Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Review

From known knowns to unknown unknowns: synthetic biology paths to antimicrobial discovery

Siqiang Chen[1,2], Hao Xiang[1,2,3], Ying Chen[1,2], Helge B. Bode[4,5,6] and Yi-Ming Shi[1,2,3]

'Structurally known' and 'Structurally unknown' have long served as a binary classification of natural products. However, with the emergence of synthetic biology (SynBio), the boundaries of this simple classification are being redrawn, because SynBio expands both what we can see and what we can build. To reflect this change, we propose a SynBio-centered framework that maps antimicrobial natural products along two axes, chemical novelty and engineering accessibility, yielding four quadrants: (i) known knowns, well-characterized natural products accessed mainly through pathway refactoring and producer strain optimization; (ii) unknown knowns, engineerable scaffolds with established diversification routes but uncertain chemical and functional outcomes; (iii) known unknowns, architectures with precedent but limited realization due to missing transferable engineering routes, often because the relevant coupling logic has not yet been converted into a practical engineering handle; (iv) unknown unknowns, unanticipated scaffolds that need to be recognized through minimal chemical, genetic, or mechanistic anchors before engineering. Since these quadrants are dynamic rather than fixed, compounds may shift across the framework as discovery, route-building, and outcome predictability improve. This two-axis map highlights complementary SynBio strategies, from optimizing existing antibiotics to accessing first-in-class entities, which may help guide efforts to combat current resistance challenges.

Addresses
[1] State Key Laboratory of Quantitative Synthetic Biology, Center for Synthetic Biochemistry, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2] Max Planck–Chinese Academy of Sciences Center for Synthetic Biochemistry, Shenzhen, China
[3] University of Chinese Academy of Sciences, Beijing, China
[4] Department of Natural Products in Organismic Interactions, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany
[5] Center for Synthetic Microbiology (SYNMIKRO) & Department of Chemistry, Philipps University of Marburg, Marburg, Germany
[6] Max Planck–Chinese Academy of Sciences Center for Synthetic Biochemistry, Marburg, Germany

Corresponding author: Shi, Yi-Ming (ym.shi@siat.ac.cn)

Current Opinion in Microbiology 2026, **91**:102579

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

Edited by **Kenan Bozhuyuk**

Available online xxxx

https://doi.org/10.1016/j.mib.2026.102759

Introduction

Traditionally, natural product discovery has relied on a binary classification: compounds are either structurally known (fully characterized by nuclear magnetic resonance (NMR) or crystallography) or structurally unknown (detectable only as bioactivity signals, mass spectrometry (MS) features, or genome sequence-based predictions). While conceptually convenient, this binary framework severely constrains bioactive compound discovery and limits access to truly unknown scaffolds in nature, because it prioritizes chemical visibility and ignores the genetic accessibility and engineering potential of biosynthetic pathways.

Additionally, ecological pressures tend to favor molecules that provide survival advantages. In parallel, biosynthetic repertoires are often constrained by enzyme specificity and by the availability of suitable metabolic intermediates, which can further compress the set of scaffolds that nature presents. Therefore, the chemical diversity we have observed so far likely represents only a narrow slice of space that is theoretically accessible.

Approximately 80% of the chemical space occupied by published natural products is represented by fewer than 6500 scaffolds [1], underlining that the explored chemical diversity is surprisingly limited. A particularly striking example is cyclic pentapeptides: theoretically, considering the 20 standard proteinogenic amino acids, around 640 000 unique cyclic pentapeptides could be

formed after accounting for rotational symmetry. However, current natural products databases document only 163 cyclic pentapeptides, grouped into 30 structural classes based on amino acid residue compositions [2]. Yet beyond these constraints, the potential 'nature-like' chemical space (i.e. molecules inspired by natural products but extending beyond the limits of natural biosynthesis) remains vast and largely unexplored. For antimicrobial discovery, this compression of scaffold diversity often leads to rediscovery and concentrates antimicrobial candidates into a limited number of existing mechanistic classes. Resistance within those classes can accumulate and reduce their clinical value. This makes expanding into novel chemotypes and modes of action particularly important.

Synthetic biology (SynBio) reshapes discovery by expanding the boundaries of both the 'Known' and the 'Unknown'. Starting from established scaffolds, it employs biosynthetic gene cluster (BGC) engineering and modular recombination to access latent diversity, while *de novo* pathway design and *in vitro* expression systems push further into unexplored chemical space, thereby helping to overcome natural biochemical and ecological constraints. Even small structural changes, such as a dimethylallylation of plantaricin [3], can significantly boost antimicrobial activity, and such changes are easily achieved with *in vitro* biocatalysis. To harness this potential and move beyond traditional constraints, we suggest a new SynBio-centered classification framework defined on two practical dimensions: chemical novelty (compared to existing structures) and ease of engineering (how easily pathways can be manipulated). Throughout this Review, we use SynBio as the primary term, while recognizing its overlap with engineering biology.

The vocabulary of 'Known knowns', 'Known unknowns', and 'Unknown unknowns' was originally popularized by Donald Rumsfeld and later adopted in natural products research, where it was used to distinguish established natural products, genome-inferred cryptic metabolites, and molecules whose existence had not yet been anticipated [4]. This logic was subsequently applied to fungal genome mining and later extended to include 'Unknown knowns' [5], while also clarifying that these knowledge states are distinct from the separate question of whether a BGC is cryptic or silent [6].

Integrating the terminology (known/unknown) and two practical dimensions (chemical novelty/engineering accessibility), we organize natural product discovery into four categories within a 2 × 2 framework (Figure 1a) inspired by the Johari Window (a model of what is known vs. unknown to self and others). In our version, chemical novelty (X-axis) reflects whether a scaffold is based on established structural precedent (known) or lacks clear precedent (unknown). Engineering accessibility (Y-axis) reflects whether a scaffold can be reached and diversified through established SynBio workflows (known) or if it has no initial implementation route (unknown). In the quadrant labels below, the first term refers to chemical novelty (X), and the second to engineering accessibility (Y), that is, (X, Y). These four quadrants serve as a roadmap, highlighting what is known and what is missing for a given scaffold, which SynBio strategy is best suited for that knowledge state, and how it can be shifted along the accessibility and novelty axes (Figure 1b and Table 1).

Four quadrants: natural product definition driven by SynBio

i) Known knowns (chemical novelty known, engineering accessibility known): well-characterized natural product structures that form the foundation of current knowledge and are primarily accessed through pathway refactoring and producer strain optimization.

ii) Unknown knowns (chemical novelty unknown, engineering accessibility known): new derivatives generated from legacy scaffolds using established engineering methods, such as pathway-level reprogramming and enzyme tailoring. While these modifications are technically accessible, their chemical results are not fully predictable from existing structural examples and may include by-products, mixtures, or unexpectedly reconfigured analogs. These chemical changes can also impact biological activity.

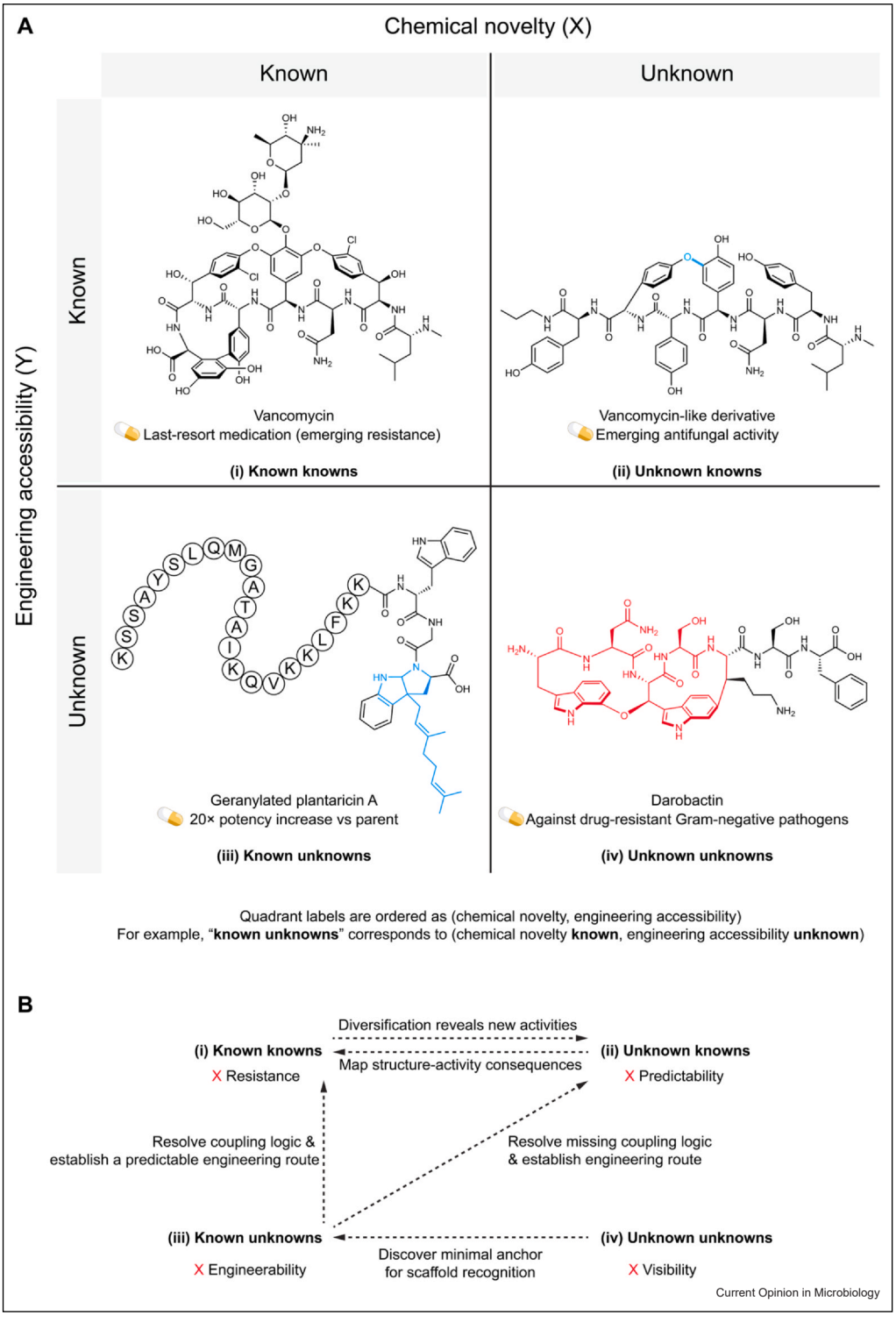
iii) Known unknowns (chemical novelty known, engineering accessibility unknown): molecular architectures that fall within established structural precedent but whose realization remains beyond current engineering practices. Robust workflows for building the required biosynthetic interactions are still lacking. Cross-biosynthetic class integrations are a typical example, and *in vitro* reconstitution can serve as a practical platform for establishing such non-native crosstalk.

iv) Unknown unknowns (chemical novelty unknown, engineering accessibility unknown): molecular architectures that fall outside established structural precedent and lack a defined engineering pathway initially. Such cases are generally approached through discovery rather than design. Only after obtaining minimal chemical or biosynthetic anchors can prioritization, pathway inference, and the development of an engineering route begin.

Known knowns: from last resort to resistance

We use vancomycin (**1**) as an example because it is well understood on both axes: its structure and its mode of targeting lipid II are clearly characterized. It is also

Figure 1



A SynBio-centered four-quadrant framework for natural product discovery and engineering. **(a)** Natural product scaffolds are organized by chemical novelty (X-axis) and engineering accessibility (Y-axis), with representative examples shown for each quadrant. Colored substructures indicate the structural element most relevant to the quadrant assignment of each example. Blue, engineered motif. Red, unprecedented scaffold-defining motif. **(b)** Dynamic relationships among the four quadrants. Dashed arrows indicate transitions between quadrants enabled by discovery or engineering, and red cross marks indicate the bottleneck in each quadrant.

Table 1

Representative examples for the four-quadrant map.

Four quadrants	Examples discussed herein	Additional representative examples
Known knowns	Vancomycin [7]	Erythromycin [47], rapamycin [48,49], and avermectins [50]
Unknown knowns	Engineered vancomycin derivatives [9–12]	Stambomycin- [51] and lacunalide-like derivatives [52]
Known unknowns	Geranylated plantaricin A [3]	Dioxanopeptins [16], rufomycin [17], goadvionin [18], lipoavitide [19], and microvionin [20]
Unknown unknowns	Darobactins [27,28] and dynobactin [29]	Polytheonamides [33], altemicidin [34], NK13650s [53], arsinothricin [54], and bisenarsan [54]

suitable for systematic pathway refactoring and producer optimization. Furthermore, vancomycin illustrates the progression from traditional last-resort antibiotic to clearly defined resistance issues and finally to SynBio-enabled engineering solutions to broaden its antimicrobial spectrum (Figure 2). Approved by the FDA in 1958, vancomycin is a glycopeptide antibiotic with three aromatic crosslinks that form a cup-shaped topology [7]. This structure gives it a high affinity for the D-alanyl-D-alanine dipeptide at the C-terminus of the lipid-linked peptidoglycan pentapeptide (lipid II). This binding prevents the pentapeptide from engaging transglycosylases and D,D-transpeptidases, which are crucial enzymes for cell wall synthesis (Figure 2a) [8].

The natural vancomycin producer *Amycolatopsis orientalis* can use biodiesel-derived crude glycerol (a by-product of biodiesel manufacturing) as a cost-effective carbon source, and optimizing its concentration along with other medium components triples the yield. Non-native producer *A. keratiniphila*, which has strong keratin-degrading ability and industrial robustness, offers a promising platform for large-scale vancomycin production, especially after a CRISPR/Cas9-mediated genome editing system boosted yield by redirecting metabolism away from a competing pathway. Heterologous expression in engineered *Streptomyces* species and *E. coli* further enhances the potential for high-yield, diversified vancomycin synthesis by overcoming the limitations of native hosts.

From the late 1980s into the 1990s, vancomycin-resistant enterococci (VRE) emerged, with VanA-type resistance being the most prevalent, alongside VanB and VanC. The *vanA* gene cluster reprograms lipid II biosynthesis, replacing the pentapeptide terminus from D-alanyl-D-alanine to D-alanyl-D-lactate (Figure 2b). This single-atom change (amide NH → ester O) eliminates a critical hydrogen-bond donor at the lipid II pentapeptide terminus and thereby weakens vancomycin's interaction, reducing its binding affinity ~1000-fold. Clinically, VanA VRE exhibits MICs ≥ 512 μg/mL, versus ~1 μg/mL for susceptible isolates.

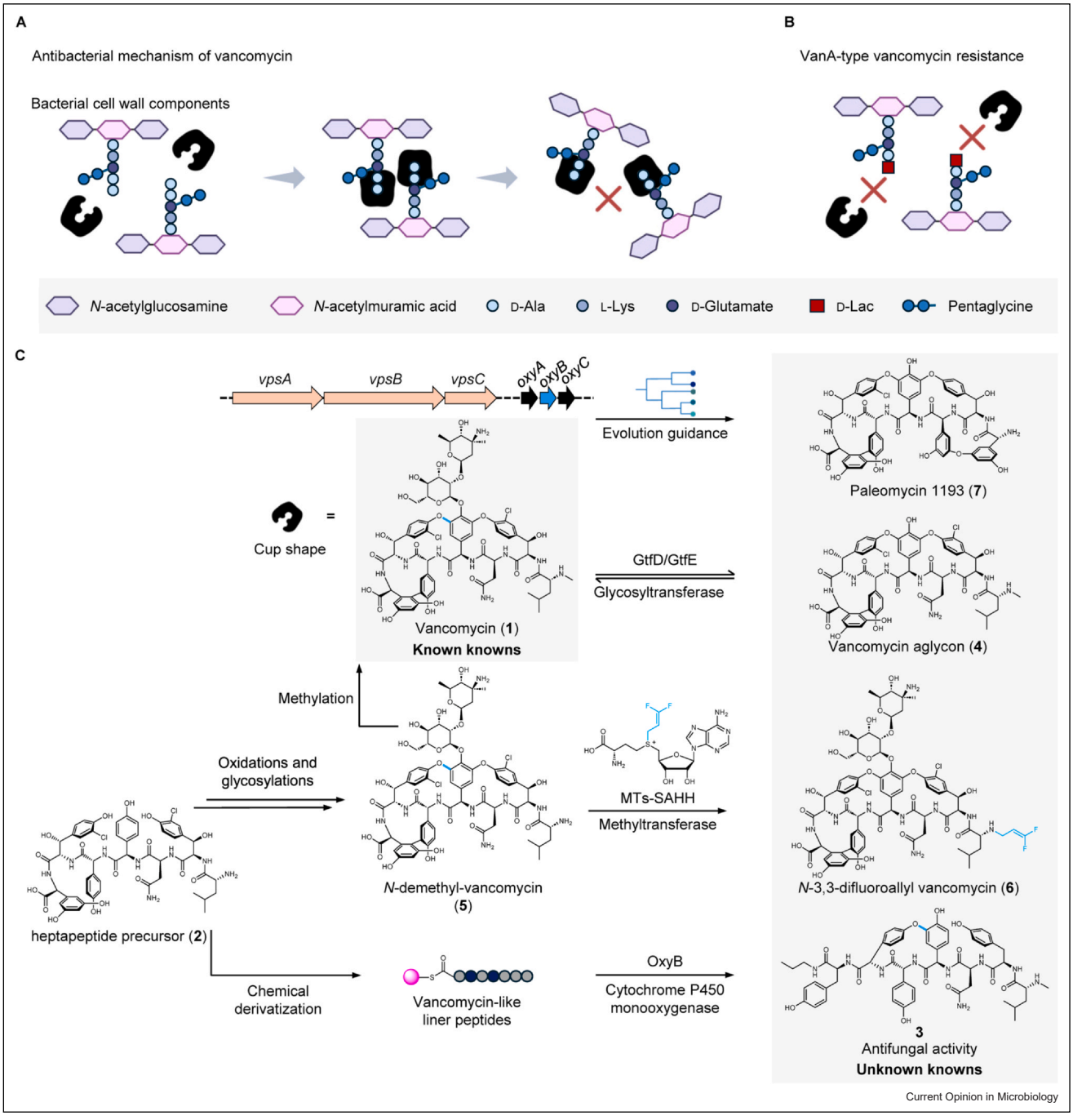
Vancomycin (and many other clinically validated antibiotics) sits in the most mature corner of the four-quadrant roadmap, as its biosynthetic pathway and mode of action are well established, and the pathway is amenable to iterative tuning. In this quadrant, the central task is resistance-driven engineering, namely, adapting a validated scaffold class to regain or extend function.

Unknown knowns: engineerable pathways into uncharted scaffold space

In this quadrant, the starting scaffold and its engineering logic are already well understood, but the chemical and functional outcomes of diversification are not yet predictable. Vancomycin provides a useful example: its glycopeptide biosynthetic logic is sufficiently tractable to support pathway-level reprogramming, enzyme-guided scaffold editing, and late-stage derivatization, thereby enabling systematic exploration beyond the native chemical space while remaining compatible with glycopeptide biosynthesis (Figure 2c).

A primary design lever is to exploit the oxidative crosslinks that shape vancomycin's cup-like pocket for lipid II. The cytochrome P450 monooxygenases OxyABC install diaryl-ether and biaryl linkages that are synthetically challenging yet enzymatically programmable. Systematic probing of OxyB with modified heptapeptide precursors (e.g. **2**) revealed broad substrate tolerance. OxyB installs diverse crosslinks to furnish macrocycles of varying size and composition. Notably, alternative OxyB products (e.g. **3**) showed unexpected antifungal properties against *Saccharomyces cerevisiae* (IC_{50} = 0.7–7.1 μM), while vancomycin displayed IC_{50} > 40 μM under the same conditions, highlighting that rewiring macrocycle connectivity may unlock phenotypes absent from the native scaffold [9]. Glycosyltransferase editing offers an orthogonal handle that can be combined with OxyB-driven macrocyclization. Reversible glycosyltransferases from glycopeptide and aryl C-glycoside pathways catalyze bidirectional sugar-aglycone exchange, enabling rapid construction of >70 differentially glycosylated variants spanning vancomycin and the derivatives thereof (e.g. **4**) [10,11]. Cofactor engineering offers a complementary route for site-defined late-stage edits. An engineered human methionine adenosyltransferase (hMAT2A I322A) produces fluorinated SAM analogs *in situ* from ATP and fluorinated methionine. This supports 3,3-difluoroallylation and fluorobenzylation of

Figure 2



Known knowns and unknown knowns exemplified by vancomycin. **(a)** Antibacterial mechanism of vancomycin through binding to the D-alanyl-D-alanine terminus of lipid II. **(b)** VanA-type resistance, in which the target is remodeled from D-alanyl-D-alanine to D-alanyl-D-lactate, thereby weakening vancomycin binding. **(c)** Representative engineering entry points around the vancomycin scaffold that access unknown knowns space, including evolutionary reconstruction, glycosyltransferase editing, methyltransferase-enabled late-stage derivatization, and OxyB-mediated macrocycle rewiring.

N-demethyl-vancomycin (**5**) with conversions up to 99% in cell-free or cellular formats. This allows precise tuning of electronic and lipophilic properties at defined positions of the scaffold [12].

Evolutionary reconstruction offers an alternative design pathway. Phylogenomics and ancestral sequence reconstruction pinpointed an ancestral lipid-II-targeting glycopeptide, paleomycin (**7**), and the allowable non-

ribosomal peptide synthetase (NRPS) modifications that shaped modern glycopeptides. Rebuilding ancestral-like NRPS systems within a glycopeptide-producing chassis demonstrated that specific backbone changes are functionally tolerated and can produce active glycopeptides *in vivo*. This approach thus makes evolutionary pathways practical design rules for creating new glycopeptide chemotypes [13].

In this quadrant, the practical goal is to use established biosynthetic techniques, such as oxidative crosslink rewiring, glycoediting, or cofactor-enabled late-stage modification, to push validated scaffolds into structurally and functionally unexplored territory. Therefore, the unknown is less about how to engineer the scaffold and more about what new activities may arise once that engineering becomes feasible.

Known unknowns: precedented architectures, lacking engineering routes

Not all precedented chemotypes are equally easy to structurally engineer. Beyond the reprogrammable space of legacy scaffolds, there is a second quadrant called 'Known unknowns', where the target architecture is already validated by natural examples, but SynBio routes for diversifying are still missing. These gaps often stem from unresolved biosynthetic crosstalk that hasn't yet been translated into engineering principles. This is because, in such systems: 1) the enzyme responsible for the coupling step might be only partly identified; 2) intermediate recognition and transfer between different enzymatic systems can be highly dependent on the context; and 3) the timing of the coupling step relative to scaffold assembly is often unclear.

These bottlenecks most often occur in lower-frequency biosynthetic hybrids (NRPS–terpene [14,15], NRPS–polyunsaturated fatty acid synthase (PUFAS) [16], NRPS–ribosomally synthesized and post-translationally modified peptide (RiPP) [17], RiPP–fatty acid synthase (FAS) [18,19], and RiPP–polyketide synthase (PKS)/FAS [20]) rather than in more common hybrid classes like PKS–NRPS [21–24] and PKS–terpene [25,26] systems, which have abundant natural examples and better-established engineering platforms. At present, experimentally tractable examples in this quadrant come mainly from peptide-related hybrid systems, so the cases below are intended to be illustrative rather than comprehensive.

A practical consequence is that progress in this quadrant is unlikely to begin with full-pathway reconstruction. Instead, complex architectures will first need to be reduced to tractable coupling problems that can be tested outside the full native pathway. For example, in late-stage transformations, *in vitro* reconstitution provides a practical entry point to evaluate donor scope, acceptor tolerance, and coupling selectivity before broader pathway integration is attempted. NRPS–terpene hybrids are still rare in nature, with flavunoidine and CJ-12662 among the few examples [14,15]. A practical approach for engineering is not to rebuild the entire upstream pathway initially but to focus on the late-stage terpene–peptide coupling step as the key transferable unit for diversification. In this context, RiPP–terpene systems serve as a useful conceptual model. Although RiPP and NRPS pathways differ fundamentally in their upstream assembly processes, both produce peptide-like acceptors that can be diversified through late-stage tailoring chemistry. Therefore, *in vitro* coupling assays provide a practical platform for testing whether terpene-appending modules can be repurposed across different biosynthetic systems.

A useful operational example comes from RiPP prenyltransferases, which can function as accessible late-stage terpene-installation modules. Cyanobactin prenyltransferases are notable for broad peptide-sequence tolerance, and the distinct prenyltransferase PalQ further illustrates that extended prenyl donors can be introduced onto Trp-containing peptide substrates. The geranylated plantaricin A (**9**) displayed potent antimicrobial activity with an IC_{50} value of 6.3 nM, an ~18-fold enhancement compared to the unmodified peptide (IC_{50} = 114.7 nM) (Figure 3b) [3].

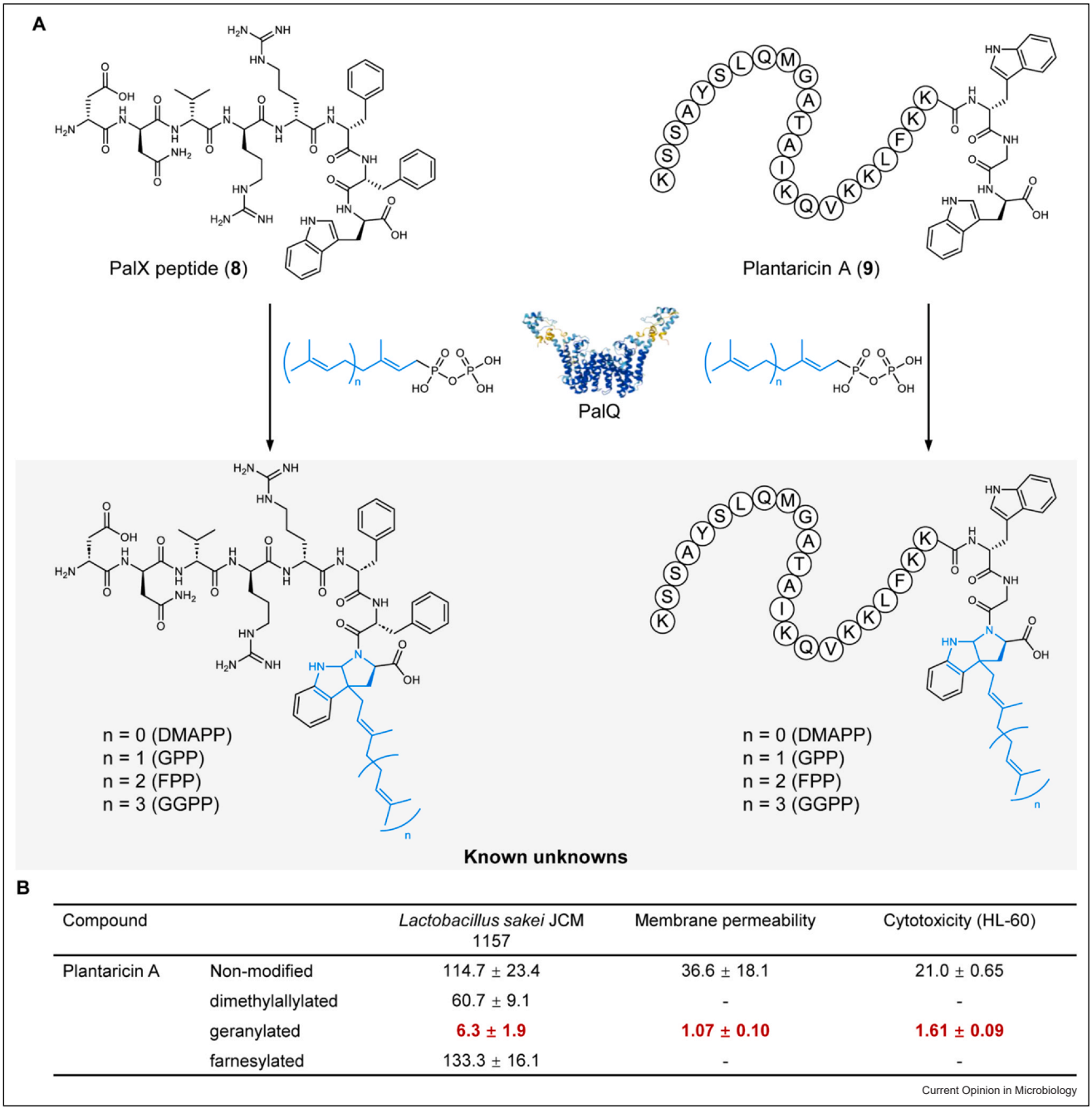
Taken together, known unknowns are chemotypes that already exist in nature but are hard to diversify systematically. The main obstacle is usually not scaffold validation, but the lack of a practical engineering handle: the key coupling enzyme, the appropriate intermediate, or the timing of the coupling step remains unclear. As a result, shifting beyond this quadrant will likely begin with simplified reconstitution systems that isolate one coupling event at a time, as shown by late-stage terpene installation in peptides using substrate-tolerant prenyltransferases. These lessons may later help enable the engineering of rarer hybrid architectures.

Unknown unknowns: scaffolds beyond current prediction

'Known unknowns' are mostly limited by how they are implemented, while 'Unknown unknowns' are mainly limited by how easy they are to discover, since the underlying architectures evade prior expectations and must first be identified before they can be engineered. Operationally, the route into this quadrant is discovery-first: minimal anchors, like a gene-level handle, a resistance clue, or a mechanistic foothold, are used to identify the scaffold and then convert it into an engineerable program.

Darobactins and dynobactins exemplify this quadrant with an unexpected RiPP bond topology. As a natural example of a completely new architecture, darobactin A (**10**) is a bicyclic RiPP heptapeptide with unusual C–C

Figure 3



B

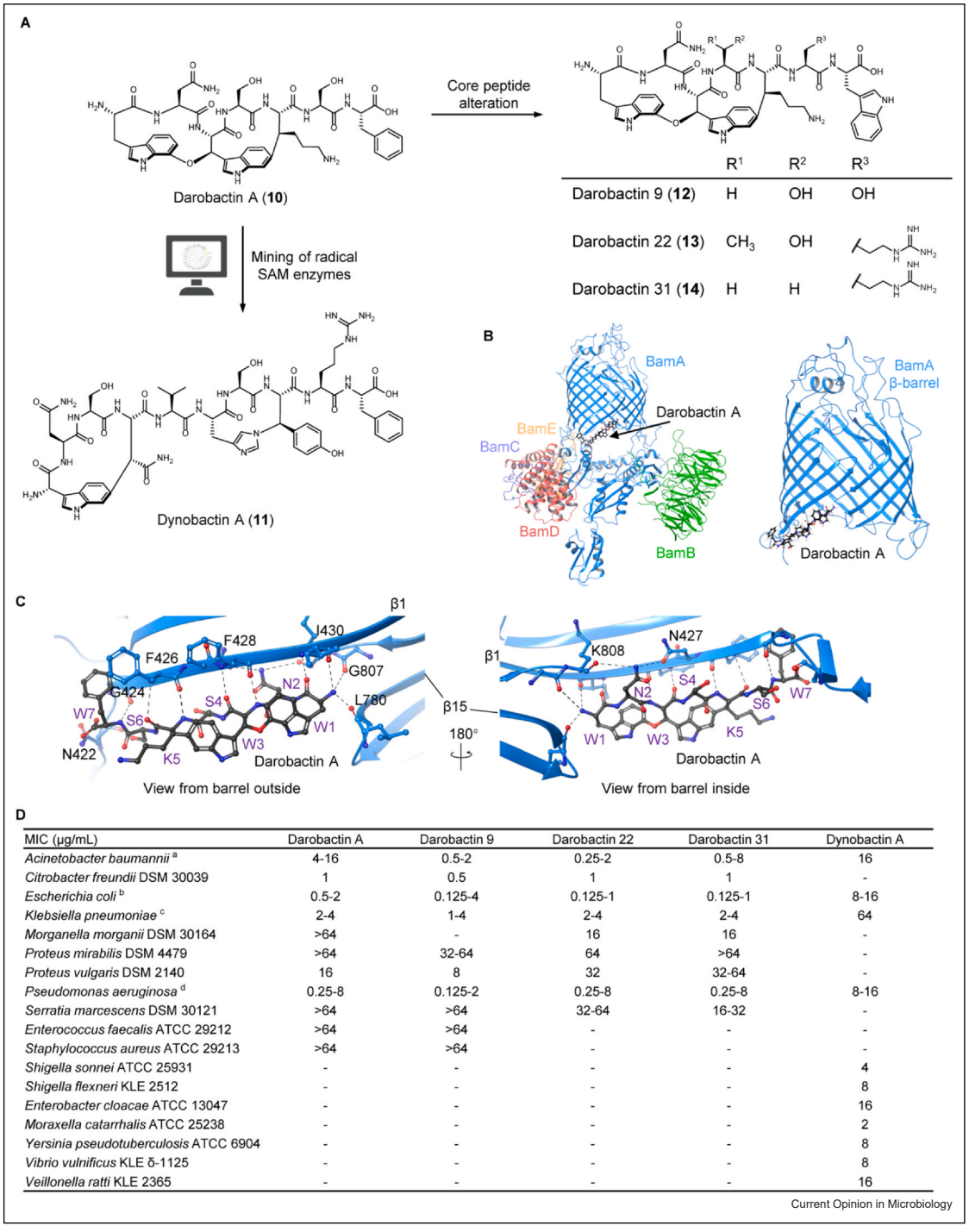
Compound		*Lactobacillus sakei* JCM 1157	Membrane permeability	Cytotoxicity (HL-60)
Plantaricin A	Non-modified	114.7 ± 23.4	36.6 ± 18.1	21.0 ± 0.65
	dimethylallylated	60.7 ± 9.1	-	-
	geranylated	**6.3 ± 1.9**	**1.07 ± 0.10**	**1.61 ± 0.09**
	farnesylated	133.3 ± 16.1	-	-

Current Opinion in Microbiology

Known unknowns exemplified by PalQ-catalyzed late-stage terpene installation in peptides. **(a)** PalQ-catalyzed prenylation of the C-terminal tryptophan in PalX and plantaricin A, highlighting donor scope from DMAPP to GGPP. **(b)** Antimicrobial activity comparison of unmodified and prenylated plantaricin A derivatives. Values are reported as IC_{50} values for antimicrobial activity (nM) and cytotoxicity (μM), and as EC_{50} values for membrane permeability (μM).

and ether crosslinks (Figure 4) [27,28]. This bond topology would have been hard to predict based only on traditional RiPP biosynthetic logic. The small *dar* BGC was not found using typical BGC annotation tools (e.g. antiSMASH) and was instead discovered through a targeted search of the core peptide sequence. The bicyclic crosslink is necessary for antibacterial activity and BamA binding. BamA is a target on the outer membrane of Gram-negative bacteria and is rarely addressed by natural product antibiotics.

Figure 4



MIC (μg/mL)	Darobactin A	Darobactin 9	Darobactin 22	Darobactin 31	Dynobactin A
Acinetobacter baumannii [a]	4-16	0.5-2	0.25-2	0.5-8	16
Citrobacter freundii DSM 30039	1	0.5	1	1	-
Escherichia coli [b]	0.5-2	0.125-4	0.125-1	0.125-1	8-16
Klebsiella pneumoniae [c]	2-4	1-4	2-4	2-4	64
Morganella morganii DSM 30164	>64	-	16	16	-
Proteus mirabilis DSM 4479	>64	32-64	64	>64	-
Proteus vulgaris DSM 2140	16	8	32	32-64	-
Pseudomonas aeruginosa [d]	0.25-8	0.125-2	0.25-8	0.25-8	8-16
Serratia marcescens DSM 30121	>64	>64	32-64	16-32	-
Enterococcus faecalis ATCC 29212	>64	>64	-	-	-
Staphylococcus aureus ATCC 29213	>64	>64	-	-	-
Shigella sonnei ATCC 25931	-	-	-	-	4
Shigella flexneri KLE 2512	-	-	-	-	8
Enterobacter cloacae ATCC 13047	-	-	-	-	16
Moraxella catarrhalis ATCC 25238	-	-	-	-	2
Yersinia pseudotuberculosis ATCC 6904	-	-	-	-	8
Vibrio vulnificus KLE δ-1125	-	-	-	-	8
Veillonella ratti KLE 2365	-	-	-	-	16

Unknown unknowns exemplified by darobactins and dynobactin. **(a)** Chemical structures of engineered darobactin analogs and dynobactin A. **(b)** Cryo-EM structure of the BAM complex bound to darobactin A. Darobactin A binds at the lateral gate of BamA and adopts a β-strand-like conformation. **(c)** Close-up view of the hydrogen-bonding interactions between darobactin A and BamA strand β1. **(d)** MICs of darobactins and dynobactin A against a panel of pathogens. [a]DSM 30007, DSM 30008, and NCTC 13301 were tested for darobactins; ATCC 19606 was tested for dynobactin A. [b]ATCC 25922, BW25113, JW0451-2 Δ*acrB*, and K12 Δ*tolC* were tested for darobactins; K-12 MG1655 and AR350 were tested for dynobactin A. [c]DSM 30104 was tested for darobactins; AR347 was tested for dynobactin A. [d]PAO1, PA14, PA14 Δ*mexAB*, DSM 1117, and NCTC 13437 were tested for darobactins; PAO1 and PA14 were tested for dynobactin A.

Expanding on the family, a computational search for genes only distantly related to the *dar* operon led to the discovery of dynobactin A (**11**) from *Photorhabdus australis*, another novel peptide antibiotic with two unlinked rings [29]. Dynobactins also target BamA, but their antibacterial activity is more dependent on outer-membrane permeability. Guided by the cryo-EM structures of the BamA–darobactin complex, which provided an interaction blueprint, core-peptide diversification rapidly converted darobactins from a one-off discovery into an engineerable series, yielding analogs with enhanced potency and broader Gram-negative coverage, such as darobactins 9 (**12**), *22* (**13**), and 31 (**14**) [30–32]. This progression highlights a key feature of the quadrant: once an unexpected scaffold is made visible through a minimal anchor, structural and mechanistic insights can swiftly turn it from a hidden discovery into an engineerable series.

Other antibacterial examples reinforce the same principle. Polytheonamides [33] and altemicidins [34] emerged from biosynthetic pathways that would have been difficult to predict from traditional paradigms, and only after elucidation of their pathways did related biosynthetic systems become recognizable more broadly across bacteria [35–37].

In this quadrant, the primary bottleneck is discoverability: the relevant structures and biosynthetic rules fall outside current predictive frameworks, making them difficult to predict or capture with rule-based genome mining tools. Once a minimal anchor is obtained, these initially opaque discoveries can be translated into engineerable programs, enabling both systematic diversification and more reliable future recognition of related chemotypes.

Conclusion and future perspectives

SynBio is transforming the discovery of antimicrobial natural products by broadening both what we can see and what we can build. Viewed through the two axes of chemical novelty and engineering accessibility, the four quadrants outlined here offer a practical map of current discovery and engineering states: legacy scaffolds that can be refined repeatedly (known knowns), engineerable scaffolds with uncertain results (unknown knowns), validated architectures that are still challenging to implement (known unknowns), and unexpected chemotypes that must be discovered before design can begin (unknown unknowns). This framework is not meant to be a static classification but a tool to link discovery logic with engineering strategy and to show how initially isolated discoveries can be converted into tractable SynBio platforms.

An important limitation of the framework is that quadrant boundaries are not always clear-cut (Figure 1b). In practice, the distinction between unknown knowns and known unknowns may blur as engineering toolkits mature, especially when a previously context-dependent transformation becomes accessible. The key question is therefore not whether a compound permanently belongs to one quadrant, but which bottleneck dominates at a given time: uncertain outcome, uncertain implementation, or uncertain discoverability. Quadrant assignments are thus time-stamped and dynamic. An unknown unknown can become a known unknown once a scaffold is recognized. A known unknown can shift toward unknown knowns or known knowns, once the relevant coupling logic is mechanistically understood and a transferable engineering route is established. An unknown known can eventually become a known known as its chemical outcome becomes predictable and reproducible.

Looking forward, artificial intelligence and machine learning (AI/ML)-assisted pathway discovery and design are likely to reshape both axes of this map [38,39]. On the discoverability front [40,41], learning-based models can move beyond rule-based genome mining by recognizing atypical precursor peptides (e.g. DeepRiPP [42], TrRiPP [43], and NeuRiPP [44]), noncanonical tailoring logic, weakly conserved pathway architectures, and metabolomic patterns that would otherwise remain invisible, thereby expanding access to unknown unknowns. On the engineering side [45,46], AI/ML may increasingly support substrate prediction, enzyme-function assignment, structure-guided mutagenesis, pathway refactoring, and design-build-test-learn (DBTL) prioritization. Thus, these could reduce the distance from known unknowns to engineerable routes, and from unknown knowns to predictable outcome spaces. In this way, AI/ML should be seen not just as a discovery tool but as a bridge connecting hidden chemistry to actionable SynBio programs.

Another priority will be to make both axes more practical by adding semi-quantitative indicators. Chemical novelty can be estimated by distance to known scaffolds (e.g. Tanimoto similarity of fingerprints) and by MS/MS fragmentation patterns that do not match those in spectral libraries. Engineering accessibility is harder to standardize across hosts and pathway classes. Still, it can be estimated by the number of DBTL iterations or engineering steps needed to produce the target product, the achievable titer and yield under similar conditions, and modular compatibility after swapping.

Viewed through this lens, the framework serves best as a guide to the next step: identifying which scaffolds need discovery, which require route-building, and which are ready for systematic diversification. As these transitions become more predictable, the search for first-in-class antimicrobials should expand and become more intentional.

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: Yi-Ming Shi reports that financial support was provided by the National Key R&D Program of China. Yi-Ming Shi reports that financial support was provided by the National Natural Science Foundation of China. Helge B. Bode reports that financial support was provided by the Max Planck Society. The other authers 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

This work was supported by the National Key R&D Program of China (Grant no. 2024YFA0917000 to YMS), the National Natural Science Foundation of China (Grant no. 32470066 to YMS), the Chinese Academy of Sciences President's International Fellowship Initiative (PIFI, grant no. 2025PY0007), and the Max Planck Society (to HBB). The collaboration and discussions underlying this work benefited from the Chinese Academy of Sciences 2025 Youth International Academic Salon ("Sino-German Youth Academic Salon in Synthetic Biology").

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. Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG: **Retrospective analysis of natural products provides insights for future discovery trends**. *Proc Natl Acad Sci USA* 2017, **114**:5601-5606.

2. •• Nainala VC, Rajan K, Kanakam SRS, Sharma N, Weißenborn V, Schaub J, Steinbeck C: **COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database**. *Nucleic Acids Res* 2024, **53**:D634-D643.
This paper presents a large, curated open database of nearly 700 000 natural products with standardized structures, taxonomy, and metadata accessible via web interface and API. It has quickly become a key reference resource for dereplication, cheminformatics, and training data-driven or AI models for natural product discovery.

3. • Ozawa H, Miyata A, Hayashi S, Miyoshi N, Kato K, Ito S, Fujinami D: **Expanding the chemical space of antimicrobial peptides via enzymatic prenylation**. *J Am Chem Soc* 2025, **147**:25642-25651.
This study shows that enzymatic prenylation can increase antimicrobial peptide potency by up to 18-fold by strengthening peptide–membrane interactions and promoting membrane destabilization. It also establishes the prenyltransferase PalQ as an engineerable platform for site-selective lipidation of diverse peptides, greatly expanding the accessible chemical space for antimicrobial peptide engineering.

4. Seyedsayamdost MR, Clardy J: **Natural products and synthetic biology**. *ACS Synth Biol* 2014, **3**:745-747.

5. Gilchrist CLM, Li H, Chooi YH: **Panning for gold in mould: can we increase the odds for fungal genome mining?** *Org Biomol Chem* 2018, **16**:1620-1626.

6. Hoskisson Paul A, Seipke Ryan F: **Cryptic or silent? The known unknowns, unknown knowns, and unknown unknowns of secondary metabolism**. *mBio* 2020, **11**:e02642-20.

7. McCormick MH, McGuire JM, Pittenger GE, Pittenger RC, Stark WM: **Vancomycin, a new antibiotic. I. Chemical and biologic properties**. *Antibiot Annu* 1955, **3**:606-611.

8. Taku A, Hirsch TM, Fan DP: **Dissociation and reconstitution of membranes synthesizing the peptidoglycan of *Escherichia coli*. Lipid dependence of the synthetic enzymes**. *J Biol Chem* 1980, **255**:2848-2854.

9. Forneris CC, Nguy AKL, Seyedsayamdost MR: **Mapping and exploiting the promiscuity of OxyB toward the biocatalytic production of vancomycin aglycone variants**. *ACS Catal* 2020, **10**:9287-9298.

10. Fu X, Albermann C, Jiang J, Liao J, Zhang C, Thorson JS: **Antibiotic optimization via in vitro glycorandomization**. *Nat Biotechnol* 2003, **21**:1467-1469.

11. Zhang C, Griffith BR, Fu Q, Albermann C, Fu X, Lee I-K, Li L, Thorson JS: **Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions**. *Science* 2006, **313**:1291-1294.

12. Ding W, Zhou M, Li H, Li M, Qiu Y, Yin Y, Pan L, Yang W, Du Y, Zhang X, *et al*.: **Biocatalytic fluoroalkylation using fluorinated S-adenosyl-l-methionine cofactors**. *Org Lett* 2023, **25**:5650-5655.

13. • Hansen MH, Adamek M, Iftime D, Petras D, Schuseil F, Grond S, Stegmann E, Cryle MJ, Ziemert N: **Resurrecting ancestral antibiotics: unveiling the origins of modern lipid II targeting glycopeptides**. *Nat Commun* 2023, **14**:7842.
This paper uses ancestral sequence reconstruction to identify and reconstitute an ancestral tetracyclic glycopeptide, paleomycin, with lipid II–targeting activity, and shows how modern vancomycin-type antibiotics evolved from it, providing design rules for engineering new glycopeptide antibiotics.

14. Yee DA, Kakule TB, Cheng W, Chen M, Chong CTY, Hai Y, Hang LF, Hung YS, Liu N, Ohashi M, *et al*.: **Genome mining of alkaloidal terpenoids from a hybrid terpene and nonribosomal peptide biosynthetic pathway**. *J Am Chem Soc* 2020, **142**:710-714.

15. Cheng W, Chen M, Ohashi M, Tang Y: **Biosynthesis of terpenoid-pyrrolobenzoxazine hybrid natural product CJ-12662**. *Angew Chem Int Ed* 2022, **61**:e202116928.

16. Chen Y, Chen Y, Xiang H, Luo C, Duan J, Hu K, Zheng X, Liu J, Xue Y, Shi YM: **Lipid-donor-anchored genome mining uncovers dioxanopeptins, antibacterial lipopeptides with a 1,3-dioxane functionalized polyunsaturated lipid tail**. *Chem Sci* 2026, https://doi.org/10.1039/d6sc00003g

17. Padva L, Zimmer L, Gullick J, Zhao Y, Sasi VM, Schittenhelm RB, Jackson CJ, Cryle MJ, Crüsemann M: **Ribosomal pentapeptide nitration for non-ribosomal peptide antibiotic precursor biosynthesis**. *Chem* 2025, **11**:102438.

18. Kozakai R, Ono T, Hoshino S, Takahashi H, Katsuyama Y, Sugai Y, Ozaki T, Teramoto K, Teramoto K, Tanaka K, *et al*.: **Acyltransferase that catalyses the condensation of polyketide and peptide moieties of goadvionin hybrid lipopeptides**. *Nat Chem* 2020, **12**:869-877.

19. Ren H, Huang C, Pan Y, Dommaraju SR, Cui H, Li M, Gadgil MG, Mitchell DA, Zhao H: **Non-modular fatty acid synthases yield distinct N-terminal acylation in ribosomal peptides**. *Nat Chem* 2024, **16**:1320-1329.

20. Wiebach V, Mainz A, Siegert MJ, Jungmann NA, Lesquame G, Tirat S, Dreux-Zigha A, Aszodi J, Le Beller D, Süssmuth RD: **The anti-staphylococcal lipolanthines are ribosomally synthesized lipopeptides**. *Nat Chem Biol* 2018, **14**:652-654.

21. •• Präve L, Kuttenlochner W, Tabak WWA, Langer C, Kaiser M, Groll M, Bode HB: **Bioengineering of syrbactin megasynthetases for immunoproteasome inhibitor production**. *Chem* 2024, **10**:3212-3223.
This study uses the XUT strategy to stepwise reassemble a PKS–NRPS hybrid and generate a rationally modified syrbactin derivative with immunoproteasome-inhibitory activity. It is important because it shows that even highly complex megasynthetases can be engineered in a modular way to access tailored bioactive products.

22. •• 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 establishes a broadly applicable strategy for modular PKS engineering by identifying two conserved motifs as robust cut sites for assembling functional hybrid pathways. It is important because it turns PKS recombination into a more general and transferable design principle, enabling custom production of polyketides across cis-AT PKS, trans-AT PKS, and even PKS–NRPS systems.

23. Awakawa T, Fujioka T, Zhang L, Hoshino S, Hu Z, Hashimoto J, Kozone I, Ikeda H, Shin-Ya K, Liu W, *et al.*: **Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution**. *Nat Commun* 2018, **9**:3534.

24. Zhang L, Wang C, Chen K, Zhong W, Xu Y, Molnár I: **Engineering the biosynthesis of fungal nonribosomal peptides**. *Nat Prod Rep* 2023, **40**:62-88.

25. Itoh T, Tokunaga K, Matsuda Y, Fujii I, Abe I, Ebizuka Y, Kushiro T: **Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases**. *Nat Chem* 2010, **2**:858-864.

26. Quan Z, Awakawa T: **Recent developments in the engineered biosynthesis of fungal meroterpenoids**. *Beilstein J Org Chem* 2024, **20**:578-588.

27. Imai Y, Meyer KJ, Iinishi A, Favre-Godal Q, Green R, Manuse S, Caboni M, Mori M, Niles S, Ghiglieri M, *et al.*: **A new antibiotic selectively kills Gram-negative pathogens**. *Nature* 2019, **576**:459-464.

28. Kaur H, Jakob RP, Marzinek JK, Green R, Imai Y, Bolla JR, Agustoni E, Robinson CV, Bond PJ, Lewis K, *et al.*: **The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase**. *Nature* 2021, **593**:125-129.

29. Miller RD, Iinishi A, Modaresi SM, Yoo BK, Curtis TD, Lariviere PJ, Liang L, Son S, Nicolau S, Bargabos R, *et al.*: **Computational identification of a systemic antibiotic for Gram-negative bacteria**. *Nat Microbiol* 2022, **7**:1661-1672.

30. Groß S, Panter F, Pogorevc D, Seyfert CE, Deckarm S, Bader CD, Herrmann J, Müller R: **Improved broad-spectrum antibiotics against Gram-negative pathogens via darobactin biosynthetic pathway engineering**. *Chem Sci* 2021, **12**:11882-11893.

31. Seyfert CE, Porten C, Yuan B, Deckarm S, Panter F, Bader CD, Coetzee J, Deschner F, Tehrani K, Higgins PG, *et al.*: **Darobactins exhibiting superior antibiotic activity by Cryo-EM structure guided biosynthetic engineering**. *Angew Chem Int Ed* 2023, **62**:e202214094.

32. Kany AM, Fries F, Seyfert CE, Porten C, Deckarm S, Chacón Ortiz M, Dubarry N, Vaddi S, Große M, Bernecker S, *et al.*: **In vivo activity profiling of biosynthetic darobactin D22 against critical gram-negative pathogens**. *ACS Infect Dis* 2024, **10**:4337-4346.

33. Hamada T, Matsunaga S, Yano G, Fusetani N: **Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, *Theonella swinhoei***. *J Am Chem Soc* 2005, **127**:110-118.

34. Takahashi A, Kurasawa S, Ikeda D, Okami Y, Takeuchi T: **Altemicidin, a new acaricidal and antitumor substance. I. Taxonomy, fermentation, isolation and physico-chemical and biological properties**. *J Antibiot* 1989, **42**:1556-1561.

35. Paoli L, Ruscheweyh HJ, Forneris CC, Hubrich F, Kautsar S, Bhushan A, Lotti A, Clayssen Q, Salazar G, Milanese A, *et al.*: **Biosynthetic potential of the global ocean microbiome**. *Nature* 2022, **607**:111-118.

36. Rust M, Helfrich EJN, Freeman MF, Nanudorn P, Field CM, Rückert C, Kündig T, Page MJ, Webb VL, Kalinowski J, *et al.*: **A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli***. *Proc Natl Acad Sci USA* 2020, **117**:9508-9518.

37. Barra L, Awakawa T, Shirai K, Hu Z, Bashiri G, Abe I: **β-NAD as a building block in natural product biosynthesis**. *Nature* 2021, **600**:754-758.

38. •• Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, Martin NI, Meijer D, Terlouw BR, Biermann F, Blin K, *et al.*: **Artificial intelligence for natural product drug discovery**. *Nat Rev Drug Discov* 2023, **22**:895-916.

This review outlines how artificial intelligence can accelerate natural product drug discovery by integrating omics-based mining, structure elucidation, bioactivity prediction, and *de novo* design. It is important because it defines a roadmap for linking natural product discovery with modern AI-driven drug design, while also highlighting the central need for large, high-quality datasets.

39. Meijer D, Beniddir MA, Coley CW, Mejri YM, Öztürk M, van der Hooft JJJ, Medema MH, Skiredj A: **Empowering natural product science with AI: leveraging multimodal data and knowledge graphs**. *Nat Prod Rep* 2025, **42**:654-662.

40. Medema MH, de Rond T, Moore BS: **Mining genomes to illuminate the specialized chemistry of life**. *Nat Rev Genet* 2021, **22**:553-571.

41. •• Zhang Y, Zong J, Liu Y, Zhou K, Shi H, Yin W-B, Liu L, Chen Y: **Progress on targeted discovery of microbial natural products based on the predictions of both structure and activity**. *Nat Prod Rep* 2025, **42**:1651-1663.

This highlight review summarizes recent methods that move beyond purely activity-guided or structure-guided discovery by jointly predicting both the structure and the bioactivity of microbial natural products. It is important because it frames a new trend toward more targeted and efficient discovery workflows that better connect genome mining with functional prioritization.

42. Merwin NJ, Mousa WK, Dejong CA, Skinnider MA, Cannon MJ, Li H, Dial K, Gunabalasingam M, Johnston C, Magarvey NA: **DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products**. *Proc Natl Acad Sci USA* 2020, **117**:371-380.

43. Gao Y, Zhong Z, Zhang D, Zhang J, Li YX: **Exploring the roles of ribosomal peptides in prokaryote-phage interactions through deep learning-enabled metagenome mining**. *Microbiome* 2024, **12**:94.

44. de Los Santos ELC: **NeuRiPP: neural network identification of RiPP precursor peptides**. *Sci Rep* 2019, **9**:13406.

45. Rodrigues T, Reker D, Schneider P, Schneider G: **Counting on natural products for drug design**. *Nat Chem* 2016, **8**:531-541.

46. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583-589.

47. Wiley PF, Gerzon K, Flynn EH, Sigal MV Jr., Weaver O, Quarck UC, Chauvette RR, Monahan R: **Erythromycin. X. Structure of erythromycin**. *J Am Chem Soc* 1957, **79**:6062-6070.

48. Vézina C, Kudelski A, Sehgal SN: **Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle**. *J Antibiot* 1975, **28**:721-726.

49. Sehgal SN, Baker H, Vézina C: **Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization**. *J Antibiot* 1975, **28**:727-732.

50. Burg RW, Miller BM, Baker EE, Birnbaum J, Currie SA, Hartman R, Kong YL, Monaghan RL, Olson G, Putter I, *et al.*: **Avermectins, new family of potent anthelmintic agents: producing organism and fermentation**. *Antimicrob Agents Chemother* 1979, **15**:361-367.

51. Su L, Hôtel L, Paris C, Chepkirui C, Brachmann AO, Piel J, Jacob C, Aigle B, Weissman KJ: **Engineering the stambomycin modular polyketide synthase yields 37-membered mini-stambomycins**. *Nat Commun* 2022, **13**:515.

52. •• Mabesoone MFJ, Leopold-Messer S, Minas HA, Chepkirui C, Chawengrum P, Reiter S, Meoded RA, Wolf S, Genz F, Magnus N, *et al.*: **Evolution-guided engineering of *trans*-acyltransferase polyketide synthases**. *Science* 2024, **383**:1312-1317.

This paper identifies a conserved motif as a fusion site that defines exchange units in trans-AT PKS and enables the construction of functional chimeric or truncated assembly lines. It establishes a general design rule for modular engineering of trans-AT PKS.

53. Yee DA, Niwa K, Perlatti B, Chen M, Li Y, Tang Y: **Genome mining for unknown–unknown natural products**. *Nat Chem Biol* 2023, **19**:633-640.

54. Hoshino S, Ijichi S, Asamizu S, Onaka H: **Insights into arsenic secondary metabolism in actinomycetes from the structure and biosynthesis of bisenarsan**. *J Am Chem Soc* 2023, **145**:17863-17871.