ACS Synthetic Biology

pubs.acs.org/synthbio
Review

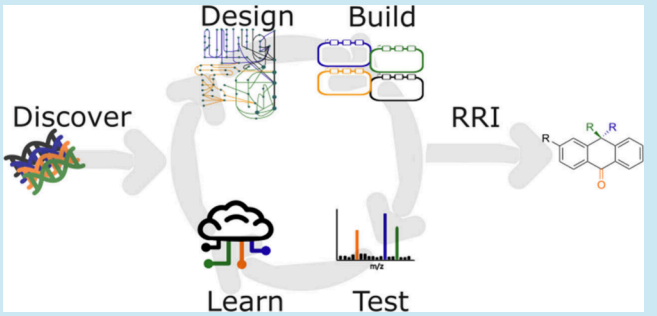
Synthetic Biology of Natural Products Engineering: Recent Advances Across the Discover–Design–Build–Test–Learn Cycle

Jonathan Foldi,[#] Jack A. Connolly,[#] Eriko Takano,[*] and Rainer Breitling

Cite This: *ACS Synth. Biol.* 2024, 13, 2684–2692
Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Advances in genome engineering and associated technologies have reinvigorated natural products research. Here we highlight the latest developments in the field across the discover–design–build–test–learn cycle of bioengineering, from recent progress in computational tools for AI-supported genome mining, enzyme and pathway engineering, and compound identification to novel host systems and new techniques for improving production levels, and place these trends in the context of responsible research and innovation, emphasizing the importance of anticipatory analysis at the early stages of process development.



KEYWORDS: *natural products, biosynthetic gene clusters, synthetic biology, genome mining, strain engineering, machine learning*

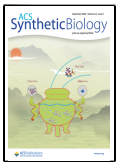
INTRODUCTION

Natural products (NPs), also known as secondary metabolites, form the basis for many products of biotechnology, including antimicrobials, chemotherapeutic agents, and pesticides.[1] This has been the case throughout human history, with the use of natural remedies being documented as far back as the Mesopotamians in 2600 BCE.[2,3] These natural sources are still just as relevant, as from 1981 to 2019, 64% of antimicrobials (excluding vaccines) either were NP-derived, contained an NP pharmacore, or were synthetic NP mimics.[4] Since the "Golden Age" of antibiotic discovery in the mid-20th century, the discovery of new antibiotics has dropped off dramatically.[5] The need for novel antimicrobials remains paramount, as antimicrobial resistance is one of the major global health challenges for this century, already causing an estimated 4.95 million deaths globally in 2019 and predicted to increase substantially in the future.[7,8] The varied structures of NPs, along with the high levels of stereochemistry relative to synthetic compounds, are highlighted in Figure 1.

In order to address this issue, significant research was directed at NP discovery and engineering after advances in genomics had reopened the field as a viable alternative to high-throughput screens of combinatorial chemical libraries.[9] Underlying much of this genomic revolution is the fact that the genes which encode the biosynthetic machinery for the production of a natural product are typically colocated in close proximity in the genome of a bacterium or fungus, forming biosynthetic gene clusters (BGCs) that serve as functional evolutionary units for the horizontal transfer of biosynthetic capabilities. This has allowed a relatively straightforward linking of natural products to genomic data that has made it easier to avoid the duplication issues which frequently arose from screening-based discovery methods, thus streamlining the discovery process.[10] These efforts have led to high numbers of unique natural products being characterized, with four widely used and steadily growing NP repositories from across academia and industry that contain only deduplicated data (Table 1). It is worth highlighting the Collection of Open Natural Products (COCONUT), which was created specifically to address the overproliferation of NP databases, many of which are insufficiently maintained after their initial publication.[11]

However, a large fraction of the biosynthetic potential remains underexplored in its biochemical detail: when analyzed at the level of gene cluster families (GCFs), collections of BGCs sharing a similar architecture and expected to produce similar NPs, it is estimated that fewer than 3% of GCFs have had their biosynthesis routes experimentally characterized and documented in a standardized way.[15] This can be due to a variety of challenges, the most important being the limited scalability of the experimental elucidation of biosynthetic pathways; however, even a well-established NP like paclitaxel, which is used as a chemotherapeutic agent for thousands of patients, had its biosynthetic pathway elucidated in its entirety

Received: May 31, 2024
Revised: August 9, 2024
Accepted: August 9, 2024
Published: August 20, 2024



https://doi.org/10.1021/acssynbio.4c00391
ACS Synth. Biol. 2024, 13, 2684–2692

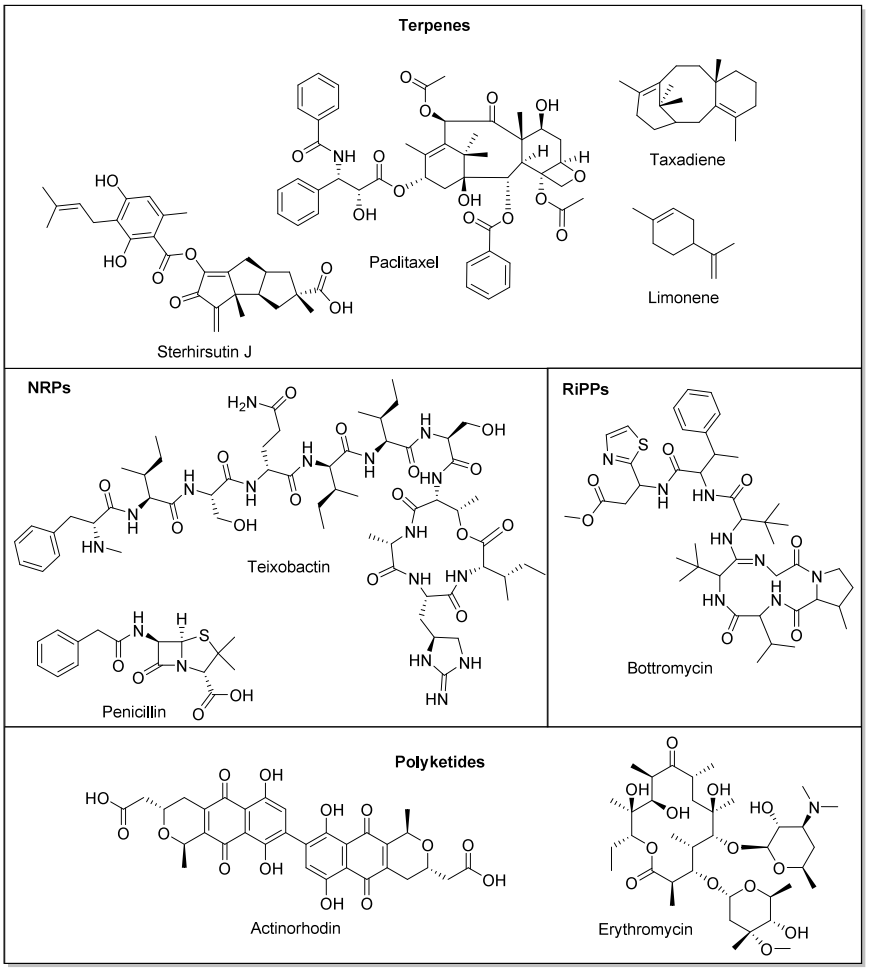


Figure 1. Example compounds from four of the major classes of NPs: terpenes, nonribosomal peptides (NRPs), ribosomally synthesized and post-translationally modified peptides (RiPPs), and polyketides.

Table 1. Repositories for Natural Products

name	number of NP entries	notes	ref
Natural Products Atlas	32,552 NPs	only microbial NPs	12
SuperNatural III	449,058 NPs and derivatives	includes NPs and derivatives of NPs	13
Dictionary of Natural Products	>340,000 NPs	only commercially available	
COCONUT	>400,000 NPs	compilation of 54 open-access NP databases	14

only very recently due to its biosynthesis genes not being clustered, as can often be the case for such plant-derived NPs.[16]

A typical route to NP discovery is the following: a BGC of interest, identified through genome mining, is endogenously or heterologously expressed *in vivo*, and its associated NP is purified and characterized through spectral analysis.[17,18] Our review reframes this standard workflow into an adapted Discover−Design−Build−Test−Learn cycle to highlight the existing and future opportunities for synthetic biology in this field. Additionally, we review recent progress in the development of responsible research and innovation practices that are commonly applied in synthetic biology research and could make a valuable contribution in natural product development (Figure 2).

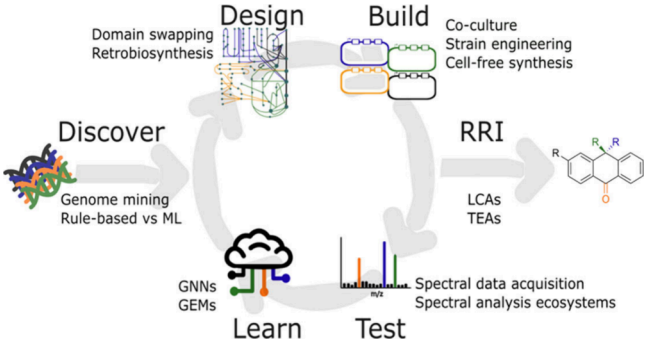


Figure 2. Diagram showing the Discover−Design−Build−Test−Learn cycle for NP engineering, with a few key aspects of each stage listed. GNN, graph neural network; GEM, genome-scale metabolic model; LCA, life cycle analysis; TEA, techno-economic assessment.

DISCOVER

A necessary first step in most NP engineering workflows is the use of genome mining to discover the genes responsible for the biosynthesis of the NPs. In microorganisms, most NPs are coded for by BGCs which can be highly variable in the number of genes they contain.[19] The number of BGCs present in a genome can also vary, with some microorganisms having over 80 BGCs in their genomes.[20] Recent advances in long-read

https://doi.org/10.1021/acssynbio.4c00391
ACS Synth. Biol. 2024, 13, 2684−2692

Table 2. Software Tools for Genome Mining and Analysis and Repositories of Annotated BGCs

name	type of tool	organisms searched	notes	ref
antiSMASH	genome mining	bacteria, fungus, archaea, plant	flagship genome mining tool for NPs; different flavors depending on the organism/biome of choice	26
PRISM 4	genome mining	bacteria	emphasis on predicting chemical structures of predicted BGC products	30
EvoMining	genome mining	bacteria	focus on overlooked enzyme classes through an evolutionary lens	31
GECCO	genome mining	bacteria	uses a specific subset of ML, conditional random field, which makes it more interpretable than neural network "black box" tools	28
SanntiS	genome mining	bacteria	uses an artificial neural network and aims to better identify less-characterized BGCs	29
BiG-SLiCE	phylogenetic analysis	bacteria, archaea	generates gene cluster families from BGCs	32
CORASON	phylogenetic analysis	bacteria	shows evolutionary relationships between BGCs within a gene cluster family using a multilocus phylogeny	33
MIBiG	repository	bacteria, archaea, fungi, plants	frequently updated and validated by subject matter experts	34
antiSMASH-DB	repository	bacteria, archaea, fungi	not as well curated as MIBiG; depends on the quality of the antiSMASH analysis of a genome	35

sequencing and culturing have offered means to navigate the estimated 99% of microbial species which are not culturable under standard laboratory conditions.[21] The novel antibiotic teixobactin was discovered in this manner using iChip technology to culture *Eleftheria terrae*, which previously had not been culturable.[22] Even though this drug has yet to live up to the initial high expectations, the iChip technology continues to be used for discovering novel BGCs.[23] This technique utilizes a small device with hundreds of pores which sequester microorganisms to grow individually while exchanging nutrients with their natural environment *via* diffusion through a semipermeable membrane before being brought back to the lab for testing. Environments which harbor these unculturable microorganisms, such as deserts and wastewater treatment plants, have recently been shown to harbor a wide swath of microorganisms which can now be sequenced and exploited.[24,25]

To discover the BGCs hidden in these genomes, a suite of tools can be used. AntiSMASH,[26] the most widely used BGC discovery tool, can annotate sequences from a variety of kingdoms and output a list of BGCs and the corresponding NP classes to which their expected products belong. Different flavors of antiSMASH account for differences in BGCs across kingdoms. For example, plantiSMASH accounts for several distinct aspects of plant BGCs relative to microbial BGCs (unique enzyme families, unclustered enzymatic pathways, higher variability in intergenic distances) to enable more sensitive detection.[27] The antiSMASH ecosystem also includes tools for interrogating samples from different biomes, such as rhizoSMASH and gutSMASH, which enable BGC discovery in the rhizosphere and gut microbiome, respectively. These tools allow for the detection of BGCs across a wide variety of samples and potential use cases, highlighting the varied potential of natural products across several different sectors. One limitation, however, is that antiSMASH is a rule-based tool and thus does not perform as well as ML-based tools in discovering novel BGCs and unclustered pathways, in contrast, *e.g.*, with GECCO, which found almost twice as many BGCs as antiSMASH in the proGenomes2 database, though a direct comparison has not yet been done using the newest release of antiSMASH.[28] Similarly, SanntiS, another ML-based tool, can detect more novel BGCs than antiSMASH and was shown to have the highest precision−recall performance of these three genome mining tools on a frequently used "9 genomes" dataset.[29] Phylogenetic analysis of BGCs can be performed with tools such as BiG-SLICE and CORASON, which can create GCFs to enable comparison across clusters. These pieces of software are listed in Table 2, along with frequently used repositories of annotated BGCs.

DESIGN

Understanding the biosynthetic pathway of an NP is critical to designing efficient biosynthesis strategies, in addition to engineering changes to the product. The modular nature of many NP biosynthesis pathways allows for domain swapping as a popular method of engineering.[36,37] This approach can help in understanding the functions of specific enzymes in biosynthesis[38] or to create novel compounds by mixing enzymes from different biosynthesis pathways in order to generate novel NP derivatives.[39,40] This strategy was recently used in a high-throughput screen for the synthesis of the nonribosomal peptide pyoverdine, in which over 1000 unique domains were substituted into the biosynthetic pathway, resulting in the identification of 16 unique NPs.[41] An alternative approach using domain swapping created a fluorescent biosensor reporting on solubility of swapped modules to find functional recombinant polyketide synthases (PKSs).[42] This approach offered an exciting new means of a high-throughput screen for soluble PKS variants, although it should be noted that solubility alone does not mean that such fusions will be functional.

These strategies make use of the natural biosynthesis pathways for NPs, but when a desired compound cannot be produced by an existing natural pathway, tools for retrobiosynthesis can be used to design a pathway. Retrobiosynthesis is a key tool in pathway engineering which takes a chemical compound of interest and breaks it into smaller constituent parts in order to build a bottom-up metabolic pathway to produce the end product using metabolites of the chassis strain.[43,44] BioNavi-NP is a retrobiosynthesis software tool that focuses on NP pathway prediction. Compared with general metabolic engineering tools such as RetroPathRL, BioNavi-NP had a 13% higher pathway hit rate accuracy on the LASER test dataset and was performed in almost an order of magnitude less time.[45] It is able to accurately predict the biosynthesis pathways for diverse NPs, including large structures such as sterhirsutin J.

For the selection of individual enzymes to catalyze the reactions within a pathway, tools such as Selenzyme can be used, which returns a ranked list of candidate enzymes, informed by parameters such as their phylogenetic distance from the host organism.[46] This approach provides a key benefit

https://doi.org/10.1021/acssynbio.4c00391
ACS Synth. Biol. 2024, 13, 2684−2692

Table 3. Recent Examples Highlighting Different Build Strategies

build strategy	product	conditions	ref
coculturing	phenylpropene	tripartite coculture of *E. coli* to spread metabolic burden across organisms	75
coculturing	(*S*)-norcoclaurine	*Scheffersomyces stipitis* produces shikimate, which *S. cerevisiae* converts to (*S*)-norcoclaurine	71
strain engineering	armeniaspirols	knockout of competing polyketide cluster in *Streptomyces* sp. A793 and introduction of heterologous fatty acid synthase to ease bottleneck of precursors from primary metabolism	76
strain engineering	citramalate	transposon-guided copy number engineering of *Issatchenkia orientalis*, which can grow at low pH	77
strain engineering	pamamycin	pamamycin-resistant *Streptomyces albus* J1074 underwent transcriptional engineering by screening of a promoter library	78
cell-free synthesis	salivaricin B	one-pot reaction of engineered *E. coli* extract with chaperone proteins, precursor peptides, and tailoring enzymes	79

of being able to propose candidate enzymes for reactions for which no matching enzyme is known yet. Optimization of the proposed enzymes for the target reaction can utilize the latest insights from protein folding algorithms, with AlphaFold serving as a prime example.[47] The protein-modeling approach from these algorithms has opened up the ability to rationally design mutants based on the proposed structures of proteins in biosynthesis pathways which have not previously been experimentally determined.[48]

BUILD

A continuing challenge in the field of NPs is that the majority of BGCs are not endogenously expressed in standard laboratory conditions, limiting efforts to characterize the NPs they synthesize.[49,50] This has led to heterologous expression being a widely used technique to investigate BGCs and their corresponding NPs. This has especially been the case with the *Streptomyces* genus, which is the most frequently studied bacterial genus for NPs due to the high number of BGCs it encodes.[51,52] Specific strains of *Streptomyces*, such as *Streptomyces albus* due to its small genome, are frequently used as heterologous hosts for BGCs from environmental *Streptomyces* strains which are not cultivatable in the laboratory.[53] This work has been aided by advances in CRISPR-Cas9 systems, which allow for rapid genetic engineering of these strains, and has recently has expanded to multiplexed base editing systems as well.[54] In a promising recent study, the authors used coevolutionary analysis to identify genes coevolving with polyketide NP production and found a gene cluster encoding for the cofactor pyrroloquinoline quinone (PQQ), which is strongly linked to polyketide synthesis.[55] Overexpression of the *pqq* operon could be introduced to 11 *Streptomyces* species and caused an increase in the production of numerous known NPs and, more importantly, a large number of previously uncharacterized NPs.

Advances in gene editing have also allowed the exploitation of cross-kingdom heterologous hosts, such as yeast strains being used for expression of plant BGCs due to the relative similarity of their subcellular compartmentations.[56,57] The model yeast *Saccharomyces cerevisiae* has often been used for the heterologous expression of plant BGCs, with several examples of medicinal NPs being synthesized.[58,59] Nonmodel yeasts can also be leveraged as heterologous hosts; for instance, *Yarrowia lipolytica* is popular because of its high capacity for lipid storage, though as with all engineered hosts there is a significant metabolic burden from accumulation of target compounds.[60] Genetic engineering advances promise to alleviate some of these issues, as a multiplex base-editing system was recently shown to have higher editing efficiency than wild-type CRISPR-Cas systems in *Y. lipolytica* and thus promises increased control in strain engineering to optimize expression.[61] These genetic engineering strategies increase the breadth of compounds the yeast can produce heterologously, a list that already includes important commercial terpenoids such as limonene and taxadiene.[62]

Genetic engineering is not the only mechanism for activating silent BGCs. Modifying specific environmental triggers including temperature,[63] pH,[64] and quorum sensing[65] has been shown to induce synthesis and allow for the characterization of novel products. The cellular environment can also be manipulated biologically by using coculture techniques to induce expression.[66] In this experimental design, which mimics the natural environmental conditions of microbial interactions, the two participating species are frequently categorized as inducer and producer strains, with the former secreting a molecule which instigates the production of an NP by the latter.[67] A surge of new insights into these interactions has been provided by recent advances in the study of microbiomes, and similar techniques of coculturing and genetic engineering can be applied to further NP engineering efforts.[68−70] Coculturing also can take advantage of the compartmentalization of pathway steps in different microorganisms, thus taking advantage of different strains' metabolic capacities[71,72] (Table 3). When needed, this interstrain compartmentalization can be forced, *e.g.*, by restricting the producer strain within hydrogel beds through which the inducer's products can diffuse.[73] Recent modeling work has shown that it should be possible to stably maintain highly burdensome biosynthetic pathways distributed across the members of a community through "dynamic division of labor" by horizontal gene transfer.[74]

Though not yet viable on industrial production scales, cell-free systems offer an intriguing means to circumvent issues inherent with cellular expression, such as diversion of energy toward cellular maintenance and unknown genetic regulation networks.[80,81] A key benefit for natural product engineering which cell-fee systems provide is the ability to utilize certain cofactors and byproducts whose expression or accumulation can be toxic to the cell, such as the analogs of *S*-adenosylmethionine used to synthesize caffeine.[82] Large and complex synthesis proteins, such as nonribosomal peptide synthetases, can also be too metabolically burdensome for some cells but can be used in cell-free systems.[80,83] Cell-free systems can also enable quicker benchtop production of certain NPs, such as thiopeptide scaffolds.[84] One example that highlights this speed utilized the cellular extract of an engineered strain of *E. coli* which lacked several peptidases and protease genes, whose products had inhibited the cell-free synthesis of the target peptide.[79] This approach, termed unified

https://doi.org/10.1021/acssynbio.4c00391
ACS Synth. Biol. 2024, 13, 2684−2692

biocatalysis, relied on a mix of precursor peptides, relevant tailoring enzymes, bespoke chaperone proteins, and cellular extract in a one-pot reaction. It yielded an increase in the target RiPP titers in a few hours, which also enabled researchers to produce several derivatives of the target compound. Despite this promise, it is important to acknowledge that a lack of thorough documentation of the components in some of these systems leads to the field remaining inscrutable at times.[85,86]

TEST

After an expression system for an NP has been built, its expression must be tested by characterizing the produced chemical compound. The field of NP research relies heavily on chemical analysis in order to characterize and quantify the production of the desired compounds. Tandem mass spectrometry (MS^2) is frequently used for this purpose, to isolate and identify chemical compounds. However, this method is often challenged by the fact that many of the most interesting products are being characterized for the first time at the time of discovery, and thus, authentic standards—by definition—are not yet available. Even for previously characterized NPs, the analysis in a new producer can be complicated by the fact that many spectral reference libraries can be poor at annotating natural products, as they often exhibit a bias toward industrial chemicals, primary metabolites, and other commercially available compounds.[87] Hence, NMR methods, which require larger amounts of pure sample, remain indispensable.[88,89] Advances in NMR analysis using ML have greatly decreased the amount of time needed to derive structural predictions.[90]

General recent advances in the field of metabolomics research have been made through concerted efforts on integrating data from different technologies and are now becoming available for NP research. MZMine3 is a tool for analyzing hybrid MS datasets that can integrate the analysis of raw spectral data from several different MS platforms.[91] This fills the gap of large-scale MS^2-based annotations, including applications for MS imaging analysis. This is highly relevant for NPs, which are often produced in spatially segregated regions of bacterial colonies, as illustrated by a recent study using MALDI-MS imaging in *in vitro* experiments using a bespoke membrane on an agar plate to obtain clear spatial data while reducing the background of smaller molecules.[92]

Collaboration in the NP field has yielded many key community tools for MS analysis, such as the Global Natural Products System (GNPS).[93] GNPS generates molecular networks for user-submitted MS^2 data that can be compared to >490,000 user-submitted MS^2 spectra in its MassIVE database.[94,95] There are approximately 50 software tools incorporated into the GNPS ecosystem, *e.g.*, microbeMASST, which can connect MS^2 data to specific organisms.[87] The use of FastSearch, originally a tool for proteomics, enables microbeMASST to search the GNPS/MassIVE repository at a rate orders of magnitude higher than MASST.[96] Its outputs include taxonomic trees which connect the spectral data to the specific organisms in which matching spectral data have previously been detected, though the resolution for the taxonomy is not always at the strain level. MS2LDA has recently made its Mass2Motifs workflow, a tool which outputs biochemically relevant substructures, compatible with the GNPS ecosystem.[97,98] A thorough review of these and other spectral analysis approaches can be found in the work of Zdouc and colleagues.[99]

LEARN

A key aspect of ML being used for NP research lies in the ability to learn from the many databases of known chemical structures of NPs and other chemicals. These datasets have been used as inputs for graph neural network (GNN) models which read in the chemical structures as graphs with atoms as nodes and atomic bonds as the edges connecting nodes. When combined with the results of high-throughput antibiotic activity screens, this approach has led to the discovery of antibiotic activity for several compounds that had not previously been known to have antibiotic activity, though not all of these have been NPs. Halicin[6] and abaucin[100] are two highly publicized drugs which had previously been characterized in the Drug Repurposing Hub[101] and ZINC15[102] libraries yet were only recently found to display antimicrobial properties after analysis using GNN models. Similar strategies should be applicable to NP research, and indeed, several related efforts have already discovered NPs which have displayed antibiotic activity against antibiotic-resistant strains.[103,104] The vast amount of structural data for known NP molecules can also be leveraged to develop synthetic NP-like molecules, but additional work will be required to design bespoke (bio)chemicals and create bioactive neo-NPs on demand.[105]

The increasingly abundant omics data reported in the literature can be used to generate improved genome-scale metabolic models (GEMs) of chassis organisms.[106] There are now GEMs available for many industrially relevant microorganisms tailored to specific use cases, such as production of a target NP or the use of a specific feedstock, as detailed in a thorough review by Han and colleagues.[107] The use of GEMs for NP engineering can involve testing knockouts of specific genes *in silico* in order to maximize target NP production *in vitro*, as was done to optimize oleanolic acid production in *S. cerevisiae*.[108] These approaches are highly beneficial when attempting to rationally engineer nonmodel organisms, as shown by recent work in the development of a GEM for *Vibrio natriegens*, which is an attractive industrial bacterium due to its high growth rate.[109]

RESPONSIBLE RESEARCH AND INNOVATION

As shown in the preceding sections, NP research can rely heavily on the concepts and methods of synthetic biology to produce and characterize its products. Synthetic biology itself has promised to enable a greener and more sustainable alternative to several industrially relevant chemical processes.[110−112] Fulfilling this promise will require careful techno-economic analysis (TEA) and full life-cycle assessment (LCA),[113,114] which are not yet available for most NPs but will be a critical step when upscaling NP production to an industrial level.[115]

Early examples of TEA and LCA in the NP domain include a recent application to different photobioreactors to determine which non-open-air bioreactor could work best for microalgae producing ketocarotenoids.[116] Another recent paper focused on comparing different corn-based feedstocks for shikimic acid bioproduction and concluded that a key factor was the economics of alternate uses of each feedstock.[113] Work by Zhao and colleagues on different synthesis pathways of vanillin production, which identified electricity usage as the major determinant of environmental burden, highlights the need for more data to best estimate large-scale production parameters

https://doi.org/10.1021/acssynbio.4c00391
ACS Synth. Biol. 2024, 13, 2684−2692

as well as the substantial context dependence of any such analysis, *e.g.*, as a result in dramatic differences of the modes of electricity generation.[117]

An impediment to using LCA or TEA can be a lack of information to guide the decision-making process at an early stage.[118] For some NP production pathways, product yield is still too small to utilize industrially relevant parameters in the production system, and NP production, especially for novel antimicrobials, is often not performed at an industrial level due to the shrinking number of major pharmaceutical companies investing in antibiotics development.[119] Nevertheless, it is possible to develop production systems already at the laboratory proof-of-concept stage in such a way that future large-scale applications are taken into account, as shown in a recent study using electrodialysis in the purification of the high-value antimicrobial peptide nisin and evaluating its benefits in a circular-economy framework.[120] To facilitate the early anticipation of potential impacts of synthetic biology and to guide the choice between alternative production methods, the recently published Early Rapid Sustainability Assessment (ERSA) seeks to minimize the *a priori* information needed to perform such predictions.[121] It introduces a framework that enables early-stage research to conduct a risk analysis for future scale-ups of an experimental workflow. This process has been tested across a range of NP fermentation processes and presents an early-stage tool for approximating the environmental impacts of experiments. ERSA thus provides an important tool for conscientiously developing NP fermentation processes while they are still on the scale of academic research and before technology lock-ins have occurred.

AUTHOR INFORMATION

Corresponding Author

Eriko Takano − *Manchester Institute of Biotechnology, Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, United Kingdom;* orcid.org/0000-0002-6791-3256; Email: eriko.takano@manchester.ac.uk

Authors

Jonathan Foldi − *Manchester Institute of Biotechnology, Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, United Kingdom*

Jack A. Connolly − *Manchester Institute of Biotechnology, Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, United Kingdom;* orcid.org/0000-0003-3643-5578

Rainer Breitling − *Manchester Institute of Biotechnology, Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, United Kingdom;* orcid.org/0000-0001-7173-0922

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.4c00391

Author Contributions

#J.F. and J.A.C. contributed equally. J.F.: Conceptualization, Investigation, Writing—original draft, review and editing. J.A.C.: Investigation, Writing—review and editing. E.T. and R.B.: Funding acquisition, Conceptualization, Project administration, Supervision, Writing—review and editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J.F. was supported by the EPSRC Center for Doctoral Training in BioDesign Engineering (EP/S022856/1). J.A.C., E.T., and R.B. were supported by UK Research and Innovation (NE/T010959/1) through the Signals in the Soil Program. E.T. and R.B. were also supported by BBSRC (BB/X012573/1).

REFERENCES

(1) Demain, A. L.; Fang, A. The natural functions of secondary metabolites. *Adv. Biochem. Eng./Biotechnol.* **2000**, *69*, 1−39.

(2) Nakanishi, K. A Brief History of Natural Products Chemistry. In *Comprehensive Natural Products Chemistry*; Barton, D. H. R., Nakanishi, K., Meth-Cohn, O., Eds.; Pergamon Press, 1999; pp 1−31. DOI: 10.1016/B978-0-08-091283-7.09008-1.

(3) Cragg, G. M.; Newman, D. J. Natural Products: A continuing source of novel drug leads. *Biochim. Biophys. Acta, Gen. Subj.* **2013**, *1830* (6), 3670−3695.

(4) Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770−803.

(5) Boyd, N. K.; Teng, C.; Frei, C. R. Brief overview of approaches and challenges in new antibiotic development: A focus on drug repurposing. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 684515.

(6) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; et al. A deep learning approach to antibiotic discovery. *Cell* **2020**, *180* (4), 688−702.

(7) Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Bisignano, C.; Wool, E.; et al. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* **2022**, *399* (10325), 629−655.

(8) *Antimicrobial Resistance*. World Health Organization, 2023. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

(9) Li, J. W.-H.; Vederas, J. C. Drug discovery and natural products: End of an era or an endless frontier? *Science* **2009**, *325* (5937), 161−165.

(10) Ortholand, J.-Y.; Ganesan, A. Natural products and combinatorial chemistry: Back to the future. *Curr. Opin. Chem. Biol.* **2004**, *8* (3), 271−280.

(11) Sorokina, M.; Steinbeck, C. Review on natural products databases: Where to find data in 2020. *J. Cheminf.* **2020**, *12* (1), 20.

(12) van Santen, J. A.; Poynton, E. F.; Iskakova, D.; McMann, E.; Alsup, T. A.; Clark, T. N.; et al. The Natural Products Atlas 2.0: A database of microbially-derived natural products. *Nucleic Acids Res.* **2022**, *50* (D1), D1317−D1323.

(13) Gallo, K.; Kemmler, E.; Goede, A.; Becker, F.; Dunkel, M.; Preissner, R.; Banerjee, P. SuperNatural 3.0—a database of natural products and natural product-based derivatives. *Nucleic Acids Res.* **2023**, *51* (D1), D654−D659.

(14) Sorokina, M.; Merseburger, P.; Rajan, K.; Yirik, M. A.; Steinbeck, C. COCONUT online: collection of open natural products database. *J. Cheminf.* **2021**, *13* (1), 2.

(15) Gavriilidou, A.; Kautsar, S. A.; Zaburannyi, N.; Krug, D.; Müller, R.; Medema, M. H.; Ziemert, N. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **2022**, *7* (5), 726−735.

(16) Zhang, Y.; Wiese, L.; Fang, H.; Alseekh, S.; Perez De Souza, L.; Scossa, F.; Molloy, J. J.; Christmann, M.; Fernie, A. R. Synthetic biology identifies the minimal gene set required for paclitaxel biosynthesis in a plant chassis. *Mol. Plant* **2023**, *16* (12), 1951−1961.

(17) Hibi, G.; Shiraishi, T.; Umemura, T.; Nemoto, K.; Ogura, Y.; Nishiyama, M.; Kuzuyama, T. Discovery of type II polyketide synthase-like enzymes for the biosynthesis of cispentacin. *Nat. Commun.* **2023**, *14* (1), 8065.

(18) Liu, W.; Tian, X.; Huang, X.; Malit, J. J. L.; Wu, C.; Guo, Z.; Tang, J.-W.; Qian, P.-Y. Discovery of P450-modified sesquiterpenoids levinoids A-D through global genome mining. *J. Nat. Prod.* **2024**, *87* (4), 876−883.
(19) Jensen, P. R. Natural products and the gene cluster revolution. *Trends Microbiol.* **2016**, *24* (12), 968−977.
(20) Belknap, K. C.; Park, C. J.; Barth, B. M.; Andam, C. P. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* **2020**, *10* (1), 2003.
(21) Agustinho, D. P.; Fu, Y.; Menon, V. K.; Metcalf, G. A.; Treangen, T. J.; Sedlazeck, F. J. Unveiling microbial diversity: Harnessing long-read sequencing technology. *Nat. Methods* **2024**, *21*, 954.
(22) Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; et al. A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, *517* (7535), 455−459.
(23) Vitorino, I. R.; Lobo-da-Cunha, A.; Vasconcelos, V.; Lage, O. M. *Aeoliella straminimaris* sp. nov., a novel member of the phylum Planctomycetota with an unusual filamentous structure. *Int. J. Syst. Evol. Microbiol.* **2023**, *73* (4), 005850.
(24) Li, S.; Lian, W.-H.; Han, J.-R.; Ali, M.; Lin, Z.-L.; Liu, Y.-H.; et al. Capturing the microbial dark matter in desert soils using culturomics-based metagenomics and high-resolution analysis. *npj Biofilms Microbiomes* **2023**, *9* (1), 67.
(25) Zhang, Y.; Wang, Y.; Tang, M.; Zhou, J.; Zhang, T. The microbial dark matter and "Wanted List" in worldwide wastewater treatment plants. *Microbiome* **2023**, *11* (1), 59.
(26) Blin, K.; Shaw, S.; Augustijn, H. E.; Reitz, Z. L.; Biermann, F.; Alanjary, M.; et al. antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **2023**, *51* (W1), W46−W50.
(27) Kautsar, S. A.; Suarez Duran, H. G.; Blin, K.; Osbourn, A.; Medema, M. H. plantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res.* **2017**, *45* (W1), W55−W63.
(28) Carroll, L. M.; Larralde, M.; Fleck, J. S.; Ponnudurai, R.; Milanese, A.; Cappio, E.; Zeller, G. Accurate *de novo* identification of biosynthetic gene clusters with GECCO. *bioRxiv* **2021**, DOI: 10.1101/2021.05.03.442509.
(29) Sanchez, S.; Rogers, J. D.; Rogers, A. B.; Nassar, M.; McEntyre, J.; Welch, M.; Hollfelder, F.; Finn, R. D. Expansion of Novel Biosynthetic Gene Clusters from Diverse Environments Using SanntiS. *bioRxiv* **2023**, DOI: 10.1101/2023.05.23.540769.
(30) Skinnider, M. A.; Johnston, C. W.; Gunabalasingam, M.; Merwin, N. J.; Kieliszek, A. M.; MacLellan, R. J.; et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.* **2020**, *11* (1), 6058.
(31) Sélem-Mojica, N.; Aguilar, C.; Gutiérrez-García, K.; Martínez-Guerrero, C. E.; Barona-Gómez, F. EvoMining reveals the origin and fate of natural product biosynthetic enzymes. *Microb. Genomics* **2019**, *5* (12), e000260.
(32) Kautsar, S. A.; van der Hooft, J. J. J.; de Ridder, D.; Medema, M. H. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *GigaScience* **2021**, *10* (1), giaa154.
(33) Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **2020**, *16* (1), 60−68.
(34) Terlouw, B. R.; Blin, K.; Navarro-Muñoz, J. C.; Avalon, N. E.; Chevrette, M. G.; Egbert, S.; et al. MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **2023**, *51* (D1), D603−D610.
(35) Blin, K.; Shaw, S.; Medema, M. H.; Weber, T. The antiSMASH database version 4: Additional genomes and BGCs, new sequence-based searches and more. *Nucleic Acids Res.* **2024**, *52* (D1), D586−D589.
(36) Skellam, E.; Rajendran, S.; Li, L. Combinatorial biosynthesis for the engineering of novel fungal natural products. *Commun. Chem.* **2024**, *7* (1), 89.
(37) Su, L.; Hôtel, L.; Paris, C.; Chepkirui, C.; Brachmann, A. O.; Piel, J.; Jacob, C.; Aigle, B.; Weissman, K. J. Engineering the stambomycin modular polyketide synthase Yields 37-membered mini-stambomycins. *Nat. Commun.* **2022**, *13* (1), 515.
(38) Yang, X.-L.; Friedrich, S.; Yin, S.; Piech, O.; Williams, K.; Simpson, T. J.; Cox, R. J. Molecular basis of methylation and chain-length programming in a fungal iterative highly reducing polyketide synthase. *Chem. Sci.* **2019**, *10* (36), 8478−8489.
(39) Cummings, M.; Peters, A. D.; Whitehead, G. F. S.; Menon, B. R. K.; Micklefield, J.; Webb, S. J.; Takano, E. Assembling a plug-and-play production line for combinatorial biosynthesis of aromatic polyketides in *Escherichia coli*. *PLoS Biol.* **2019**, *17* (7), No. e3000347.
(40) Pang, B.; Graziani, E. I.; Keasling, J. D. Acyltransferase domain swap in modular type I polyketide synthase to adjust the molecular gluing strength of rapamycin. *Tetrahedron Lett.* **2022**, *112*, 154229.
(41) Messenger, S. R.; McGuinniety, E. M. R.; Stevenson, L. J.; Owen, J. G.; Challis, G. L.; Ackerley, D. F.; Calcott, M. J. Metagenomic domain substitution for the high-throughput modification of nonribosomal peptides. *Nat. Chem. Biol.* **2024**, *20* (2), 251−260.
(42) Englund, E.; Schmidt, M.; Nava, A. A.; Klass, S.; Keiser, L.; Dan, Q.; Katz, L.; Yuzawa, S.; Keasling, J. D. Biosensor guided polyketide synthases engineering for optimization of domain exchange boundaries. *Nat. Commun.* **2023**, *14* (1), 4871.
(43) Delépine, B.; Duigou, T.; Carbonell, P.; Faulon, J.-L. RetroPath2.0: A retrosynthesis workflow for metabolic engineers. *Metab. Eng.* **2018**, *45*, 158−170.
(44) Yu, T.; Boob, A. G.; Volk, M. J.; Liu, X.; Cui, H.; Zhao, H. Machine learning-enabled retrobiosynthesis of molecules. *Nat. Catal.* **2023**, *6* (2), 137−151.
(45) Zheng, S.; Zeng, T.; Li, C.; Chen, B.; Coley, C. W.; Yang, Y.; Wu, R. Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP. *Nat. Commun.* **2022**, *13* (1), 3342.
(46) Stoney, R. A.; Hanko, E. K. R.; Carbonell, P.; Breitling, R. SelenzymeRF: Updated enzyme suggestion software for unbalanced biochemical reactions. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 5868−5876.
(47) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493−500.
(48) Thomas, F.; Kayser, O. Improving CBCA synthase activity through rational protein design. *J. Biotechnol.* **2023**, *363*, 40−49.
(49) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R. A natural product chemist's guide to unlocking silent biosynthetic gene clusters. *Annu. Rev. Biochem.* **2021**, *90* (1), 763−788.
(50) Li, L. Next-generation synthetic biology approaches for the accelerated discovery of microbial natural products. *Eng. Microbiol.* **2023**, *3* (1), 100060.
(51) Chandra, G.; Chater, K. F. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol Rev.* **2014**, *38* (3), 345−379.
(52) Caicedo-Montoya, C.; Manzo-Ruiz, M.; Ríos-Estepa, R. Pan-genome of the genus *Streptomyces* and prioritization of biosynthetic gene clusters with potential to produce antibiotic compounds. *Front. Microbiol.* **2021**, *12*, 677558.
(53) Kallifidas, D.; Jiang, G.; Ding, Y.; Luesch, H. Rational engineering of *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene clusters. *Microb. Cell Fact.* **2018**, *17* (1), 25.
(54) Whitford, C. M.; Gren, T.; Palazzotto, E.; Lee, S. Y.; Tong, Y.; Weber, T. Systems analysis of highly multiplexed CRISPR-base editing in Streptomycetes. *ACS Synth. Biol.* **2023**, *12* (8), 2353−2366.
(55) Wang, X.; Chen, N.; Cruz-Morales, P.; Zhong, B.; Zhang, Y.; Wang, J.; Xiao, Y.; et al. Elucidation of genes enhancing natural product biosynthesis through co-evolution analysis. *Nat. Metab.* **2024**, *6*, 933.
(56) Perrot, T.; Marc, J.; Lezin, E.; Papon, N.; Besseau, S.; Courdavault, V. Emerging trends in production of plant natural

products and new-to-nature biopharmaceuticals in yeast. *Curr. Opin. Biotechnol.* **2024**, *87*, 103098.
(57) Liu, Y.; Zhao, X.; Gan, F.; Chen, X.; Deng, K.; Crowe, S. A.; et al. Complete biosynthesis of QS-21 in engineered yeast. *Nature* **2024**, *629* (8013), 937−944.
(58) Wu, S.; Malaco Morotti, A. L.; Wang, S.; Wang, Y.; Xu, X.; Chen, J.; Wang, G.; Tatsis, E. C. Convergent gene clusters underpin hyperforin biosynthesis in St John's wort. *New Phytol.* **2022**, *235* (2), 646−661.
(59) Forman, V.; Luo, D.; Geu-Flores, F.; Lemcke, R.; Nelson, D. R.; Kampranis, S. C.; Staerk, D.; Møller, B. L.; Pateraki, I. A gene cluster in *Ginkgo biloba* encodes unique multifunctional cytochrome P450s that initiate ginkgolide biosynthesis. *Nat. Commun.* **2022**, *13* (1), 5143.
(60) Sun, M.-L.; Gao, X.; Lin, L.; Yang, J.; Ledesma-Amaro, R.; Ji, X.-J. Building *Yarrowia lipolytica* cell factories for advanced biomanufacturing: challenges and solutions. *J. Agric. Food Chem.* **2024**, *72* (1), 94−107.
(61) Ganesan, V.; Monteiro, L.; Pedada, D.; Stohr, A.; Blenner, M. High-efficiency multiplexed cytosine base editors for natural product synthesis in *Yarrowia lipolytica*. *ACS Synth. Biol.* **2023**, *12* (10), 3082−3091.
(62) Ma, Y.; Zu, Y.; Huang, S.; Stephanopoulos, G. Engineering a universal and efficient platform for terpenoid synthesis in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (1), No. e2207680120.
(63) Liu, Y.; Zhang, T.; Lu, X.; Ma, B.; Ren, A.; Shi, L.; Jiang, A.; Yu, H.; Zhao, M. Membrane fluidity is involved in the regulation of heat stress induced secondary metabolism in *Ganoderma lucidum*. *Environ. Microbiol.* **2017**, *19* (4), 1653−1668.
(64) Kim, H.; Kim, J.-Y.; Ji, C.; Lee, D.; Shim, S. H.; Joo, H.-S.; Kang, H.-S. Acidonemycins A-C, glycosylated angucyclines with antivirulence activity produced by the acidic culture of *Streptomyces indonesiensis*. *J. Nat. Prod.* **2023**, *86* (8), 2039−2045.
(65) Gonzales, M.; Jacquet, P.; Gaucher, F.; Chabrière, É.; Plener, L.; Daudé, D. AHL-based quorum sensing regulates the biosynthesis of a variety of bioactive molecules in bacteria. *J. Nat. Prod.* **2024**, *87* (4), 1268−1284.
(66) Boruta, T.; Ścigaczewska, A.; Bizukojć, M. Investigating the stirred tank bioreactor co-cultures of the secondary metabolite producers *Streptomyces noursei* and *Penicillium rubens*. *Biomolecules* **2023**, *13* (12), 1748.
(67) Selegato, D. M.; Castro-Gamboa, I. Enhancing chemical and biological diversity by co-cultivation. *Front. Microbiol.* **2023**, *14*, 1117559.
(68) Arnold, J.; Glazier, J.; Mimee, M. Genetic engineering of resident bacteria in the gut microbiome. *J. Bacteriol.* **2023**, *205* (7), No. e00127-23.
(69) Jing, J.; Garbeva, P.; Raaijmakers, J. M.; Medema, M. H. Strategies for tailoring functional microbial synthetic communities. *ISME J.* **2024**, *18* (1), wrae049.
(70) Gasparek, M.; Steel, H.; Papachristodoulou, A. Deciphering mechanisms of production of natural compounds using inducer-producer microbial consortia. *Biotechnol. Adv.* **2023**, *64*, 108117.
(71) Gao, M.; Zhao, Y.; Yao, Z.; Su, Q.; Van Beek, P.; Shao, Z. Xylose and shikimate transporters facilitates microbial consortium as a chassis for benzylisoquinoline alkaloid production. *Nat. Commun.* **2023**, *14* (1), 7797.
(72) Peng, H.; Darlington, A. P. S.; South, E. J.; Chen, H.-H.; Jiang, W.; Ledesma-Amaro, R. A molecular toolkit of cross-feeding strains for engineering synthetic yeast communities. *Nat. Microbiol.* **2024**, *9* (3), 848−863.
(73) Zhao, R.; Sengupta, A.; Tan, A. X.; Whelan, R.; Pinkerton, T.; Menasalvas, J.; et al. Photobiological production of high-value pigments via compartmentalized co-cultures using Ca-alginate hydrogels. *Sci. Rep.* **2022**, *12* (1), 22163.
(74) Hamrick, G. S.; Maddamsetti, R.; Son, H.-I.; Wilson, M. L.; Davis, H. M.; You, L. Programming dynamic division of labor using horizontal gene transfer. *ACS Synth. Biol.* **2024**, *13* (4), 1142−1151.
(75) Brooks, S. M.; Marsan, C.; Reed, K. B.; Yuan, S.-F.; Nguyen, D.-D.; Trivedi, A.; Altin-Yavuzarslan, G.; Ballinger, N.; Nelson, A.; Alper, H. S. A tripartite microbial co-culture system for de novo biosynthesis of diverse plant phenylpropanoids. *Nat. Commun.* **2023**, *14* (1), 4448.
(76) Heng, E.; Lim, Y. W.; Leong, C. Y.; Ng, V. W. P.; Ng, S. B.; Lim, Y. H.; Wong, F. T. Enhancing armeniaspirols production through multi-level engineering of a native *Streptomyces* producer. *Microb. Cell Fact.* **2023**, *22* (1), 84.
(77) Wu, Z.-Y.; Sun, W.; Shen, Y.; Pratas, J.; Suthers, P. F.; Hsieh, P.-H.; Dwaraknath, S.; et al. Metabolic engineering of low-pH-tolerant non-model yeast, *Issatchenkia orientalis*, for production of citramalate. *Metab. Eng. Commun.* **2023**, *16*, No. e00220.
(78) Eckert, N.; Rebets, Y.; Horbal, L.; Zapp, J.; Herrmann, J.; Busche, T.; Müller, R.; Kalinowski, J.; Luzhetskyy, A. Discovery and overproduction of novel highly bioactive pamamycins through transcriptional engineering of the biosynthetic gene cluster. *Microb. Cell Fact.* **2023**, *22* (1), 233.
(79) Liu, W.-Q.; Ji, X.; Ba, F.; Zhang, Y.; Xu, H.; Huang, S.; Zheng, X.; et al. Cell-free biosynthesis and engineering of ribosomally synthesized lanthipeptides. *Nat. Commun.* **2024**, *15* (1), 4336.
(80) Bogart, J. W.; Cabezas, M. D.; Vögeli, B.; Wong, D. A.; Karim, A. S.; Jewett, M. C. Cell-free exploration of the natural product chemical space. *ChemBioChem* **2021**, *22* (1), 84−91.
(81) Moore, S. J.; Lai, H.-E.; Li, J.; Freemont, P. S. *Streptomyces* cell-free systems for natural product discovery and engineering. *Nat. Prod. Rep.* **2023**, *40* (2), 228−236.
(82) Ditzel, A.; Zhao, F.; Gao, X.; Phillips, G. N. Utilizing a cell-free protein synthesis platform for the biosynthesis of a natural product, caffeine. *Synth. Biol.* **2023**, *8* (1), ysad017.
(83) Dinglasan, J. L. N.; Sword, T. T.; Barker, J. W.; Doktycz, M. J.; Bailey, C. B. Investigating and optimizing the lysate-based expression of nonribosomal peptide synthetases using a reporter system. *ACS Synth. Biol.* **2023**, *12* (5), 1447−1460.
(84) Fleming, S. R.; Bartges, T. E.; Vinogradov, A. A.; Kirkpatrick, C. L.; Goto, Y.; Suga, H.; Hicks, L. M.; Bowers, A. A. Flexizyme-enabled benchtop biosynthesis of thiopeptides. *J. Am. Chem. Soc.* **2019**, *141* (2), 758−762.
(85) Wagner, L.; Jules, M.; Borkowski, O. What remains from living cells in bacterial lysate-based cell-free systems. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 3173−3182.
(86) Rasor, B. J.; Chirania, P.; Rybnicky, G. A.; Giannone, R. J.; Engle, N. L.; Tschaplinski, T. J.; Karim, A. S.; Hettich, R. L.; Jewett, M. C. Mechanistic insights into cell-free gene expression through an integrated -omics analysis of extract processing methods. *ACS Synth. Biol.* **2023**, *12* (2), 405−418.
(87) Zuffa, S.; Schmid, R.; Bauermeister, A.; P. Gomes, P. W.; Caraballo-Rodriguez, A. M.; El Abiead, Y.; et al. microbeMASST: A taxonomically informed mass spectrometry search tool for microbial metabolomics data. *Nat. Microbiol.* **2024**, *9* (2), 336−345.
(88) Junot, C.; Pinu, F. R.; Van Der Hooft, J. J. J.; Moco, S. Editorial: NMR-based metabolomics. *Front. Mol. Biosci.* **2023**, *10*, 1337566.
(89) Wishart, D. S.; Cheng, L. L.; Copié, V.; Edison, A. S.; Eghbalnia, H. R.; Hoch, J. C.; et al. NMR and metabolomics—a roadmap for the future. *Metabolites* **2022**, *12* (8), 678.
(90) Kim, H. W.; Zhang, C.; Reher, R.; Wang, M.; Alexander, K. L.; Nothias, L.-F.; et al. DeepSAT: Learning molecular structures from nuclear magnetic resonance data. *J. Cheminf.* **2023**, *15* (1), 71.
(91) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **2023**, *41* (4), 447−449.
(92) Müller, W. H.; McCann, A.; Arias, A. A.; Malherbe, C.; Quinton, L.; De Pauw, E.; Eppe, G. Imaging metabolites in agar-based bacterial co-cultures with minimal sample preparation using a DIUTHAME membrane in surface-assisted laser desorption/ionization mass spectrometry. *ChemistrySelect* **2022**, *7* (18), No. e202200734.

(93) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; et al. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* **2016**, *34* (8), 828−837.
(94) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **2020**, *15* (6), 1954−1991.
(95) Leao, T. F.; Clark, C. M.; Bauermeister, A.; Elijah, E. O.; Gentry, E. C.; Husband, M.; Oliveira, M. F.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Quick-start infrastructure for untargeted metabolomics analysis in GNPS. *Nat. Metab.* **2021**, *3* (7), 880−882.
(96) Batsoyol, N.; Pullman, B.; Wang, M.; Bandeira, N.; Swanson, S. P-Massive: A real-time search engine for a multi-terabyte mass spectrometry database. In *SC22: International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE: Dallas, TX, USA, 2022; pp 1−15. DOI: 10.1109/SC41404.2022.00014.
(97) Rogers, S.; Ong, C. W.; Wandy, J.; Ernst, M.; Ridder, L.; Van Der Hooft, J. J. J. Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. *Faraday Discuss.* **2019**, *218*, 284−302.
(98) Wang, M.; Nothias, L.-F.; van der Hooft, J. MS2LDA substructure discovery. GNPS Documentation. https://ccms-ucsd.github.io/GNPSDocumentation/ms2lda/.
(99) Zdouc, M. M.; Van Der Hooft, J. J. J.; Medema, M. H. Metabolome-guided genome mining of RiPP natural products. *Trends Pharmacol. Sci.* **2023**, *44* (8), 532−541.
(100) Liu, G.; Catacutan, D. B.; Rathod, K.; Swanson, K.; Jin, W.; Mohammed, J. C.; et al. Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nat. Chem. Biol.* **2023**, *19* (11), 1342−1350.
(101) Corsello, S. M.; Bittker, J. A.; Liu, Z.; Gould, J.; McCarren, P.; Hirschman, J. E.; et al. The drug repurposing hub: A next-generation drug library and information resource. *Nat. Med.* **2017**, *23* (4), 405−408.
(102) Sterling, T.; Irwin, J. J. ZINC 15 - Ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324−2337.
(103) Rahman, A. S. M. Z.; Liu, C.; Sturm, H.; Hogan, A. M.; Davis, R.; Hu, P.; Cardona, S. T. A machine learning model trained on a high-throughput antibacterial screen increases the hit rate of drug discovery. *PLoS Comput. Biol.* **2022**, *18* (10), No. e1010613.
(104) Swanson, K.; Liu, G.; Catacutan, D. B.; Arnold, A.; Zou, J.; Stokes, J. M. Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. *Nat. Mach. Intell.* **2024**, *6* (3), 338−353.
(105) Tay, D. W. P.; Yeo, N. Z. X.; Adaikkappan, K.; Lim, Y. H.; Ang, S. J. 67 million natural product-like compound database generated via molecular language processing. *Sci. Data* **2023**, *10* (1), 296.
(106) Griesemer, M.; Navid, A. Uses of multi-objective flux analysis for optimization of microbial production of secondary metabolites. *Microorganisms* **2023**, *11* (9), 2149.
(107) Han, Y.; Tafur Rangel, A.; Pomraning, K. R.; Kerkhoven, E. J.; Kim, J. Advances in genome-scale metabolic models of industrially important fungi. *Curr. Opin. Biotechnol.* **2023**, *84*, 103005.
(108) Zhang, N.; Li, X.; Zhou, Q.; Zhang, Y.; Lv, B.; Hu, B.; Li, C. Self-controlled *in silico* gene knockdown strategies to enhance the sustainable production of heterologous terpenoid by *Saccharomyces cerevisiae*. *Metab. Eng.* **2024**, *83*, 172−182.
(109) Coppens, L.; Tschirhart, T.; Leary, D. H.; Colston, S. M.; Compton, J. R.; Hervey, W. J.; Dana, K. L.; Vora, G. J.; Bordel, S.; Ledesma-Amaro, R. *Vibrio natriegens* genome-scale modeling reveals insights into halophilic adaptations and resource allocation. *Mol. Syst. Biol.* **2023**, *19* (4), No. e10523.
(110) Kovelakuntla, V.; Meyer, A. S. Rethinking sustainability through synthetic biology. *Nat. Chem. Biol.* **2021**, *17* (6), 630−631.
(111) Jo, C.; Zhang, J.; Tam, J. M.; Church, G. M.; Khalil, A. S.; Segrè, D.; Tang, T.-C. Unlocking the magic in mycelium: Using synthetic biology to optimize filamentous fungi for biomanufacturing and sustainability. *Mater. Today Bio* **2023**, *19*, 100560.
(112) Kim, G. B.; Choi, S. Y.; Cho, I. J.; Ahn, D.-H.; Lee, S. Y. Metabolic engineering for sustainability and health. *Trends Biotechnol.* **2023**, *41* (3), 425−451.
(113) Jadidi, Y.; Frost, H.; Das, S.; Liao, W.; Draths, K.; Saffron, C. M. Comparative life cycle assessment and technoeconomic analysis of biomass-derived shikimic acid production. *ACS Sustainable Chem. Eng.* **2023**, *11* (33), 12218−12229.
(114) Romero-Suarez, D.; Keasling, J. D.; Jensen, M. K. Supplying plant natural products by yeast cell factories. *Curr. Opin. Green Sustainable Chem.* **2022**, *33*, 100567.
(115) Holland, C.; Shapira, P. Building the bioeconomy: A targeted assessment approach to identifying biobased technologies, challenges and opportunities in the UK. *Eng. Biol.* **2024**, *8* (1), 1−15, DOI: 10.1049/enb2.12030.
(116) Onorato, C.; Rösch, C. Comparative life cycle assessment of astaxanthin production with *Haematococcus pluvialis* in different photobioreactor technologies. *Algal Res.* **2020**, *50*, 102005.
(117) Zhao, X.; Chen, J.; Meng, X.; Li, L.; Zhou, X.; Li, J.; Bai, S. Environmental profile of natural biological vanillin production via life cycle assessment. *J. Cleaner Prod.* **2021**, *308*, 127399.
(118) Matthews, N. E.; Stamford, L.; Shapira, P. Aligning sustainability assessment with responsible research and onnovation: Towards a framework for constructive sustainability assessment. *Sustainable Prod. Consumption* **2019**, *20*, 58−73.
(119) Allen, A. As big pharma loses interest in new antibiotics, infections are only growing stronger. *KFF Health News*, July 15, 2022. https://kffhealthnews.org/news/article/big-pharma-new-antibiotics-infections-growing-stronger.
(120) Rulence, A.; Perreault, V.; Thibodeau, J.; Firdaous, L.; Fliss, I.; Bazinet, L. Nisin purification from a cell-free supernatant by electrodialysis in a circular economy framework. *Membranes* **2024**, *14* (1), 2.
(121) McCarthy, A.; Holland, C.; Shapira, P. The development and testing of an early, rapid sustainability assessment tool for responsible innovation in engineering biology. *SocArXiv* **2024**, DOI: 10.31235/osf.io/edhjx.

https://doi.org/10.1021/acssynbio.4c00391
ACS Synth. Biol. 2024, 13, 2684−2692