



Cite this: *Nat. Prod. Rep.*, 2025, 42, 1489

## Engineering modular enzyme assembly: synthetic interface strategies for natural products biosynthesis applications

Gahyeon Kim,<sup>†a</sup> Dukwon Lee,<sup>†a</sup> Ji Hun Kim,<sup>†a</sup> Seong Do Kim,<sup>†b</sup> Hongki Kim,<sup>†b</sup> Jae Heon Kim,<sup>†b</sup> Sung Sun Yim,<sup>ab</sup> Soo-Jin Yeom,<sup>c</sup> Jay D. Keasling<sup>def</sup> and Byung-Kwan Cho<sup>ib\*ab</sup>

Covering: 2020 to 2025

Natural products remain indispensable sources of therapeutic and bioactive compounds, yet traditional discovery strategies are constrained by compound rediscovery. Modular biosynthetic enzymes, such as type I polyketide synthases (PKSs) and type A non-ribosomal peptide synthetases (NRPSs), offer promising platforms for combinatorial biosynthesis owing to their programmable architectures. However, practical implementation is frequently limited by inter-modular incompatibility and domain-specific interactions. This review highlights recent advances in modular enzyme assembly enabled by synthetic interfaces—including cognate docking domains, synthetic coiled-coils, SpyTag/SpyCatcher, and split inteins—which function as orthogonal, standardized connectors to facilitate post-translational complex formation. These interfaces support rational investigations into substrate specificity, module compatibility, and pathway derivatization as well as general enzyme clustering applications beyond PKS and NRPS systems. Synthetic interfaces can be integrated with computational tools to support a more systematic and scalable framework for modular enzyme engineering by providing predictive insights into domain compatibility and interface design. These approaches within iterative design-build-test-learn workflows can accelerate the programmable assembly of biosynthetic systems and expand the accessible chemical space for natural products.

Received 11th April 2025

DOI: 10.1039/d5np00027k

rsc.li/npr

- |   |  |
|---|--|
| <ol style="list-style-type: none"> <li>1. Introduction</li> <li>2. Fundamentals of PKS and NRPS architecture</li> <li>3. Modular assembly and engineering of PKS and NRPS</li> <li>3.1 Evolutionary functional boundaries of PKS and NRPS modules</li> <li>3.2 Engineering with cognate interfaces: orthologous interaction modules</li> <li>3.2.1 OIM-based engraftment strategies</li> <li>3.2.2 Expanding the utility of OIMs</li> </ol> | <ol style="list-style-type: none"> <li>3.3 Beyond orthologous interfaces: xenologous interaction modules in synthetic biology</li> <li>3.3.1 Synthetic coiled-coils</li> <li>3.3.2 SpyTag/SpyCatcher system</li> <li>3.3.3 Split inteins</li> <li>3.4 Functional applications of synthetic interfaces in modular enzyme engineering</li> <li>4. <i>In silico</i> retrobiosynthesis design for modular natural product pathways</li> <li>4.1 Retrobiosynthesis prediction toolkits for PKS/NRPS</li> <li>4.2 Retrobiosynthesis prediction toolkits for NRPS</li> <li>5. Conclusion</li> <li>6. Data availability</li> <li>7. Author contributions</li> <li>8. Conflicts of interest</li> <li>9. Acknowledgements</li> <li>10. References</li> </ol> |
|---|--|

<sup>a</sup>Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Republic of Korea. E-mail: bcho@kaist.ac.kr

<sup>b</sup>Graduate School of Engineering Biology, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Republic of Korea

<sup>c</sup>School of Biological Sciences and Technology, Chonnam National University, Gwangju, 61186, Republic of Korea

<sup>d</sup>Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA, USA

<sup>e</sup>Joint BioEnergy Institute, Emeryville, CA, USA

<sup>f</sup>Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>†</sup> These authors contributed equally.



## 1. Introduction

Natural products derived from microbial secondary metabolites have provided a rich reservoir of bioactive compounds, significantly contributing to pharmaceutical, agricultural, and biotechnological applications. Despite their immense potential, traditional natural product discovery frequently faces the challenge of compound rediscovery, often yielding known metabolites and limiting the identification of novel structures.<sup>1</sup> Moreover, the escalating resistance issues related to antibiotics, herbicides, and other bioactive agents have created a pressing need for diverse derivatives or structurally novel bioactive compounds.<sup>2–4</sup> Consequently, innovative methodologies are urgently required to efficiently access new and diverse chemical spaces. One promising solution to this issue lies in modular enzyme-based combinatorial biosynthesis. Modular enzymes, such as polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), exhibit remarkable structural flexibility,

theoretically enabling the incorporation of various functional groups into target molecules without significant structural or functional limitations.<sup>5</sup>

However, despite this potential, practical laboratory implementation faces several technical hurdles, including module incompatibility and limitations imposed by intrinsic enzymatic specificity, exemplified by the restrictive selectivity of ketosynthase gatekeeper domains.<sup>6</sup> Moreover, another critical challenge arises from truncated mRNA transcripts due to the substantial size of biosynthetic gene clusters (BGCs).<sup>7</sup> To mitigate these challenges, recent efforts have focused on engineering modular interfaces to enable coordinated module swapping and enhance compatibility among enzyme modules. Such strategies incorporate synthetic interfaces, including docking domains (DDs), communication-mediating (COM) domains, which are naturally derived but can be synthetically repurposed across non-cognate contexts, as well as engineered interaction modules such as synthetic coiled-coils, SpyTag/SpyCatcher system, and split inteins.



**Ji Hun Kim**

*received his B.S. degree in Chemical and Biological Engineering from Seoul National University in 2020. He is currently working as a graduate student under the supervision of Prof. Byung-Kwan Cho at KAIST. He is interested in systems and synthetic biological approach to Streptomyces engineering including synthetic bio-part development, base editor application, and secondary metabolite production enhancement.*



**Seong Do Kim**

*received his B.S. degree in Biotechnology and Sociology from Yonsei University, South Korea in 2024. He is currently pursuing graduate studies under the supervision of Prof. Byung-Kwan Cho at KAIST. His research focuses on developing automation pipelines for systems and synthetic biology application in Streptomyces, with particular interest in high throughput biosynthetic gene cluster capturing and chimeric*

*BGC assembly.*



**Gahyeon Kim**

*received her B.S. degree in Systems Biotechnology from Chung-Ang University, South Korea in 2021. She is currently pursuing her PhD in the Department of Biological Sciences at Korea Advanced Institute of Science and Technology (KAIST) under the supervision of Prof. Byung-Kwan Cho. Her research focuses on genome mining of Streptomyces and modular engineering of polyketide synthases to enable the*

*discovery and development of novel natural products.*



**Dukwon Lee**

*earned his BA and PhD in Food Science and Biotechnology from Seoul National University. His doctoral research focused on the structural analysis of the Bacillus cereus spore coat protein CotE using cryo-electron microscopy. He has led and contributed to structural and mechanistic studies of bacterial proteins through Cryo-EM and X-ray crystallography. Now transitioning into synthetic biology, he applies structure-guided protein engineering to terminal deoxynucleotidyl transferase and megaenzyme such as PKS and NRPS. His research integrates modeling, mutagenesis, and biophysical analysis to develop novel synthetic biology applications.*



The exploration and development of such synthetic interfaces are highly relevant in the context of synthetic biology, an innovative field that integrates engineering principles into biological research. A key goal of synthetic biology involves the establishment of standardized parts, modularity, and universality. Unlike traditional protein engineering approaches, which tend to produce highly specialized solutions with limited transferability, synthetic biology emphasizes the development of standardized components capable of performing consistent tasks across various biological systems and engineering contexts.<sup>8</sup> To achieve this, synthetic biology employs diverse toolkits categorized by their functional targets. At the transcription level, standardized genetic parts include inducible systems, promoters, and terminators, facilitating precise control of gene expression timing and strength.<sup>9,10</sup> At the translation level, tools such as ribosome-binding sites and codon optimization strategies enable enhanced protein synthesis efficiency and accuracy.<sup>11</sup> However, current synthetic biology toolkits predominantly focus on transcriptional and translational regulation, leaving a gap in tools available for protein association control. Therefore, the development and characterization of protein interaction domains as standardized biological components are critically important.

The integration of synthetic interfaces with modular enzyme assembly offers significant advantages, providing enhanced modularity, structural versatility, and assembly efficiency. Furthermore, advancements in AI-driven enzyme engineering have emerged, promising to further streamline enzyme optimization, predict functional compatibilities, and facilitate the design of novel enzyme modules. These advancements are increasingly organized within a rational engineering framework known as design-build-test-learn (DBTL) cycle for synthetic biology, which enables iterative improvement of modular biosynthetic systems (Fig. 1). In the design step, a desired compound, either an existing bioactive molecule or a newly

designed structure, is deconstructed to identify suitable biosynthetic modules for its synthesis. This includes determining which PKS or NRPS domains are compatible with the intended substrates and how to configure them into a functional module architecture capable of producing the target molecule. The build step enables automation-assisted combinatorial construction of modular enzyme assemblies using well-characterized parts. During the test step, chimeric constructs are expressed, and their functionality is quantified through analytic methods. Finally, in the learn step, AI-assisted linker optimization and modular enzyme design integrate experimental outcomes to improve subsequent design cycles. Beyond realizing retobiosynthesis from desired scaffolds, the DBTL cycle can further advance toward derivatization by enabling systematic recombination of modules and synthetic interfaces. This iterative optimization facilitates the rational exploration of chemical space and supports the generation of novel bioactive derivatives.

## 2. Fundamentals of PKS and NRPS architecture

In 1997, the structural elucidation of the actinorhodin synthase acyl carrier protein from *Streptomyces coelicolor* marked the first resolved structure of a PKS.<sup>12</sup> This milestone opened the door to structural and mechanistic investigations of megasynthases such as PKSs and NRPSs, which are characterized by their exceptionally large molecular sizes. The sheer size and complexity of these enzymes posed substantial analytical challenges, often limiting studies to isolated domains. Despite their complexity, these architectures are biologically intriguing and modular, with catalytic units acting in an assembly-line manner. Such modularity offers compelling opportunities for engineering novel biosynthetic pathways.<sup>13</sup>

6-Deoxyerythronolide B synthase (DEBS) from *Streptomyces erythraeus* is one of the most thoroughly studied PKSs for



**Hongki Kim**

received his B.S. degree in Life Sciences from Handong Global University in 2024. He is currently working in the laboratory of Professor Byung-Kwan Cho at KAIST, focusing on the identification and application of *Streptomyces* biosynthetic gene clusters (BGCs), including antibiotics and therapeutically valuable secondary metabolites. In particular, he is developing a *Streptomyces*-based cell-free protein synthesis platform.

With this system, he aims to enable the rapid and high-throughput expression of cryptic or intractable BGCs that are difficult to access using conventional approaches. In doing so, he aims to expand the synthetic biology toolbox for genome mining and natural product discovery.

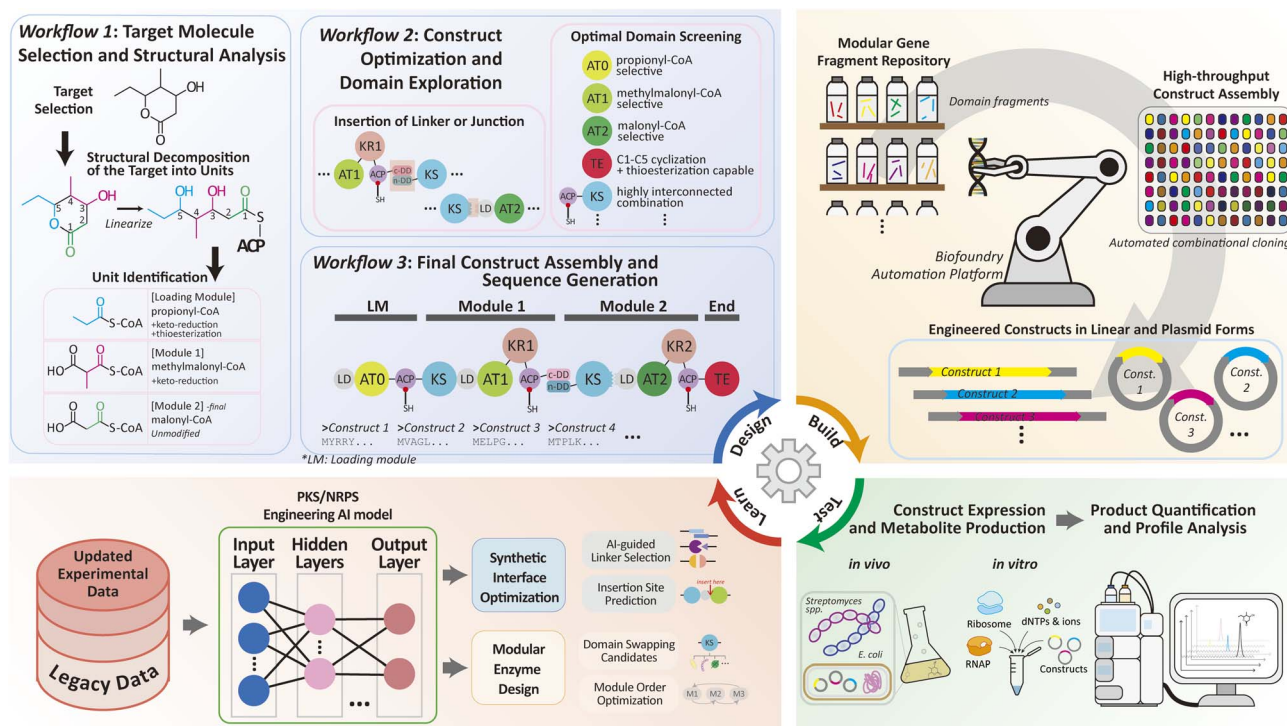


**Jae Heon Kim**

received his B.S. degree in Computer Science from New York University in 2020. He then earned an M.S. degree in Information Systems from Yonsei University, where his research focused on AI applications in systems biology. His work included graph neural networks for molecular property prediction, time-series forecasting for public health, and fine-tuning domain-specific large language models using retrieval-

augmented generation. He is a PhD student under the supervision of Professor Byung-Kwan Cho at KAIST, developing AI and bioinformatics tools to support *Streptomyces* research in systems biology and natural product biosynthesis.





**Fig. 1** Design-build-test-learn cycle for modular PKS/NRPS engineering. The DBTL cycle provides an integrated framework for engineering modular enzyme assemblies to produce targeted natural products. In the Design phase (blue panels), target molecules are structurally decomposed into biosynthetic units, guiding the identification of functional domains and modules for assembly. During the build phase (orange panel), modular gene fragments from a prepared repository are combinatorially assembled *via* automation into diverse linear and plasmid constructs. In the test phase (green panel), engineered constructs are heterologously expressed, and resulting metabolites are quantified and characterized to determine biosynthetic efficacy. Data collected from these steps feed into the learn phase (red panel), employing AI-based linker optimization and graph neural network (GNN)-based modular design strategies. Iterative incorporation of these insights refines subsequent DBTL cycles, progressively enhancing modular enzyme assembly designs for optimized biosynthesis.

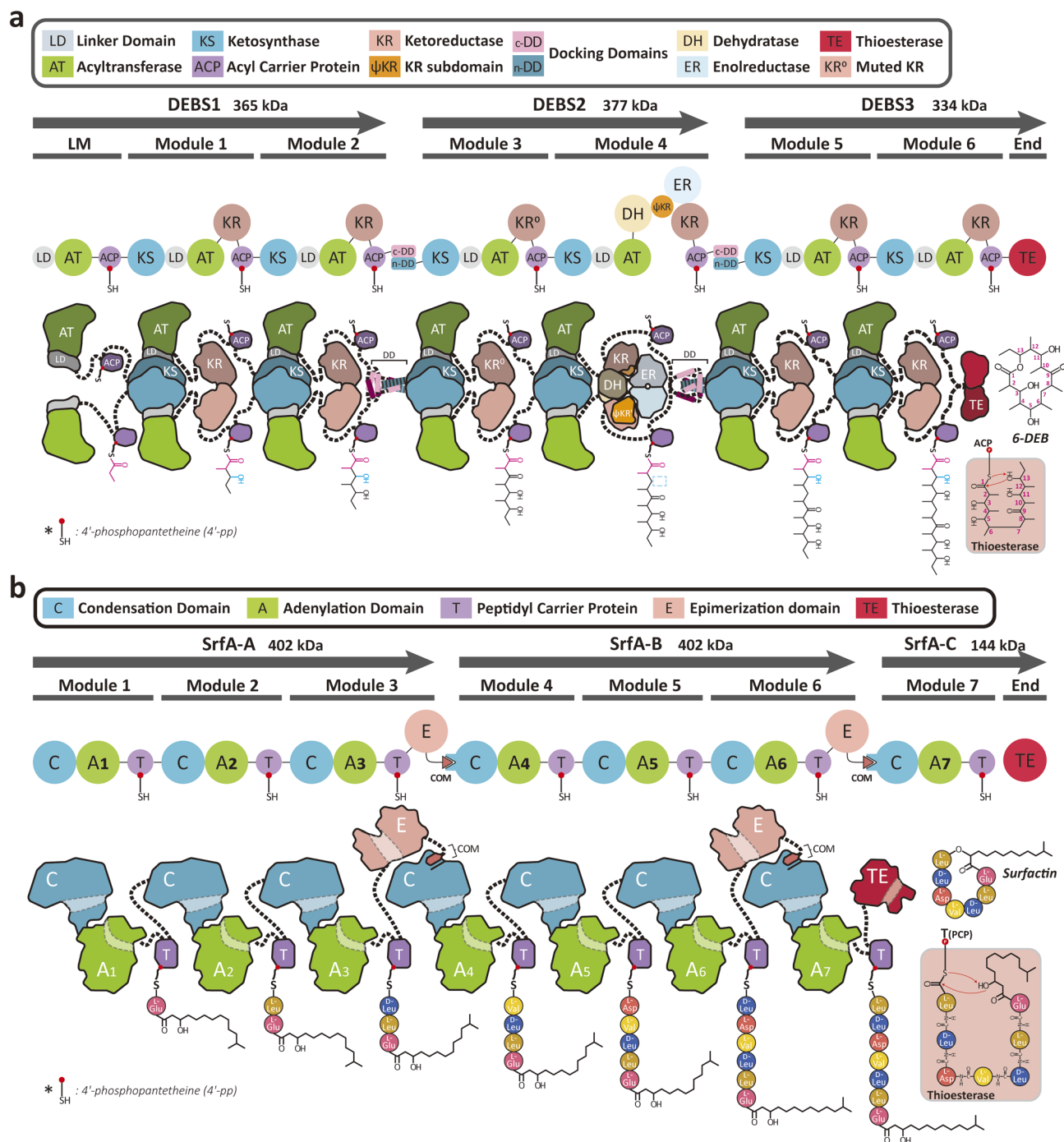
understanding the architecture and mechanism of modular type I PKSs. DEBS has been extensively investigated because it produces the macrolide precursor of erythromycin, a clinically important antibiotic.<sup>14</sup> The significance of DEBS lies not only in its pharmaceutical relevance but also in its highly organized modular architecture, which exemplifies the assembly line strategy of type I PKSs.<sup>15</sup> This feature makes DEBS an ideal system for illustrating the principles of modularity and domain organization in PKS biosynthesis. DEBS exemplifies the assembly-line logic of type I PKSs, comprising eight modules that sequentially mediate the biosynthesis of the macrolide precursor to erythromycin. These modules-covering loading, elongation, and termination steps-are distributed across three polypeptides, yet maintain functional continuity *via* DDs at their N- and C-termini. As illustrated in Fig. 2a, each elongation module incorporates a unique set of catalytic activities, enabling structural variation at every cycle. This modular repetition combined with functional variability underlies the remarkable chemical diversity observed in polyketides (PKs).<sup>16,17</sup>

At the domain level, each module consists of a defined arrangement of catalytic units typically organized as KS-AT-modifying domains-ACP. These domains act in concert to facilitate chain elongation and functional group introduction in a processive, ordered manner. Although the identity and presence of modifying domains vary between modules, the overall

architecture is conserved, providing a robust framework for rational reprogramming of PKS systems.<sup>16,17</sup> The repetitive KS-AT-modifying domains-ACP pattern is a hallmark of type I PKSs, illustrating their assembly line-like biosynthetic mechanism (Fig. 2a). In this system, each module functions as a dedicated processing station, analogous to a unit on a conveyor belt, where it adds a single malonyl-derived extension unit to the growing PK chain before passing it along to the next module.<sup>18</sup> By strategically rearranging modules, replacing them, or introducing new catalytic units, researchers can engineer diverse structural and chemical properties into PK products. This adaptability has enabled the creation of novel bioactive compounds through domain swapping, module excision, and recombination strategies, highlighting the practical versatility of type I PKSs.<sup>19,20</sup>

All catalytic steps are orchestrated around the KS dimer axis, allowing for the efficient and accurate biosynthesis of complex organic compounds such as 6-DEB. The dimerization property of PKS modules is also crucial in the function of DDs, which mediate inter-protein interactions.<sup>21</sup> In DEBS, two DDs are present at the interfaces between modules 1–2 and modules 2–3. These DDs facilitate stable protein docking by tightly interlocking  $\alpha$ -helices, a feature essential for maintaining structural integrity (Fig. 2a).<sup>22</sup> The C-terminal DD helix of the preceding module and the N-terminal DD helix of the succeeding module intertwine in a module-specific manner, preventing misalignment and





**Fig. 2** Sequential modular architecture of megasynth(et)ases DEBS and SrfA. (a) Graphical representation of *S. erythraeus* DEBS PKS. Each module of DEBS carries out specific chemical transformations, and individual domains within modules catalyze distinct reactions. To clearly depict the repetitive modular structure and avoid confusion due to the large size of the megasynthase, identical domain types are represented consistently with the same colors throughout the figure. Gene names (DEBS1-3), protein sizes, and modular boundaries are indicated clearly. Domains are schematically represented by simplified lines and circles under the module representation. The two-dimensional arrangement of structure-reflecting domain diagrams is represented under the simplified domain arrangement reflects the actual structural organization and illustrates the progressive transfer of synthesized chemical intermediates attached to the 4'-phosphopantetheine (4'-pp) arm of ACP. In the chemical structures, newly synthesized segments within each module are highlighted in magenta, and segments modified in the current module are colored cyan. The final thioesteration reaction site is specifically depicted in a red "Thioesterase" box. Carbon atoms in chemical structures are numbered accordingly. (b) Graphical representation of *B. subtilis* surfactin synthase SrfA, a NRPS. Visual representation follows the same scheme used in panel (a), except that amino acids are depicted as colored circles instead of chemical structures.



ensuring precise module-to-module communication. This module-selective docking mechanism makes DDs particularly suitable targets for protein engineering.<sup>23</sup>

Although recent advances have markedly enhanced our understanding of NRPS modular structures—particularly through high-resolution studies of full mono- and di-module systems—the field continues to explore the broader structural organization of type A NRPSs.<sup>24–27</sup> Surfactin synthetase (SrfA) from *Bacillus subtilis* stands out as a representative example, offering valuable insights into the architecture and mechanism of modular NRPSs.<sup>28,29</sup> While modular representations of SrfA have been proposed, a universally accepted full structural model is still lacking, underscoring the continued challenges in elucidating the spatial organization and inter-domain dynamics of NRPS systems.<sup>30</sup> Nonetheless, the accumulated structural and functional studies of SrfA provide a foundational framework for understanding NRPS modularity, domain architecture, and catalytic mechanisms.<sup>31</sup>

Like PKSs, type A NRPSs exhibit modular organization consisting of sequential catalytic domains arranged typically in a condensation–adenylation–thiolation (C–A–T) pattern, along with specialized domains such as epimerization (E) and thioesterase (TE) domains (Fig. 2b).<sup>28,32</sup> Whereas, unlike PKSs, which elongate PK intermediates, NRPS systems incorporate amino acids to produce diverse peptides. Despite these mechanistic distinctions, both systems employ analogous assembly-line logic, sequentially extending and modifying intermediates. In modular NRPSs, inter-protein communication is mediated by specialized communication-mediating (COM) domains, which enable efficient substrate transfer and coordinated catalysis across module boundaries.

In SrfA, this intermodular communication critically depends on COM regions located at the terminal ends of each protein subunit (Fig. 2b). Although SrfA lacks integrated modifying domains within its C–A–T core, it incorporates modifications *via* separate domains, most notably the C-terminal E domain.<sup>33</sup> The E domain not only catalyzes L-to-D amino acid conversion but also contributes structurally by stabilizing COM-mediated docking.<sup>25,27,34</sup> This COM-driven mechanism ensures precise inter-protein alignment and functional continuity in SrfA.

Unlike the DD in PKSs, which forms an intertwined helix pair between separate modules, the COM domain in NRPSs constitutes an asymmetric hybrid interface. It consists of a short  $\alpha$ -helix segment (COM-donor, COM<sup>D</sup>) from the upstream module (typically appended to the E domain) and a shallow surface pocket (COM-acceptor, COM<sup>A</sup>) embedded within the C domain of the downstream module. Due to the inherent mutual specificity between COM<sup>D</sup> and COM<sup>A</sup>, it is more appropriate to consider the COM interface as a cognate domain pair, rather than two independent binding sites. This distinction is particularly relevant when repurposing or engineering NRPS modules for synthetic applications.

Interaction domains such as DD in PKSs and COM domains in NRPSs exhibit notable structural diversity across different biosynthetic systems. These domains are not uniform in architecture; instead, they comprise multiple structural classes that differ in helical organization, interaction interfaces, and

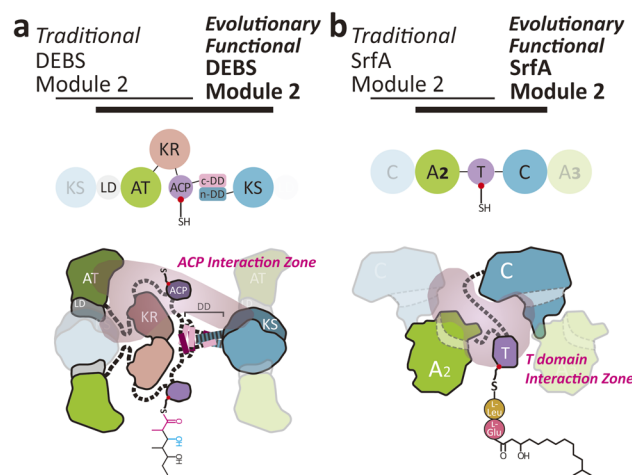
specificity-conferring residues.<sup>35</sup> This diversity enables selective and directional intermodular communication and provides a broad repertoire of interchangeable elements for synthetic design. As illustrated in Fig. 2a, even within a single biosynthetic pathway such as DEBS, the structural features of DDs between subunit interfaces (*e.g.*, DEBS1-DEBS2 *vs.* DEBS2-DEBS3) can differ, underscoring their functional flexibility. Understanding and leveraging this variability is essential for expanding the design space of engineered assembly lines.

These modular biosynthetic systems, despite architectural differences, operate through a conserved conveyor belt-like logic wherein each module executes a defined reaction step.<sup>32</sup> This principle underpins the programmability of both PKS/NRPS systems and offers a conceptual basis for modular engineering, which guide the rational design of chimeric enzymes for the biosynthesis of novel bioactive compounds.

### 3. Modular assembly and engineering of PKS and NRPS

#### 3.1 Evolutionary functional boundaries of PKS and NRPS modules

Recent advances have proposed revisiting PKS/NRPS module boundaries based on evolutionary and structural organization, reshaping our understanding of these enzymatic assembly



**Fig. 3** Evolutionary functional reinterpretation of modular boundaries in PKS/NRPS systems. Modular boundaries in PKS/NRPS systems can be more accurately interpreted by centering on the carrier protein domains (ACP or T), which mediate critical intermodular substrate transfer. In both systems, conventional domain-based boundaries (top horizontal lines) are contrasted with reorganized boundaries that emphasize the spatial and functional autonomy of the carrier domain. Domains outside the proposed functional units are shown with reduced opacity. The magenta-highlighted regions represent the interaction zones surrounding the carrier domains, reflecting their pivotal roles in modular communication and catalytic coordination. (a) PKS module 2 from the DEBS system, reorganized as an evolutionary functional unit centered on the ACP domain and its adjacent DDs. (b) NRPS module 2 from the SrfA system, reorganized as an evolutionary functional unit centered on the T domain and its interface with the upstream C domain.



lines.<sup>29</sup> Traditionally, PKS modules were defined as starting from the KS domain and terminating at the ACP domain, with ACP serving as the final domain responsible for transferring intermediates to the KS domain of the next module (Fig. 3a).<sup>16</sup> However, evolutionary and bioinformatic analyses revealed that processing domains including KR, DH, and ER consistently cluster and move genetically with the KS domain immediately following the ACP, rather than with the preceding KS. Therefore, ACP domains are better interpreted as internal carrier units, reflecting their role within an evolutionary functional module (intermediate domains) rather than terminal domains.<sup>5,36,37</sup>

This updated boundary definition, rooted in evolutionary module architecture, clarifies longstanding challenges in PKS engineering.<sup>38</sup> Previously, engineering efforts frequently involved swapping or modifying KS domains positioned downstream of ACP, disrupting precise interactions and reducing substrate transfer efficiency.<sup>39</sup> By adopting the updated boundary, where ACP domains internally transfer intermediates directly to the subsequent KS domain within the same module, domain recognition becomes stable and consistent, significantly enhancing biosynthetic efficiency and product fidelity (Fig. 3a).<sup>40</sup>

Analogous structural insights extend to NRPS systems. Previously, NRPS modules were similarly defined, placing the T domain at the module's terminus.<sup>28</sup> Yet recent engineering data indicate that, like PKS systems, NRPS modules can similarly be reorganized into evolutionary functional units with the T domain at the center.<sup>41</sup> In this configuration, the T domain captures the amino acid monomer, undergoes modifications exclusively within its own module, and transfers the fully processed intermediate directly to its module's C domain, which is now positioned as the downstream boundary (Fig. 3b). Consequently, intermediate transfer and peptide elongation are confined and stabilized by interactions occurring strictly within each module. This modular restructuring prevents unintended interactions and backward progression, reflecting a strictly enforced biosynthetic directionality.

Ultimately, redefining the module boundary based on evolutionary and structural criteria has significant practical implications. Notably, employing the updated boundaries has already proven beneficial, exemplified by successful engineering efforts where modules designed according to new boundaries yielded previously unattainable synthetic products, such as homoauoreothin.<sup>5,36</sup> Thus, adopting evolutionary functional boundaries provides structurally and biologically coherent frameworks for rational engineering for future engineering approaches, enhancing our ability to generate novel PK and peptide derivatives with predictable and efficient outcomes.

### 3.2 Engineering with cognate interfaces: orthologous interaction modules

In light of the modular architecture of PKS/NRPS systems, DDs-short amino acid motifs situated at the C- and N- termini of subunits that mediate the assembly of PKS/NRPS megasynth(et)ases-have emerged as powerful tools for enabling the

combinatorial engineering of biosynthetic assembly lines. In this section, we refer to DD systems as orthologous interaction modules (OIMs)-cognate motif pairs that originate from PKS/NRPS lineages but are repurposed to engineer PKS/NRPS systems (Fig. 4a and b).

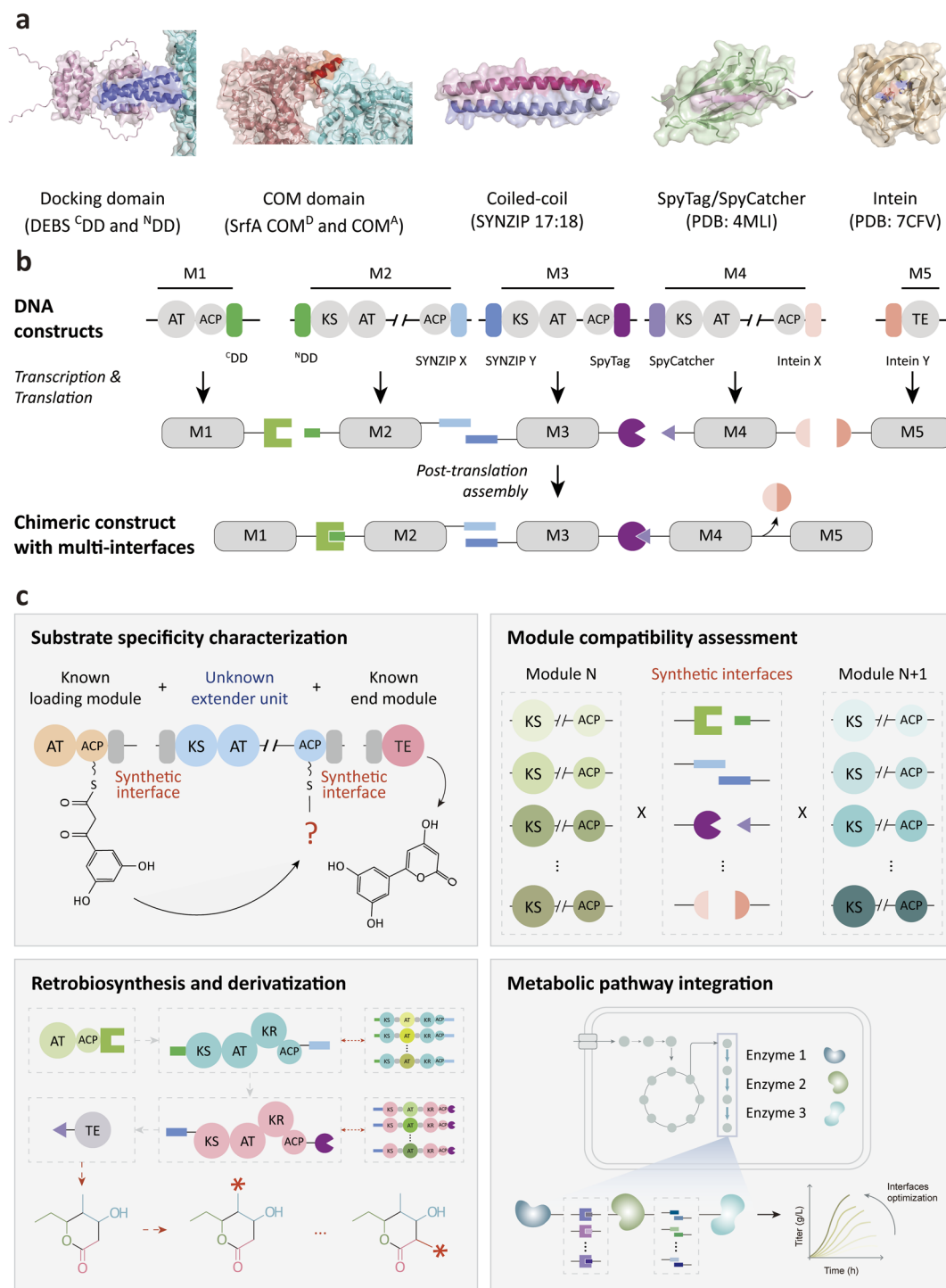
**3.2.1 OIM-based engraftment strategies.** Prior to the past five years, most research regarding OIM-mediated PKS/NRPS engineering was concerned with investigation, primarily focusing on elucidating the structural motifs of various DDs in natural PKS/NRPS systems, with the aim of understanding their mechanisms of action. Collectively, these efforts culminated in the development of a structure-based classification system for PKS/NRPS DDs, providing detailed molecular insights into their functions.<sup>35</sup> Notably, these classifications also represent evolutionary barriers that prevent crosstalk between OIMs originating from distinct evolutionary lineages, further highlighting their inherent orthogonality.<sup>28</sup> In parallel, efforts were directed towards assessing the modularity of these domains through engraftment into model

PKS/NRPS systems.<sup>22,42,43</sup> These studies led to the establishment of various model OIM systems, most notably the 6-DEBS OIMs.

Building upon these foundational findings, research efforts over the past five years have shifted toward expanding the applicability of OIMs, aiming to transform them from narrowly optimized, target-specific tools into robust and versatile platforms for engineering a broad range of systems, including non-model PKS/NRPS pathways and entirely unrelated metabolic pathways. These advancements can be broadly classified into three categories: employing OIM engraftment to generate truncated variants of native PKS/NRPS systems, enhance the productivity of PKS/NRPS systems, and assemble chimeric enzymes.

Using OIMs to truncate native PKS/NRPS systems into mini-synthase variants is a key strategy in biosynthetic pathway engineering. This approach not only enables the generation of diverse chemical derivatives of the original product, but also serves as an effective means to deconstruct large PKS/NRPS systems into smaller, more manageable units for detailed characterization of module compatibility and downstream engineering endeavours.<sup>44</sup> In this context, both endogenous OIMs from the target PKS/NRPS system and heterologous OIMs can be employed to construct novel inter-modular interfaces that may not naturally occur in the absence of these elements. Notable applications include the mini-stambomycin and mini-azalomycin F variants with the aid of endogenous OIMs.<sup>44,45</sup> When applied with alternative modification strategies such as domain swapping and rational site-directed mutagenesis, OIM modification can further broaden the scope of derivatization.<sup>44,45</sup> This strategy has also been extended to NRPS systems, as demonstrated by the engineering of plipastatin NRPS systems using native OIMs to produce shortened module variants for product derivatization.<sup>46</sup> A key consideration when applying this strategy to NRPS systems, however, is the inherent promiscuity exhibited by the COM domain families of NRPS OIMs. Given the technical difficulties associated with enhancing the fidelity of COM domains, a practical guideline





**Fig. 4** Synthetic interfaces for modular enzyme assembly and biosynthetic applications. (a) Representative structural models of synthetic interfaces used for modular assembly: DD (DEBS1-<sup>CDD</sup>/DEBS2-<sup>NDD</sup>, AlphaFold3), COM (srfAA-COM<sup>D</sup>/srfAB-COM<sup>A</sup>, AlphaFold3), coiled-coil (SYNZIP 17/18, AlphaFold3), SpyTag/SpyCatcher (PDB: 4MLI), and intein (PDB: 7CFV) pair. (b) Synthetic interfaces are genetically fused to modular PKS/NRPS proteins at the N- or C-termini, enabling post-translational assembly of functional multi-enzyme constructs. Distinct interface types (green: DD; blue: SYNZIP; purple: SpyTag/SpyCatcher; pink: intein) mediate the formation of chimeric biosynthetic assemblies from individually transcribed and translated modules. (c) Functional applications of synthetic interfaces. These modules facilitate the characterization of substrate specificity by connecting known and unknown modules, allowing inference of substrate preference based on product profiles. They also support combinatorial testing for module compatibility, construction of synthetic pathways through retrobiosynthetic logic, and scaffold derivatization via domain/module substitutions. Beyond PKS/NRPS systems, synthetic interfaces also enable modular clustering in metabolic pathway engineering to enhance flux and control.



will be to leverage their promiscuity to expand the structural diversity of derivative compounds.<sup>28</sup> Alternatively, the promiscuous interaction modules of the NRPS system may be exchanged with more stringent NRPS OIMs, such as the *Xenorhabdus*-derived SLiM-βhD domain pairs.<sup>47,48</sup>

OIM engraftment may also be utilized to enhance both the stability and productivity of existing PKS/NRPS systems. Among the available options, the well-characterized DEBS OIMs stand out as a prominent and extensively adopted tool for this strategy. Beyond substituting native interaction domains, DEBS OIMs have also been used to segment intact PKS genes into smaller modules, effectively replacing covalent inter-modular linkages with non-covalent interfaces while preserving the functionality of the entire system. This approach, often used in *in vitro* reconstitution experiments, is particularly useful for enabling the expression of large PKS genes in heterologous hosts such as *E. coli* and the production of the target product from split modules.<sup>49–51</sup> Similar strategies have also been implemented using less-characterized OIMs. For instance, OIMs derived from salinomycin and stigmatellin have been applied to split a wide array of unstable PKS systems, including butenyl-spinosyn, salinomycin, avermectin, and epothilone, into stably expressed module parts.<sup>7</sup> This approach was similarly applied to the xefoampeptide-producing NRPS from *Xenorhabdus bovienii*, demonstrating its utility across both PKS/NRPS systems.<sup>52</sup>

OIMs may also be used to construct interfaces between heterologous modules, enabling the production of designer molecules. The construction of bi-modular chimeric PKS libraries using subunits from pikromycin synthase and venemycin synthase exemplifies this approach.<sup>38</sup> Furthermore, OIMs may also be employed to alter the starter units of PKS systems by facilitating the exchange of the loading modules, as demonstrated by the use of curacin-derived OIMs to enable the production of alkyne-tagged polyketides.<sup>53</sup>

Alternatively, there has been growing interest in extending the utility of OIMs toward broader synthetic biology applications such as metabolic engineering, artificial enzyme assembly, and synthetic protein scaffolds. Because OIM pairs retain their interaction capabilities independently of their native biosynthetic contexts, they can be repurposed as orthogonal protein connectors. This potential has been quantitatively demonstrated through fusion to split GFP variants, demonstrating that OIMs can function as modular tools for general enzymatic assembly.<sup>54</sup> Notably, the utility of OIM-mediated metabolic engineering was demonstrated in the synthetic tri-modular reconstitution of the maytansinol biosynthetic pathway, where OIM engraftment enhanced the interaction between constituent enzymes.<sup>55</sup> In a similar note, when model OIM pairs from different evolutionary lineages were grafted into the astaxanthin biosynthetic pathway to facilitate a tri-modular enzyme assembly in *Escherichia coli*, a 2.4-fold increase in astaxanthin production was observed.<sup>56</sup> This result demonstrated that OIMs from distinct structural families-typically orthogonal-can be implemented simultaneously within a single metabolic pathway, supporting their potential use in assembling longer and more complex

biosynthetic pathways, demonstrating the transferability of these systems. These findings highlight the exciting potential that OIMs bring to the field of protein assembly. Given the current scarcity of reliable tools for programmable enzymatic organization, tapping into the vast natural diversity of DDs presents a promising solution.<sup>56</sup> As a point of reference, the type I cis-AT PKS family alone is predicted to encompass over 1600 distinct clusters, representing a rich and largely untapped reservoir of candidate interaction domains for synthetic biology applications.<sup>17</sup>

**3.2.2 Expanding the utility of OIMs.** All in all, these trends reflect a concerted effort to expand the available OIM toolkit and diversify the repertoire of engineering strategy that can be employed using these tools-signalling a shift from traditional PKS/NRPS engineering to synthetic biology applications. However, the use of OIMs in such a manner underscores the need to further develop them into standardized bioparts. Bioparts, by definition, are standardized components that can be assembled seamlessly into larger biological devices.<sup>57</sup> As such, their activity is typically well-characterized and quantitatively measured, with clearly defined specifications for their operational parameters. These specifications are often documented and shared through open-access registries to facilitate widespread use and reproducibility across the synthetic biology community. Synthetic promoter libraries-a staple in the synthetic biologist's toolbox-clearly exemplifies these principles. They provide a palette of promoters with well-defined expression strengths, allowing engineers to precisely fine tune the expression levels of all constituent genes within a pathway to orchestrate a coordinated biological event.<sup>58</sup>

To develop such a library, a multi-step approach is required. First, OIMs must be identified and selected from natural PKS/NRPS systems as candidates for characterization. Once selected, these domains must be cloned into a variety of biosynthetic contexts whose products have already been characterized. The engineered constructs must then be introduced into a heterologous expression host to evaluate whether the presence of the OIMs restore or enable production of the desired compound. Functional parameters derived from these experiments can be used to quantitatively characterize their utility as bioparts. These data should be systematically catalogued in a centralized, open-access database to facilitate reuse and interoperability. Furthermore, the sequences of validated DDs can be rationally modified through techniques such as site-directed mutagenesis to generate a library of OIM bioparts with varying interaction strengths. This would enable a synthetic biologist to design more complex circuits at the post-translational level by fine-tuning and orchestrating protein-protein interactions. Continuous expansion of the library is essential through collaboration and crowdsourcing data, ensuring it evolves with new innovations and applications in synthetic biology. Together, this comprehensive process will create a robust, versatile collection of DDs that can be leveraged to design and optimize efficient, modular biosynthetic pathways for diverse applications in synthetic biology.



### 3.3 Beyond orthologous interfaces: xenologous interaction modules in synthetic biology

As efforts in synthetic biology continue to advance toward modular and programmable biosynthetic systems, an increasing number of protein–protein interaction platforms have been adapted or newly developed to facilitate enzyme assembly beyond native PKS/NRPS systems. In this section, we refer to these systems, including synthetic coiled-coils, peptide-based covalent systems, and self-splicing inteins, as xenologous interaction modules, which originate outside of PKS/NRPS biosynthetic lineages but can be repurposed to mimic or improve natural domain interfaces (Fig. 4a and b). These tools represent an expansion of the interaction toolkit for reconfigurable and customizable biosynthetic engineering.

**3.3.1 Synthetic coiled-coils.** Coiled-coil motifs have emerged as valuable tools, particularly in the context of modular enzyme engineering. These structural motifs, composed of intertwined  $\alpha$ -helices, provide a flexible framework for assembling complex enzymatic systems like PKSs/NRPSs.<sup>59</sup> Unlike natural DDs, which are often limited by specificity and compatibility constraints between non-cognate modules, synthetic coiled-coils provide a rationally engineered alternative with tunability and orthogonality, making them attractive for reprogramming biosynthetic pathways and engineering modular enzyme assemblies.<sup>60</sup> Heterospecific synthetic coiled-coil peptides, called SYNZIPs, were introduced in the early 2010s. SYNZIP toolbox comprises a library of orthogonal interaction pairs designed to form stable, heterodimeric interactions without undesired cross-reactivity.<sup>61</sup> Their rationally engineered interfaces allow for predictable, high-affinity interactions, while also maintaining structural and functional integrity across different biological systems, positioning them as a highly reliable ‘plug-and-play’ solution within the synthetic biology toolkit.<sup>59</sup>

The feasibility of SYNZIPs has been demonstrated in various systems, including the reconstitution of DEBS-derived bimodular PKSs and the modularization of fungal NRPSs (Fig. 4b).<sup>56</sup> For instance, SYNZIP-mediated interactions successfully circumvent constraints imposed by KS domains in PKSs, enabling the assembly of non-native modules.<sup>62</sup> Also, coiled-coils have proven instrumental in DNA-templated enzyme systems, where they synergize with other domains, such as zinc fingers, to position enzymes strategically along DNA templates, enhancing spatially confined catalysis and biosynthetic efficiency.<sup>63</sup> Furthermore, coiled-coil-mediated assembly has been leveraged to create hybrid biosynthetic pathways, enabling the production of novel bioactive compounds with enhanced structural diversity.<sup>64,65</sup> By integrating coiled-coil-based enzyme engineering with combinatorial biosynthesis approaches, researchers aim to expand the chemical space of natural product derivatives, potentially leading to the discovery of new pharmaceuticals.

**3.3.2 SpyTag/SpyCatcher system.** Beyond synthetic coiled-coil systems like SYNZIPs, covalent interaction strategies have also been explored for modular enzyme engineering. The SpyTag/SpyCatcher system, derived from *Streptococcus pyogenes*,

forms an irreversible covalent bond between a short peptide (SpyTag) and a protein domain (SpyCatcher) (Fig. 4a and b).<sup>66</sup> This system has proven effective in stabilizing enzyme complexes at low concentrations, as the covalent linkage prevents subunit dissociation and enhances structural rigidity.<sup>67</sup> It has been applied to stabilize split NRPS modules and create stapled NRPS architectures that show enhanced catalytic efficiency and product yields.<sup>68</sup> Also, this system facilitates directed substrate transfer through enzyme proximity, and enables modular designs that are robust across different expression contexts.<sup>69</sup> Furthermore, to enhance orthogonality, engineered SpyTag/SpyCatcher mutant libraries have been developed that allow multiple, mutually orthogonal reactivity profiles.<sup>70</sup> These variants enable parallel covalent labelling and selective enzyme assembly, broadening the utility of SpyTag/SpyCatcher for complex biosynthetic applications.

Comparative studies of SYNZIPs and SpyTag/SpyCatcher systems highlight complementary strengths but also reveal common engineering challenges that must be addressed for optimal functionality. Both systems can impose structural rigidity on the engineered interface, potentially hindering the enzymatic activity by restricting conformational dynamics essential for substrate channelling and catalytic turnover.<sup>76</sup> For SYNZIPs, excessive  $\alpha$ -helical length or tightly packed structures can restrict conformational flexibility. To address this, flexible glycine–serine linkers have been introduced between the SYNZIP and enzymatic domains, restoring the required mobility. Additionally, truncation of SYNZIP sequences, particularly SZ1/SZ2 pair, has been shown to improve production yield, in some cases by over 50-fold, by reducing steric hinderance while preserving binding affinity.<sup>77</sup> Similarly, while SpyTag/SpyCatcher offers the advantage of irreversible covalent bonding, this covalency may introduce rigidity, limiting the structural plasticity required during enzymatic turnover. To overcome this, hinge-like sequences have been incorporated adjacent to ligation points to mimic the flexibility of natural linker regions.<sup>77</sup> Thus, regardless of the binding modality, non-covalent or covalent, interface flexibility emerges as a critical determinant of successful modular reconstitution. Otherwise, SYNZIPs and SpyTag/SpyCatcher systems show complementary features. SYNZIPs offer orthogonal and reversible interactions ideal for contexts requiring conditional control. In contrast, SpyTag/SpyCatcher excels in permanent stabilization of multi-enzyme complexes, ensuring robustness in dynamic or heterologous environments. Strategic selection or combinatorial use of these systems, including with DDs or COM domains, opens new avenues for hierarchical design of reconfigurable biosynthetic pathways.

**3.3.3 Split inteins.** In addition to coiled-coil and covalent tag systems, inteins have emerged as another promising component of the protein–protein interaction toolbox in synthetic biology (Fig. 4b).<sup>78,79</sup> Inteins are naturally occurring protein segments capable of self-excising and ligating their flanking sequences (exteins) through a process known as protein splicing.<sup>80</sup> Particularly, split inteins, expressed as two separate fragments, can associate *in vivo* to mediate trans-splicing, making them valuable for post-translational protein



assembly in a modular fashion.<sup>81</sup> While inteins have not yet been widely applied to PKS/NRPS engineering, a proof-of-concept study has demonstrated their use in selectively labelling an NRPS module (TycA) *via* protein trans-splicing, highlighting their potential in this space.<sup>82</sup> Furthermore, split inteins have been used in systems such as split-intein circular ligation of proteins and peptides (SICLOPPS) method to generate cyclic peptides within microbial hosts, and in conditional assembly strategies where enzyme activity can be gated by intein-mediated splicing.<sup>83,84</sup> Additionally, engineered inteins have demonstrated environmental responsiveness, such as activation by pH or redox conditions, further suggesting their potential utility for programmable control of enzyme complex formation.<sup>81</sup> Moreover, split intein pairs exhibit mutual orthogonality, with up to 15 distinct pairs showing minimal cross-reactivity in both *in vitro* and *in vivo* systems, greatly expanding their potential for multiplexed or combinatorial applications.<sup>75</sup> In addition to inteins, asparaginyl endopeptidases (AEPs) such as OaAEP1 have also been applied to NRPS systems, catalyzing site-specific peptide bond formation between protein fragments.<sup>24,85</sup> AEPs offer a chemoenzymatic strategy for post-translational module assembly and have enabled ligation of split NRPS domains for structural and mechanistic studies. Their ability to mediate covalent ligation without external cofactors and with precise specificity is shared by inteins, which makes them an attractive orthogonal alternative.

Future directions include AI-assisted design of xenologous interaction modules with enhanced context-specific functionality, high-throughput screening of interaction libraries under variable expression conditions, and the development of switchable interface systems responsive to environmental cues. By expanding the synthetic biology toolbox, coiled-coil-based and covalent interaction systems are poised to accelerate the rational engineering of enzyme assemblies for the discovery and production of novel natural products.

### 3.4 Functional applications of synthetic interfaces in modular enzyme engineering

The synthetic interfaces introduced in the previous sections, including orthologous and xenologous interaction modules serve not only as structural connectors but also critical functional tools in modular enzyme assembly (Table 1). These interfaces act as standardized connectors that enable systematic exploration of biosynthetic designs through combinatorial assembly (Fig. 4c). Their orthogonality, post-translational assembly properties, and tunability enable flexible

reconfiguration of enzyme modules, supporting broader applications in functional characterization, PKS/NRPS reprogramming, and general pathway construction.

One application lies in substrate specificity characterization (Fig. 4c). In modular systems, the identity of extender unit incorporated by an extender module remains largely unknown, posing a bottleneck to module selection and engineering. By linking a validated loading module to an uncharacterized extender module using stable and orthogonal interfaces, the resulting product can reveal substrate preference. Using non-cognate or poorly matched interaction domains in this context often leads to instability in module assembly or significantly reduced product titers.<sup>35</sup> Thus, employing orthogonal and validated linkers is critical for accurate characterization.

Module compatibility assessment is another key step in combinatorial biosynthesis (Fig. 4c). Even with known substrate specificities, modules originating from different biosynthetic contexts often display suboptimal interaction due to incompatible domain–domain interfaces. A well-documented challenge is the inability of non-cognate KS domains to accept and extend the intermediate from upstream modules, which frequently leads to failed biosynthetic outputs.<sup>44,62</sup> To address this, compatibility can be improved through rational engineering of domain boundaries, such as replacing KS domains in acceptor modules to match the specificity of upstream partners. In one example, cassette replacement of a poorly active KS domain with a compatible one restored activity in previously inactive PKS modules, providing a basis for compatibility mapping in combinatorial assembly.<sup>62</sup> This strategy facilitates the rational design of synthetic pathways by delineating functional interface rules.

Beyond characterization, synthetic interfaces can be employed in retrobiosynthetic pathway construction and scaffold derivatization (Fig. 4c). Once compatibility constraints are clarified, synthetic pathways can be built by recombining modules to match retrobiosynthetic logic derived from a desired molecule. In this context, orthogonal synthetic interfaces become indispensable, as multiple module boundaries must be reliably bridged within a single construct. Furthermore, modular constructs can be iteratively refined to introduce structural diversity through domain swaps, enabling derivatization of bioactive scaffolds.<sup>86</sup> This combinatorial approach expands the accessible chemical space beyond natural diversity and accelerates analog discovery.

Finally, synthetic interfaces are not limited to PKS/NRPS systems (Fig. 4c). Their applicability extends to general metabolic pathway engineering, where they enable the spatial organization of unstructured enzyme cascades. For example, the

**Table 1** Comparative structural and functional features of interaction domains and modules for modular enzyme engineering

Feature	Docking domains	Synthetic coiled-coils	SpyTag/SpyCatcher system	Split inteins
Interaction type	Non-covalent bond <sup>71</sup>	Non-covalent bond <sup>23</sup>	Covalent bond <sup>66</sup>	Protein splicing <sup>72</sup>
Affinity ( $K_D$ )	1–130 $\mu\text{M}$ <sup>21,43,73</sup>	<10 nM (validated pairs) <sup>59</sup>	Irreversible <sup>74</sup>	Irreversible <sup>72</sup>
Orthogonality	Low <sup>21</sup>	High <sup>59</sup>	High <sup>70</sup>	High <sup>75</sup>
Structural flexibility	High <sup>22</sup>	Limited <sup>76</sup>	Limited <sup>76</sup>	Moderate <sup>75</sup>



mPKSeal strategy, which utilizes DDs derived from type I cis-AT PKSs, has been applied to assemble heterologous metabolite enzymes in *E. coli* for astaxanthin production to enable physical clustering of otherwise dispersed metabolic enzymes.<sup>56</sup> This multienzyme assembly prompted spatial co-localization and improved overall product yields. As discussed in the previous section, other synthetic interfaces such as SpyTag/SpyCatcher and split inteins have also been applied to spatially organize enzymes, offering enhanced enzyme-to-enzyme transfer and conditional pathway activation.<sup>69,81</sup> This has been demonstrated in engineered metabolic pathways where synthetic scaffolds or tag-mediated tethering improved pathway yield and stability, even outside of PKS/NRPS contexts.

While synthetic interfaces offer new routes for reconstituting modular enzyme assemblies, several studies have shown that proofreading functions embedded in catalytic domains play a decisive role in determining the success of engineered pathways. In PKSs, KS domains selectively accept properly processed intermediates and reject non-cognate or misprocessed substrates, effectively acting as gatekeepers.<sup>6,39</sup> Similarly, in NRPSs, C domains enforce acceptor substrate specificity at the elongation step, often preventing chain extension when paired with incompatible upstream adenylation domains.<sup>87,88</sup>

These fidelity mechanisms safeguard the accuracy of native biosynthesis but also impose significant constraints on modular recombination strategies. Misprimed intermediates can stall the assembly line, and improperly matched domains may result in inactive constructs or undesired shunt products. For example, engineered stambomycin PKS systems with redesigned docking interfaces exhibited reduced yields due to KS-domain substrate rejection and premature intermediate off-loading by TE domains, despite successful module fusion. To address these issues, auxiliary enzymes, such as type II TEs in NRPS systems and trans-acting acyl hydrolases in PKSs, have been employed to remove aberrant intermediates and restore biosynthetic flow.<sup>89,90</sup>

Additional strategies focus on co-designing module boundaries and domain compatibility. Domain recombination at structurally conserved splice sites, along with orthogonal assembly platforms such as SYNZIP- or intein-mediated systems, have shown promise in mitigating gatekeeping effects.<sup>91</sup> Together, these findings underscore the importance of integrating both structural interface design and substrate-specific proofreading considerations when engineering modular biosynthetic pathways. As these biosynthetic systems continue to evolve through expanded orthogonal libraries, tunable binding affinities, and AI-assisted linker optimization, they are expected to play an increasingly integral role to the rational design of modular biosynthesis.

## 4. *In silico* retrobiosynthesis design for modular natural product pathways

The modular nature of PKS/NRPS renders them particularly amenable to retrobiosynthetic tracking based on structural information of the target compound. In many cases,

biosynthetic pathways have been successfully inferred without the aid of *in silico* tools by predicting the product types, incorporated monomers, required biosynthetic domains, and relevant tailoring steps.<sup>92–96</sup> However, such rational predictions remain limited when applied to structurally complex molecules lacking known analogs. Accordingly, the development of computational toolkits capable of retrieving compatible PKS/NRPS modules from large databases and inferring feasible modular architectures is essential for systematically identifying suitable enzymatic building blocks for constructing novel biosynthetic pathways.

To date, several computational platforms have been developed for PKS/NRPS systems. These platforms address the inherent complexity of megasynth (et)ases, which consist of multiple interacting domains, by computationally deconstructing target molecules into biosynthetically plausible modules.<sup>86,97</sup> The recent integration of structural modeling tools, including AlphaFold2 and ColabFold, further enhances intra-module interface prediction and module compatibility.<sup>98,99</sup> Emerging resources such as generating retrobiosynthetic analysis for polyketides and nonribosomal peptides (GRAPE) and gene and reaction linker for informed clusters (GARLIC) implement retrobiosynthetic logic to map target metabolites to gene clusters.<sup>86</sup> Also, the platforms like ClusterCAD provide user-friendly databases and design algorithms for modular PKS/NRPS engineering.<sup>97</sup> These innovations address the limitations of structure-guided rational approaches, enabling more comprehensive reconstruction of modular biosynthetic pathways, accelerating the creation of novel secondary metabolites *via* synthetic biology.

### 4.1 Retrobiosynthesis prediction toolkits for PKS/NRPS

The first *in silico* retrobiosynthesis for PKS/NRPS systems were GRAPE and GARLIC. GRAPE deconstructs PK and non-ribosomal peptide (NRP) molecules in reverse, using SMILES-based inputs including complex tailoring reactions to trace back biosynthetic logic through known chemical bonds. GRAPE is also capable of analyzing hybrid PK-NRP structures by disassembling polyketide-extended amino acids and identifying polyketide chain patterns through carbon backbone analysis and oxidation state inference. Complementarily, GARLIC heuristically aligns predicted monomer units from tools like PRISM with those derived from GRAPE, accounting for factors such as gene-product degeneracy, module colinearity, and structural diversity within BGCs.

Whereas GRAPE/GARLIC primarily focused on identifying putative BGCs capable of producing a given PK or NRP structure, the development of ClusterCAD marked a shift toward designing new biosynthetic pathways for *de novo* compound production.<sup>97</sup> The initial release, ClusterCAD 1.0, was tailored specifically for type I modular PKSs and enabled rational design of chimeric PKSs by recombining modules from different origins. Given a target polyketide structure, the toolkit identifies a biosynthetically related intermediate from known PKS modules, which serve as a 'truncated' starter PKS for engineering. From this point, ClusterCAD suggests which modules



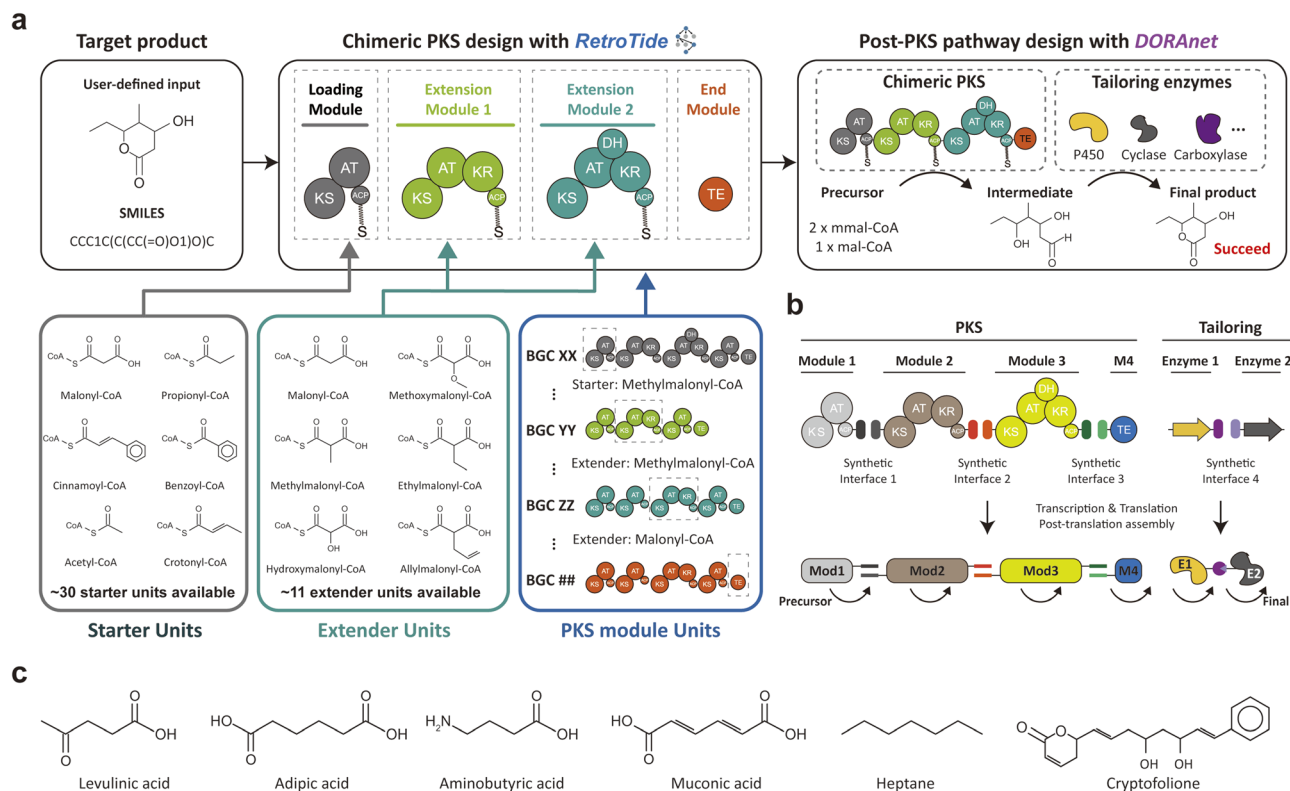


Fig. 5 Application in natural product synthesis. (a) Schematic overview of the BioPKS pipeline integrating *RetroTide* and *DORAnet* for the reconstruction of chimeric PKS pathways and subsequent post-PKS enzymatic modifications. The pipeline accepts a target molecule's SMILES input (e.g., 6-ethyl-4-hydroxy-5-methyloxan-2-one), then predicts suitable loading, extension, and end modules based on available starter units, extender units, and PKS module building blocks. PKS modules are systematically assembled to generate the desired PK backbone. mmal-CoA: methylmalonyl-CoA, mal-CoA: malonyl-CoA. (b) Application of synthetic interfaces to facilitate efficient assembly of chimeric PKS modules, as well as effective recruitment and co-localization of post-PKS tailoring enzymes. (c) Representative examples of PKs synthesized through the integrated BioPKS pipeline. This demonstrates retrobiosynthetic design ranging from simple PKs assembled solely by engineered PKS modules, to complex antibiotic structures requiring integration of chimeric PKSs and additional post-PKS tailoring enzymes.

must be replaced to achieve the production of the desired structure and recommends 'donor modules' based on sequence similarity to the starter PKS. Recent advancements in ClusterCAD 2.0 have further expanded its capabilities to include PKS-NRPS hybrids and NRPS systems, as well as a broader range of starter units.<sup>100</sup> Using ClusterCAD, researchers successfully constructed a variety of chimeric PKSs based on the Rimocidin biosynthetic system, leading to the production of structurally diverse diols and alcohols.<sup>101</sup> These examples highlight the growing utility of PKS-based retrobiosynthetic approaches in enabling the rational development of innovative biomaterials.

A notable addition is the BioPKS pipeline, a newly released framework that enables rational assembly of PKS modules with monofunctional enzymes, facilitating the production of structurally novel polyketides from biosynthetic gene clusters (Fig. 5a). Owing to this pipeline, from simple polyketides (only by chimeric PKSs) to complicated antibiotics (by both chimeric PKSs and post-PKS enzymes), their retrobiosynthetic pathways are successfully predicted (Fig. 5c).<sup>102</sup> Synthetic interfaces can be applied to efficiently construct multi-protein chimeric enzymes, and also to co-localize post-PKS tailoring enzymes for productivity enhancement (Fig. 5b).<sup>103</sup>

## 4.2 Retrobiosynthesis prediction toolkits for NRPS

In the context of NRPS biosynthesis, several retro-biosynthetic toolkits have been developed: rBAN, Nerpa, and BioCAT. The first toolkit specifically developed for NRPS systems was Retro-Biosynthetic analysis of NRPS (rBAN), which retrieves monomeric structures from SMILES input of the target compound.<sup>104</sup> rBAN simulates the retrobiosynthesis of NRPS to predict the required enzymatic machinery and prioritize candidate BGCs for novel compound discovery. It has been successfully integrated with the Norine database, the primary curated repository for NRPS. For NRPS not included in Norine, rBAN can infer their monomeric composition and, when novel monomers are identified, automatically connect to PubChem to annotate them. The incorporation of the Kendrick formula predictor module further allows the estimation of mass-to-charge ratios from mass spectrometry data, enabling structural prediction and automated expansion of the NRPS chemical space with high-quality annotations.

By leveraging rBAN to deduce monomeric structures and using antiSMASH to retrieve NRPS BGC information, subsequent tools such as Nerpa have enabled the linking of target NRP structures with their corresponding biosynthetic gene



clusters. Nerpa demonstrated successful structure-to-gene alignment for 117 BGCs across bacterial genomes.<sup>105</sup> Building upon this framework, BioCAT was introduced with enhanced alignment sensitivity and broader coverage, surpassing both GARLIC and Nerpa in its ability to associate NRPs with their likely producer organisms. BioCAT's primary aim is to maximize the number of accurate NRP-BGC pairings, prioritizing sensitivity over specificity. Importantly, BioCAT accounts for non-classical biosynthetic logic, including type B and type C NRPS pathways, thus offering greater flexibility as a retro-biosynthetic prediction toolkit.<sup>33</sup> Despite the availability of these powerful *in silico* platforms, there has yet to be a reported example where these tools have been applied to guide the combinatorial engineering of NRPS systems for the development of novel bioactive compounds.

Given the modular architecture of PKS/NRPS systems and the growing availability of protein engineering strategies, such as the use of DDs and synthetic coiled-coil motifs, these retro-biosynthetic toolkits are expected to play a critical role in accelerating the rational design of synthetic PKS/NRPSs. Their integration into synthetic biology workflows holds strong potential for the discovery and development of structurally novel and functionally diverse bioactive molecules.

To access a broader chemical space *via* PKS/NRPS pathways, future strategies must integrate megasynth(et)ases-based design tools with existing monofunctional enzyme-centric retrosynthesis platforms. Such integration would enable the generation of diverse PK- and NRP-derivatives which transcend the limitations inherent to module- or domain-only recombination strategies. Moreover, current design tools are often restricted by narrow training datasets, limiting their generalizability. Open-source platforms capable of incorporating both public and user-defined datasets will be critical to improving predictive performance and facilitating broader application.

## 5. Conclusion

The engineering of PKS/NRPS systems has entered an unprecedented era of opportunity, fueled by decades of structural and functional characterization.<sup>106</sup> Recent advances in precise modular boundary definitions, orthologous/xenologous interfaces, and retrobiosynthetic prediction tools have collectively expanded our capability to rationally re-design these complex biosynthetic pathways, marking a clear transition to the realm of synthetic biology.

However, to fully realize PKS/NRPS synthetic biology, understanding the natural constraints governing module-module interactions is essential. These constraints, including structural compatibility of interaction domains, their functional performance across different biosynthetic contexts, and compatibility between inter-modular substrate transfer, serve as foundational principles guiding engineering strategies. A particularly critical constraint lies in the intrinsic proofreading mechanisms embedded within gatekeeper domains.<sup>6,39</sup> While these fidelity filters ensure biosynthetic precision in native systems, they often reject non-cognate intermediates during engineered recombination, limiting the flexibility of modular

design. Current limitations are evident from the approximately 50% success rate observed when bimodular biosynthesis was attempted without prior design considerations.<sup>107</sup> Before achieving true 'plug-and-play' recombination of PKS/NRPS modules to generate designer molecules, both the functional parameters of each biopart and the constraints governing molecule compatibility must be systematically quantified and clearly defined.

The most straightforward approach to addressing this challenge lies in nature itself. Recombining naturally occurring modules and interaction domains in a combinatorial fashion enables systematic quantification of their interactions and elucidation of design principles governing successful module-module communication. This approach requires high-throughput experimentation to capture the diversity of possible combinations. The ClusterCAD database exemplifies this potential. This repository of PKS/NRPS clusters with well-characterized module intermediates contains, for type I PKS alone, 531 loading modules (LM), 2515 elongation modules (EM), and 208 termination modules (END) with TE domains. When incorporating just five distinct synthetic interfaces at each junction of chimeric (LM)-(EM1)-(EM2)-(END) configuration, the number of possible combinations approaches  $\sim 10^{13}$ .<sup>100</sup> This design space expands further when including putative type I PKSs from additional databases, which collectively list over 1600 additional entries.<sup>17</sup> While these numbers exceed practical experimental scale, they demonstrate the vast design space available for modular PKS/NRPS synthetic biology.

Exploring even a fraction of this immense combinatorial space requires experimental capabilities far beyond conventional laboratory methods. Biofoundry provides the necessary infrastructure, offering high-throughput automation and standardized workflows for efficient execution, tracking, and analysis of large-scale combinatorial designs (Fig. 1). These facilities enable massively parallel experiments by integrating design, build, and test workflows through automated protocols. Each workflow comprises multiple unit operations-from PCR and transformation to colony selection-executed by specialized equipment including liquid handlers and automated thermocyclers.<sup>107</sup> The build phase particularly benefits from automated DNA assembly workflows incorporating PCR, purification, ligation and transformation-operations requiring precise coordination for reproducible outcomes at scale. International initiatives like the Global Biofoundry Alliance demonstrate the advanced capabilities of these platforms.<sup>108</sup> The Edinburgh Genome Foundry, for example, performs up to 2000 DNA assembly reactions weekly using methods such as Golden Gate and Gibson assembly, while iBioFAB produces 1000 TALEN constructs daily at less than \$3 per construct.<sup>109,110</sup> These examples highlight the feasibility of implementing automated biofoundry platforms for the systematic exploration of modular assembly strategies at a scale necessary for comprehensive PKS/NRPS engineering.

Early progress in this direction is already evident. Researchers have demonstrated combinatorial PKS library construction involving 120 plasmids ranging from 7 to 14 kb in size, assembled from 4 to 7 DNA fragments.<sup>111</sup> However,



published experimental data also reveals a significant bottleneck in applying *in vivo* molecular cloning approaches to PKS/NRPS modular engineering at biofoundry scale. In one study, researchers generated 882 plasmid requiring amplification of 502 unique DNA fragments through 706 PCR reactions, each incorporating 60-bp homologous overlaps for yeast-based recombination. Of these, only 623 fragments were successfully confirmed by capillary electrophoresis, enabling theoretical assembly of 715 out of the 882 plasmids. Following yeast transformation and survival screening, the final recovery rate of mutation-free plasmid constructs was approximately 14% of total colonies, highlighting limitations in throughput and fidelity.

Cell-free protein synthesis (CFPS) integrated with automated biofoundry workflows offers a promising solution to these challenges (Fig. 1). Recent studies have shown that CFPS can achieve protein expression levels comparable to *in vivo* systems when using the same DNA constructs.<sup>112–114</sup> By employing PCR-amplified linear templates, CFPS circumvents the need for yeast-based cloning, transformation, colony picking, and full validation of coding sequences, thereby accelerating the build phase and reducing resource intensity. Additionally, CFPS-enabled screening allows for rapid functional assessment of engineered constructs. Emerging sequential-phase CFPS systems further enhance capabilities by enabling protein expression followed by precursor-to-product conversion in single-well reactions, effectively creating miniaturized biocatalytic testbeds for megasynth(et)ase functionality.

The integration of biofoundry platforms represents a crucial advancement in PKS/NRPS engineering, enabling experimental throughput at unprecedented scale. Building upon this foundation, computational approaches can further amplify biofoundry capabilities by guiding experimental design and interpreting results. AI-driven methods work synergistically with biofoundry infrastructure, enhancing each phase of the DBTL cycle and maximizing the value of high-throughput experimental data.

To support modular enzyme engineering in natural product biosynthesis, AI-driven structural prediction tools have emerged as valuable complements to automated biofoundry platforms. Among them, AlphaFold offers high-resolution structural predictions of domain–domain interfaces, facilitating the rational design of synthetic connections.<sup>115</sup> While primarily developed for monomeric protein structure prediction, AlphaFold's residue-level resolution can be repurposed to inform linker placement and suggest plausible domain boundaries for modular recombination.<sup>115,116</sup> In particular, predicted folding patterns and surface accessibility maps can help identify plausible domain boundaries that may minimize structural disruption during module fusion. However, the utility of such predictions in complex, multi-domain systems like PKS or NRPS remains largely unvalidated and demands empirical verification.

RFdiffusion contributes scaffold-generation capabilities that extend beyond structure prediction to design novel interface architectures, including specialized coiled-coil or plug-and-socket configurations.<sup>117–120</sup> This approach enables synthetic

interface scaffold recommendation-predicting which interface type will best accommodate specific domain pairs while maintaining proper orientation and communication. RFdiffusion expands the design space by generating novel interface scaffolds fit to the geometric constraints of specific domain pairs.

ProteinMPNN complements structure-based models by optimizing amino acid sequences at domain junctions, supporting folding and expression compatibility.<sup>121</sup> This is particularly useful in modular engineering scenarios where non-cognate domains are recombined, and local sequence adjustments are needed to maintain interface integrity. While MPNN does not guarantee functional restoration for domains with altered substrate preferences, it assists in refining sequence contexts to promote structural coherence across engineered boundaries.

As illustrated in Fig. 1, these computational approaches specifically facilitate critical tasks such as AI-guided linker selection and precise insertion site prediction, further enhancing modular engineering efficiency. Advanced AI frameworks combining graph neural networks (GNNs) and large language models (LLMs) offer promising approaches for optimizing modular biosynthetic pathways. Looking ahead, next-generation AI models—such as those based on GNNs or LLMs—may offer new paradigms for pathway design by interpreting target chemical structures and predicting optimal module arrangements.<sup>122</sup> GNNs enable modeling of the complex relationships between PKS/NRPS domains as interconnected nodes, allowing critical interfaces such as ACP-KS connections to be evaluated for structural compatibility and assigned meaningful interaction scores. These graph-based representations enhance predictive accuracy by capturing the inherent modularity of biosynthetic pathways, enabling systematic assessment of potential domain arrangements and their likely functional outcomes. Complementing this structural analysis, LLMs, leveraging biochemical data, may assist in translating target product specifications into suitable modular enzyme arrangements. These predictive systems have the potential to interpret target chemical structures (*e.g.*, from SMILES), automatically identifying thioesterization sites, distinguishing initial starter units from subsequently added extender units, and systematically recognizing required modifications and their corresponding enzymatic domains.

The implementation of such frameworks follows an iterative learning cycle. Initial linker designs predicted by AI ensembles are experimentally validated, with performance data feeding back into model training pipelines. This recursive approach creates self-improving systems where each design cycle enhances model accuracy. Crucially, as experimental databases expand through biofoundry-scale testing in the build and test cycle, AI models transition from few-shot learning scenarios to increasingly sophisticated prediction algorithms capable of addressing complex domain interface problems in PKS/NRPS engineering efforts.

The integration of AI-driven design with biofoundry implementation provides new opportunities in PKS/NRPS engineering, enhancing our ability to analyze domain interactions. These computational approaches generate interfaces with



diverse interaction properties, allowing researchers to select connections with precisely calibrated strengths for specific module pairs. While natural systems may exhibit complex interdependencies between adjacent domains, computational frameworks can provide insights into designing interfaces that function more independently, simplifying the engineering process. As experimental data accumulates, AI systems evolve from data-limited models to sophisticated predictive tools leveraging thousands of experimental outcomes, progressively enhancing our understanding of these complex biosynthetic interfaces and accelerating the discovery and research of natural products.

## 6. Data availability

No primary research results, datasets, software, or code have been included, and no new data were generated or analysed as part of this review. Therefore, data availability statements are not applicable for this article.

## 7. Author contributions

Byung-Kwan Cho conceived and supervised the manuscript. Gahyeon Kim, Dukwon Lee, Ji Hun Kim, Seong Do Kim, Hongki Kim, Jae Heon Kim, Sung Sun Yim, Soo-Jin Yeom, Jay D. Keasling, and Byung-Kwan Cho wrote the manuscript. All authors proofread the entire manuscript and provided suggestions for improvement on all sections.

## 8. Conflicts of interest

There are no conflicts to declare.

## 9. Acknowledgements

This research was supported by the Bio&Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2021-NR056596, RS-2021-NR-056566, and RS-2024-00352229). This research was also supported by a grant of the Korea-US Collaborative Research Fund (KUCRF), funded by the Ministry of Science and ICT and Ministry of Health & Welfare, Republic of Korea (grant number: RS-2024-00468410).

## 10. References

- D. A. Dias, S. Urban and U. Roessner, *Metabolites*, 2012, **2**, 303–336.
- F. Prestinaci, P. Pezzotti and A. Pantosti, *Pathog. Global Health*, 2015, **109**, 309–318.
- A. M. Fuentefria, B. Pippi, D. F. Dalla Lana, K. K. Donato and S. F. de Andrade, *Lett. Appl. Microbiol.*, 2018, **66**, 2–13.
- R. Y. Qu, B. He, J. F. Yang, H. Y. Lin, W. C. Yang, Q. Y. Wu, Q. X. Li and G. F. Yang, *Pest Manage. Sci.*, 2021, **77**, 2620–2625.
- S. Hwang, N. Lee, S. Cho, B. Palsson and B.-K. Cho, *Front. Mol. Biosci.*, 2020, **7**, 87.
- D. Yi, M. A. Wakeel and V. Agarwal, *Biochemistry*, 2024, **63**, 2240–2244.
- Y. Liu, C. Song, Q. Cui, H. Sun, C. Jiang, R. Guo, R. He, Z. Li, J. Luan and H. Wang, *Nat. Commun.*, 2025, **16**, 774.
- S. A. Benner and A. M. Sismour, *Nat. Rev. Genet.*, 2005, **6**, 533–543.
- R. Moore, A. Chandradas and L. Bleris, *ACS Synth. Biol.*, 2014, **3**, 708–716.
- H. Redden, N. Morse and H. S. Alper, *FEMS Yeast Res.*, 2015, **15**, 1–10.
- H. M. Salis, E. A. Mirsky and C. A. Voigt, *Nat. Biotechnol.*, 2009, **27**, 946–950.
- M. P. Crump, J. Crosby, C. E. Dempsey, J. A. Parkinson, M. Murray, D. A. Hopwood and T. J. Simpson, *Biochemistry*, 1997, **36**, 6000–6008.
- D. E. Cane, C. T. Walsh and C. Khosla, *Science*, 1998, **282**, 63–68.
- K. Farzam, T. A. Nessel and J. Quick, in *StatPearls [Internet]*, StatPearls Publishing, 2023.
- D. P. Cogan, K. Zhang, X. Li, S. Li, G. D. Pintilie, S.-H. Roh, C. S. Craik, W. Chiu and C. Khosla, *Science*, 2021, **374**, 729–734.
- S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. Narayan, K. Håkansson, D. H. Sherman and J. L. Smith, *Nature*, 2014, **510**, 512–517.
- A. Nivina, K. P. Yuet, J. Hsu and C. Khosla, *Chem. Rev.*, 2019, **119**, 12524–12547.
- J. L. Smith, G. Skiniotis and D. H. Sherman, *Curr. Opin. Struct. Biol.*, 2015, **31**, 9–19.
- M. Klaus and M. Grninger, *Nat. Prod. Rep.*, 2018, **35**, 1070–1081.
- K. Kudo, T. Hashimoto, J. Hashimoto, I. Kozono, N. Kagaya, R. Ueoka, T. Nishimura, M. Komatsu, H. Suenaga and H. Ikeda, *Nat. Commun.*, 2020, **11**, 4022.
- T. J. Buchholz, T. W. Geders, F. E. Bartley III, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.
- R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay and K. J. Weissman, *Chem. Biol.*, 2003, **10**, 723–731.
- M. Klaus, A. D. D'Souza, A. Nivina, C. Khosla and M. Grninger, *ACS Chem. Biol.*, 2019, **14**, 426–433.
- A. Pistofidis, P. Ma, Z. Li, K. Munro, K. Houk and T. M. Schmeing, *Nature*, 2025, **638**, 270–278.
- G. W. Heberlig, J. J. La Clair and M. D. Burkart, *Nature*, 2025, **638**, 261–269.
- C. M. Fortinez, K. Bloudoff, C. Harrigan, I. Sharon, M. Strauss and T. M. Schmeing, *Nat. Commun.*, 2022, **13**, 548.
- J. Wang, D. Li, L. Chen, W. Cao, L. Kong, W. Zhang, T. Croll, Z. Deng, J. Liang and Z. Wang, *Nat. Commun.*, 2022, **13**, 592.
- M. A. Marahiel, *Nat. Prod. Rep.*, 2016, **33**, 136–140.
- A. S. Brown, M. J. Calcott, J. G. Owen and D. F. Ackerley, *Nat. Prod. Rep.*, 2018, **35**, 1210–1228.
- A. Tanovic, S. A. Samel, L.-O. Essen and M. A. Marahiel, *Science*, 2008, **321**, 659–663.



- 31 K. D. Patel, M. R. MacDonald, S. F. Ahmed, J. Singh and A. M. Gulick, *Nat. Prod. Rep.*, 2023, **40**, 1550–1582.
- 32 M. A. Skiba, F. P. Maloney, Q. Dan, A. E. Fraley, C. C. Aldrich, J. L. Smith and W. C. Brown, *Methods Enzymol.*, 2018, **604**, 45–88.
- 33 R. D. Sussmuth and A. Mainz, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 3770–3821.
- 34 C. D. Fage, S. Kosol, M. Jenner, C. Oster, A. Gallo, M. Kaniusaite, R. Steinbach, M. Staniforth, V. G. Stavros and M. A. Marahiel, *ACS Catal.*, 2021, **11**, 10802–10813.
- 35 H. G. Smith, M. J. Beech, J. R. Lewandowski, G. L. Challis and M. Jenner, *J. Ind. Microbiol. Biotechnol.*, 2021, **48**, kuab018.
- 36 A. T. Keatinge-Clay, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 4658–4660.
- 37 L. Zhang, T. Hashimoto, B. Qin, J. Hashimoto, I. Kozono, T. Kawahara, M. Okada, T. Awakawa, T. Ito and Y. Asakawa, *Angew. Chem., Int. Ed.*, 2017, **56**, 1740–1745.
- 38 T. Miyazawa, M. Hirsch, Z. Zhang and A. T. Keatinge-Clay, *Nat. Commun.*, 2020, **11**, 80.
- 39 M. Hirsch, B. J. Fitzgerald and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2021, **16**, 2515–2526.
- 40 D. A. Vander Wood and A. T. Keatinge-Clay, *Proteins: Struct., Funct., Bioinf.*, 2018, **86**, 664–675.
- 41 K. A. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert and H. B. Bode, *Nat. Chem.*, 2018, **10**, 275–281.
- 42 R. S. Gokhale, S. Y. Tsuji, D. E. Cane and C. Khosla, *Science*, 1999, **284**, 482–485.
- 43 S. Y. Tsuji, D. E. Cane and C. Khosla, *Biochemistry*, 2001, **40**, 2326–2331.
- 44 L. Su, L. Hotel, C. Paris, C. Chepkirui, A. O. Brachmann, J. Piel, C. Jacob, B. Aigle and K. J. Weissman, *Nat. Commun.*, 2022, **13**, 515.
- 45 G. Zhai, Y. Zhu, G. Sun, F. Zhou, Y. Sun, Z. Hong, C. Dong, P. F. Leadlay, K. Hong, Z. Deng, F. Zhou and Y. Sun, *Nat. Commun.*, 2023, **14**, 612.
- 46 L. Gao, W. Ma, Z. Lu, J. Han, Z. Ma, H. Liu and X. Bie, *Synth. Syst. Biotechnol.*, 2022, **7**, 1173–1180.
- 47 X. Cai, L. Zhao and H. B. Bode, *Org. Lett.*, 2019, **21**, 2116–2120.
- 48 X. Cai, L. Zhao and H. B. Bode, *ACS Synth. Biol.*, 2023, **12**, 203–212.
- 49 K. P. Yuet, C. W. Liu, S. R. Lynch, J. Kuo, W. Michaels, R. B. Lee, A. E. McShane, B. L. Zhong, C. R. Fischer and C. Khosla, *J. Am. Chem. Soc.*, 2020, **142**, 5952–5957.
- 50 D. Yi, D. Niroula, W. R. Gutekunst, J. E. Loper, Q. Yan and V. Agarwal, *ACS Chem. Biol.*, 2022, **17**, 1351–1356.
- 51 D. Yi and V. Agarwal, *ACS Chem. Biol.*, 2023, **18**, 1060–1065.
- 52 C. Kegler and H. B. Bode, *Angew. Chem., Int. Ed. Engl.*, 2020, **59**, 13463–13467.
- 53 W. B. Porterfield, N. Poenateetai and W. Zhang, *iScience*, 2020, **23**, 100938.
- 54 J. L. Meinke, A. J. Simon, D. T. Wagner, B. R. Morrow, S. You, A. D. Ellington and A. T. Keatinge-Clay, *ACS Synth. Biol.*, 2019, **8**, 2017–2024.
- 55 Z. Li, Z. Zhu, G. Xu, L. Wei, J. Liu, H. Wang, C. Lu, Y. Li, D. Zhu and Y. Shen, *ACS Catal.*, 2024, **14**, 8062–8072.
- 56 X. Sun, Y. Yuan, Q. Chen, S. Nie, J. Guo, Z. Ou, M. Huang, Z. Deng, T. Liu and T. Ma, *Nat. Commun.*, 2022, **13**, 5541.
- 57 K. L. Garner, *Essays Biochem.*, 2021, **65**, 791–811.
- 58 J. A. J. Arpino, E. J. Hancock, J. Anderson, M. Barahona, G. V. Stan, A. Papachristodoulou and K. Polizzi, *Microbiology (Reading)*, 2013, **159**, 1236–1253.
- 59 K. E. Thompson, C. J. Bashor, W. A. Lim and A. E. Keating, *ACS Synth. Biol.*, 2012, **1**, 118–129.
- 60 C. Negron and A. E. Keating, *J. Am. Chem. Soc.*, 2014, **136**, 16544–16556.
- 61 A. W. Reinke, R. A. Grant and A. E. Keating, *J. Am. Chem. Soc.*, 2010, **132**, 6025–6031.
- 62 M. Klaus, L. Buyachuihan and M. Grininger, *ACS Chem. Biol.*, 2020, **15**, 2422–2432.
- 63 H.-M. Huang, P. Stephan and H. Kries, *Cell Chem. Biol.*, 2021, **28**, 221–227.
- 64 K. A. Bozhueyuek, J. Watzel, N. Abbood and H. B. Bode, *Angew. Chem., Int. Ed.*, 2021, **60**, 17531–17538.
- 65 N. Abbood, T. Duy Vo, J. Watzel, K. A. Bozhueyuek and H. B. Bode, *Chem.–Eur. J.*, 2022, **28**, e202103963.
- 66 B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E690–E697.
- 67 Y. Wang, Y. Chang, R. Jia, H. Sun, J. Tian, H. Luo, H. Yu and Z. Shen, *Process Biochem.*, 2020, **95**, 260–268.
- 68 S. Huang, F. Ba, W. Q. Liu and J. Li, *Biotechnol. Bioeng.*, 2023, **120**, 793–802.
- 69 M. Y. Ali, Q. Chang, Y. Su, J. Wu, Q. Yan, L. Yin, Y. Zhang and Y. Feng, *ACS Appl. Bio Mater.*, 2021, **4**, 3027–3034.
- 70 Y. Liu, D. Liu, W. Yang, X.-L. Wu, L. Lai and W.-B. Zhang, *Chem. Sci.*, 2017, **8**, 6577–6582.
- 71 A. Miyanaga, F. Kudo and T. Eguchi, *Nat. Prod. Rep.*, 2018, **35**, 1185–1209.
- 72 N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, **5**, 446–461.
- 73 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, **20**, 1340–1351.
- 74 A. H. Keeble, P. Turkki, S. Stokes, I. N. Khairil Anuar, R. Rahikainen, V. P. Hytönen and M. Howarth, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 26523–26533.
- 75 F. Pinto, E. L. Thornton and B. Wang, *Nat. Commun.*, 2020, **11**, 1529.
- 76 L. Buyachuihan, Y. Zhao, C. Schelhas and M. Grininger, *ACS Chem. Biol.*, 2023, **18**, 1500–1509.
- 77 N. Abbood, J. Effert, K. A. Bozhueyuek and H. B. Bode, *ACS Synth. Biol.*, 2023, **12**, 2432–2443.
- 78 H. Wang, L. Wang, B. Zhong and Z. Dai, *Front. Bioeng. Biotechnol.*, 2022, **10**, 810180.
- 79 S. Anastassov, M. Filo and M. Khammash, *Biotechnol. Adv.*, 2024, 108349.
- 80 C. J. Noren, J. Wang and F. B. Perler, *Angew. Chem., Int., Ed.*, 2000, **39**, 450–466.
- 81 N. I. Topilina and K. V. Mills, *Mobile DNA*, 2014, **5**, 1–14.
- 82 T. Kurpiers and H. D. Mootz, *ChemBioChem*, 2008, **9**, 2317–2325.



- 83 A. Tavassoli, *Curr. Opin. Chem. Biol.*, 2017, **38**, 30–35.
- 84 G. Deschuyteneer, S. P. Garcia, B. Michiels, B. Baudoux, H. Degand, P. Morsomme and P. Soumillion, *ACS Chem. Biol.*, 2010, **5**, 691–700.
- 85 A. Pistofidis and T. M. Schmeing, *RSC Chem. Biol.*, 2025, **6**, 590–603.
- 86 C. A. Dejong, G. M. Chen, H. Li, C. W. Johnston, M. R. Edwards, P. N. Rees, M. A. Skinnider, A. L. Webster and N. A. Magarvey, *Nat. Chem. Biol.*, 2016, **12**, 1007–1014.
- 87 M. Kaniusaite, J. Tailhades, E. A. Marschall, R. J. Goode, R. B. Schittenhelm and M. J. Cryle, *Chem. Sci.*, 2019, **10**, 9466–9482.
- 88 M. J. Calcott, J. G. Owen and D. F. Ackerley, *Nat. Commun.*, 2020, **11**, 4554.
- 89 F. Pourmasoumi, S. De, H. Peng, F. Trottmann, C. Hertweck and H. Kries, *ACS Chem. Biol.*, 2022, **17**, 2382–2388.
- 90 K. Jensen, H. Niederkrüger, K. Zimmermann, A. L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Pöplau, C. Kohlhaas and C. A. Townsend, *Chem. Biol.*, 2012, **19**, 329–339.
- 91 N. Abbood, L. Prëve, K. A. Bozhueyuek and H. B. Bode, in *Non-Ribosomal Peptide Biosynthesis and Engineering: Methods and Protocols*, Springer, 2023, pp. 219–234.
- 92 A. A. Arishi, Z. Shang, E. Lacey, A. Crombie, D. Vuong, H. Li, J. Bracegirdle, P. Turner, W. Lewis, G. R. Flematti, A. M. Piggott and Y. H. Chooi, *Chem. Sci.*, 2024, **15**, 3349–3356.
- 93 N. E. Avalon, A. E. Murray, H. E. Daligault, C.-C. Lo, K. W. Davenport, A. E. Dichosa, P. S. Chain and B. J. Baker, *Front. Chem.*, 2021, **9**, 802574.
- 94 L. Chen, X. Wang, Y. Zou and M. C. Tang, *Org. Lett.*, 2024, **26**, 3597–3601.
- 95 J. Courtial, J.-J. Helesbeux, H. Oudart, S. Aligon, M. Bahut, B. Hamon, G. N'guyen, S. Pigné, A. G. Hussain and C. Pascouau, *Sci. Rep.*, 2022, **12**, 8155.
- 96 S. Zhao, C. Lu, H. Wang, Y. Li and Y. Shen, *Org. Lett.*, 2023, **25**, 6954–6958.
- 97 C. H. Eng, T. W. H. Backman, C. B. Bailey, C. Magnan, H. García Martín, L. Katz, P. Baldi and J. D. Keasling, *Nucleic Acids Res.*, 2018, **46**, D509–D515.
- 98 A. A. Nava, J. Roberts, R. W. Haushalter, Z. Wang and J. D. Keasling, *ACS Synth. Biol.*, 2023, **12**, 3148–3155.
- 99 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat. Methods*, 2022, **19**, 679–682.
- 100 X. B. Tao, S. LaFrance, Y. Xing, A. A. Nava, H. G. Martin, J. D. Keasling and T. W. Backman, *Nucleic Acids Res.*, 2023, **51**, D532–D538.
- 101 Q. Dan, Y. Chiu, N. Lee, J. H. Pereira, B. Rad, X. Zhao, K. Deng, Y. Rong, C. Zhan, Y. Chen, S. Cheong, C. Li, J. W. Gin, A. Rodrigues, T. R. Northen, T. W. H. Backman, E. E. K. Baidoo, C. J. Petzold, P. D. Adams and J. D. Keasling, *Nat. Catal.*, 2025, **8**, 147–161.
- 102 Y. Chainani, J. Diaz, M. Guilarte-Silva, V. Blay, Q. Zhang, W. Sprague, K. E. Tyo, L. J. Broadbelt, A. Mukhopadhyay and J. D. Keasling, *bioRxiv*, 2024, preprint, DOI: [10.1101/2024.11.04.621673](https://doi.org/10.1101/2024.11.04.621673).
- 103 H. Lee, W. C. DeLoache and J. E. Dueber, *Metab. Eng.*, 2012, **14**, 242–251.
- 104 E. Ricart, V. Leclere, A. Flissi, M. Mueller, M. Pupin and F. Lisacek, *J. Cheminform.*, 2019, **11**, 13.
- 105 O. Kunyavskaya, A. M. Tagirdzhanov, A. M. Caraballo-Rodriguez, L. F. Nothias, P. C. Dorrestein, A. Korobeynikov, H. Mohimani and A. Gurevich, *Metabolites*, 2021, **11**, 693.
- 106 A. Iram, Y. Dong and C. Ignea, *Curr. Opin. Biotechnol.*, 2024, **87**, 103143.
- 107 A. Stephenson, L. Lastra, B. Nguyen, Y.-J. Chen, J. Nivala, L. Ceze and K. Strauss, *ACS Synth. Biol.*, 2023, **12**, 3156–3169.
- 108 N. Hillson, M. Caddick, Y. Cai, J. A. Carrasco, M. W. Chang, N. C. Curach, D. J. Bell, R. Le Feuvre, D. C. Friedman and X. Fu, *Nat. Commun.*, 2019, **10**, 2040.
- 109 T. Tang, L. Fu, E. Guo, Z. Zhang, Z. Wang, C. Ma, Z. Zhang, J. Zhang, J. Huang and T. Si, *Chin. Sci. Bull.*, 2021, **66**, 300–309.
- 110 R. Chao, J. Liang, I. Tasan, T. Si, L. Ju and H. Zhao, *ACS Synth. Biol.*, 2017, **6**, 678–685.
- 111 A. A. Nava, A. L. Fear, N. Lee, P. Mellinger, G. Lan, J. McCauley, S. Tan, N. Kaplan, G. Goyal and R. C. Coates, *ACS Synth. Biol.*, 2023, **12**, 3506–3513.
- 112 J. Shin and V. Noireaux, *ACS Synth. Biol.*, 2012, **1**, 29–41.
- 113 J. Chappell, K. Jensen and P. S. Freemont, *Nucleic Acids Res.*, 2013, **41**, 3471–3481.
- 114 Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux and R. M. Murray, *ACS Synth. Biol.*, 2014, **3**, 387–397.
- 115 J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard and J. Bambrick, *Nature*, 2024, **630**, 493–500.
- 116 C. Elfmann and J. Stülke, *Nucleic Acids Res.*, 2023, **51**, W404–W410.
- 117 J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte and L. F. Milles, *Nature*, 2023, **620**, 1089–1100.
- 118 A. Lauko, S. J. Pellock, K. H. Sumida, I. Anishchenko, D. Juergens, W. Ahern, J. Jeung, A. Shida, A. Hunt and I. Kalvet, *Science*, 2025, eadu2454.
- 119 J. Min, X. Rong, J. Zhang, R. Su, Y. Wang and W. Qi, *J. Chem. Theory Comput.*, 2024, **20**, 532–550.
- 120 W. Yang, D. R. Hicks, A. Ghosh, T. A. Schwartze, B. Coventry, I. Goreshnik, A. Allen, S. F. Halabiya, C. J. Kim and C. S. Hinck, *Nat. Commun.*, 2025, **16**, 2001.
- 121 J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. Wicky, A. Courbet, R. J. de Haas and N. Bethel, *Science*, 2022, **378**, 49–56.
- 122 J. Xu, Z. Wu, M. Lin, X. Zhang and S. Wang, *arXiv*, 2024, preprint, arXiv:2406.01032, DOI: [10.48550/arXiv.2406.01032](https://doi.org/10.48550/arXiv.2406.01032).

