1 by conversion of L-leucine to *N*-acetyl-L-leucine was determined by a previously described protocol.[47]

Chemical analysis of metabolites.

General. Metabolite samples were routinely measured on an Agilent 1290 infinity II UHPLC coupled with an Agilent 6130 single quadrupole mass spectrometer (positive ionisation mode) using method B (Table S14). Metabolite preparation was conducted on a 1260 and 1200 HPLC system. HR-MS-MS spectra of identified compounds were recorded on a Q Exactive Plus mass spectrometer (Thermo Scientific). NMR spectra were measured on a Bruker Avance III 600 MHz spectrometer at 300 K using d_6-DMSO as solvent and internal standard (δ_H 2.50 ppm and δ_C 39.5 ppm).

Precursor-directed biosynthesis and metabolite quantification. After 7 days of cultivation in 100 mL AMM with 5 mM respective amino acids, mycelia were harvested and the culture broth was extracted three times with an equal volume of ethyl acetate. After solvent evaporation to dryness, the residue was solved in 5 mL methanol and 10 µl were subjected to UHPLC-MS analysis (method B, Table S14). Metabolites were quantified by integration of the area under the curve (AUC) of the extracted ion chromatograms (EIC). Quantification was assayed using a calibration curve from an authentic standard of **1** ranging from

0.005 to 5 mg mL $^{-1}$. Metabolite amounts are given as ratios relative to **1**.

Advanced Marfey´s analysis. Compound hydrolysis and amino acid derivatization was carried out as described.[6, 48] In brief, 0.1 mg of **6** or **7** were hydrolysed in 6 M HCl at 100° C overnight. The hydrolysate was neutralized (6 M KOH), evaporated to dryness and dissolved in 100 µL H_2O. In a total reaction volume of 100 µl, 25 µL of the hydrolysate (approx. 1 µM) were derivatized with 15 mM 1-fluoro-2,4-dinitrophenyl-5-L-leucine-amid (L-FDLA). For authentic standards, 10 µL of L- or D-configured amino acids (100 mM) were used. Finally, reaction was stopped by addition of 25 µL methanol and measured by UHPLC-MS with method C (Table S14). Retention times of respective coupling products were determined from extracted ion chromatograms (EIC).

Metabolite extraction and isolation.

Extraction and isolation of 6 and 7. Ten flasks, each containing 1 L AMM medium amended with 100 mM glucose, 20 mM ammonium nitrate and 8 mM methionine, were inoculated with *M. alpina* and cultivated for 7 days. Freeze-dried, ground mycelium was extracted three times using a mixture of methanol, butanol and DMSO (12:12:1, 400 mL). Culture broth was extracted three times using the same amount of ethyl acetate. Extracts were pooled, dried under vacuum and solved in 50 mL methanol. For initial separation, crude extracts were submitted to preparative HPLC using method D (Table S14). Fractions containing **6** and **7** were then transferred to a semipreparative HPLC and a separation from **1** as main contaminant was achieved by method E (Table S14). Method F (Table S14) was applied to separate pure **6** and **7** from **2** and **3** using a methanol gradient. Final separation of the isomers **6** (9 mg) and **7** (16 mg) was accomplished on a C_{18} reverse phase column using an acetonitrile gradient (method G, Table S14).

Extraction and isolation of 23. 23 was produced in 15 flasks, each containing 200 mL AMM medium and 5 mM 4-bromo-phenylalanin, inoculated with *M. alpina* as described above. Mycelium was harvested and extracted as described for **6** and **7**. After the first preparative separation step by HPLC (method D, Table S14), fractions containing **23** were submitted to further purification by method E (Table S14) yielding 12.8 mg pure **23**.

Chemical synthesis.

Synthesis and purification of 26. 12.8 mg of **23** (13.6 µmol) were solved in 100 µL reaction solvent (EtOH : water = 7 : 3). 2 eq. NaN_3, 0.1 eq. CuI (catalyst), 0.1 eq. *N,N'*-dimethylethylendiamin (DMEA, ligand) and 0.2 eq. sodium ascorbate were added to a final reaction volume of 400 µL as described elsewhere.[45] After incubation at 95°C for 1 h under argon atmosphere, the reaction was evaporated under vacuo, resolved in methanol and **26** was separated using method E (Table S14), yielding 40 mg (85%). The structure of **26** was verified by HR-MS/MS (Tables S6 and S7, Figure S44).

Synthesis of malpinin-coupled 5-FAM dye (27). Synthesis was pursued by mixing 8 µmol (7.25 mg) **26** and 17.5 µmol (7 mg) 5-FAM-alkyne (Jena Bioscience) in 1.2 mL reaction solvent (EtOH : water = 7 : 3). 5 µL 0.5 M $CuSO_4$ and 15 µL 0.5 M sodium ascorbate (both Jena Bioscience) were added immediately and after a reaction time of 12 h. After 18 h at RT and mild agitation (20 rpm) the reaction mixture was dried by lyophilisation and resolved in methanol. Separation was performed by HPLC using method E (Table S14). 1 mg was recovered and submitted to NMR and HR-MS/MS structure verification. (Table S6-S7, Figure S45).

Physicochemical and antimicrobial properties of 6 and 7.

Surface tension was determined by the ring tear off methods using the De Noüy ring tensiometer (Krüss Processor Tensiometer K12, Krüss, Hamburg, Germany) in a concentration range from 1000 to 1.95 µg/mL as described previously.[2] Antimicrobial activity testing was carried out by agar diffusion tests.[2] Ciprofloxacin dissolved in water (5 µg mL $^{-1}$) and amphotericin B in DMSO/methanol (1:1) (10 µg ml $^{-1}$) served as controls.

Statistical analysis.

Statistical analysis was carried out using GraphPad Prism 7 software. Pearson correlation was calculated assuming Gaussian distribution with a confidence interval of 95% and a significance level of 5%.

Conclusions

The current study is one of the rare cases, in which a complete determination of substrate specificities of an NRPS was carried out, which in turn led to the identification of a wealth of new metabolites (Figure S49). Promiscuity-based broadening of product diversity is a common strategy in nature to switch the biosynthesis from one compound to another with a higher selective advantage. In synthetic biology, this can be adapted: The cell-free biosynthesis of pharmaceutical natural products provides a sustainable and cost-effective alternative to traditional metabolite extraction or partially feasible chemical synthesis. In the past ten years, tremendous progress has been made to reconstitute and engineer NRPS peptide biosynthesis *in vitro* by combining several NRPS modules by either an inter-domain peptide linker[49, 50] or a DNA template as binding platform[51]. In both cases, promiscuous adenylation domains enable the incorporation of diverse substrates with a single enzymatic domain set or facilitate the incorporation of non-natural, but click-enable amino acids. Hence, the highly promiscuous domains of NRPS from early diverging fungi bear good prospects in combinatory enzyme engineering and enables a flexibly altered nonribosomal peptide design.

Author Contributions

The manuscript was written by HK and MG. JMW and AS performed the experimental work, with guidance and supervision from HK and MG. SZ performed cloning and the full-length gene sequencing of *malA*. SJ carried out biphasic measurements on malpinins. MG envisioned and managed the project and coordinated manuscript writing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We are grateful to Heike Heinecke and Andrea Perner (both from Hans Knöll Institute [HKI], Jena, Germany) for their technical assistance in recording NMR and HR-MS/MS spectra, respectively. We thank Kerstin Voigt (Jena Microbial Resource Collection, Jena, Germany) for antimicrobial testings. Hajo Kries gratefully acknowledges a fellowship from the Daimler und Benz Foundation and financial support from the DFG (grant number 441781663) and the Fonds der Chemischen Industrie (FCI).

Notes and references

The sequences of the *malA* gene from *M. alpina* and *M. amoeboidea* are deposited under Genbank accession numbers MW984675 and MW984676, respectively.

1. H. Kikukawa, E. Sakuradani, A. Ando, S. Shimizu and J. Ogawa, *J. Adv. Res.*, 2018, **11**, 15-22.
2. F. Baldeweg, P. Warncke, D. Fischer and M. Gressler, *Org. Lett.*, 2019, **21**, 1444-1448.
3. N. Koyama, S. Kojima, T. Fukuda, T. Nagamitsu, T. Yasuhara, S. Omura and H. Tomoda, *Org. Lett.*, 2010, **12**, 432-435.
4. N. Koyama, S. Kojima, K. Nonaka, R. Masuma, M. Matsumoto, S. Omura and H. Tomoda, *J. Antibiot. (Tokyo)*, 2010, **63**, 183-186.
5. T. Kamiyama, T. Umino, N. Nakayama, Y. Itezono, T. Satoh, Y. Yamashita, A. Yamaguchi and K. Yokose, *J. Antibiot. (Tokyo)*, 1992, **45**, 424-427.
6. J. M. Wurlitzer, A. Stanišic, I. Wasmuth, S. Jungmann, D. Fischer, H. Kries and M. Gressler, *Appl. Environ. Microbiol.*, 2020, **87**, e02051-02020.
7. S. D. Copley, *Curr. Opin. Struct. Biol.*, 2017, **47**, 167-175.
8. H. Kaljunen, S. H. Schiefelbein, D. Stummer, S. Kozak, R. Meijers, G. Christiansen and A. Rentmeister, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 8833-8836.
9. A. Stanišić and H. Kries, *ChemBioChem*, 2019, **20**, 1347-1356.
10. S. Meyer, J. C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth and E. Dittmann, *Cell Chem. Biol.*, 2016, **23**, 462-471.
11. H. Kaljunen, S. H. Schiefelbein, D. Stummer, S. Kozak, R. Meijers, G. Christiansen and A. Rentmeister, *Angew. Chem.*, 2015, **54**, 8833-8836.
12. L. G. Otten, M. L. Schaffer, B. R. Villiers, T. Stachelhaus and F. Hollfelder, *Biotechnol. J.*, 2007, **2**, 232-240.
13. D. J. Wilson and C. C. Aldrich, *Anal. Biochem.*, 2010, **404**, 56-63.
14. A. Stanišić, A. Hüsken and H. Kries, *Chem. Sci.*, 2019, **10**, 10395-10399.
15. L. Wang, W. Chen, Y. Feng, Y. Ren, Z. Gu, H. Chen, H. Wang, M. J. Thomas, B. Zhang, I. M. Berquin, Y. Li, J. Wu, H. Zhang, Y. Song, X. Liu, J. S. Norris, S. Wang, P. Du, J. Shen, N. Wang, Y. Yang, W. Wang, L. Feng, C. Ratledge, H. Zhang and Y. Q. Chen, *PLoS One*, 2011, **6**, e28319.
16. L. Wagner, B. Stielow, K. Hoffmann, T. Petkovits, T. Papp, C. Vagvolgyi, G. S. de Hoog, G. Verkley and K. Voigt, *Persoonia*, 2013, **30**, 77-93.
17. K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2019, **47**, W81-W87.
18. H. Kikukawa, E. Sakuradani, A. Ando, T. Okuda, M. Ochiai, S. Shimizu and J. Ogawa, *J. Biotechnol.*, 2015, **208**, 63-69.
19. B. Diez, S. Gutierrez, J. L. Barredo, P. van Solingen, L. H. van der Voort and J. F. Martin, *J. Biol. Chem.*, 1990, **265**, 16358-16365.
20. W. Kallow, J. Kennedy, B. Arezi, G. Turner and H. von Döhren, *J. Mol. Biol.*, 2000, **297**, 395-408.
21. J. F. Martin, *J. Antibiot. (Tokyo)*, 2000, **53**, 1008-1021.
22. M. Wittmann, U. Linne, V. Pohlmann and M. A. Marahiel, *FEBS J.*, 2008, **275**, 5343-5354.
23. H. J. Imker, D. Krahn, J. Clerc, M. Kaiser and C. T. Walsh, *Chem. Biol.*, 2010, **17**, 1077-1083.
24. L. Zhong, X. Diao, N. Zhang, F. Li, H. Zhou, H. Chen, X. Bai, X. Ren, Y. Zhang, D. Wu and X. Bian, *Nat. Commun.*, 2021, **12**, 296.
25. C. Greco, B. T. Pfannenstiel, J. C. Liu and N. P. Keller, *ACS Chem. Biol.*, 2019, **14**, 1121-1128.
26. C. Liu, A. Minami, T. Ozaki, J. Wu, H. Kawagishi, J. I. Maruyama and H. Oikawa, *J. Am. Chem. Soc.*, 2019, **141**, 15519-15523.
27. C. Mehner, D. Müller, A. Krick, S. Kehraus, R. Löser, M. Gütschow, A. Maier, H. F. Fiebig, R. Brun and G. M. König, *Eur. J. Org. Chem.*, 2008, **10**, 1732-1739.
28. I. Pergament and S. Carmeli, *Tetrahedron Lett.*, 1994, **35**, 8473-8476.
29. P. N. Leao, A. R. Pereira, W. T. Liu, J. Ng, P. A. Pevzner, P. C. Dorrestein, G. M. König, V. M. Vasconcelos and W. H. Gerwick, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 11183-11188.
30. E. N. Zainuddin, R. Jansen, M. Nimtz, V. Wray, M. Preisitsch, M. Lalk and S. Mundt, *J. Nat. Prod.*, 2009, **72**, 1373-1378.
31. A. Mavaro, A. Abts, P. J. Bakkes, G. N. Moll, A. J. Driessen, S. H. Smits and L. Schmitt, *J. Biol. Chem.*, 2011, **286**, 30552-30560.
32. M. R. Levengood, C. C. Kerwood, C. Chatterjee and W. A. van der Donk, *ChemBioChem*, 2009, **10**, 911-919.
33. W. H. Chen, K. Li, N. S. Guntaka and S. D. Bruner, *ACS Chem. Biol.*, 2016, **11**, 2293-2303.
34. S. Wang, Q. Fang, Z. Lu, Y. Gao, L. Trembleau, R. Ebel, J. H. Andersen, C. Philips, S. Law and H. Deng, *Angew. Chem.*, 2021, **60**, 3229-3237.
35. P. Marfey, *Carlsberg Res. Commun.*, 1984, **49**, 591-596.
36. S. Caboche, V. Leclère, M. Pupin, G. Kucherov and P. Jacques, *J. Bacteriol.*, 2010, **192**, 5143-5150.

37. T. Sano, T. Usui, K. Ueda, H. Osada and K. Kaya, *J. Nat. Prod.*, 2001, **64**, 1052-1055.
38. M. A. Rashid, K. R. Gustafson, J. L. Boswell and M. R. Boyd, *J. Nat. Prod.*, 2000, **63**, 956-959.
39. J. Kobayashi, T. Nakamura and M. Tsuda, *Tetrahedron*, 1996, **52**, 6355-6360.
40. G. R. Pettit, B. E. Toki, J. P. Xu and D. C. Brune, *J. Nat. Prod.*, 2000, **63**, 22-28.
41. G. Christiansen, B. Philmus, T. Hemscheidt and R. Kurmayer, *J. Bacteriol.*, 2011, **193**, 3822-3831.
42. T. K. Shishido, J. Jokela, A. Humisto, S. Suurnakki, M. Wahlsten, D. O. Alvarenga, K. Sivonen and D. P. Fewer, *Mar. Drugs*, 2019, **17**.
43. D. L. Niquille, I. B. Folger, S. Basler and D. Hilvert, *J. Am. Chem. Soc.*, 2021, **143**, 2736-2740.
44. H. Kries, R. Wachtel, A. Pabst, B. Wanner, D. Niquille and D. Hilvert, *Angew. Chem.*, 2014, **53**, 10105-10108.
45. J. Andersen, U. Madsen, F. Björkling and X. F. Liang, *Synlett*, 2005, **1**, 2209-2213.
46. R. W. Barratt, G. B. Johnson and W. N. Ogata, *Genetics*, 1965, **52**, 233-246.
47. F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker and M. A. Marahiel, *Chem. Biol.*, 2010, **17**, 872-880.
48. E. H. Hashimoto, H. Kato, Y. Kawasaki, Y. Nozawa, K. Tsuji, E. Y. Hirooka and K. Harada, *Chem. Res. Toxicol.*, 2009, **22**, 391-398.
49. M. Kaniusaite, R. J. A. Goode, J. Tailhades, R. B. Schittenhelm and M. J. Cryle, *Chem. Sci.*, 2020, **11**, 9443-9458.
50. K. A. J. Bozhueyuek, J. Watzel, N. Abbood and H. B. Bode, *bioRxiv*, 2020, DOI: 10.1101/2020.05.06.080655, 2020.2005.2006.080655.
51. H. M. Huang, P. Stephan and H. Kries, *Cell Chem. Biol.*, 2021, **28**, 221-227.

8 MANUSCRIPT VII

Sulfonium Acids Loaded onto an Unusual Thiotemplate Assembly Line Construct the Cyclopropanol Warhead of a Burkholderia Virulence Factor

Felix Trottmann, Keishi Ishida, Jakob Franke, **Aleksa Stanišić**, Mie Ishida-Ito, Hajo Kries, Georg Pohnert, and Christian Hertweck

Published manuscript: *Angew. Chem. Int. Ed.* 2020, 59, 13511–13515
doi: 10.1002/anie.202003958

Summary:

Malleicyprols from pathogenic *Burkholderia pseudomallei* were identified as important bacterial virulence factors, yet the biosynthetic origin of their cyclopropanol warhead has remained enigmatic. Here, we find that sulfonium acids, dimethylsulfoniumpropionate (DMSP) and gonyol, known as osmolytes are key intermediates en route to the cyclopropanol unit. We uncover a specialized pathway to DMSP involving a rare prokaryotic SET-domain methyltransferase for a cryptic methylation, and show that, before being transformed into gonyol, DMSP is loaded onto the NRPS-PKS hybrid assembly line by an adenylation domain dedicated to zwitterionic starter units.

The candidate is:

☐ First author ☐ Second author ☐ Corresponding author ☒ Coauthor

Estimated authors' contributions:

Author	Conception	Data analysis	Experimental	Writing	Provision of the material
FT	80 %	90 %	75 %	60 %	
KI			10 %	2.5 %	
JF	5 %		2.5 %	5 %	
AS		10 %	7.5 %	2.5 %	
MI			5 %	2.5 %	
HK	5 %			5 %	
GP				2.5 %	
CH	10 %			20 %	

Supplementary datasets can be found under:
https://onlinelibrary.wiley.com/doi/10.1002/ange.202003958 and are included in the enclosed CD-ROM.

GDCh Zuschriften Angewandte Chemie

Check for updates

Biosynthesis

Zitierweise: *Angew. Chem. Int. Ed.* **2020**, *59*, 13511–13515
Internationale Ausgabe: doi.org/10.1002/anie.202003958
Deutsche Ausgabe: doi.org/10.1002/ange.202003958

Sulfonium Acids Loaded onto an Unusual Thiotemplate Assembly Line Construct the Cyclopropanol Warhead of a *Burkholderia* Virulence Factor

*Felix Trottmann, Keishi Ishida, Jakob Franke, Aleksa Stanišić, Mie Ishida-Ito, Hajo Kries, Georg Pohnert, and Christian Hertweck**

Abstract: *Pathogenic bacteria of the Burkholderia pseudomallei group cause severe infectious diseases such as glanders and melioidosis. Malleicyprols were identified as important bacterial virulence factors, yet the biosynthetic origin of their cyclopropanol warhead has remained enigmatic. By a combination of mutational analysis and metabolomics we found that sulfonium acids, dimethylsulfoniumpropionate (DMSP) and gonyol, known as osmolytes and as crucial components in the global organosulfur cycle, are key intermediates en route to the cyclopropanol unit. Functional genetics and in vitro analyses uncover a specialized pathway to DMSP involving a rare prokaryotic SET-domain methyltransferase for a cryptic methylation, and show that DMSP is loaded onto the NRPS-PKS hybrid assembly line by an adenylation domain dedicated to zwitterionic starter units. Then, the megasynthase transforms DMSP into gonyol, as demonstrated by heterologous pathway reconstitution in E. coli.*

[*] F. Trottmann, Dr. K. Ishida, Dr. M. Ishida-Ito, Prof. Dr. C. Hertweck
Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology (HKI)
Beutenbergstrasse 11a, 07745 Jena (Germany)
E-mail: christian.hertweck@leibniz-hki.de
Homepage: https://www.leibniz-hki.de

A. Stanišić, Dr. H. Kries
Junior Research Group Biosynthetic Design of Natural Products
Leibniz Institute for Natural Product Research and Infection Biology (HKI)
Beutenbergstrasse 11a, 07745 Jena (Germany)

Prof. Dr. J. Franke
Institute of Botany, Leibniz University Hannover
30419 Hannover (Germany)

Prof. Dr. G. Pohnert
Institute for Inorganic and Analytical Chemistry
Friedrich Schiller University Jena
07743 Jena (Germany)

Prof. Dr. C. Hertweck
Natural Product Chemistry, Faculty of Biological Sciences
Friedrich Schiller University Jena
07743 Jena (Germany)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202003958.

Burkholderia pseudomallei group pathogens cause lethal infections that are difficult to treat.[1] The best-known members of this pathogen complex are *Burkholderia mallei*, which causes the zoonotic disease glanders, and *B. pseudomallei*, the causative agent of melioidosis. The low infective dose needed and the possibility of infections through inhalation have led to the classification of *B. mallei* and *B. pseudomallei* as biological warfare agents[2] and a threat to global health.[3] Since infections by these notorious pathogens are difficult to treat,[4] novel therapeutic approaches such as anti-virulence strategies are needed. As prerequisite to disarming pathogens, it is essential to understand their virulence factors and the biosynthetic pathways involved.[5] For *B. pseudomallei* and related pathogens such as the less virulent model organism *B. thailandensis*, various macromolecular[6] and low-molecular-weight virulence factors[3,7] have been identified. Notably, all pathogens of the *B. pseudomallei* complex share a gene locus coding for an unusual modular thiotemplate assembly line with components of modular non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). Gene inactivation experiments unequivocally linked this gene cluster, named *bur*, to pathogenicity in nematode[8] and mouse models.[9] However, the first metabolite associated to this pathway—burkholderic acid[10] *syn.* malleilactone[8] (**1**)—did not exhibit any activity explaining the phenotypes observed in the infection models.

Recently, we found **1** to be the inactivated form of the true virulence factor, a highly reactive, cyclopropanol-substituted congener named bis-malleicyprol (**2a**, Figure 1 A) formed from the monomer malleicyprol (**2b**).[11] Nematode and toxicity assays showed dramatically increased activity of **2a** compared to **1**, implicating the cyclopropanol warhead in virulence.[11] Thus, understanding its biosynthesis would set the basis for antivirulence strategies. According to stable-isotope labeling experiments and gene knockouts, the NRPS-PKS hybrid enzyme BurA assembles the cyclopropanol-containing fragment of **2b**, followed by dimerization to **2a**, which opens to form the inactive propanone-substituted unit of **1**, from a yet unknown methionine (**3**)-derived C3 building block and malonyl-CoA (Figure 1 B).[10] Yet, structures and biotransformations of the precursors loaded onto BurA have remained a riddle. Here we decipher the biogenetic origin of the rare cyclopropanol warhead of malleicyprol and show that a set of zwitterionic sulfonium acids initiates biosynthesis that play key roles in global sulfur cycling.

To elucidate the biogenetic origin of the malleicyprol warhead we first focused on the NRPS-PKS hybrid enzyme

GDCh Zuschriften Angewandte Chemie

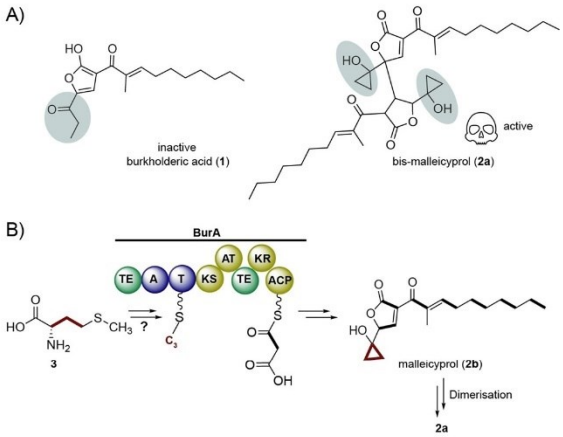


Figure 1. A) Structures of burkholderic acid (**1**) and bis-malleicyprol (**2a**) featuring a cyclopropanol warhead. B) Isotope-labeling studies[10] suggest acetate and methionine (**3**) as precursors to malleicyprol (**2b**), the monomer[11] of **2a**.

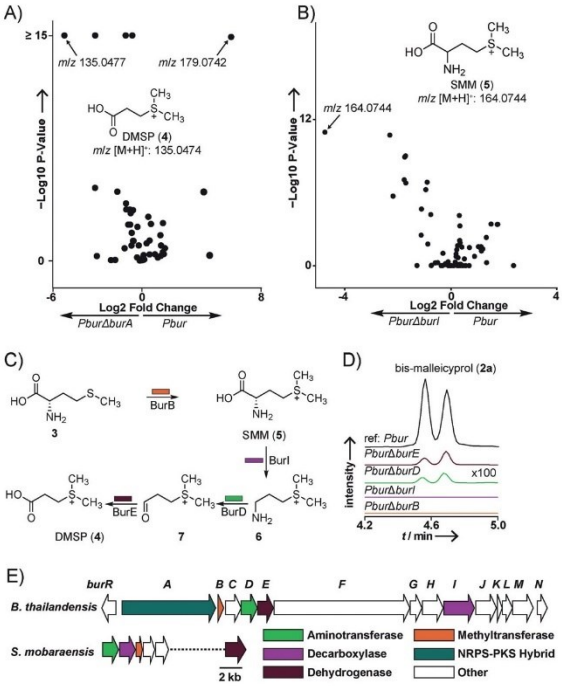


Figure 2. Identification of sulfonium intermediates in malleicyprol biosynthesis. A) Volcano plot analysis comparing pellet extracts of *B. thailandensis* variants *Pbur* and *PburΔburA*. B) Volcano plot analysis comparing supernatant extracts of *B. thailandensis* variants *Pbur* and *PburΔburI* (Figure S2). C) Proposed DMSP biosynthesis in *B. thailandensis*. D) Production of two bis-malleicyprol (**2a**, *m/z* 611.3589) diastereomers in gene inactivation mutants of *B. thailandensis Pbur* monitored by UHPLC-MS (EIC in negative ion mode). E) Genomic alignment of the DMSP assembly line from *S. mobaraensis* to the *bur* biosynthetic gene cluster.

BurA and its unknown starter unit. Specifically, we tested for accumulation of intermediates in a mutant of the malleicyprol overexpressing strain *B. thailandensis Pbur* lacking a functional copy of *burA* (*B. thailandensis PburΔburA*[10]). Candidates for the sulfur-containing pathway intermediates, however, could not be detected by routine HR-LCMS-based metabolic profiling. Only a comparative metabolomics analysis (*Pbur* vs. *PburΔburA*) searching for highly polar metabolites revealed the elusive methionine-derived starter unit. Compound **4** with *m/z* 135 accumulates in *B. thailandensis PburΔburA* cells (Figure 2A). Based on its exact mass (*m/z* 135.0474; $[M+H]^+$) we deduced the molecular formula ($C_5H_{11}O_2S$) of **4**. By comparison with authentic reference compounds we identified **4** as the zwitterionic compound dimethylsulfoniopropionate (DMSP; *m/z* 135.0474; $[M+H]^+$; Supporting Information, Figure S1). This finding is intriguing since DMSP plays a pivotal role in the marine organosulfur cycle, serving as osmolyte for marine algae and as abundant carbon and sulfur source for bacteria.[12] It is the precursor of the climate-relevant gas dimethylsulfide that is emitted at remarkably high amounts of $>10^7$ tons per year into the atmosphere.[13] Despite its wide distribution, DMSP has thus far not been implicated as a building block in natural product biosynthesis.

To confirm DMSP as a precursor of malleicyprols, we performed stable-isotope labeling experiments by chemically complementing suitable block mutants. Therefore, we required insight into the molecular basis of DMSP formation in *B. thailandensis*. By analogy to one of the established bacterial DMSP biosynthesis pathways[12] (see Ref. [14] for alternative routes to DMSP) methionine would undergo *S*-methylation to form *S*-methylmethionine (**5**), decarboxylation, transamination and oxidation (Figure 2C). In silico analysis of the *bur* gene locus revealed candidate genes for a methyltransferase (BurB), a decarboxylase (BurI), a transaminase (BurD), and a dehydrogenase (BurE) (Figure 2E). Comparison of the deduced protein sequences to the recently published bacterial DMSP biosynthetic machinery in *Streptomyces mobaraensis*[14b] showed BurD and BurE to be almost identical with their orthologues (96% and 98%), whereas BurB and BurI are only distantly related to their *S. mobaraensis* counterparts (35% and 48%).

To disrupt malleicyprol production, we individually inactivated each of the four putative DMSP biosynthesis genes in *B. thailandensis Pbur* by homologous crossover and replacement with a resistance cassette. The Δ*burD* and Δ*burE* mutants are still capable of producing the malleicyprol complex, albeit in reduced amounts (Figure 2D). Possibly, unspecific housekeeping enzymes take over transamination of sulfonium amine **6** and oxidation of the instable aldehyde **7**.[14b] In contrast, the two main malleicyprol diastereomers are absent in the Δ*burB* (methyltransferase) and Δ*burI* (decarboxylase) mutants (Figure 2D).

Closer inspection of culture extracts from *PburΔburI* revealed enrichment of metabolite **5** (*m/z* 164.0744 in positive ion mode) that is identical to the unusual charged amino acid *S*-methylmethionine **5** (SMM; Figure 1D and Figure S3).

Since **5** cannot be detected in the mutant lacking *burB*, its gene product BurB must act in the biosynthetic pathway upstream of DMSP. BurB's role in DMSP formation is remarkable as it belongs to class V of the methyltransferase superfamilies containing a SET-domain.[15] SET-domain-containing methyltransferases are well studied in eukaryotes and regulate gene expression through histone lysine methylation.[15] Yet, little is known about prokaryotic SET-domain methyltransferases, and natural product modifying methyltransferases have exclusively been found in class I (Rossmann-like fold) or class III (tetrapyrrole methylase).[16] To corroborate the function of BurB, we cloned and overexpressed *burB* in *Escherichia coli* and purified the His6-tagged protein via Ni-affinity. Incubation of purified BurB with L-methionine (**3**) and *S*-adenosylmethionine (SAM) generated SMM (Figure 3A), which was detected after derivatization with **8** to compound **9** (Figure 3A). Thus, BurB represents a novel *C-S* bond forming enzyme in secondary metabolism.[17]

Having established the key steps to DMSP in *B. thailandensis* and with suitable null producers at hand, we performed chemical complementation. Therefore, we synthesized $^{13}C_3$-labeled DMSP from $^{13}C_3$ acrylic acid (**10**, Figure 3C). Supplementation of the *Pbur*Δ*burI* mutant with $^{13}C_3$-DMSP not only restored production of the malleicyprols (see Figure S4) but also enriched ^{13}C in their cyclopropanol residue (Figure 3B). These results unequivocally confirm DMSP as a key intermediate in the formation of the malleicyprol warhead.

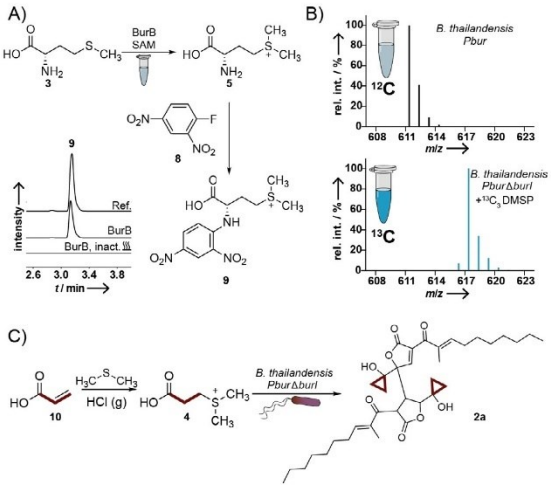


Figure 3. DMSP is a precursor of the cyclopropanol warhead of malleicyprol. A) Transformation of methionine (**3**) to *S*-methylmethionine (**5**; SMM) by BurB and derivatization with **8**; HR-LCMS detection of derivatized SMM; EIC *m/z* 330.0754 in positive ion mode; top: SMM derivatized with FDNB, middle: methylation of methionine with BurB, bottom: heat-inactivated BurB. B) Mass spectra of native bis-malleicyprol (top) and of ^{13}C-enriched bis-malleicyprol (bottom) C) Synthesis of ^{13}C-labeled DMSP and subsequent complementation of *B. thailandensis PburΔburI* leads to incorporation of the C_3 unit into **2**.

Since DMSP accumulates in the Δ*burA* mutant, we reasoned that this unusual zwitterionic substrate would be activated and loaded onto the *bur* assembly line. In NRPS, adenylation (A) domains select, activate and load amino acids onto the assembly line and are thus regarded as gatekeepers.[18] As bioinformatic substrate predictors[19] failed on BurA-A we generated a homology model of the A domain. In this way we noted the absence of the conserved aspartate (D235 in GrsA-A) present in all α-amino acid activating A domains.[18] Consequently, BurA-A was expected to select a non-canonical starter unit lacking the α-amino group usually bound by this aspartate. According to homology modeling, BurA-A shares important binding pocket features with the A-domain ATRR-A activating glycine betaine.[20] In both ATRR-A and BurA-A, the loop carrying conserved D235 (GrsA-A) has been replaced with a shorter loop carrying hydrophobic residues (Figure 4A; Figure S5). In another position, both binding pockets have an acidic residue (D606 in BurA-A) well placed to interact with a positively charged substrate moiety such as the sulfonium group of DMSP.

To verify whether BurA-A selects and activates DMSP we cloned and expressed the gene fragment for the A domain (*burA-A*) in *E. coli* BL21(DE3) and purified the His6-tagged protein via Ni-affinity. We probed the activity of BurA-A with the MesG/hydroxylamine assay, which monitors pyrophosphate released during substrate adenylation in a coupled photometric assay (Figure 4B).[21] From a panel of carboxylic acids, BurA-A shows highest activity for DMSP. In stark contrast, neither a mixture of the proteinogenic amino acids, **5** nor 3-(methylthio)propionic acid (**11**) are activated (Figure 4C). Apparently, the positively charged sulfonium group enables substrate binding. Replacing this group with protonated nitrogen in **12**, thus maintaining the positive charge, reduces activity more than four-fold. In substrate saturation kinetics with DMSP, we determined a k_{cat} of 2.4 min^{-1} and a K_M of 0.15 mM (Figure S7). Altogether, these results indicate that DMSP is the preferred substrate of BurA-A, making it the first adenylation domain that incorporates the osmolyte DMSP into a natural product assembly line.

An in silico analysis of the modular architecture of BurA suggested that DMSP, once loaded onto the thiolation (T) domain, would be elongated through Claisen condensation with malonyl-ACP, and the resulting β-keto intermediate transformed into the corresponding alcohol by the ketoreductase (KR) domain (Figure 4E). To identify the downstream product of DMSP we heterologously reconstituted the assembly line. Therefore, we cloned and expressed *burA* in *E. coli*. Only when we employed IPTG for induction and supplemented DMSP, we detected production of compound **13** with *m/z* 179.0742 (positive ion mode) in XAD16 extracts (Figure 4D). The presence of the same species in *B. thailandensis Pbur* (Figure 1B) indicates that compound **13** is not an artifact generated in *E. coli* but actually formed in the intact *bur* pathway. By HR-LCMS and MS^2 comparison with an authentic reference we found that **13** is identical to the sulfonium acid gonyol.[22] The structure of **13** agrees with our in silico prediction of the biosynthetic steps mediated by BurA. However, the discovery of this sulfonium intermediate

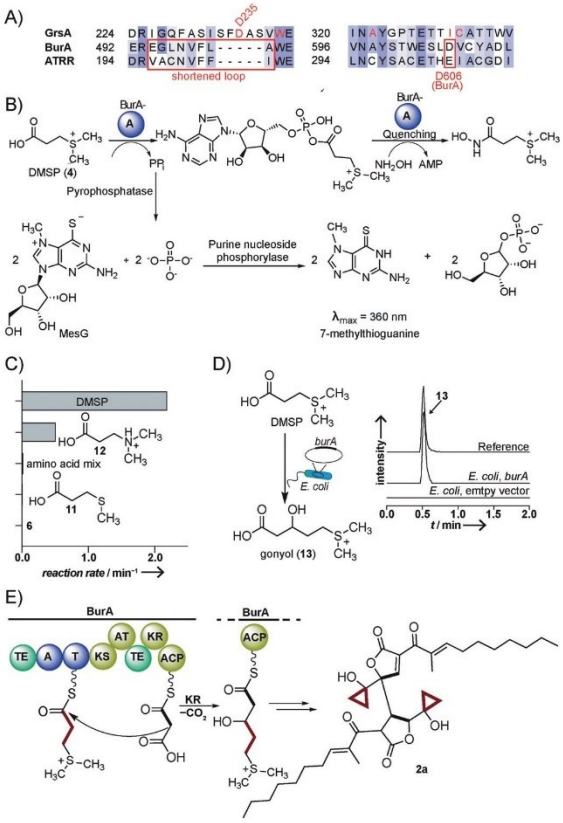


Figure 4. A) Sequence alignment of BurA-A with the canonical A domain GrsA and the glycine betaine accepting A domain ATRR. B) Concept of the MesG/hydroxylamine A domain assay C) Specificity profile of BurA-A; amino acid mix: all 20 proteinogenic amino acids (Figure S6). D) Heterologous production of gonyol in *E. coli*; UHPLC-MS monitoring: EIC (*m/z* 179.0736) in positive ion mode; top, synthetic reference, middle, culture extracts of *E. coli Rosetta2* (DE3) expressing *burA* on plasmid pHIS8; bottom, *E. coli Rosetta2* (DE3) with empty pHIS8. E) Loading of DMSP onto BurA leads to production of gonyol and malleicyprols.

of the malleicyprol assembly line is surprising, as **13** has been reported as a dominant zwitterion in the marine dinoflagellate *Gonyaulax polyedra*.[22] Moreover, **13** is widely distributed as minor osmolyte in several phytoplankton groups.[23] Previous studies have identified DMSP and acetate as the precursors of **13** in *G. polyedra*,[22,24] but the enzymes involved in gonyol biosynthesis have remained unknown. We now report BurA as the first enzyme involved in a gonyol biosynthetic pathway, and unexpectedly, it is a modular NRPS-PKS hybrid. We reason that the sulfonium group represents a leaving group, likely as dimethylsulfide that enables cyclopropanol formation downstream of BurA. The enzymes and mechanisms involved in the cyclization are the subject of ongoing studies.

In summary, we have uncovered crucial steps in the biosynthetic pathway to the virulence factor malleicyprol employed by animal and human pathogenic *Burkholderia* species. Our findings have broad implications for ecology and synthetic biology. We describe BurB as a new *C-S* bond-forming enzyme in secondary metabolism that mediates a cryptic methylation to form the sulfonium group of the DMSP precursor. A role of this zwitterionic sulfonium acid in bacterial secondary metabolism is new. In contrast, DMSP is widely distributed in marine life, and metagenomics of known DMSP methyltransferase genes show that bacteria are significant producers of DMSP in marine environments.[25] Our discovery of a new methyltransferase associated with DMSP biosynthesis and the identification of a PKS-NRPS hybrid as a gonyol synthetase allows for genomics-based identification of ecologically relevant producer strains. From a biosynthetic perspective DMSP is noteworthy as a novel PKS primer unit,[26] and the sulfonium-accepting adenylation domain BurA-A is an important addition to the synthetic biology toolbox that opens up new possibilities for engineering polyketides and nonribosomal peptides.

Acknowledgements

We thank A. Perner for LC-MS measurements. Financial support by the DFG (SFB 1127 ChemBioSys, and Leibniz Award to C.H.) and the European Regional Development Fund (ERDF) (MassNat) is gratefully acknowledged. J.F. acknowledges financial support by the SMART BIOTECS alliance between the Technische Universität Braunschweig and the Leibniz Universität Hannover, supported by the Ministry of Science and Culture (MWK) of Lower Saxony, Germany. H.K. received a fellowship from the Daimler and Benz foundation.

Conflict of interest

The authors declare no conflict of interest.

Keywords: Biosynthesis · DMSP · Mass spectrometry · NRPS · Virulence factors

[1] G. C. Whitlock, D. M. Estes, A. G. Torres, *FEMS Microbiol. Lett.* **2007**, *277*, 115–122.
[2] J. Gilad, I. Harary, T. Dushnitsky, D. Schwartz, Y. Amsalem, *Isr. Med. Assoc. J.* **2007**, *9*, 499–503.
[3] J. B. Biggins, H.-S. Kang, M. A. Ternei, D. DeShazer, S. F. Brady, *J. Am. Chem. Soc.* **2014**, *136*, 9484–9490.
[4] D. Dance, *Int. J. Antimicrob. Agents* **2014**, *43*, 310–318.
[5] R. Thänert, O. Goldmann, A. Beineke, E. Medina, *Nat. Commun.* **2017**, *8*, 14268.
[6] a) A. Cruz-Migoni, G. M. Hautbergue, P. J. Artymiuk, P. J. Baker, M. Bokori-Brown, C.-T. Chang, M. J. Dickman, A. Essex-Lopresti, S. V. Harding, N. M. Mahadi, L. E. Marshall, G. W. Mobbs, R. Mohamed, S. Nathan, S. A. Ngugi, C. Ong, W. F. Ooi, L. J. Partridge, H. L. Phillips, M. F. Raih, S. Ruzheinikov, M. Sarkar-Tyson, S. E. Sedelnikova, S. J. Smither, P. Tan, R. W. Titball, S. A. Wilson, D. W. Rice, *Science* **2011**, *334*, 821–824; b) W. J. Wiersinga, H. S. Virk, A. G. Torres, B. J. Currie, S. J.

Peacock, D. A. B. Dance, D. Limmathurotsakul, *Nat. Rev. Dis. Primers* **2018**, *4*, 17107.

[7] a) H. Yang, C. D. Kooi, P. A. Sokol, *Infect. Immun.* **1993**, *61*, 656–662; b) A. F. Alice, C. S. López, C. A. Lowe, M. A. Ledesma, J. H. Crosa, *J. Bacteriol.* **2006**, *188*, 1551–1566; c) J. Franke, K. Ishida, M. Ishida-Ito, C. Hertweck, *Angew. Chem. Int. Ed.* **2013**, *52*, 8271–8275; *Angew. Chem.* **2013**, *125*, 8429–8433; d) J. Franke, K. Ishida, C. Hertweck, *J. Am. Chem. Soc.* **2014**, *136*, 5599–5602; e) J. Franke, K. Ishida, C. Hertweck, *Chem. Eur. J.* **2015**, *21*, 8010–8014; f) R.-R. Wong, C. Kong, S.-H. Lee, S. Nathan, *Sci. Rep.* **2016**, *6*, 27475; g) F. Trottmann, J. Franke, K. Ishida, M. García-Altares, C. Hertweck, *Angew. Chem. Int. Ed.* **2019**, *58*, 200–204; *Angew. Chem.* **2019**, *131*, 206–210.

[8] a) J. B. Biggins, M. A. Ternei, S. F. Brady, *J. Am. Chem. Soc.* **2012**, *134*, 13192–13195; b) J. R. Klaus, J. Deay, B. Neuenswander, W. Hursh, Z. Gao, T. Bouddhara, T. D. Williams, J. Douglas, K. Monize, P. Martins, C. Majerczyk, M. R. Seyedsayamdost, B. R. Peterson, M. Rivera, J. R. Chandler, *J. Bacteriol.* **2018**, *200*, e00008-18.

[9] a) M. G. Gutierrez, D. R. Yoder-Himes, J. M. Warawa, *Front. Cell. Infect. Microbiol.* **2015**, *5*, 78; b) M. G. Moule, N. Spink, S. Willcocks, J. Lim, J. A. Guerra-Assunção, F. Cia, O. L. Champion, N. J. Senior, H. S. Atkins, T. Clark, G. J. Bancroft, J. Cuccui, B. W. Wren, *Infect. Immun.* **2016**, *84*, 701–710.

[10] a) J. Franke, K. Ishida, C. Hertweck, *Angew. Chem. Int. Ed.* **2012**, *51*, 11611–11615; *Angew. Chem.* **2012**, *124*, 11779–11783; b) Note: Met accumulation in media leads to increased malleilactone production: A. Li, D. Mao, A. Yoshimura, P. C. Rosen, W. L. Martin, E. Gallant, M. Wühr, M. R. Seyedsayamdost, *mBio* **2020**, *11*, e03210-19.

[11] F. Trottmann, J. Franke, I. Richter, K. Ishida, M. Cyrulies, H.-M. Dahse, L. Regestein, C. Hertweck, *Angew. Chem. Int. Ed.* **2019**, *58*, 14129–14133; *Angew. Chem.* **2019**, *131*, 14267–14271.

[12] M. A. Moran, B. P. Durham, *Nat. Rev. Microbiol.* **2019**, *17*, 665–678.

[13] A. J. Kettle, M. O. Andreae, *J. Geophys. Res.* **2000**, *105*, 26793–26808.

[14] a) A. R. J. Curson, B. T. Williams, B. J. Pinchbeck, L. P. Sims, A. B. Martínez, P. P. L. Rivera, D. Kumaresan, E. Mercadé, L. G. Spurgin, O. Carrión, S. Moxon, R. A. Cattolico, U. Kuzhiumparambil, P. Guagliardo, P. L. Clode, J.-B. Raina, J. D. Todd, *Nat. Microbiol.* **2018**, *3*, 430–439; b) C. Liao, F. P. Seebeck, *Angew. Chem. Int. Ed.* **2019**, *58*, 3553–3556; *Angew. Chem.* **2019**, *131*, 3591–3594.

[15] a) R. Alvarez-Venegas, *Front. Genet.* **2014**, *5*, 65–65; b) C. Nwasike, S. Ewert, S. Jovanovic, S. Haider, S. Mujtaba, *Ann. N. Y. Acad. Sci.* **2016**, *1376*, 18–28.

[16] D. K. Liscombe, G. V. Louie, J. P. Noel, *Nat. Prod. Rep.* **2012**, *29*, 1238–1250.

[17] K. L. Dunbar, D. H. Scharf, A. Litomska, C. Hertweck, *Chem. Rev.* **2017**, *117*, 5521–5577.

[18] a) G. L. Challis, J. Ravel, C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211–224; b) F. Kudo, A. Miyanaga, T. Eguchi, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 515–536; c) T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493–505.

[19] a) C. N. Boddy, *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 443–450; b) M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362–W367.

[20] Y. Hai, A. M. Huang, Y. Tang, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 10348–10353.

[21] B. P. Duckworth, D. J. Wilson, C. C. Aldrich, *Methods Mol. Biol.* **2016**, *1401*, 53–61.

[22] H. Nakamura, K. Fujimaki, O. Sampei, A. Murai, *Tetrahedron Lett.* **1993**, *34*, 8481–8484.

[23] B. Gebser, G. Pohnert, *Mar. Drugs* **2013**, *11*, 2168–2182.

[24] H. Nakamura, T. Jin, M. Funahashi, K. Fujimaki, O. Sampei, A. Murai, T. Roenneberg, J. W. Hastings, *Tetrahedron* **1997**, *53*, 9067–9074.

[25] L. Zha, Y. Jiang, M. T. Henke, M. R. Wilson, J. X. Wang, N. L. Kelleher, E. P. Balskus, *Nat. Chem. Biol.* **2017**, *13*, 1063–1065.

[26] B. T. Williams, K. Cowles, A. Bermejo Martínez, A. R. J. Curson, Y. Zheng, J. Liu, S. Newton-Payne, A. J. Hind, C.-Y. Li, P. P. L. Rivera, O. Carrión, J. Liu, L. G. Spurgin, C. A. Brearley, B. W. Mackenzie, B. J. Pinchbeck, M. Peng, J. Pratscher, X.-H. Zhang, Y.-Z. Zhang, J. C. Murrell, J. D. Todd, *Nat. Microbiol.* **2019**, *4*, 1815–1825.

[27] B. S. Moore, C. Hertweck, *Nat. Prod. Rep.* **2002**, *19*, 70–99.

Manuscript received: March 17, 2020
Accepted manuscript online: April 21, 2020
Version of record online: May 29, 2020

9 Discussion

Engineering nonribosomal synthetases offers an attractive, scalable and sustainable route towards tailored peptides. The progress in this direction has been impeded by incomplete mechanistic understanding of how these complex enzymes operate. The development of reliable engineering strategies builds upon the insight into the factors determining substrate specificity and kinetic bottlenecks within the assembly line. A comprehensive overview of adenylation assays and the role of the A-domain in NRPS engineering has been laid out in Manuscript I. The problem of the tedious and time-consuming measurement of A-domain activity has been addressed in Manuscript II by developing the HAMA specificity assay for adenylating enzymes. This assay provides specificity profiles recorded under competition conditions making it superior over previously employed methods in this purpose. Next, HAMA has been used to probe the functional

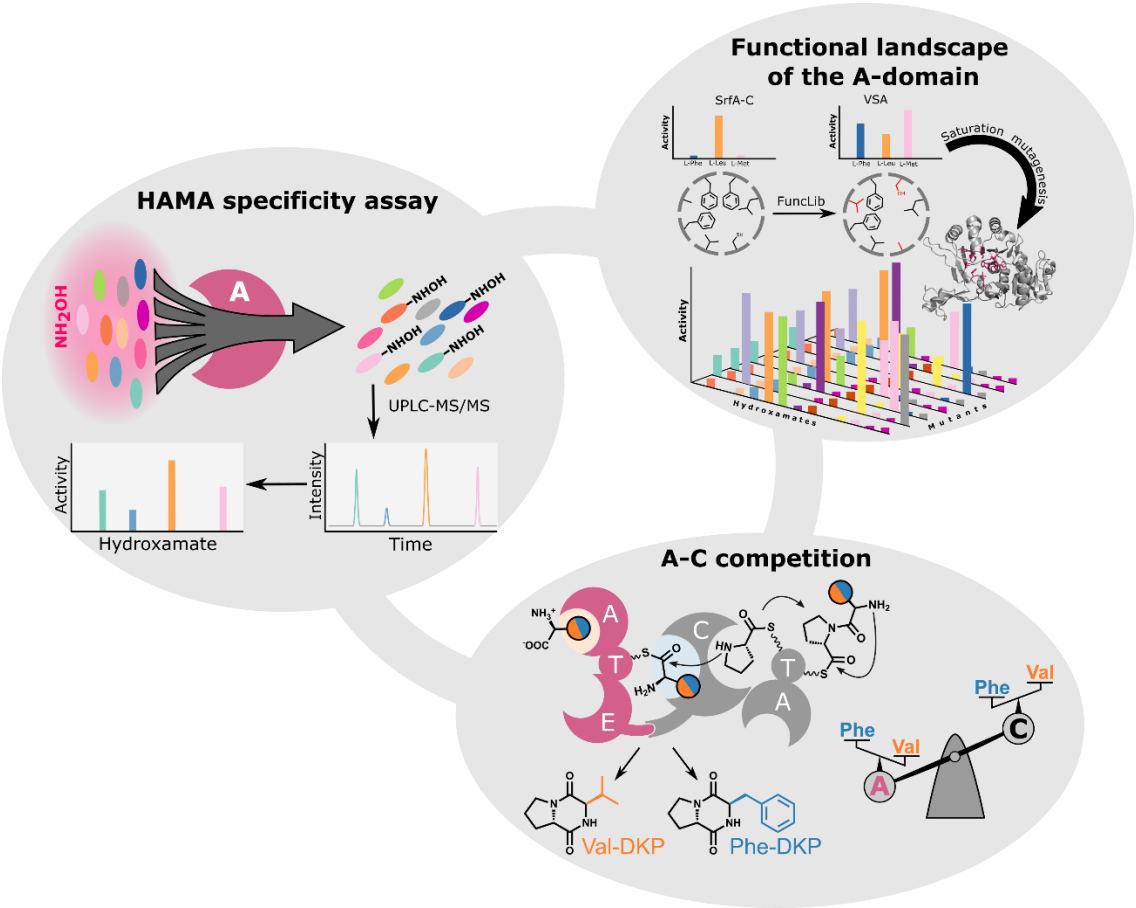


Figure 7. Core projects of this thesis.

space of the A-domain in manuscript III and characterize novel fungal A-domains in *Mortierella alpina* in Manuscript V and VI. A thorough kinetic dissection of the NRPS mechanism has been described in Manuscript IV by taking advantage of an engineered, dimodular NRPS system with opposing A- and C-domain specificity. Experimental determination of individual rate constants and nonlinear kinetic modelling have been employed to demonstrate the power of the A-domain to override the selectivity barrier imposed by the C-domain.

9.1 Development of the specificity assay for A-domains

The growing pool of identified NRPS sequences provides a powerful training dataset for sequence-based algorithms for the prediction of A-domain specificity.[41,45] However, the experimental characterization of A-domains has lagged behind. Classical assays for adenylation activity measuring pyrophosphate released during adenylation cannot be readily transferred for use in A-domains of NRPSs. Activated amino acids remain tightly bound in the binding pocket of the A-domain awaiting the second, thiolation half-reaction. Once excised from their native enzymatic scaffold, A-domains lose the ability to execute multiple turnovers, with aminoacyl adenylate acting as a strong inhibitor.[101–104] As pyrophosphate release assays suffer from product inhibition, the pyrophosphate exchange instead probes ^{32}P-ATP synthesis in the reverse reaction in the presence of excess pyrophosphate, which has been the gold standard for measuring A-domain activity for a long time. However, its discontinuous assay format and tedious sample processing have been major drawbacks.

As numerous engineering attempts clearly illustrate, repurposing NRPSs is not a trivial task. Even if the desired change in specificity is achieved, a loss of catalytic efficiency is likely to require iterative rounds of mutagenesis to rescue the activity of the enzyme. Only one substrate at a time can be analysed with the pyrophosphate exchange assay, requiring at least 20 separate reactions for determining the specificity of a single enzyme variant towards all proteinogenic amino acids.[69,99] Additionally, these conditions do not reflect the situation in the cell where multiple substrates compete for binding. Specificity profiles of A-domains are typically generated by long incubation of excised enzymes with high concentration of individual substrates. As there is no substrate competition, preferred substrates will cause quick and complete exchange while side activities continue to accumulate, resulting in specificity profiles appearing falsely promiscuous. HAMA completely resolves this issue by operating under competition conditions while preventing substrate depletion.

9.1.1 Specificity constants of adenylation drive the formation of amino acid hydroxamates

In Manuscript II, I have taken advantage of hydroxylamine quenching to develop a powerful, quick and robust assay for determining substrate specificity of A-domains in a single reaction. Hydroxylamine was shown to be nucleophilic enough to react with aminoacyl adenylate bound to the A-domain, releasing AMP and a resulting amino acid hydroxamate.[93] This reaction has already been adapted for determining A-domain specificity through spectrophotometric detection of coloured complexes between Fe^{3+} and hydroxamate.[142] However, this method suffers from poor sensitivity and low throughput. We have taken a step further by allowing the enzyme to process an equimolar mixture of substrates in the presence of hydroxylamine and quantifying the amino acid hydroxamates formed. Hydroxamates of 19 proteinogenic and three nonproteinogenic (phenyl-glycine, β-phenylalanine and pipecolic acid) amino acids were synthesized and characterized (Manuscript II, Supplementary Information). These extremely polar compounds are separated with hydrophilic interaction chromatography (HILIC) and detected by ESI-MS/MS, achieving limits of quantitation in the nanomolar range. Additionally, it is possible to distinguish isomers of the substrate, for instance D- and L-forms, using isotopic labels. As expected from kinetic theory, under competition conditions, the amounts of hydroxamates formed in HAMA correlate well with specificity constants (k_{cat}/K_M) from saturation kinetics of the adenylation step (Manuscript II, Figure 2a). This observation confirms that hydroxamate formation indeed informs about the adenylation step. While determining complete Michaelis Menten kinetics for each substrate would require days of work, a complete HAMA specificity profile can be generated in less than one hour.

9.1.2 HAMA is generally applicable

To demonstrate that HAMA can be used in different experimental settings with a wide range of proteins, we proceeded to apply it on a panel of characterized and uncharacterized NRPS modules, a multimodular NRPS and engineered variants. In all cases, specificity profiles confirmed the expected substrate incorporation with marginal side activities, typical for highly specific wild type enzymes. HAMA can be further extended to the analysis of aminoacyl-tRNA synthetases, key enzymes from primary metabolism catalysing a similar reaction as NRPS A-domains using tRNA as a nucleophile instead of a T-domain (Manuscript II, Figure 3b). A meaningful specificity profile was obtained even for the heterologously expressed tetramodular protein GrsB (Manuscript II, Figure 5). The ability to distinguish multiple A-domains of a multimodular NRPS, if the specificities are different, will be advantageous in cases where cutting out modules harms protein integrity.

In addition to confirming the specificities of known NRPS modules, HAMA has been utilized for the characterization of A-domains from newly identified NRPS clusters. An uncharacterized A-T didomain from jessenipeptin BGC ($JesA1_{AT}$) predicted to activate L-Thr was expressed and assayed (Manuscript II, Figure 3a). Despite the poor expression level and the presence of contaminants in the protein samples, the amount of the active protein was enough to detect the formation of L-Thr hydroxamate, thus confirming postulated specificity. This highlights an additional advantage of HAMA's highly specific detection method over PP_i release-based adenylation assays, which are often plagued with interferences arising from phosphate or amino acid contaminations. Additionally, a simple purification step over nickel affinity beads efficiently diminishes background activity which could potentially arise from alternative intracellular adenylating enzymes such as AARSs.

HAMA has proven to be a valuable tool for characterizing A-domains from various sources including fungi. While prevalent in Ascomycota, early diverging fungi are not considered prolific producers of secondary metabolites. Nevertheless, the cyclic peptides malpicyclin (MpcA) and malpibaldin (MpbA) have been isolated from cultures of *Mortierella alpina*. Heterologous expression of two adenylation domains from predicted BGCs enabled the experimental characterization of specificity (Manuscript V, Figure 5). HAMA profile confirmed the postulated incorporation of L-Arg by the third module of MpcA with no detectable side activities. In contrast, MpbA3 showed promiscuous activity towards aromatic amino acids. While the identity of activated amino acids corresponds to residues found at that position in the product, the peptide ratios did not exactly match HAMA profiles. Such discrepancies may be caused by downstream specificity filters or differences in cytoplasmatic substrate availability. When interpreting HAMA profiles of A-domains that are naturally promiscuous, it is important to also consider the intracellular substrate concentrations which can vary by several orders of magnitude.[145] Facile detection of promiscuity is an important strength of HAMA that has enabled the discovery of dozens of novel NRPs in *Mortierella*, and will improve our understanding of NRPS biology and evolution (Manuscript VI).

Beside characterization and profiling of A-domains from newly discovered NRPS clusters, HAMA can be utilised also for engineering. In this setting, it is essential to detect promiscuous, side activities towards alternative substrates which can be orders of magnitude lower than the wild type. For instance, HAMA has been employed for the directed evolution of sdV-GrsA, a chimeric NRPS module suffering a ~1000-fold loss of activity compared to its progenitor GrsA from gramicidin S synthetase (Manuscript II, Figure 4 and Manuscript IV). By reverting subdomain residues back to their original identity in GrsA, small and focused libraries of single mutants have been designed (Manuscript IV, Supplementary Information). Characterization of mutants with HAMA revealed a range of substrate specificities, despite the low activity.

9.2 Specificity landscape of the A-domain

To date, the investigation of A-domain specificity has been limited to sequence analysis of specificity code residues or finding routes towards novel specificities. However, due to the lack of suitable assays, it was not possible to thoroughly study the influence of mutations on A-domain substrate specificity. With HAMA as a straightforward assay for determining complete specificity profiles of A-domains, we proceeded to use it to investigate how mutational pressure affects substrate selection. This approach required adapting HAMA to screening format in 96 well plates. Using affinity purified His_6-proteins trapped on magnetic beads, without the need for elution, complete specificity profiles with 19 substrates have been generated for hundreds of enzyme variants per day.

Using HAMA screening, we have generated a promiscuous variant of the SrfAC A-domain. We show that a small, focused library of triple mutants can be sufficient to relax the wild type specificity of A-domain of SrfAC (Manuscript III, Figure 2). Three out of 8 specificity code residues were selected for simultaneous randomization. We took advantage of FuncLib to filter out unproductive residue combinations in the binding pocket thus dramatically reducing the size of the library for screening. Indeed, FuncLib randomization resulted in 50% of library members showing detectable activity. From the pool of 210 mutants, we selected VSA showing enhanced stability, unaffected catalytic efficiency and expanded substrate repertoire with respect to the parent SrfAC. This outcome is remarkable considering the severe reductions of A-domain activity often suffered upon site directed mutagenesis of the binding pocket.[38,117,118,146,147] In addition to the wild-type substrate L-Leu, VSA binds and activates L-Phe and L-Met at nearly equal rates. This is likely a consequence of the F702S mutation creating more space in the binding pocket for accommodation of bulky L-Phe and L-Met side chains. A similar single W239S (GrsA numbering) mutation at the bottom of the binding pocket of GrsA was sufficient to accommodate the propargyl- group of the noncognate propargyl-tyrosine substrate.[127] Surprisingly, VSA is also more stable compared to the wild type SrfAC (Manuscript III, Figure 2C). Natural proteins are typically not optimized to be exceptionally stable outside of their native context[148] which is why protein stabilization can be routinely achieved by directed evolution.[149,150] Additionally, stabilizing mutations are located distal from the active site[151–153] while gain of function mutations are expected to have the opposite, destabilizing effect.[154,155] In contrast, VSA shows both expanded substrate scope and increased thermostability.

The promiscuous VSA variant was a good starting point for further exploring the specificity landscape in single mutational steps, because specificity changes are more visible when several products are above the detection limit. Therefore, we generated site-saturation mutagenesis libraries at 15 individual positions of the VSA A-domain and determined a complete specificity profile for each variant. This protocol allowed the most

thorough assessment of A-domain substrate flexibility yet. We have revealed a remarkable flexibility of adenylation specificity, at least within a broad range of apolar substrates (Manuscript III, Figure 3). No mutants had detectable activity towards polar or charged substrates, which is unsurprising considering the phylogenetic distance between A-domains with apolar and polar or charged substrates. In the SrfAC scaffold, a larger number of mutations seems necessary to bridge the gap towards the activation of the polar and charged amino acids. However, the gap between apolar and charged substrates is smaller, or absent, in other scaffolds. In a *Planktothrix* A-domain, unusual bispecificity for apolar Tyr and charged Arg has been observed.[119] Additionally, 23 % of single mutants were above 50 % of VSA activity levels, illustrating the high mutational tolerance of the SrfAC A-domain. A striking finding is the invariability of 4 out of 8 specificity code positions. This may arise as a consequence of epistasis which would condition the beneficial effects of acquired mutations on others fixed at earlier stages of evolution. The contribution of epistasis to enzyme evolution has been debated.[156–159] While pairwise epistasis has only around 5 % incidence at the level of the whole protein,[160,161] this frequency can dramatically increase up to 80 % when it comes to activity-enhancing mutations.[159,162–164] Nevertheless, the deleterious effects of mutations at these positions are likely a culprit for chronically unsuccessful rational mutagenesis of A-domain specificity code. We show that variability at three positions can result in dramatically different specificity profiles, ranging from almost completely specific (V660L, G728M) to a remarkable S702F mutant activating 7 different substrates (Manuscript III, Figures 4 and 5). Full coverage of the sequence space allows us to identify trends directed towards individual substrate groups. Aromatic amino acids are favoured by Ala mutation at positions V660 and S702 while Gly at V760 confers the ability to activate D-amino acid. Small amino acids are favoured by Ala, Leu or Met mutation at G728. An interesting observation is the high prevalence of variants activating L-Met, a substrate rarely encountered in natural NRP structures.[77,165] We show that second shell residues do not play a significant role in substrate selection, generally affecting only activity but to a much lower extent the specificity. Although second shell residues can enhance the accuracy of sequence-based algorithms for the prediction of A-domain specificity, their importance is likely more prominent at later evolutionary stages during the refinement and optimization of the acquired activity.

The potential of engineering adenylating enzymes is best illustrated with the example of amino acid tRNA synthetases (aaRS) which were engineered to activate dozens of nonproteinogenic substrates for ribosomal protein synthesis.[166] Considering that their essential role made them one of the most substrate-specific enzyme classes in the cellular repertoire and that they catalyse essentially the same reaction as NRPS A-domains, I expect that at least a similar degree of functional diversification will be achievable for the latter. For the first time, we here demonstrate the great functional

flexibility of the SrfAC A-domain in terms of substrate specificity and show that even a minimal evolutionary step can result in vastly different specificity profiles. This is enabled by HAMA as an indispensable tool for quantifying the contribution and relative importance of individual active site residues for specificity. Since A-domain evolution is tightly coupled to substrate specificity, we envision that binding pocket mutagenesis will be most efficient for specificity switches towards structurally related substrates. Less conservative changes will likely require domain transplantation or extensive directed evolution campaigns. Continuing work on A-domains of different specificities and phylogenetic origins will prove essential to identify the engineering routes causing minimal disturbances to the enzyme and having the highest likelihood for success. When large datasets of specificity data are collected over several screening rounds, HAMA can map structure-function relationships of A-domains in unprecedented detail. This information will possibly prove useful to train machine learning algorithms to predict mutational trajectories towards substrates of interest.

9.3 Crosstalk between A and C-domain in a chimeric NRPS

Engineering of biosynthetic assembly lines often suffers from low product titres. It has been suggested that functional A-domain is not sufficient to efficiently incorporate a noncognate substrate into the peptide due to specificity filters at subsequent biosynthetic steps. Lack of straightforward ways to measure the acceptance of alternative substrates by downstream modules is one of the main bottlenecks in the engineering process. It is now generally accepted that a second specificity filter acts at the level of the C-domain, interfering with successful A-domain engineering.[55,91,129,136] While the stereospecificity of the acceptor and the donor site is well established, the importance of the side chain specificity is more ambiguous.[59,61,115,167] Sequence analysis of the C-domain is able to predict the stereochemistry of the substrate, but the existence of a proofreading mechanism based on side-chain identity is not well established.[62] In Manuscript IV, we have taken advantage of an engineered bimodular NRPS system with conflicting A and C-domain specificities to investigate the relative contribution of both domains to the kinetics of product formation. First module sdV-GrsA contains a chimeric A-domain generated by subdomain swapping, showing promiscuous adenylation, while the second module GrsB1 maintains a wild type specificity. We employ nonlinear kinetic modelling to probe the condensation reaction and explain the unusual time-dependent inversion of the formation of two products.

9.3.1 Subdomain swapping can be fixed with binary mutations

By transplanting the substrate-binding “subdomain” of the A-domain from L-Val-specific GrsB2 into L-Phe-specific GrsA from gramicidin S synthetase, Kries et al. have generated

chimeric sdV-GrsA able to incorporate both Val and Phe but with reduced catalytic efficiency.[99] Aiming to improve the unstable and promiscuous sdV-GrsA, a short directed evolution experiment has been performed (Manuscript IV, Supplementary Information). We hypothesized that structural disturbance caused by subdomain swap can be minimized by reverting subdomain residues back to their original counterparts in GrsA. This approach allows only two possibilities at each position: "swapped" and "wild type" to keep library size small. Informed by structural data of progenitor GrsA and a homology model of sdV-GrsA, five positions at the A-T interface were selected in the first round and activity was detected by screening for Val-Pro DKP formation. Targeting the region where T-domain interacts with the A-domain yielded only a small improvement (Manuscript IV, Supplementary Figure 3). In a following step, buried regions of the subdomain at the border to the surrounding GrsA scaffold were targeted. We envisioned that these regions would cause clashes and decrease protein stability. Twelve residues were reversed to their GrsA identities, with four of them showing improved peptide formation over sdV-GrsA (Manuscript IV, Supplementary Figure 4). Finally, in a third step, beneficial mutations from two rounds were recombined resulting in four mutants with improved activities (Manuscript IV, Supplementary Figure 5). Although the mutational effects were not additive, the most active mutant (STAP) shows 6-fold higher peptide formation than sdV-GrsA. Beside improved peptide formation, assaying mutants under competition with L-Val and L-Phe also revealed the changes in specificity. The MS mutant, for example, shows a 10-fold increase in specificity towards Val-Pro DKP formation (Manuscript IV, Supplementary Figure 5). Notably, the activity of mutants was positively correlated with protein yield, suggesting that mutations improve protein folding or stability. Differences observed between mutants may arise as a consequence of improved adenylation or thiolation partial reactions of the A-domain. By experimentally dissecting these two steps, we pinpointed the origins of improvements to the adenylation reaction (Manuscript IV, Figure 3).

Transplanting fragments of A-domains can be an effective way to change A-domain specificity, especially if the homology of donor and acceptor A-domain is taken into account.[120] Resulting structural impairments are likely to cause losses of activity or protein yield. However, the small size of the subdomain (~100 aa) provides an opportunity to mutate a limited number of residues without the need for a comprehensive screening procedure. We demonstrate that small, focused libraries of binary mutations can achieve significant improvements in terms of activity (STAP) and substrate specificity (MS). Introduction of binary mutations has already been used to improve the enantioselectivity of limonene epoxide hydrolase.[168] Owing to the small number of combinations, this procedure is inexpensive and may be used to quickly improve the activity of engineered A-domains, especially when the assay throughput is limited.

9.3.2 A-domain can overrule C-domain specificity

The engineered combination of promiscuous A-domain and native C-domain specificity in sdV-GrsA/GrsB1 has created a unique opportunity to study the translation of domain into product specificity. When incubated with large concentrations of competing substrates, formation of DKP products by sdV-GrsA and GrsB1 can be expected to be constant. However, the incorporation ratio varies over the course of the reaction. Initially, *D*F-DKP is produced at 3-fold higher rate than *D*V-DKP. Surprisingly, the preference later inverts and *D*V-DKP becomes the main product (Manuscript IV, Figure 4a). One possible explanation could be the slowing of the peptide formation rate due to substrate depletion. However, as substrates are added at 1 mM and DKPs are formed at nM levels, this possibility can be excluded. We explain the observed inversion by a dynamic loading state of sdV-GrsA's T-domain. Namely, in the initial phase, sdV-GrsA activates and loads Val with a slight preference, according to the specificity of the A-domain. Subsequently, being a wild type substrate, D-Phe-loaded sdV-GrsA will be preferentially deacylated by GrsB1. Released sdV-GrsA will reincorporate Val/Phe ratios dictated by the A-domain, however the fraction of Val-loaded sdV-GrsA will increase over time, forcing the GrsB1 C-domain to interact with the D-Val loaded population of sdV-GrsA. As GrsB1 condenses more D-Val, the *D*V-DKP formation accelerates until the steady state is reached defined by balanced deacylation of Val- and Phe-loaded sdV-GrsA.

In the above hypothesis, we assume the C-domain to prefer the native substrate D-Phe, but confirming this hypothesis would require measuring the selectivity of the C-domain donor site towards D-Phe and D-Val. Since condensation reaction is difficult to measure experimentally, we have taken advantage of Dynafit, a modelling software which uses enzyme kinetic data to perform nonlinear least-squares regression.[169] This software fits experimental progress data to differential equations derived from hypothetical reaction mechanisms and extracts the defined rate constants. The quality of fit indicates which reaction mechanism best reflects reality. A simple two-step model consisting of one loading and one condensation step has been sufficient to describe the time-courses with three sdVGrsA variants and various substrate concentrations. Surprisingly, although the mutants differ only at the A-domain, the ratio of condensation constants for Val over Phe is 5-8 times lower than in sdV-GrsA. This discrepancy may indicate an impact of A-domain mutations on C-domain selectivity. It has already been suggested that A-domain activity can be affected by the presence of the C-domain but not vice versa.[69] We speculate that interventions at the A-domain level can disturb the conformation changes required for the transfer of the substrate to the C-domain.

Substrates loaded on the T-domain are racemized in the E-domain before the condensation reaction. Since the C-domain of GrsB1 belongs to a $^{D}C_{L}$ family, stereoselectivity of the acceptor site ensures the incorporation of D-Phe into the peptide.

However, this specificity filter seems to be more permissive when L-Val is activated and racemized as indicated by higher rate of *L*V-DKP formation compared to the *L*F-DKP (Manuscript IV, Figure 4e,f). Analysis of kinetic data obtained with an inactivated E-domain result in higher condensation rate constants for L-Val than for L-Phe indicating that stereoselectivity can be influenced by the side chain identity of the loaded amino acid (Manuscript IV, Supplementary Table 2).

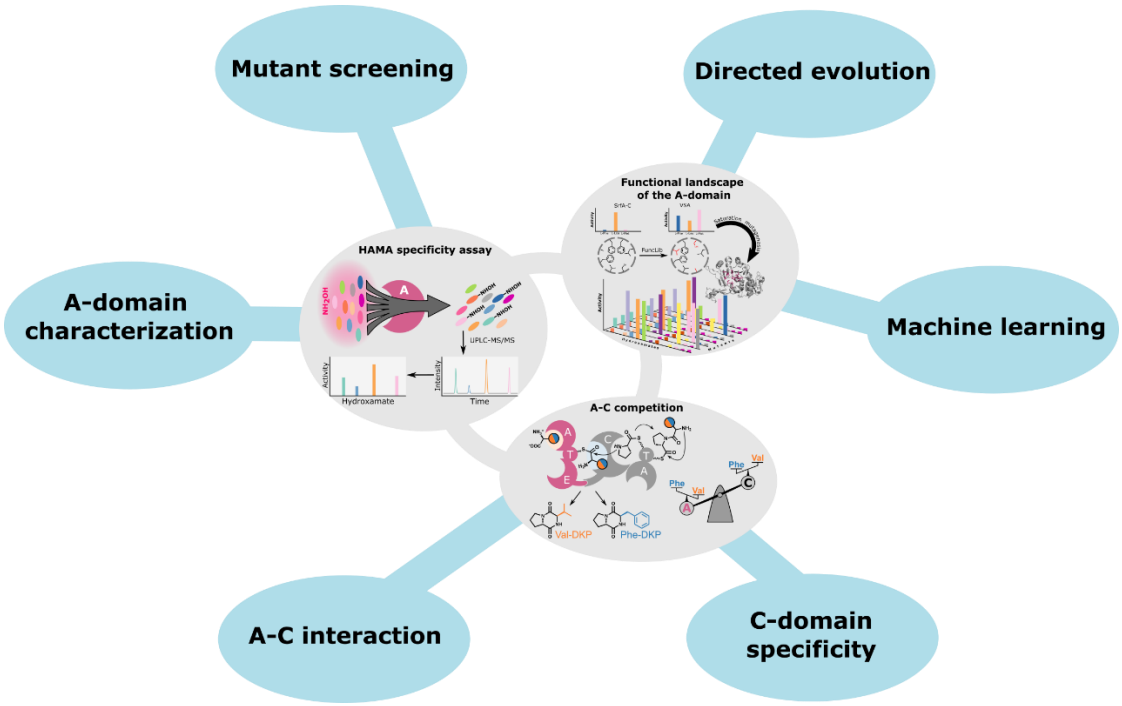
Mismatch of A- and C-domain specificity in engineered sdV-GrsA/GrsB1 system offered a unique opportunity for detailed investigation of the rules governing the NRPS mechanism. This is especially relevant in context of ongoing debate about the importance of C-domain proofreading for the NRPS engineering. It is generally considered that the acceptor site of the C-domain shows strict substrate specificity towards the aminoacyl-loaded T-domain. On the other hand, donor site of C-domains of elongation modules is considered less stringent towards peptidyl-intermediates. Although in our system it is the donor site of the GrsB1 C-domain that is exposed to a noncognate substrate, it can be considered generally applicable considering that sdV-GrsA is an initiation module offering aminoacyl-T-domain for condensation. Moreover, if the donor site does show some degree of specificity, it is likely towards the C-terminal amino acid residue in the peptidyl-intermediate. In sdV-GrsA/GrsB1, the preference of the donor site of the C-domain towards wild type substrate Phe causes the accumulation of Val-loaded sdV-GrsA, resulting in an unusual time-dependent change of Val- and Phe-DKP ratios. The resulting excess of Val-loaded sdV-GrsA eventually overrules the C-domain specificity of GrsB1, making *D*V-DKP the main product.

The complex NRPS mechanism consisting of five individual catalytic steps can be described with a surprisingly simple, two-step kinetic model. Ratios of Val/Phe acylation constants are well determined and clearly reflect the experimental data with radiolabelled substrates (Manuscript IV, Figure 4d). The dominance of the A-domain over C-domain is illustrated by a simulation of a hypothetical two-module system with different adenylation and condensation rates (Manuscript IV, Figure 5). Naturally, being a rate limiting step in NRPS machinery, changes in the condensation rates of the C-domain can influence the rate of the DKP formation. However, the ratio of the two DKP products depends exclusively on the acylation ratio dictated by the substrate selection of the A-domain. According to our model, the dominance of the A-domain over product specificity will be broken when hydrolysis of the T-domain thioester catalysed by type II thioesterases[72,76] constantly resets the loading state of the T-domain.

Another important consideration for NRPS engineering is the relative importance of A- and C-domain for catalysis. If the evolutionary pressure is strong enough, wild type A-domains will typically attain high specificity, with several orders of magnitude in catalytic efficiency separating the two most preferred substrates.[143] In contrast, the <16-fold specificity for Phe over Val determined for the C-domain donor

site of GrsB1 indicates that the C-domain selectivity filter presents significant but manageable obstacle to NRPS engineering. Moreover, even a small preference of the A-domain towards Val, as found in sdV-GrsA, will eventually cause the accumulation of Val-loaded module, thus determining the product ratio alone. It is likely that the main challenge of NRPS engineering is to maintain the high activity of the A-domain without disturbing the interdomain communication within the assembly line. In cases where C-domain proofreading is encountered, I expect that minor interventions relaxing the specificity will be sufficient to restore peptide formation to wild type rates. Our results demonstrate the value of kinetic modelling for probing complex nonribosomal synthetase mechanisms and strengthen the role of the A-domain as a decisive factor in substrate selection and incorporation into the natural product.

10 PERSPECTIVES



The known portfolio of natural products is only a fraction of the diversity available in nature. For millions of years, in a still ongoing process, biosynthetic pathways have been honed and perfected by natural evolution to fulfil various functions. In the case of nonribosomal peptide synthetases and polyketide synthetases, nature has taken advantage of the inherent potential for diversification of modular architectures. Once adjusted to a specific function, individual parts are shuffled and recombined between different systems.[121] Over the last 50 years, the same strategy has been utilized in the laboratory with various degrees of success. Although success stories for NRPS engineering are reported more and more frequently, they remain isolated examples and a robust and general approach which would be widely applicable is still lacking. The challenges we are facing are rooted in our lack of understanding of mechanisms and dynamics of these complex enzymes. Recently, large advances in structural biology of NRPSs have raised questions about the boundaries between domains and modules which are essential for identifying efficient recombination points.[89,92,167] Of particular relevance is the issue of specificity filters in the assembly line which can limit the processing of modified peptide intermediates. A-domains are well established as a main decision points about the identity of the incorporated substrate. Although A-domains are most thoroughly studied NRPS units, their reprogramming remains a challenge.

This thesis establishes HAMA as a robust and straightforward assay for determining the specificity of the A-domain. Current experimental setting allows the simultaneous detection of the activity towards 20 substrates. In principle, this is still far from the theoretical analytical limits imposed by the UPLC-MS/MS detection method. The substrate panel can be further expanded by including additional hydroxamate standards. HAMA will find the most obvious use in the characterization of heterologously expressed A-domains, engineering intermediates and deorphanization of newly discovered BGCs. Using HAMA, we are now able to probe A-domain activity at an unprecedented detail. As proof of concept, we have mapped an expansive functional landscape of the A-domain of SrfA-C and unravelled a stunning flexibility in specificity. When expanded A-domains from phylogenetically distant clusters, HAMA profiling of mutant libraries will enable us to retrace the evolutionary trajectories of substrate specificity - indispensable information for engineering and understanding NRPS. In directed evolution, enzyme engineers typically aim to minimize size while maximising functional diversity of libraries. This library design process requires knowledge of the residues with largest influence on activity and specificity. By pinpointing residues most tolerant to mutations, HAMA can enable the design of libraries that are more focused and streamlined, thus greatly reducing the experimental effort and resources required for screening. In cases where specificity shifts to distantly related substrates are required, subdomain swapping can be employed analogously to natural recombination events, followed by a rescue of activity by directed evolution.[170] Considering the recent findings that specificity code combinations found in nature do not necessarily exhaust the complete functional space of the A-domain[171], a possibility is created for the utilization of machine learning and pattern recognition methods for the prediction of productive enzyme variants. For this purpose, it will be essential to obtain detailed and high-quality training datasets generated with HAMA.

Effective A-domain engineering alone will likely not be sufficient to generate custom made peptides. Downstream specificity filters have to be taken into account for maintaining high product titres. C-domains are by far the least studied NRPS domains due to their internal location and the complexity of their substrates. In our work on sdV-GrsA:GrsB1 system we aimed to shed light on this issue through the use of nonlinear kinetic modelling. We describe the mechanism by which A-domain overrules the specificity of the C-domain donor site through the accumulation of T-domain loaded intermediates. This stalling will be particularly relevant in the context biosynthetic systems encoding type II thioesterases. The hydrolysis of the acyl-T domain can counteract A-domain engineering and shift the balance of peptide products. Considering the lack of a distinct binding pocket in the C-domain for the side chain of aminoacyl-T-domains, future studies will be required to pinpoint the mechanism by which C-domains

hinder the incorporation of noncognate substrates. Our data suggest that these constraints are of a modest magnitude (~16-fold), compared to the analogous differences at the level of the A-domain. It is tempting to speculate that only a few mutations at yet to be identified, key locations in the C-domain will be sufficient to relax specificity and restore rates to wild type levels. Compelling differences in condensation constants between sdV-GrsA mutants paired with GrsB1 suggest the possible influence of the A-domain on C-domain specificity. To test this phenomenon, *in trans* acylation of the T-domain with promiscuous PPtase can be utilized to bypass the A-domain in order to isolate the condensation step. In this context, A-domains should be modified by mutagenesis to probe a possible influence on the condensation reaction. The development of fast and reliable assays for condensation activity will prove essential to tackle the issue of C-domain specificity and activity. Ideally, a screening method with a condensation-related output signal should be coupled to a high-throughput sorting method such as FACS. This would enable the experimental investigation of binding sites, dynamics, and mutational sensitivity of the C-domain, one of the main enigmas of NRPS enzymology.

The success of NRPS engineering strongly depends on the availability of adequate screening assays. Work on ribosomal code expansion has resulted in the incorporation of numerous nonnatural substrates by targeting aaRS, the ribosomal counterpart of A-domains. I anticipate A-domain screening guided by HAMA to replicate these successes with nonribosomal A-domains to routinely make variants of important bioactive peptides. A-domain engineering combined with the investigation of the A/C-domain interplay, as outlined in this thesis, will open sustainable routes towards tailored drugs in the future.

11 LITERATURE

1. Gross, M. The paradoxical evolution of agriculture. *Curr. Biol.* **23**, R667–R670 (2013).
2. Latinne, A. *et al.* Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **11**, 4235 (2020).
3. Zaffiri, L., Gardner, J. & Toledo-Pereyra, L. H. History of Antibiotics. From Salvarsan to Cephalosporins. *J. Investig. Surg.* **25**, 67–77 (2012).
4. Provenzani, A. *et al.* Multidrug-resistant gram-negative organisms: a review of recently approved antibiotics and novel pipeline agents. *Int. J. Clin. Pharm.* **42**, 1016–1025 (2020).
5. 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).
6. Karaiskos, I., Lagou, S., Pontikis, K., Rapti, V. & Poulakou, G. The 'Old' and the 'New' antibiotics for MDR Gram-negative pathogens: For whom, when, and how. *Front. Public Heal.* **7**, 1–25 (2019).
7. Clardy, J., Fischbach, M. A. & Walsh, C. T. New antibiotics from bacterial natural products. *Nat. Biotechnol.* **24**, 1541–1550 (2006).
8. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
9. O'Shea, R. & Moser, H. E. Physicochemical properties of antibacterial compounds: Implications for drug discovery. *J. Med. Chem.* **51**, 2871–2878 (2008).
10. Silver, L. L. Are natural products still the best source for antibacterial discovery? The bacterial entry factor. *Expert Opin. Drug Discov.* **3**, 487–500 (2008).
11. Kim, E., Moore, B. S. & Joon Yoon, Y. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **11**, 649–659 (2015).
12. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35**, 833–844 (2017).
13. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509–523 (2015).
14. Brickner, S. J., Barbachyn, M. R., Hutchinson, D. K. & Manninen, P. R. Linezolid (ZYVOX), the First Member of a Completely New Class of Antibacterial Agents

for Treatment of Serious Gram-Positive Infections. *J. Med. Chem.* **51**, 1981–1990 (2008).

15. Livermore, D. M. Quinupristin/dalfopristin and linezolid: where, when, which and whether to use? *J. Antimicrob. Chemother.* **46**, 347–350 (2000).
16. Dang, T. & Süssmuth, R. D. Bioactive Peptide Natural Products as Lead Structures for Medicinal Use. *Acc. Chem. Res.* **50**, 1566–1576 (2017).
17. Giessen, T. W. & Marahiel, M. A. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett.* **586**, 2065–2075 (2012).
18. 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.
19. Molinero, A. A., Kingston, D. G. I. & Reed, J. W. Biosynthesis of Antibiotics of the Virginiamycin Family, 6. Biosynthesis of Virginiamycin S 1. *J. Nat. Prod.* **52**, 99–108 (1989).
20. Landman, D., Georgescu, C., Martin, D. A. & Quale, J. Polymyxins Revisited. *Clin. Microbiol. Rev.* **21**, 449–465 (2008).
21. Hedges, J. B. & Ryan, K. S. Biosynthetic Pathways to Nonproteinogenic α-Amino Acids. *Chem. Rev.* **120**, 3161–3209 (2020).
22. Wohlleben, W., Stegmann, E. & Süssmuth, R. D. B. T.-M. in E. Chapter 18 Molecular Genetic Approaches to Analyze Glycopeptide Biosynthesis. in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part A: Overview Articles and Peptides* vol. 458 459–486 (Academic Press, 2009).
23. Bode, H. B. *et al.* Structure Elucidation and Activity of Kolossin A, the Pentadecapeptide Product of a Giant Nonribosomal Peptide Synthetase. *Angew. Chemie Int. Ed.* **54**, 10352–10355 (2015).
24. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259–9264 (2014).
25. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**, 490–504 (2002).
26. Quadri, L. E. N. *et al.* Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carder protein domains in peptide synthetases. *Biochemistry* **37**, 1585–1595 (1998).
27. Hahn, M. & Stachelhaus, T. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15585–90 (2004).

28. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
29. Reimer, J. M., Aloise, M. N., Harrison, P. M. & Martin Schmeing, T. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239–242 (2016).
30. Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.* **291**, 22559–22571 (2016).
31. Izoré, T. & Cryle, M. J. The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep.* **35**, 1120–1139 (2018).
32. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science (80-.).* **321**, 659–663 (2008).
33. Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **25**, 783-793.e4 (2017).
34. Gulick, A. M. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827 (2009).
35. Tan, X. F. *et al.* Structure of the adenylation-peptidyl carrier protein didomain of the Microcystis aeruginosa microcystin synthetase McyG. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 873–881 (2015).
36. Caetano-Anollés, G., Kim, H. S. & Mittenthal, J. E. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci.* **104**, 9358 LP – 9363 (2007).
37. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
38. 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).
39. 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).
40. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide

synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).

41. Röttig, M. *et al.* NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, 362–367 (2011).
42. Li, M. H. T., Ung, P. M. U., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D. H. Automated genome mining for natural products. *BMC Bioinformatics* **10**, 185 (2009).
43. Skinnider, M. A. *et al.* Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* **43**, 9645–9662 (2015).
44. Blin, K. *et al.* AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).
45. Chevrette, M. G., Aicheler, F., Kohlbacher, O., Currie, C. R. & Medema, M. H. SANDPUMA: Ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics* **33**, 3202–3210 (2017).
46. Schomer, R. A. & Thomas, M. G. Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase. *Biochemistry* **56**, 5380–5390 (2017).
47. Baltz, R. H. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol.* **38**, 1747–1760 (2011).
48. 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).
49. Owen, J. G. *et al.* Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem. Biol.* **23**, 1–12 (2016).
50. Keating, T. a, Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–6 (2002).
51. Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **97**, 2651–2674 (1997).
52. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.* **425**, 3137–3150 (2013).
53. De Crécy-Lagard, V., Marlière, P. & Saurin, W. Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C. R. Acad. Sci. III.* **318**, 927–936 (1995).

54. Bergendahl, V., Linne, U. & Marahiel, M. a. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **269**, 620–629 (2002).

55. Marshall, C. G., Hillson, N. J. & Walsh, C. T. Catalytic Mapping of the Vibriobactin Biosynthetic Enzyme VibF. *Biochemistry* **41**, 244–250 (2002).

56. Roche, E. D. & Walsh, C. T. Dissection of the EntF Condensation Domain Boundary and Active Site Residues in Nonribosomal Peptide Synthesis. *Biochemistry* **42**, 1334–1344 (2003).

57. 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**, 781–92 (2007).

58. 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).

59. Belshaw, P. J. Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science (80-.).* **284**, 486–489 (1999).

60. Clugston, S. L., Sieber, S. a, Marahiel, M. A. & Walsh, C. T. Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. *Biochemistry* **42**, 12095–104 (2003).

61. 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).

62. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007).

63. 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) Preprint doi:10.21203/rs.3.rs-125509/v1.

64. Peypoux, F. & Michel, G. Controlled biosynthesis of Val7- and Leu7-surfactins. *Appl. Microbiol. Biotechnol.* **36**, 515–517 (1992).

65. Ruttenberg, M. A. & Mach, B. Studies on Amino Acid Substitution in the Biosynthesis of the Antibiotic Polypeptide Tyrocidine. *Biochemistry* **5**, 2864–2869 (1966).

66. 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).

67. 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).

68. Degen, A., Mayerthaler, F., Mootz, H. D. & Di Ventura, B. Context-dependent activity of A domains in the tyrocidine synthetase. *Sci. Rep.* **9**, 5119 (2019).

69. 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).

70. Li, R., Oliver, R. A. & Townsend, C. A. Identification and Characterization of the Sulfazecin Monobactam Biosynthetic Gene Cluster. *Cell Chem. Biol.* **24**, 24–34 (2017).

71. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. a. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **273**, 22773–22781 (1998).

72. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* **99**, 14083–14088 (2002).

73. Shi, C., Miller, B. R., Alexander, E. M., Gulick, A. M. & Aldrich, C. C. Design, Synthesis, and Biophysical Evaluation of Mechanism-Based Probes for Condensation Domains of Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **15**, 1813–1819 (2020).

74. Horsman, M. E., Hari, T. P. a & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).

75. Kohli, R. M., Walsh, C. T. & Burkart, M. D. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**, 658–661 (2002).

76. Yeh, E., Kohli, R. M., Bruner, S. D. & Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBioChem* **5**, 1290–1293 (2004).

77. Caboche, S., Leclère, V., Pupin, M., Kucherov, G. & Jacques, P. Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. *J. Bacteriol.* **192**, 5143–5150 (2010).

78. Walsh, C. T., O'Brien, R. V & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed. Engl.* **52**, 7098–124 (2013).

79. Labby, K. J., Watsula, S. G. & Garneau-Tsodikova, S. Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **32**, 641–653 (2015).

80. Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7**, 623–642 (2000).

81. Roongsawang, N., Washio, K. & Morikawa, M. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. *Int. J. Mol. Sci.* **12**, 141–172 (2010).

82. Zaleta-Rivera, K. *et al.* A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins. *Biochemistry* **45**, 2561–2569 (2006).

83. Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **24**, 735–749 (2007).

84. von Döhren, H. A survey of nonribosomal peptide synthetase (NRPS) genes in Aspergillus nidulans. *Fungal Genet. Biol.* **46 Suppl 1**, S45-52 (2009).

85. Kaniusaite, M., Goode, R. J. A., Tailhades, J., Schittenhelm, R. B. & Cryle, M. J. Exploring modular reengineering strategies to redesign the teicoplanin non-ribosomal peptide synthetase. *Chem. Sci.* **11**, 9443–9458 (2020).

86. Gehring, A. M., Mori, I., Perry, R. D. & Walsh, C. T. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis. *Biochemistry* **37**, 11637–11650 (1998).

87. Chen, W. H., Li, K., Guntaka, N. S. & Bruner, S. D. Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **11**, 2293–2303 (2016).

88. 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).

89. Farag, S. *et al.* Inter-Modular Linkers play a crucial role in governing the biosynthesis of non-ribosomal peptides. (2019).

90. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).

91. 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).

92. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. **11**, (2020).

93. Stanišić, A. & Kries, H. Adenylation Domains in Nonribosomal Peptide Engineering. *ChemBioChem* **20**, 1347–1356 (2019).

94. Sun, X., Li, H., Alfermann, J., Mootz, H. D. & Yang, H. Kinetics profiling of gramicidin S synthetase A, a member of nonribosomal peptide synthetases. *Biochemistry* **53**, 7983–7689 (2014).

95. Stulberg, M. P. & Novelli, G. D. [95] Amino acid-activating enzymes: Methods of assay. *Methods Enzymol.* **5**, 703–707 (1962).

96. Hansen, B. S. An improved method for assaying pyrophosphate exchange measuring Cerenkov radiation. *Anal. Biochem.* **109**, 12–17 (1980).

97. Linne, U. & Marahiel, M. a. Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases. *Methods Enzymol.* **388**, 293–315 (2004).

98. Otten, L. G., Schaffer, M. L., Villiers, B. R. M., Stachelhaus, T. & Hollfelder, F. 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 (2007).

99. Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.* **22**, 640–648 (2015).

100. Geladopoulos, T. P., Sotiroudis, T. G. & Evangelopoulos, A. E. A malachite green colorimetric assay for protein phosphatase activity. *Anal. Biochem.* **192**, 112–116 (1991).

101. McQuade, T. J. *et al.* A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal. Biochem.* **386**, 244–250 (2009).

102. Katano, H., Tanaka, R., Maruyama, C. & Hamano, Y. Assay of enzymes forming AMP + PPi by the pyrophosphate determination based on the formation of 18-molybdopyrophosphate. *Anal. Biochem.* **421**, 308–312 (2012).

103. Katano, H., Watanabe, H., Takakuwa, M., Maruyama, C. & Hamano, Y. Colorimetric determination of pyrophosphate anion and its application to adenylation enzyme assay. *Anal. Sci.* **29**, 1095–1098 (2013).

104. Kittilä, T., Schoppet, M. & Cryle, M. J. Online pyrophosphate assay for analyzing adenylation domains of nonribosomal peptide synthetases. *ChemBioChem* **17**, 576–584 (2016).

105. Webb, M. R. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4884–4887 (1992).

106. Upson, R. H., Haugland, R. P., Malekzadeh, M. N. & Haugland, R. P. A spectrophotometric method to measure enzymatic activity in reactions that generate inorganic pyrophosphate. *Anal. Biochem.* **243**, 41–45 (1996).

107. Nixon, A. E., Hunter, J. L., Bonifacio, G., Eccleston, J. F. & Webb, M. R. Purine nucleoside phosphorylase: its use in a spectroscopic assay for inorganic phosphate

and for removing inorganic phosphate with the aid of phosphodeoxyribomutase. *Anal. Biochem.* **265**, 299–307 (1998).

108. Cheng, J. *et al.* Purine nucleoside phosphorylase-catalyzed, phosphate-independent hydrolysis of 2-amino-6-mercapto-7-methylpurine ribonucleoside. *Bioorg. Chem.* **27**, 307–325 (1999).
109. Wilson, D. J. & Aldrich, C. C. A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal. Biochem.* **404**, 56–63 (2010).
110. Finking, R. *et al.* Aminoacyl adenylate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases. *ChemBioChem* **4**, 903–906 (2003).
111. Bernier, S., Akochy, P.-M. M., Lapointe, J. & Chênevert, R. Synthesis and aminoacyl-tRNA synthetase inhibitory activity of aspartyl adenylate analogs. *Bioorganic Med. Chem.* **13**, 69–75 (2005).
112. Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51**, 3252–3263 (2012).
113. Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol.* **19**, 188–198 (2012).
114. 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).
115. 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).
116. Schneider, A., Stachelhaus, T. & Marahiel, M. A. Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Mol. Gen. Genet.* **257**, 308–318 (1998).
117. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726 (2002).
118. 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).
119. Kaljunen, H. *et al.* Structural elucidation of the bispecificity of A domains as a basis for activating non-natural amino acids. *Angew. Chemie Int. Ed.* **54**, 8833–8836 (2015).

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

121. Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **10**, (2014).

122. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).

123. Arnold, F. H. Directed evolution: Bringing new chemistry to life. *Angew. Chemie - Int. Ed.* **57**, 4143–4148 (2018).

124. Zeymer, C. & Hilvert, D. Directed evolution of protein catalysts. *Annu. Rev. Biochem.* **87**, 131–157 (2018).

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

126. 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).

127. Kries, H. *et al.* Reprogramming nonribosomal peptide synthetases for 'clickable' amino acids. *Angew Chem Int Ed Engl* **53**, 10105–10108 (2014).

128. Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* **10**, 282–287 (2018).

129. Steiniger, C., Hoffmann, S. & Süssmuth, R. D. Probing Exchange Units for Combining Iterative and Linear Fungal Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* **26**, 1526-1534.e2 (2019).

130. Doekel, S. *et al.* Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in Streptomyces roseosporus. *Microbiology* **154**, 2872–2880 (2008).

131. Nguyen, K. T. *et al.* Genetically engineered lipopeptide antibiotics related to A54145 and daptomycin with improved properties. *Antimicrob. Agents Chemother.* **54**, 1404–1413 (2010).

132. Miao, V. *et al.* Genetic engineering in Streptomyces roseosporus to produce hybrid lipopeptide antibiotics. *Chem. Biol.* **13**, 269–276 (2006).

133. Baltz, R. H. Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.* **3**, 748–758 (2014).

134. 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).

135. Calcott, M. J. & Ackerley, D. F. Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol.* **15**, 1–13 (2015).

136. Kaniusaite, M. *et al.* A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. *Chem. Sci.* **10**, 9466–9482 (2019).

137. Fewer, D. P. *et al.* Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. *BMC Evol. Biol.* **7**, 183 (2007).

138. Ishida, K. *et al.* Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria. *Appl. Environ. Microbiol.* **75**, 2017–2026 (2009).

139. Höfer, I. *et al.* Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. *Chem. Biol.* **18**, 381–91 (2011).

140. Götze, S. *et al.* Structure elucidation of the syringafactin lipopeptides provides insight in the evolution of nonribosomal peptide synthetases. *Chem. Sci.* **10**, 10979–10990 (2019).

141. Ishikawa, F. *et al.* An Engineered Aryl Acid Adenylation Domain with an Enlarged Substrate Binding Pocket. *Angew. Chemie Int. Ed.* **58**, 6906–6910 (2019).

142. Hara, R., Suzuki, R. & Kino, K. Hydroxamate-based colorimetric assay to assess amide bond formation by adenylation domain of nonribosomal peptide synthetases. *Anal. Biochem.* **477**, 89–91 (2015).

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

144. Hori, K. & Kurotsu, T. Characterization of Gramicidin S synthetase aggregation substance: Control of Gramicidin S synthesis by its product, Gramicidin S. *J. Biochem.* **122**, 606–615 (1997).

145. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5**, 593–9 (2009).

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

147. Zhang, K. et al. Engineering the substrate specificity of the DhbE adenylation domain by yeast cell surface display. *Chem. Biol.* **20**, 92–101 (2013).

148. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5869–5874 (2006).

149. Xu, Z., Cen, Y.-K., Zou, S.-P., Xue, Y.-P. & Zheng, Y.-G. Recent advances in the improvement of enzyme thermostability by structure modification. *Crit. Rev. Biotechnol.* **40,** 83–98 (2020).

150. Eijsink, V. G. H., Gåseidnes, S., Borchert, T. V & van den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **22,** 21–30 (2005).

151. Rigoldi, F. et al. Thermal stabilization of the deglycating enzyme Amadoriase I by rational design. *Sci. Rep.* **8,** 3042 (2018).

152. Bednar, D. et al. FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLOS Comput. Biol.* **11,** e1004556 (2015).

153. Jones, B. J., Lim, H. Y., Huang, J. & Kazlauskas, R. J. Comparison of Five Protein Engineering Strategies for Stabilizing an α/β-Hydrolase. *Biochemistry* **56,** 6521–6532 (2017).

154. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4,** 35–37 (2008).

155. Beadle, B. M. & Shoichet, B. K. Structural Bases of Stability–function Tradeoffs in Enzymes. *J. Mol. Biol.* **321,** 285–296 (2002).

156. Ben-David, M. et al. Enzyme evolution: An epistatic ratchet versus a smooth reversible transition. *Mol. Biol. Evol.* **37,** 1133–1147 (2020).

157. Yang, G., Miton, C. M. & Tokuriki, N. A mechanistic view of enzyme evolution. *Protein Sci.* **29,** 1724–1747 (2020).

158. Ballal, A. et al. Sparse Epistatic Patterns in the Evolution of Terpene Synthases. *Mol. Biol. Evol.* **37,** 1907–1924 (2020).

159. Yang, G. et al. Higher-order epistasis shapes the fitness landscape of a xenobiotic-degrading enzyme. *Nat. Chem. Biol.* **15,** 1120–1128 (2019).

160. Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R. & Fields, S. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* **19,** 1537–1551 (2013).

161. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24,** 2643–2651 (2014).

162. Lozovsky, E. R. et al. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 12025–12030 (2009).

163. Moriuchi, R. et al. Stepwise enhancement of catalytic performance of haloalkane dehalogenase LinB towards β-hexachlorocyclohexane. *AMB Express* **4,** 1–10 (2014).

164. Miton, C. M. & Tokuriki, N. How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.* **25,** 1260–1272 (2016).

165. Sano, T., Usui, T., Ueda, K., Osada, H. & Kaya, K. Isolation of new protein phosphatase inhibitors from two cyanobacteria species, Planktothrix spp. *J. Nat. Prod.* **64,** 1052–1055 (2001).

166. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).

167. Bozhüyük, K. A., Micklefield, J. & Wilkinson, B. Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51,** 88–96 (2019).

168. Sun, Z. et al. Reshaping an Enzyme Binding Pocket for Enhanced and Inverted Stereoselectivity: Use of Smallest Amino Acid Alphabets in Directed Evolution. *Angew. Chemie - Int. Ed.* **54,** 12410–12415 (2015).

169. Kuzmič, P. DynaFit—A Software Package for Enzymology. in *Methods in Enzymology* vol. **467** 247–280 (2009).

170. 1. Baunach, M., Chowdhury, S., Stallforth, P. & Dittmann, E. The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity. *Mol. Biol. Evol.* (2021) doi:10.1093/molbev/msab015.

171. Throckmorton, K. et al. Directed Evolution Reveals the Functional Sequence Space of an Adenylation Domain Specificity Code. *ACS Chem. Biol.* **14,** 2044–2054 (2019).

12 APPENDIX

12.1 List of Abbreviations and Acronyms

AARS	Aminoacyl t-RNA synthetase
A-domain	Adenylation domain
BGC	Biosynthetic gene cluster
C-domain	Condensation domain
E-domain	Epimerization domain
FACS	Fluorescence activated cell sorting
MDR	Multidrug-resistant bacteria
MLP	MbtH-like proteins
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
PKS	Polyketide synthetase
PP_i	Pyrophosphate
PPtase	4'-phosphopantetheine transferase
P_i	Phosphate
Ppant	4'-phosphopantetheine
RiPP	Ribosomally synthesized, posttranslationally modified peptide
SAM	S-adenosyl methionine
SNAC	N-acetylcysteamine thioester
T-domain	Thiolation domain
TE-domain	Thioesterase domain
YSD	Yeast surface display

12.2 Contribution to original publications

Manuscript II

Stanišić et al (2019), *Chem. Sci.*

Contribution of the doctoral candidate to Figures that reproduce experimental data:

Figures 2, 3, 5, SI Figures 1-6	☒	100% (the data shown in this Figure come entirely from experimental work carried out by the candidate)
Figure 4	☒	0% (the data shown in this Figure are based exclusively on the work of other coauthors)

Manuscript III

In preparation for submission to *ACS Catalysis.*

Contribution of the doctoral candidate to Figures that reproduce experimental data:

Figure 2	☒	75% (FuncLib randomization and screening (A) and HAMA (B))
Figure 3	☒	50 % (binding pocket randomization (A), establishment of the screening protocol, data analysis (B-D))
Figure 4	☒	50 % (data analysis)
Figure 5	☒	30 % (data analysis (A))
Supp. Figures 1, 2	☒	100 %
Supp. Figures 3-5	☒	0 %

Manuscript IV

Submitted to *ACS Catalysis*

Contribution of the doctoral candidate to Figures that reproduce experimental data:

Figures 2, 3, 4	☒	80 % (optimization of experimental conditions)
Figure 5	☒	100 %
Supp. Figures 1, 2, 10	☒	100 %
Supp. Figures 3-9	☒	0 %

Manuscript V

Wurlitzer et al (2020) *Appl. Environ. Microbiol.*

Contribution of the doctoral candidate to Figures that reproduce experimental data:

Figure 5	☒	100 %
Figures 2, 3, 4, 6	☒	0 %

Manuscript VI

In preparation for submission to *Chem. Sci.*

Contribution of the doctoral candidate to Figures that reproduce experimental data:

Figure 2, S4, S9	☒	100 %
Figures 4, 5, 6, S1-S3, S5-S8, S10-S49	☒	0 %

Manuscript VII

Trottmann et al (2020) *Angew. Chem. Int. Ed.*

Contribution of the doctoral candidate to Figures that reproduce experimental data:

Figure 4	☒	40 % (B and C)
Figures 2, 3	☒	0 %
Figures S6, S7	☒	100 %
Figures S1-S5	☒	0 %

12.3 Curriculum vitae

Name	Aleksa Stanišić
Contact	Binswangerstr. 6, 07747, Jena +49 (0)163 768 5062 aleksa.stanisic@gmail.com
Date of birth	29.04.1991

Education

Since 11/2016	PhD candidate at Leibniz-Institute for Natural Product Research and Infection Biology e.V. – Hans-Knoell Institute, Jena (group of Dr. Hajo Kries) Title: "Mechanistic analysis of nonribosomal peptide synthetases"
2009 - 2014	Master of Science (Integrated Studies) at Faculty of Pharmacy, University of Belgrade, Serbia (grade 9.2/10)
2005 - 2009	High school "Niko Rolović" in Bar, Montenegro

Experience

02/2015 – 11/2016	Research Associate at Max Planck Institute for Chemical Ecology, Jena (group of Prof. Jonathan Gershenzon)

Publications

2020

Trottmann, F., Ishida, K., Franke, J., Stanišić, A., Ishida-Ito, M., Kries, H., Pohnert, G., & Hertweck, C. (2020). Sulfonium Acids Loaded onto an Unusual Thiotemplate Assembly Line Construct the Cyclopropanol Warhead of a Burkholderia Virulence Factor. *Angewandte Chemie (International ed. in English)*, *59*(32), 13511–13515.

Malka, O., Easson, M., Paetz, C., Götz, M., Reichelt, M., Stein, B., Luck, K., Stanišić, A., Juravel, K., Santos-Garcia, D., Mondaca, L., Springate, S., Colvin, J., Winter, S., Gershenzon, J., Morin, S., Vassão, D. (2020) Glucosylation prevents plant defense activation in phloem-feeding insects. *Nat. Chem. Biol.* 16, 1420–1426.

Wurlitzer, J. M., Stanišić, A., Wasmuth, I., Jungmann, S., Fischer, D., Kries, H., Gressler, M. (2020) Bacterial-Like Nonribosomal Peptide Synthetases Produce Cyclopeptides in the Zygomycetous Fungus Mortierella alpina. *Appl. Environ. Microbiol.* 87, AEM.02051-20.

2019 Stanišić, A. & Kries, H. Adenylation Domains in Nonribosomal Peptide Engineering. *ChemBioChem* 20, 1347–1356 (2019).

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).

Posters and talks

2019 The best poster award at the 2nd Synthetic Biology for Natural Products Conference (Puerto Vallarta, Mexico)

2018 Oral presentation at 2nd Industry Contact Forum (Jena, DE)

Poster presentation at 69th Mosbacher Kolloquium „Synthetic Biology - from Understanding to Application“ (Mosbach, DE)

Supervision and public-related activities

2020 Supervision of Master thesis “Exploring the Functional Sequence Space of Amino Acid Adenylation in SrfA-C by Hydroxamate Profiling” by Ulrich Ettelt

2019 Supervision of Master thesis “Repairing the subdomain swapped NRPS sdV-GrsA” by Annika Hüsken

Participation in the “Forsche-Schüler Tag”, HKI Jena

Participation in the “Long Night of Sciences”, HKI Jena

2018 Supervision of Master thesis “Directed evolution of a subdomain swapped nonribosomal peptide synthetase” by Shobana Murugan

Participation in the “Long Night of Sciences”, HKI Jena

13 Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbst verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Mir ist die geltende Promotionsordnung der Fakultät für Biowissenschaften der Friedrich-Schiller-Universität Jena bekannt. Personen, die mich bei den Experimenten, der Datenanalyse und der Verfassung der Manuskripte unterstützt haben, sind als Ko-Autoren auf den entsprechenden Manuskripten verzeichnet. Personen die mich bei der Verfassung der Dissertation unterstützt haben, sind in der Danksagung der Dissertation vermerkt. Die Hilfe eines Promotionsberaters wurde nicht in Anspruch genommen. Es haben Dritte weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die vorliegende Arbeit wurde in gleicher oder ähnlicher Form noch bei keiner anderen Hochschule als Dissertation eingereicht und auch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung verwendet.

Jena, den _____

_____

Aleksa Stanišić