Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Evolutionary insights and guidelines to achieve effective and high-yield non-ribosomal peptide and polyketide engineering

Mark J Calcott[1,2,3], Anna C Sang[1,2], David F Ackerley[1,2,3] and Rory F Little[1,2,3]

Engineering of non-ribosomal peptide synthetase (NRPS) and/or polyketide synthase (PKS) assembly lines to generate modified products has long offered promise to produce novel antibiotics and other bioactive molecules. However, it is only in recent years that this promise has been realised with any consistency. Key to this has been a shift away from engineering approaches informed solely by structural data, and towards strategies that incorporate insights from evolutionary principles and datasets. Such analyses have not only guided the selection of optimal recombination boundaries for substitution of key subdomains, domains or modules, but also methods for increasing engineering throughput, often trading accuracy for volume. Diverse approaches have proven successful in NRPS systems, but a consistent theme has been that recombinant assembly lines are generally impaired in terms of product yield, and a meta-analysis of published results to date indicates that no one engineering strategy is significantly best for minimising yield losses. Evolution-inspired strategies have advanced the engineering of, and product yields for, PKS systems, and further breakthroughs appear imminent. Although no 'one size fits all' solution is apparent for either NRPS or PKS engineering, this review highlights important advances in synthetic biology that will support both discovery and production of next-generation antibiotics.

Addresses
[1] School of Biological Sciences, Victoria University of Wellington, Kelburn Parade, Wellington 6012, New Zealand
[2] Te Matapihipihi - The Centre for Biodiscovery, Victoria University of Wellington, New Zealand
[3] The Maurice Wilkins Centre for Molecular Biodiscovery, Victoria University of Wellington, New Zealand

Corresponding authors: Ackerley, David F (david.ackerley@vuw.ac.nz), Little, Rory F (rory.little@vuw.ac.nz)

Current Opinion in Microbiology 2026, **91**:102755

This review comes from a themed issue on **Antimicrobials**

Edited by **Kenan Bozhuyuk**

Available online xxxx

https://doi.org/10.1016/j.mib.2026.102755

Introduction

Fleming's chance discovery of the non-ribosomal peptide penicillin in 1928 [1] was a seminal event in the modern pharmaceutical era, which not only underpinned a transformation in healthcare and slashed infection-related deaths but also ushered in the golden age of antibiotics [2]. Screening for microbial natural products became a key focus [3,4], and non-ribosomal peptides have continued to be a fruitful source of clinical drug candidates. In consequence, over 50% of drugs in clinical use today were inspired by naturally occurring non-ribosomal peptides and/or the biosynthetically related polyketides [5]. The ecological roles of these natural products are often opaque [6], but a consistent theme is that they have usually evolved in environments different to the human body and require modification to function effectively as drugs [7]. Modification of a core antibiotic scaffold may also be required to evade resistance that has arisen in response to high-level clinical use [8]. In either case, achieving desired changes can be problematic, owing to structural complexities that make total synthesis challenging or impractical [7]. A tantalising alternative, which directly produces the desired analogues, is to re-engineer target biosynthetic assembly lines by substituting key domains or modules. Such efforts have historically focused on structure-guided modification of the biosynthetic enzymes, but the past few years have witnessed a growing trend to gain insights from natural evolutionary processes and apply them to more

effectively create new biosynthetic pathways. Evolution-informed strategies avoid some of the issues that arise from making changes to highly dynamic assembly lines based on static structural snapshots, and can also account for long-range epistasis and domain–domain co-evolution effects that are difficult to infer from isolated structures. This review considers engineering approaches that harness natural evolutionary principles of diversification and selection to generate new non-ribosomal peptides or polyketides and mitigate losses in yield.

Biosynthetic logic and natural diversification of non-ribosomal peptides and polyketides

The core enzymes governing biosynthesis of each class of molecule are non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) (Type I [9] PKSs will be the focus in this review, as most PKS-engineering efforts have centred on these). NRPS modules comprise at least a thiolation (T) and an adenylation (A) domain, the latter of which recognises and adenylates a particular amino acid or related monomer, then adheres it to the free sulfhydryl moiety of a 4'-phosphopantetheine cofactor bound to the T domain. The T domain then shuttles the monomer or peptide intermediates between other catalytic domains, including condensation (C) domains, which govern amide bond formation between the monomers activated by adjacent modules. PKS modules similarly employ a 4'-phosphopantetheine cofactor (here, tethered to an acyl carrier protein (ACP)) to shuttle activated monomers or intermediates between catalytic domains. Activated monomers (typically acetate or propionate) are delivered to the ACP-linked phosphopantetheine cofactor by a cognate acyltransferase (AT) domain, and a ketosynthase (KS) domain then catalyses the condensation of extension units tethered to adjacent modules. Each module can additionally contain up to three reduction (ketoreductase (KR), dehydratase (DH), and enoylreductase (ER)) domains that reduce the β-keto position of the nascent polyketide [9,10]. The complete PKS module is defined as either *cis* or *trans*, according to whether the AT domain is integrated or a standalone protein [11]. Finally, in both NRPS and PKS systems, product release is catalysed by thioesterase (Te), reductase (R) or other specialised release domains associated with the terminal module [12].

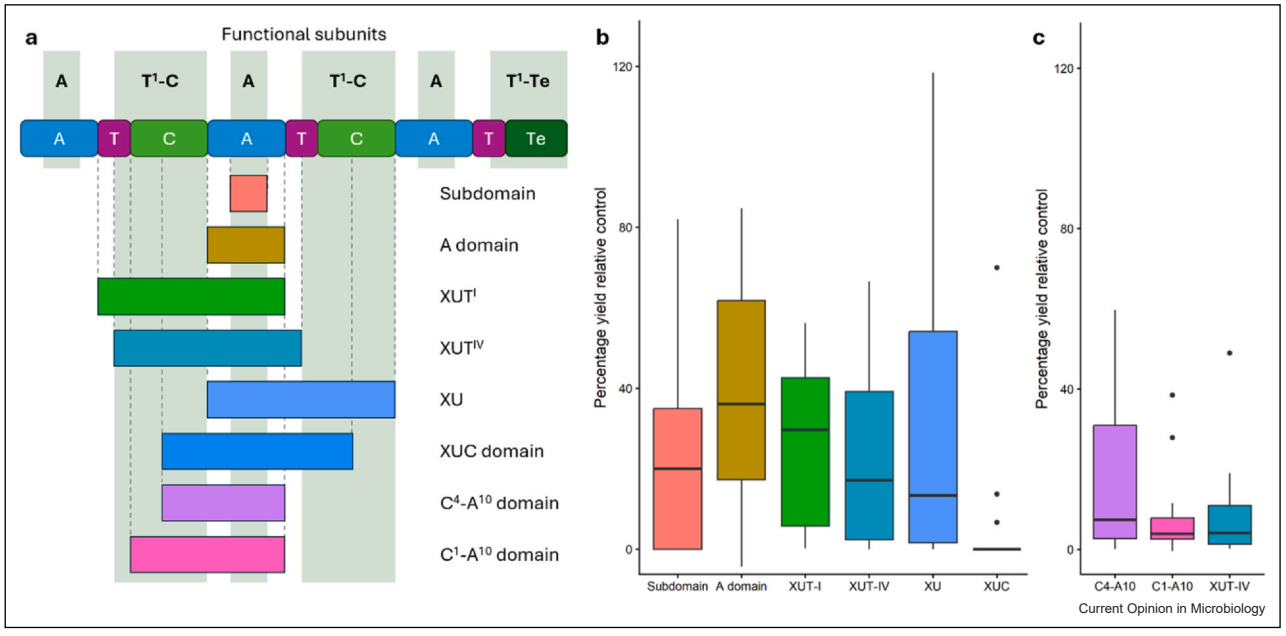
The genes encoding the core NRPS and PKS enzymes are often associated with tailoring and other accessory genes to form biosynthetic gene clusters (BGCs) [13]. Clustering provides an evolutionary advantage as it allows propagation of genes as functional blocks by both vertical and horizontal transmission [13,14]. The vast diversity of non-ribosomal peptides and polyketides in nature suggests rapid evolution by lateral transfer, and a high rate of changes via mutation and recombination events, including domain substitution and/or module duplication in the core biosynthetic genes [15], coupled to loss or gain of accessory enzymes [13]. Of note, diversification need not add complexity but can also simplify molecules. For example, ancestral state reconstruction of glycopeptide antibiotic BGCs has indicated that ancestral gene clusters likely produced compounds more complex than typical extant glycopeptides [16]. This simplification may serve to hone the cluster to produce a more optimal compound or lessen the energy burden of biosynthesis.

Evolution-informed rational domain substitution strategies

While previous work has described evolutionarily-constrained NRPS regions, a recent study provided further refinement and coined the term 'functional subunit' (Figure 1a) [17]. This term helps to define regions that demonstrate a strong association during natural recombination events, and represents both the frequency with which bordering sequences recombine and the likelihood that the substituted unit will be sufficiently functional in its new context for natural selection to retain it [18]. The first functional subunit comprises the substrate binding pocket of the A domain, and it has been suggested that partial A domain substitution, and to a lesser extent, complete A domain substitution, is a predominant mechanism of NRPS diversification in nature [18,19]. Supporting such conjecture, the region containing the A domain binding pocket has consistently been shown to be phylogenetically distinct from surrounding regions [17,18,20,21] and to cluster by A domain substrate specificity [17,18,21]. The second functional subunit includes the T–C or T–E–C (E = epimerisation) domains and extends from the core motif of the T domain [17,20] to the end of the C-domain [17,21]. This second functional subunit exhibits sequence conservation that correlates strongly with the type of domain immediately following the T domain [17,20]. However, it should be noted that there are exceptions that deviate from these functional subunits, for example, the evolution of BGCs for glycopeptide antibiotic biosynthesis has an example of both A domain binding pocket mutation and C–A domain substitution [16]. Isolated examples like this show that diverse mechanisms exist, and further research may uncover additional evolutionary trends.

Several recent studies have performed NRPS domain substitutions seeking to maintain discrete functional subunits to improve functionality and yield (summarised in Figure 1a). In particular, four recent evolution-inspired strategies that sought to alter the identity of a target non-ribosomal peptide residue via substitution of the corresponding NRPS substrate-binding pocket were:

Figure 1



Schematic summary of key NRPS domain substitution experiments. **(a)** Structure of a typical three-module NRPS showing the relative locations of the adenylation (A), thiolation (T), condensation (C), and thioesterase (Te) domains. Functional units representing the A subdomain and the region stretching from the core (T^1) motif of the T domain to the end of the C-domain are highlighted by green shading. Regions that were exchanged in different substitution experiments are shown below the NRPS, with the naming of the substituted region provided to the right and approximate recombination sites indicated using dashed lines. Broadly equivalent recombination sites from different studies (within 10–20 residues of each other) are represented by a single dashed line. **(b)** Boxplots showing the percentage yield of compound production by strains containing chimeric NRPS enzymes relative to an unmodified control, as reported in key studies that implemented different substitution strategies. The boxes span the interquartile range (IQR, 25th to 75th percentile), the horizontal line inside each box is the median, the whiskers extend to the most extreme value within 1.5 × IQR, and points show values outside that range. The y-axis has been limited to 125% for display purposes, which caused three data points to not be shown at 414.4%, 187.1% and 197.3% for the XUT^I, XUT^{IV} and XU strategies, respectively. The sample size was similar in each study (17 for subdomains [22], 16 for A domains [18], and 15 for XUT^I, 16 for XUT^{IV}, 14 for X^U and 16 for XU^C [20]). For A domain substitution, the percentage yield for four strains producing D-Phe-L-Leu diketopiperazine (DKPs) was calculated relative to a previously-established D-Phe-L-Phe DKP-producing strain [18]. It should also be noted that the study comparing XUT^I, XUT^{IV}, X^U and XU^C substitutions [20] included substitutions into GxpS without percentage yields provided, so to enable comparisons here, percentage yields were calculated relative to the yield from unmodified GxpS reported previously by the same team [23,24]. A Mann–Whitney U test was conducted to compare subdomain substitutions with every other substitution type. Only XUC domain substitutions showed a significant effect ($p = 0.0021$; Bonferroni-adjusted $\alpha = 0.05/5$). **(c)** Boxplots showing the percentage yield of each different compound analogue produced by the top variants created via each of the three high-throughput strategies considered in this review. The boxes, whiskers, and points have the same parameters used in panel b. The different substitution strategies generated 14 (C^4–A^{10} [41]), 12 (C^1–A^{10} [41]), and 15 (XUT^{IV} [42]) compound analogues, respectively. A Mann–Whitney U test was conducted to compare C^4–A^{10} domain substitutions with each other substitution type. No significant differences were detected at the 0.05 level or at the Bonferroni-corrected threshold.

1) subdomain substitution focused on the binding pocket of the A domain [22]; 2) A domain substitution that replaced the entire A domain [18]; and eXchange Unit between T domains (XUTs) that used whole modules with recombination points at either 3) the start of the T domain (XUT^I); or 4) within the T domain motif (XUT^{IV}) [20]. When evaluating relative product yields resulting from successful substitutions, as reported by the authors of each study, all methods showed reasonably similar percentage yields for the substitution strains compared to controls when examining the median and interquartile range (Figure 1b). Direct comparison of the evolution-inspired XUT domain substitutions to previous structure-based exchange unit (XU) [23] and exchange unit condensation domain (XUC) [24] methods pioneered by the same team showed that the XUT recombination sites were comparable to XU and performed substantially better than XUC [20]. However, it is important to note that all the studies considered in Figure 1b focused on different NRPS templates and assessed relatively small numbers of recombination events. Statistically powered head-to-head comparisons in the same enzymatic systems would be required to overcome stochastic variation and more robustly assess the relative effectiveness of each strategy.

Thus, although some recombination sites are clearly better than others for achieving functional and higher-yield outcomes, no one strategy is clearly 'best' for every

BGC. For example, subdomain substitution boundaries that were successful in generating modified enduracidin analogues [22] were unsuccessful in modifying pyoverdine biosynthesis [18]. Moreover, although each method has had success in different NRPS systems, product yields following individual substitution events are generally less than 50% relative to unmodified controls (Figure 1b). Such activity losses may be nearly unavoidable due to the disruption of evolutionarily optimised structural interactions between native domains, as well as the inevitable introduction of new residue sidechain clashes. These suppositions are supported by recent statistical coupling analyses of groups of amino acids that are functionally linked [21], and the use of software to predict residue clashes [18]. Overall, while the use of evolutionarily-informed NRPS fusion sites has improved outcomes overall, diminished product yields are common, and there does not appear to be any universal fusion point that can reliably avoid perturbing important functional interactions. As distant interactions can be predicted by statistical coupling and then combined with projections of likely clashes, it may be that future machine learning studies will address this. In the meantime, it may be more pragmatic to test multiple recombination strategies in any new target BGC.

Some parallels have been observed in recent PKS engineering efforts (Figure 3), although we consider that there has not yet been sufficient product yield data reported to enable meaningful comparisons between the different strategies. While modular PKS and NRPS enzymes share similar biosynthetic mechanisms, PKS engineering poses some unique challenges. For example, PKS systems typically contain more domains per module, suggesting a heightened risk that important protein–protein interactions between domains will be disrupted, and yield losses can occur due to ketosynthase domain substrate gatekeeping [25,26]. A PKS module was traditionally considered to be made up of KS–AT–X–ACP domains (where X represents the reductive domains present) [27,28], but phylogenetic analysis has shown that KS domains tend to migrate with the corresponding upstream processing domains [28,29], much like T–C or T–E–C pairings in NRPS systems. The importance of this was apparent in a study that sought to replace rapamycin PKS domains, which resulted in a variety of unexpected analogues arising from module-duplication and deletion events [30]. Notably, the recombination sites that led to these diversified BGCs were within the KS domain, consistent with the AT–X–ACP–KS domain structure implied by phylogenetic analysis. This effect, termed 'accelerated evolution', was subsequently recapitulated to deliberately induce recombination between homologous regions and create novel tylosin analogues [30], while in another study, the mid-KS region was rationally targeted to engineer chimeric stambomycin PKS enzymes [31]. Chimeric PKSs have also been constructed simply by introducing genes encoding different PKS modules into *Saccharomyces cerevisiae*, a highly recombinogenic intermediate host that proved effective in fusing modules together, most commonly within KS or AT domains [32].

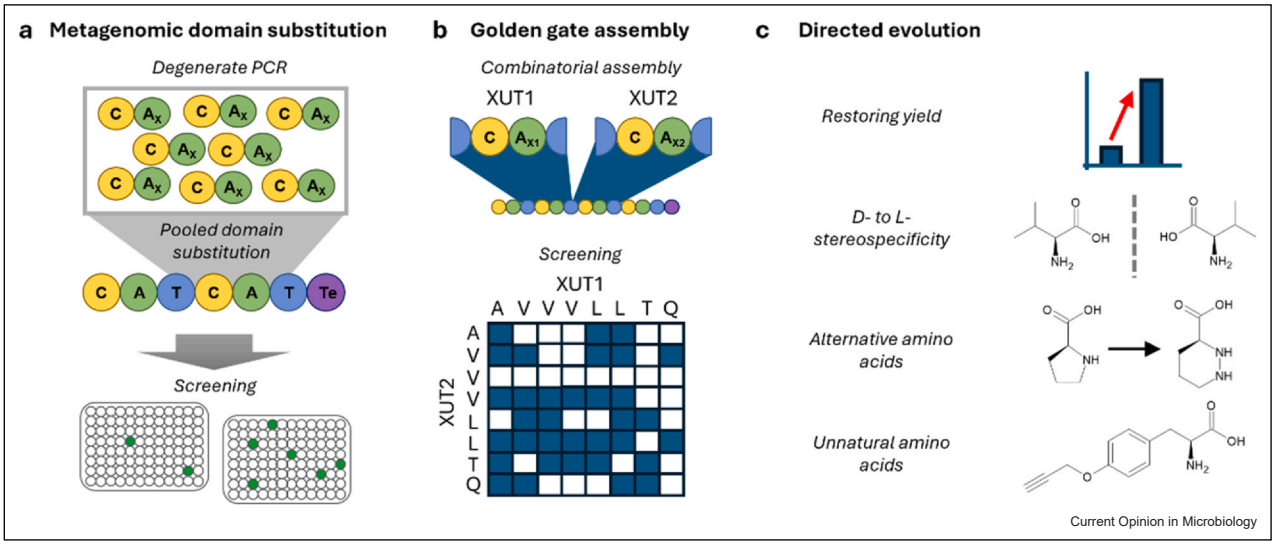
Phylogenetically defined exchange units (i.e. AT–X–ACP–KS, where the splice site is within the KS–AT linker sequence) have now been directly compared to traditional module constructs (KS–AT–X–ACP), and demonstrated to yield superior recombination outcomes in multiple experimental systems [33–36]. However, not all module substitutions are equally tolerated, even using the updated boundaries [34,36], with module skipping sometimes observed [36]. Thus, while progress has been made overall, there are clearly other aspects to consider before high-yielding chimeric PKSs can be reliably created.

Preferred PKS module boundaries to enhance functionality and product yield were adjusted slightly in 2025, based on phylogenetic analyses of the KS–AT linker sequence that indicated that these sequences form clades according to the nature of the AT domain that follows (e.g. selective for *2S*-methylmalonyl-CoA). This suggested that KS–AT pairings are an evolutionarily conserved unit, and it was further inferred that the KS–AT linker interacts physically with both the upstream KS domain and downstream AT domain [37]. Based on these inferences, a post-AT splice site was identified as an effective exchange unit boundary (i.e. X–ACP–KS–AT) that can outperform the KS–AT linker boundary [37,38]. The equivalent region of trans-AT PKSs was also found to be an effective recombination site that enabled the generation of functional trans-AT PKS chimeras (via both insertion and removal of domains) [39]. Module cut sites for the insertion of a terminal reductase (TR) domain were also identified within evolutionarily conserved amino acid motifs between domains [40]. This led to the ACP–TR domain being identified as a reliable unit for transplanting onto the end of PKS modules to reductively release the polyketide product [40]. Based on current information, we advise that researchers planning new PKS engineering experiments use these module boundaries as a starting point.

Evolution-inspired strategies focused on increased throughput

Having the ability to conduct an unfathomable number of recombination experiments in parallel, nature can afford most NRPS and PKS recombination events to be unsuccessful. Lab-based efforts do not have this luxury, but some teams have developed methods that sacrifice a degree of efficiency in exchange for substantially increased throughput relative to individual rational domain

Figure 2

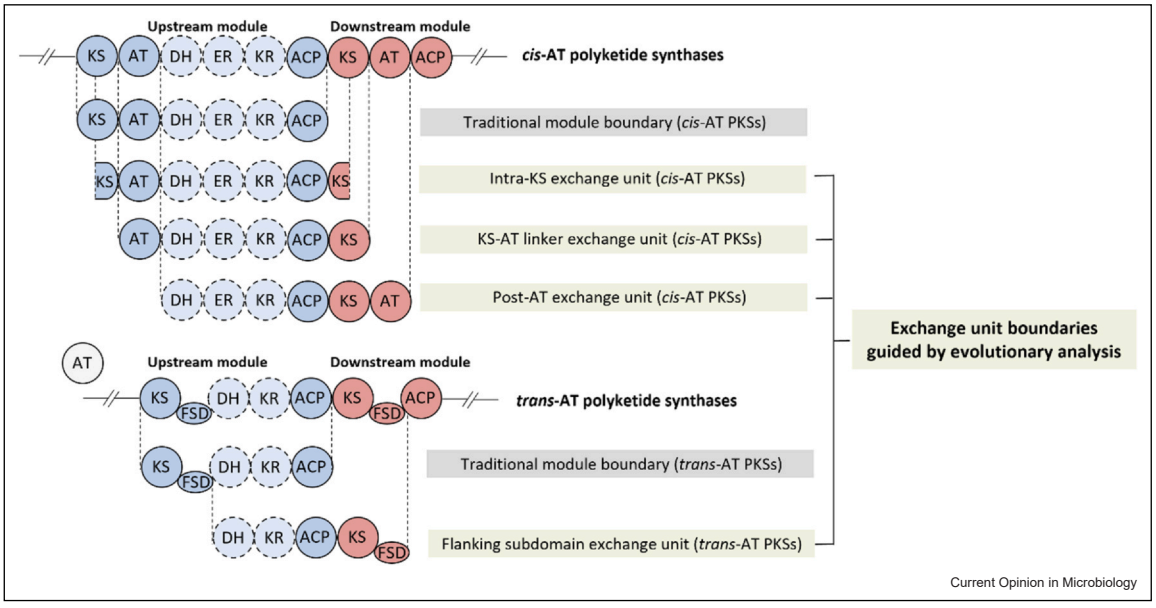


A schematic summary of high-throughput methods to engineer non-ribosomal peptide biosynthesis. **(a)** Metagenomic domain substitution amplifies pools of domains using conserved motifs and substitutes them en masse into an NRPS. Functional variants are then identified by screening [38]. **(b)** Golden Gate Assembly has utilised conserved glycine residues within the T domain motif to generate overhangs for piecing together modules. These were assembled in a combinatorial fashion, and strains containing the resulting constructs were screened for activity [42]. **(c)** Directed evolution has been used to randomly mutate DNA or specific residues to restore yield [45,46], change stereospecificity [48] or incorporate either alternative or unnatural amino acids [49,50].

substitutions. One recent approach used degenerate PCR primers to target the highly conserved C1 or C4 and A10 motifs within NRPS modules present in soil metagenomic DNA, to amplify C^1–A^{10} and C^4–A^{10} domain fragments from the thousands of bacterial genomes represented therein. This yielded over 1000 unique domain fragments that were then used to substitute the equivalent region in a *Pseudomonas aeruginosa* module involved in pyoverdine biosynthesis (Figure 2a) [41]. Screening identified over 100 unique functional variants, and 16 distinct pyoverdines were produced as major products (Figure 1c) [41]. Another high-throughput approach based on XUT^{IV} domain substitutions targeted two conserved glycine residues within the T domain as a recombination site for Golden Gate Assembly (Figure 2b) [42]. Chimeric artificial NRPS enzymes of three or four modules in length were made using different starter, elongation and termination modules, with 46 out of 96 producing the expected peptide product. An additional 44 variants were created to modify xenomicin biosynthesis by replacing two modules with one or two modules, resulting in the production of 15 peptides of the expected masses (Figure 1c) [42]. A further combinatorial approach used restriction sites strategically installed within the C/A-domain linker regions of NRPS modules involved in biosynthesis of plipastatin or gramicidin S to randomly exchange modules between the two systems [43]. Thirty-two distinct combinations of chimeric plipastatin–gramicidin S gene clusters were recovered from 50 randomly-selected transformants, and ultimately nine unique lipopeptides were recovered alongside a range of smaller peptide products [43]. Importantly, each of the examples above demonstrates how the historically low success rates for individual NRPS recombination events can be offset by enhancing throughput and selecting desired outcomes, that is, strategies that more closely resemble natural evolution.

Given the requirement for precise molecular choreography between domains, it seems likely that natural NRPS or PKS recombination events also usually yield low-activity initial outcomes, followed by an accrual of point mutations that 'massage' the newly acquired domain(s) into place [44]. This subsequent optimisation is generally difficult to recapitulate in laboratory-based recombination studies, although it was achieved in a landmark study using error-prone PCR to restore function to inactive recombinant NRPS enzymes involved in enterobactin or andrimid synthesis, for which a high-throughput product selection and screen, respectively, were available [45]. More recent work has shown that semi-rational approaches can get remarkable results with relatively few mutations required to restore activity [46]. In one case, subdomain substitution into a GrsA/GrsB1 model system that usually yields a D-Phe-L-Pro diketopiperazine generated a variant that additionally made a D-Val-L-Pro analogue [47]. Combining mutations to revert individual subdomain residues predicted to clash

Figure 3



A summary of the different PKS exchange units used in recent engineering experiments. Top: Schematic structure of a two-module cis-AT PKS, containing ketosynthase (KS), acyltransferase (AT), dehydratase (DH), enoylreductase (ER), ketoreductase (KR), and acyl carrier protein (ACP) domains. The optional reduction domains (DH, ER, and KR) are highlighted using a broken black border. The traditional biochemical PKS module definition starts at an N-terminal KS domain and ends with a C-terminal ACP domain. The different PKS exchange units identified using an evolution-guided approach, which have been used to create hybrid or engineered PKS enzymes, are then summarised (intra-KS exchange unit, KS–AT linker exchange unit, post-AT exchange unit). Bottom: The structure of a model two-module trans-AT PKS containing ketosynthase (KS), flanking subdomain (FSD), acyltransferase (AT), dehydratase (DH), ketoreductase (KR), and ACP domains. A recently identified trans-AT PKS exchange unit spans from the C-terminal end of the upstream module's FSD to the C-terminal end of the downstream module's FSD.

with the rest of the A domain not only increased dipeptide production up to sixfold, but also altered specificity towards Val or Phe depending on the mutations (Figure 2c) [47].

In a different study using the same DKP-producing system, saturation mutagenesis was applied to improve the relative yield of a side product. Selected residues in the C-domain of GrsA/GrsB1 were targeted to modulate the preference for incorporating different L-amino acids (Figure 2c), with some portability of key mutations to other NRPS systems also observed [48]. It was also found that the A domain of GrsB1 could be mutated to allow low-activity incorporation of L-piperazic acid instead of L-Pro, with additional mutagenesis restoring activity (Figure 2c) [49]. Transferring the mutations to full-length GrsB allowed production of gramicidin S containing L-piperazic acid at 15 mg/l [49].

Engineering NRPS or PKS enzymes involved in siderophore biosynthesis offers potential to mimic the enhanced throughput of natural evolution, by applying a conditional viability selection on media amended with chelating agents that prevent passive iron uptake. This strategy was recently applied for domain substitution to replace the Ser-specific A domain of the enterobactin synthetase EntF with a Leu-specific A domain, and subsequently find mutations to shift the specificity back to serine. Although it was expected that only serine-containing enterobactins would transport iron, recovered strains contained tyrosine-incorporating A domains that generated a novel tyrosine-containing enterobactin [50]. This illustrates the potential for evolutionary approaches to provide unexpected insights by avoiding preconceptions associated with rational strategies.

Another attractive selection was employed for high-throughput yeast display to engineer a C-domain variant of surfactin synthetase SrfA-C to better tolerate *O*-propargyl-L-Tyr and 10-undecynoic acid as donor substrates (Figure 2c). The donor substrates were passed from another NRPS enzyme or an artificially loaded T domain to a displayed module in which the C-domain was mutated using degenerate codons. Absence of a Te-domain caused any product to remain attached to the T domain of the displayed module, while the terminal alkynes of *O*-propargyl-L-Tyr and 10-undecynoic acid allowed attachment of a fluorescent label and selection by fluorescence-activated cell sorting [51]. Application of this display system to modify C-domains demonstrates its versatility, having previously been used to modify A-domain specificity to recognise clickable amino acids [52,53], α-hydroxy acids [54], or β-amino acids [55].

Evolution-informed strategies to improve product yields

Evolutionary insights can also be applied to overcome the losses in yield that are typically associated with NRPS and PKS BGC engineering. For example, a comparison of 720 actinobacteria genomes found that the most PKS-rich species were consistently associated with a five-gene operon for pyrroloquinoline (PQQ) biosynthesis [56]. Overexpression of the PQQ operon was found to enhance polyketide production via a proposed mechanism of increasing cytoplasmic ATP, NADH, NADPH and acyl-CoAs [56]. An overexpressed PQQ operon may therefore offset yield losses in organisms expressing chimeric PKSs. Other features that could help enhance engineered product yields include increasing the pool of available malonyl-CoA via orthogonal pathway insertion [57], splitting large PKS or NRPS genes into smaller units to enhance mRNA stability [58,59], and modifying ACP domains for maximal activation by the preferred phosphopantetheinyl transferase [60].

To mimic the stochastic and high-throughput nature of evolution, combinatorial strategies are increasingly being applied to both NRPS and PKS systems. Cloning genes into alternative plasmids with different copy numbers and identifying optimal combinations increased salivabactin and epoxomicin production two to eightfold, respectively [61,62], while precursor supplementation [61,62], linking enzymes together [62], and modifying drug efflux systems [61] further increased yield. Similarly, the use of homology-based assembly to create compatible plasmids with different gene combinations allowed rapid characterisation of diverse gene combinations, and identification of new product analogues as well as the minimal genes for heterologous expression in a new host [63]. An alternative approach of iteratively and combinatorially testing different promoters improved production of polymyxin B in a heterologous host by 40-fold [64]. As screening can be limited by compound detection, another study inserted a standalone single-module NRPS that generates the blue pigment indigoidine into diverse BGCs, which served as a proxy for expression. This allowed screening of CRISPRi libraries targeting hundreds of repressor genes, and strains were created producing daptomycin, thaxtomin A and surfactin at 1054 mg/L, 352 mg/L and 878 mg/L, respectively [65].

Conclusions

Substantial progress has been made in experimentally defining optimal fusion sites, developing high-throughput engineering methods and increasing compound yields. Although there have been attempts to define rules for engineering, given the complexity of protein–protein interactions within complex NRPS and PKS megaenzymes, it is unlikely that there is a universally optimal approach. Rather, we suggest that researchers who seek optimal outcomes should be prepared to be flexible, trialling multiple strategies in target BGCs and scaling the most productive. It is probable that domain substitutions will continue to impact product yield, but these losses may in the future be offset by the development of innovative selection strategies to enable directed evolution and advances in heterologous expression to improve yield. Moreover, methods have been developed to detect long-range interactions that are disrupted by domain recombination, but these rely on covariance within large evolutionary datasets and have not yet been applied to increase the success of individual domain substitutions.

Resolving the challenges associated with achieving domain substitution at scale is likely to prove essential to increasing the likelihood of individual studies succeeding. Addressing these challenges could also generate the large datasets required for machine learning to bridge the gap between covariance and prediction, to project the recombination strategy likely to prove optimal for each new target BGC. We anticipate a positive feedback loop as the scale of experiments increases and our knowledge of natural NRPS and PKS diversification improves, providing opportunities to further fine-tune and accurately predict outcomes. Given the looming threat of antibiotic resistance and the dwindling numbers of natural products entering clinical pipelines, achieving these goals is becoming increasingly critical.

Funding sources

Dr Mark Calcott is supported by a Health Research Council of New Zealand Hercus Fellowship (contract 23/188/A) and Dr Rory Little by an Aola Richards Postdoctoral Research Fellowship in Entomology and a Royal Society of New Zealand Marsden Fast Start grant (contract 24-VUW-096). Professor David Ackerley and Anna Sang received additional support from Research for Life (the Wellington Medical Research Foundation).

CRediT authorship contribution statement

Mark J Calcott: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing. **Anna C Sang:** Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing. **David F Ackerley:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing. **Rory F Little:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing.

Data Availability

No data were used for the research described in the article.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mark Calcott reports financial support was provided by Health Research Council of New Zealand. Rory Little reports financial support was provided by The Royal Society of New Zealand. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr Kenan Bozhüyük for the kind invitation to submit this review and for proposing the general topic area; and Thomas Skurr for assistance with the graphical abstract (created in BioRender. Skurr, T. (2026) https://BioRender.com/zpbx2rj).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Fleming A: **On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzæ***. *Br J Exp Pathol* 1929, **10**:226-236.

2. Kardos N, Demain AL: **Penicillin: the medicine with the greatest impact on therapeutic outcomes**. *Appl Microbiol Biotechnol* 2011, **92**:677-687.

3. Katz L, Baltz RH: **Natural product discovery: past, present, and future**. *J Ind Microbiol Biotechnol* 2016, **43**:155-176.

4. Hutchings MI, Truman AW, Wilkinson B: **Antibiotics: past, present and future**. *Curr Opin Microbiol* 2019, **51**:72-80.

5. Agrawal S, Acharya D, Adholeya A, Barrow CJ, Deshmukh SK: **Nonribosomal peptides from marine microbes and their antimicrobial and anticancer potential**. *Front Pharmacol* 2017, **8**:828.

6. Chevrette MG, Currie CR: **Emerging evolutionary paradigms in antibiotic discovery**. *J Ind Microbiol Biotechnol* 2019, **46**:257-271.

7. Kunakom S, Eustaquio AS: **Natural products and synthetic biology: where we are and where we need to go**. *Msystems* 2019, **4**:e00113-19.

8. MacGowan A, Macnaughton E: **Antibiotic resistance**. *Medicine* 2017, **45**:622-628.

9. Hertweck C: **The biosynthetic logic of polyketide diversity**. *Angew Chem Int Ed Engl* 2009, **48**:4688-4716.

10. Keatinge-Clay AT: **The structures of type I polyketide synthases**. *Nat Prod Rep* 2012, **29**:1050-1073.

11. Helfrich EJ, Piel J: **Biosynthesis of polyketides by trans-AT polyketide synthases**. *Nat Prod Rep* 2016, **33**:231-316.

12. Little RF, Hertweck C: **Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis**. *Nat Prod Rep* 2022, **39**:163-205.

13. Fischbach MA, Walsh CT, Clardy J: **The evolution of gene collectives: how natural selection drives chemical innovation**. *Proc Natl Acad Sci USA* 2008, **105**:4601-4608.

14. Lawrence J: **Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes**. *Curr Opin Genet Dev* 1999, **9**:642-648.

15. Medema MH, Cimermancic P, Sali A, Takano E, Fischbach MA: **A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis**. *PLoS Comput Biol* 2014, **10**:e1004016.

16. •• Hansen MH, Adamek M, Iftime D, Petras D, Schuseil F, Grond S, *et al.*: **Resurrecting ancestral antibiotics: unveiling the origins of modern lipid II targeting glycopeptides**. *Nat Commun* 2023, **14**:7842.

This publication describes ancestral reconstruction of BGCs involved in the biosynthesis of glycopeptide antibiotics. The authors traced evolution over time and reconstructed a BGC to produce an ancestral molecule. They further explored how A domain binding pocket can evolve through point mutations to accept new amino acid substrates.

17. • Li Z, Ióca LP, He R, Donia MS: **Natural diversifying evolution of nonribosomal peptide synthetases in a defensive symbiont reveals nonmodular functional constraints**. *PNAS Nexus* 2024, **3**:pgae384.

This publication further refined the concept of T-C and T-E-C functional units based on evolutionary analysis and suggested that NRPSs should be considered as comprising two main functionally-linked subunits. Evidence for this is well-supported by other evolutionary and biochemical studies.

18. Calcott MJ, Owen JG, Ackerley DF: **Efficient rational modification of non-ribosomal peptides by adenylation domain substitution**. *Nat Commun* 2020, **11**:4554.

19. Baunach M, Chowdhury S, Stallforth P, Dittmann E: **The landscape of recombination events that create nonribosomal peptide diversity**. *Mol Biol Evol* 2021, **38**:2116-2130.

20. •• Bozhüyük KAJ, Präve L, Kegler C, Schenk L, Kaiser S, Schelhas C, *et al.*: **Evolution-inspired engineering of nonribosomal peptide synthetases**. *Science* 2024, **383**:eadg4320.

This publication identified phylogenetic incongruence between T domains and the immediately downstream domain using large-scale analysis, the incongruence being strongly correlated to the downstream domain type. This guided the insightful selection of preferred recombination sites that were then used to create a series of chimeric NRPS enzymes.

21. He R, Zhang J, Shao Y, Gu S, Song C, Qian L, *et al.*: **Knowledge-guided data mining on the standardized architecture of NRPS: subtypes, novel motifs, and sequence entanglements**. *PLoS Comput Biol* 2023, **19**:e1011100.

22. Thong WL, Zhang YX, Zhuo Y, Robins KJ, Fyans JK, Herbert AJ, *et al.*: **Gene editing enables rapid engineering of complex antibiotic assembly lines**. *Nat Commun* 2021, **12**:6872.

23. Bozhuyuk KAJ, Fleischhacker F, Linck A, Wesche F, Tietze A, Niesert CP, *et al.*: **De novo design and engineering of non-ribosomal peptide synthetases**. *Nat Chem* 2018, **10**:275-281.

24. Bozhuyuk KAJ, Linck A, Tietze A, Kranz J, Wesche F, Nowak S, *et al.*: **Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains**. *Nat Chem* 2019, **11**:653-661.

25. Yi D, Wakeel MA, Agarwal V: **Gatekeeping activity of collinear ketosynthase domains limits product diversity for engineered type I polyketide synthases**. *Biochemistry* 2024, **63**:2240-2244.

26. Hirsch M, Fitzgerald BJ, Keatinge-Clay AT: **How cis-acyltransferase assembly-line ketosynthases gatekeep for processed polyketide intermediates**. *ACS Chem Biol* 2021, **16**:2515-2526.

27. Chen B, Bülbül EF, Bang S, Minas HA, Bozhüyük KA: **Toward a unified pipeline for natural product discovery: tools and strategies for NRPS and PKS pathway exploration and engineering**. *Nat Prod Rep* 2025, **43**:352-370.

28. Nivina A, Yuet KP, Hsu J, Khosla C: **Evolution and diversity of assembly-line polyketide synthases**. *Chem Rev* 2019, **119**:12524-12547.

29. Zhang L, Hashimoto T, Qin B, Hashimoto J, Kozone I, Kawahara T, *et al.*: **Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides**. *Angew Chem* 2017, **56**:1740-1745.

30. Wlodek A, Kendrew SG, Coates NJ, Hold A, Pogwizd J, Rudder S, *et al.*: **Diversity oriented biosynthesis via accelerated evolution of modular gene clusters**. *Nat Commun* 2017, **8**:1206.

31. Su L, Hôtel L, Paris C, Chepkirui C, Brachmann AO, Piel J, *et al.*: **Engineering the stambomycin modular polyketide synthase yields 37-membered mini-stambomycins**. *Nat Commun* 2022, **13**:515.

32. Chemler JA, Tripathi A, Hansen DA, O'Neil-Johnson M, Williams RB, Starks C, Park SR, Sherman DH: **Evolution of efficient modular polyketide synthases by homologous recombination**. *J Am Chem Soc* 2015, **137**:10603-10609.

33. Miyazawa T, Fitzgerald BJ, Keatinge-Clay AT: **Preparative production of an enantiomeric pair by engineered polyketide synthases**. *Chem Commun* 2021, **57**:8762-8765.

34. Miyazawa T, Hirsch M, Zhang Z, Keatinge-Clay AT: **An in vitro platform for engineering and harnessing modular polyketide synthases**. *Nat Commun* 2020, **11**:80.

35. Peng H, Ishida K, Sugimoto Y, Jenke-Kodama H, Hertweck C: **Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases**. *Nat Commun* 2019, **10**:3918.

36. • Ray KA, Lutgens JD, Bista R, Zhang J, Desai RR, Hirsch M, *et al.*: **Assessing and harnessing updated polyketide synthase modules through combinatorial engineering**. *Nat Commun* 2024, **15**:6485.

This publication describes a flexible and elegant platform for generating hybrid polyketide synthases, using the KS-AT linker splice site, and yielding diverse triketide, tetraketide, and pentaketide products. Importantly, equivalent hybrid synthases created using the traditional module boundaries were less functional.

37. •• Huang Z, Xie S, Liu R-Z, Xiang C, Yao S, Zhang L: **Plug-and-play engineering of modular polyketide synthases**. *Nat Chem Biol* 2025, **21**:1361-1367.

This study provides a side-by-side comparison of different evolution-guided splice sites to create hybrid polyketide synthases. A key outcome from this work is the identification of the conserved post-AT site, which appears to be an excellent starting point for future PKS engineering experiments.

38. Grininger M: **Design principles imprinted by evolution**. *Nat Chem Biol* 2025, **21**:1301-1303.

39. •• Mabesoone MF, Leopold-Messer S, Minas HA, Chepkirui C, Chawengrum P, Reiter S, *et al.*: **Evolution-guided engineering of trans-acyltransferase polyketide synthases**. *Science* 2024, **383**:1312-1317.

This publication brought evolution-guided engineering to trans-AT PKS enzymes. Co-evolving regions of trans-AT PKS enzymes were identified and used to identify trans-AT PKS exchange units. These sequences are homologous to the post-AT region of cis-AT PKSs, unifying the evolutionary-guided exchange unit approach for each PKS class.

40. Yao S, Xie S, Liu R-Z, Huang Z, Zhang L: **Expanding catalytic versatility of modular polyketide synthases for alcohol biosynthesis**. *Nat Chem Biol* 2025, **21**:1368-1375.

41. Messenger SR, McGuinniety EMR, Stevenson LJ, Owen JG, Challis GL, Ackerley DF, *et al.*: **Metagenomic domain substitution for the high-throughput modification of nonribosomal peptides**. *Nat Chem Biol* 2024, **20**:251-260.

•This publication describes a strategy of targeting conserved motifs with degenerate primers to amplify large pools of unknown domains for substitution into a recipient NRPS module. This permits the generation of large amounts of diversity in parallel without having to perform individual PCRs from predefined genetic material.

42. •• Podolski A, Lindeboom TA, Präve L, Dehghan M, Minas HA, Kranz J, *et al.*: **High-throughput engineering and modification of non-ribosomal peptide synthetases based on Golden Gate Assembly**. *Angew Chem Int Ed* 2025, **64**:e202508967.

This publication describes the use of Golden Gate Assembly to create libraries of NRPS variants. This method of assembly allows subunits to be combined using optimised recombination points to quickly create chimeric NRPS enzymes.

43. Jagadeesh V, Okahashi N, Matsuda F, Tsuge K, Kondo A: **Combinatorial nonribosomal peptide synthetase libraries using the SEAM-Combi-OGAB method**. *ACS Synth Biol* 2025, **14**:520-530.

44. Ackerley DF: **Cracking the nonribosomal code**. *Cell Chem Biol* 2016, **23**:535-537.

45. Fischbach MA, Lai JR, Roche ED, Walsh CT, Liu DR: **Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes**. *Proc Natl Acad Sci USA* 2007, **104**:11951-11956.

46. Stanišić A, Hüsken A, Stephan P, Niquille DL, Reinstein J, Kries H: **Engineered nonribosomal peptide synthetase shows opposite amino acid loading and condensation specificity**. *ACS Catal* 2021, **11**:8692-8700.

47. Kries H, Niquille DL, Hilvert D: **A subdomain swap strategy for reengineering nonribosomal peptides**. *Chem Biol* 2015, **22**:640-648.

48. Peng H, Schmiederer J, Chen X, Panagiotou G, Kries H: **Controlling substrate- and stereospecificity of condensation domains in nonribosomal peptide synthetases**. *ACS Chem Biol* 2024, **19**:599-606.

49. • Stephan P, Langley C, Winkler D, Basquin J, Caputi L, O'Connor SE, *et al.*: **Directed evolution of piperazic acid incorporation by a nonribosomal peptide synthetase**. *Angew Chem Int Ed Engl* 2023, **62**:e202304843.

This publication describes the use of directed evolution to modify a Pro-specific A domain to incorporate piperazic acid into gramicidin S biosynthesis. The shift in specificity was achieved by screening a relatively small library of 1200 mutants, showing that directed evolution can still be highly effective when high-throughput screening is unavailable. This is an important result, as screening can be a bottleneck for engineering secondary metabolite biosynthesis.

50. •• Conley E, Wadler CS, Bell BA, Lucier I, Haynie C, Eldred S, *et al.*: **Directed evolution of an adenylation domain alters substrate specificity and generates a new catechol siderophore in *Escherichia coli***. *Biochemistry* 2024, **63**:3126-3135.

This publication describes an A domain substitution in an enterobactin NRPS that eliminated siderophore production, followed by high-throughput selection of A domain binding pocket mutations that restored siderophore activity. The resulting product surprisingly contained Tyr instead of the naturally occurring Ser, highlighting the ability of directed evolution to achieve unexpected solutions.

51. • Folger IB, Frota NF, Pistofidis A, Niquille DL, Hansen DA, Schmeing TM, *et al.*: **High-throughput reprogramming of an NRPS condensation domain**. *Nat Chem Biol* 2024, **20**:761-769.

This publication describes the use of click chemistry to fluorescently label non-ribosomal peptides tethered to yeast cells. This enabled directed evolution that utilised flow cytometry-based screening of large libraries to increase the catalytic efficiency of C domains with alternative fatty acid donor substrates.

52. 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**:10105-10108.

53. Camus A, Gantz M, Hilvert D: **High-throughput engineering of nonribosomal extension modules**. *ACS Chem Biol* 2023, **18**:2516-2523.

54. Camus A, Truong G, Mittl PRE, Markert G, Hilvert D: **Reprogramming nonribosomal peptide synthetases for site-specific insertion of α-hydroxy acids**. *J Am Chem Soc* 2022, **144**:17567-17575.

55. Niquille DL, Hansen DA, Mori T, Fercher D, Kries H, Hilvert D: **Nonribosomal biosynthesis of backbone-modified peptides**. *Nat Chem* 2018, **10**:282-287.

56. Wang X, Chen N, Cruz-Morales P, Zhong B, Zhang Y, Wang J, *et al.*: **Elucidation of genes enhancing natural product biosynthesis through co-evolution analysis**. *Nat Metab* 2024, **6**:933-946.

57. Klass SH, Wesselkamper M, Cowan AE, Lee N, Lanclos N, Cheong S, *et al.*: **Engineering controllable alteration of malonyl-CoA levels to enhance polyketide production**. *Nat Chem Biol* 2025, **21**:1214-1225.

58. Liu Y, Song C, Cui Q, Sun H, Jiang C, Guo R, *et al.*: **Improving polyketide biosynthesis by rescuing the translation of truncated mRNAs into functional polyketide synthase subunits**. *Nat Commun* 2025, **16**:774.

59. Kegler C, Bode HB: **Artificial splitting of a non-ribosomal peptide synthetase by inserting natural docking domains**. *Angew Chem Int Ed Engl* 2020, **59**:13463-13467.

60. Li K, Cho YI, Tran MA, Wiedemann C, Zhang S, Koweek RS, *et al.*: **Strategic acyl carrier protein engineering enables functional type II polyketide synthase reconstitution in vitro**. *ACS Chem Biol* 2025, **20**:197-207.

61. Ko Y-S, Gu D, Lee A, Al Marzooqi D, Zhang W: **Engineering *Escherichia coli* for the production of the PK/NRP hybrid antibiotic salivabactin**. *ACS Sustain Chem Eng* 2025, **13**:10556-10562.

62. Messenger SR, Ackerley DF, Calcott MJ: **Heterologous expression of epoxomicin in *Escherichia coli***. *ACS Synth Biol* 2024, **13**:2702-2709.

63. Calcott MJ, Kröger J, Ackerley DF: **A rapid combinatorial assembly method for gene cluster characterisation illuminates glidobactin biosynthesis**. *Synth Syst Biotechnol* 2026, **11**:85-90.

64. Sun HZ, Li Q, Shang W, Qiao B, Xu QM, Cheng JS: **Combinatorial metabolic engineering of *Bacillus subtilis* for de novo production of polymyxin B**. *Metab Eng* 2024, **83**:123-136.

65. •• Yan H, Xin Z, Sang Z, Li X, Xie J, Wu J, *et al.*: **A rational multi-target combination strategy for synergistic improvement of non-ribosomal peptide production**. *Nat Commun* 2025, **16**:1883.
This publication describes the insertion of an indigoidine-producing NRPS into a non-ribosomal peptide-producing BGC and shows that expression of both peptides was strongly correlated. The indigoidine production provided a colourimetric output that enabled screening of libraries via CRISPR interference to identify repressor genes outside of BGCs.