frontiers | Research Topics

Recent advances in application of synthetic biology for production of bioactive compounds, volume II

Edited by
Luan Luong Chu, Jae Kyung Sohng, Dipesh Dhakal and Hanhong Bae

Published in
Frontiers in Bioengineering and Biotechnology

ISSN 1664-8714
ISBN 978-2-8325-6683-1
DOI 10.3389/978-2-8325-6683-1

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Recent advances in application of synthetic biology for production of bioactive compounds, volume II

Topic editors

Luan Luong Chu — University of Science, Vietnam National University, Hanoi, Vietnam

Jae Kyung Sohng — Sun Moon University, Republic of Korea

Dipesh Dhakal — University of Florida, United States

Hanhong Bae — Yeungnam University, Republic of Korea

Citation

Chu, L. L., Sohng, J. K., Dhakal, D., Bae, H., eds. (2025). *Recent advances in application of synthetic biology for production of bioactive compounds, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6683-1

Table of contents

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Editorial
PUBLISHED 14 November 2024
DOI 10.3389/fbioe.2024.1517610

OPEN ACCESS

EDITED AND REVIEWED BY
Jean Marie François,
Institut Biotechnologique de Toulouse (INSA), France

*CORRESPONDENCE
Luan Luong Chu,
luancl@vnu.edu.vn

RECEIVED 26 October 2024
ACCEPTED 06 November 2024
PUBLISHED 14 November 2024

CITATION
Chu LL, Sohng JK, Bae H and Dhakal D (2024) Editorial: Recent advances in application of synthetic biology for production of bioactive compounds, volume II.
Front. Bioeng. Biotechnol. 12:1517610.
doi: 10.3389/fbioe.2024.1517610

Editorial: Recent advances in application of synthetic biology for production of bioactive compounds, volume II

Luan Luong Chu[1]*, Jae Kyung Sohng[2,3], Hanhong Bae[4] and Dipesh Dhakal[5]

[1]National Key Laboratory of Enzyme and Protein Technology, University of Science, Vietnam National University, Hanoi (VNU), Hanoi, Vietnam, [2]Department of Life Science and Biochemical Engineering, SunMoon University, Asan-si, Chungnam, Republic of Korea, [3]Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan-si, Chungnam, Republic of Korea, [4]Department of Biotechnology, Yeungnam University, Gyeongsan, Gyeongbuk, Republic of Korea, [5]Department of Medicinal Chemistry, University of Florida, Gainesville, FL, United States

KEYWORDS

synthetic biology, bioactive compounds, metabolic engineering, crispr cas, engineered microbes

Editorial on the Research Topic
Recent advances in application of synthetic biology for production of bioactive compounds, volume II

Bioactive compounds are the substances that are found in medicinal plants, bacteria, fungi and marine organisms. Both natural and unnatural bioactive compounds include secondary metabolite and their derivatives, such as isoprenoids, isoflavonoids, peptide antibiotics, and glycoside derivatives of alkaloids. These compounds play a significant role in various fields including pharmaceutical and agrochemical products, cosmetics, biofuels, and food additives. The extraction and isolation of natural products from living organisms have played an important role in the production of medicines. Along with nature bioactive compounds, synthetic biology has been developed to produce natural and non-natural compounds. This Research Topic provides recent advances, emerging challenges, and prospects of synthetic biology for bioactive compounds.

Along with model microbes (*E. coli* and *Saccharomyces cerevisiae*) that have been used for the production of natural bioactive compounds, non-conventional hosts have been developed for the biosynthesis of industrial products. Rojo et al. demonstrated the advantages of *Escherichia coli* and *S. cerevisiae* compared to plant-based production systems. The yields of isoflavonoids derivatives pterocarpans and coumestans from plant species are usually low and time-consuming industrial products are required. To overcome these limitations, engineered microbes have been used as an alternative approach to boost the production titer of pterocarpans and coumestans. Giménez et al. reviewed that filamentous fungi have been developed as a novel platform in the biotechnology field. The most advantageous filamentous fungi include the ability to grow on many different substrates and plant residues, which have a key contribution to the circular bioeconomy. According to the well-known whole-genome sequence of

filamentous fungi, *Agrobacterium tumefaciens*-mediated transformation and the GoldenBraid (GB) modular cloning platform [FungalBraid (FB) (https://gbcloning.upv.es/fungal/)] have been developed to adapt to the production of bioactive compounds. The use of FB toolkit, including CRISPR/Cas system, has been easily adapted to fungi, given the interchangeability of DNA parts between GB and FB systems. Interestingly, Fathy et al. showed that microalgae including *Synechocystis* sp. PAK13 and *Chlorella variabilis* DT025 are emerging hosts for industrial production from sustainable feedstock. This is because microalgae have substantial advantages compared to traditional microbial production systems, such as high photosynthesis efficiency, metabolic versatility, rich metabolite content, CO_2 sequestration, efficient lipid production, and more. Moreover, it is demonstrated that *Synechocystis* sp. PAK13 and *C. variabilis* could increase the accumulation of malic acid, amino acids, and indole-3-acetic acid using glycine as both a carbon and nitrogen source. Intestinally, the fatty acid content significantly increased by 1.36-fold in *Chlorella* and 2.5-fold in *Synechocystis* in the presence of glycine. Glycine is a cost – effective feedstock for microalgal cultivation, offering key benefits such as a nitrogen source, enhancing photosynthesis, reducing CO_2 dependency, and increasing lipid and biomass production.

Expanding the structural diversity of bioactive compounds through biosynthetic methods has resulted in the formation of novel compounds. These compounds are expected to improve metabolic stability and biological activitives. Lu et al. efficiently obtained 22 new analogs with modified 5′-aminouridine moieties through mutational biosynthesis of *SsaM* and *SsaK*. Both genes are responsible for biosynthesis of the 5-aminouridine moiety of sansanmycin *in vivo*. As a result, SS-KK-2 exhibited better antibacterial activity against *E. coli* ΔtolC than the parent compound sansanmycin A. Moreover, SS-KK-3 not only showed significantly increased structural stability but also retained the sama anti-TB activity against *M. tuberculosis* H37Rv compared to sansanmycin A. On the other hand, the glycoside derivative demonstrated that it is a promising alternative to the production of natural and unnatural bioactive compounds. Heo et al. identified two glycosyltransferase-coding genes, *kid7* and *kid21,* in the kidamycin biosynthetic gene cluster (BGC) of *Streptomyces* sp. W2061 strain. While Kid7 first attached N, N-dimethylvancosamine to the C_{10} position of angucycline aglycone, Kid21 transferred an anglosamine moiety to C_8 of the C_{10}-glycosylated angucycline in the following step. These catalyzation resulted in di-C-glycosylated angucycline, which is an unusual C-glycosylated residue. Similarly, Yang et al. studied the glycosylation of rebaudioside D (Reb D). A selective glycosyltransferase (UGT94D1) and its mutants (UGT94D1-F119I/D188P) from *Sesamum indicum* catalyzed for glycosylation of Reb D, leading to the production of a mono β-1,6-glycosylated derivative, Reb M2. Notably, this mutant showed a 6.33-fold improvement in catalytic efficiency and produced Reb M2 with 92% yield (29.79 mg/mL). Therefore, these studies provide an efficient method for the future development of synthetic derivatives.

Gene expression is a critical strategy for producing bioactive compounds in microbial cell factories. Understanding the key enzymes of biosynthetic pathways, as well as their complex regulation, is necessary to improve the production of bioactive compounds. Huang et al. summarized the role of (E)-4-hydroxy-3-methylbut-2-enyl pyrophosphate (HMBPP) reductase (IspH) in the methylerythritol phosphate (MEP) pathway for the synthesis of isoprenoid compounds from engineered microorganisms. Moreover, novel catalytic activities and potential biological applications have been suggested as part of the significant role of IspH. Traditional methods, including chemical inducers and genetic modifications, have been successfully used to increase gene expression levels. However, the development of novel approaches for obtaining gene expression data is still required to control the timing and location of gene expression. Yamazaki et al. suggested the novel light-controlled gene expression and proposed a label-free light control approach using mid-infrared and terahertz light. This method allows for photocontrol of gene expression without any photosensitive molecular tags. It is expected that light-controlled gene expression exhibits the potential for groundbreaking innovations in the field of gene expression, contributing to medical and pharmaceutical development. Additionally, the modulation of enzyme spatial distribution is an essential strategy to increase enzyme activity and reduce the loss of intermediate metabolites. Liu et al. reviewed the development and employment of artificial scaffold systems in microbial cell factories. Synthetic scaffolds are based on proteins, nucleic acids, and various organelles. The approaches played a significant role in increasing the titers of bioactive compounds in *E. coli*, *Bacillus subtilis*, and *S. cerevisiae*.

In summary, recent research has shown the importance of synthetic biology in the production of bioactive compounds. Noticeably, Wang et al. insisted on the influence of synthetic biology from natural sciences to humanities and social sciences by introducing biosafety, biosecurity, and ethical issues to society. Along with the development of synthetic biology tools (CRISPR/Cas, artificial scaffold systems, omics technology, and machine learning (ML) platforms), ML-based synthetic biology-combined artificial intelligence (AI) has promising applications for improving the yield of bioactive compounds from industrial microbes as well as non-conventional hosts. Artificial and intelligent hosts are expected to combine carbon-fixing autotrophs and heterotrophs along with a high yield of bioactive compounds. It is believed that engineered microbes play a significant role in agricultural and medical biotechnology, with net zero greenhouse gas emissions.

Author contributions

LC: Writing–original draft, Writing–review and editing. JS: Writing–review and editing. HB: Writing–review and editing. DD: Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The editors appreciate the contribution of all authors to this Research Topic, the constructive comments of all reviewers, and editorial support from Frontiers throughout the publication process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fbioe.2022.985696

OPEN ACCESS

EDITED BY
Dipesh Dhakal,
University of Florida, United States

REVIEWED BY
Chin-Yuan Chang,
National Chiao Tung University, Taiwan
Yasushi Ogasawara,
Hokkaido University, Japan

*CORRESPONDENCE
Jae-Hyuk Jang,
jangjh@kribb.re.kr
Young-Soo Hong,
hongsoo@kribb.re.kr

SPECIALTY SECTION
This article was submitted to Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and Biotechnology

RECEIVED 04 July 2022
ACCEPTED 21 July 2022
PUBLISHED 25 August 2022

CITATION
Heo KT, Lee B, Jang J-H and Hong Y-S (2022), Elucidation of the di-c-glycosylation steps during biosynthesis of the antitumor antibiotic, kidamycin. *Front. Bioeng. Biotechnol.* 10:985696. doi: 10.3389/fbioe.2022.985696

Elucidation of the di-c-glycosylation steps during biosynthesis of the antitumor antibiotic, kidamycin

Kyung Taek Heo[1,2], Byeongsan Lee[1], Jae-Hyuk Jang[1,2]* and Young-Soo Hong[1,2]*

[1]Chemical Biology Research Center, Korea Research Institute of Bioscience and Biotechnology, Chungbuk, South Korea, [2]Department of Bio-Molecular Science, KRIBB School of Bioscience, University of Science and Technology(UST), Daejeon, South Korea

Kidamycins belong to the pluramycin family of antitumor antibiotics that contain di-C-glycosylated angucycline. Owing to its interesting biological activity, several synthetic derivatives of kidamycins are currently being developed. However, the synthesis of these complex structural compounds with unusual C-glycosylated residues is difficult. In the kidamycin-producing *Streptomyces* sp. W2061 strain, the genes encoding the biosynthetic enzymes responsible for the structural features of kidamycin were identified. Two glycosyltransferase-coding genes, *kid7* and *kid21*, were found in the kidamycin biosynthetic gene cluster (BGC). Gene inactivation studies revealed that the subsequent glycosylation steps occurred in a sequential manner, in which Kid7 first attached N,N-dimethylvancosamine to the C10 position of angucycline aglycone, following which Kid21 transferred an anglosamine moiety to C8 of the C10-glycosylated angucycline. Therefore, this is the first report to reveal the sequential biosynthetic steps of the unique C-glycosylated amino-deoxyhexoses of kidamycin. Additionally, we confirmed that all three methyltransferases (Kid4, Kid9, and Kid24) present in this BGC were involved in the biosynthesis of these amino-deoxyhexoses, N,N-dimethylvancosamine and anglosamine. Aglycone compounds and the mono-C-glycosylated compound obtained in this process will be used as substrates for the development of synthetic derivatives in the future.

KEYWORDS

kidamycin, biosynthetic gene cluster (BGC), C-glycosyltransferase, methyltransferase, angucycline, streptomyces

Abbreviations: BGC, biosynthetic gene cluster; GT, glycosyl transferase; PKS, polyketide synthase; LC-MS, liquid chromatography-mass spectrometry.

1 Introduction

Angucycline compounds of the pluramycin family are a group of naturally occurring antibiotics with antitumor activity (Hansen and Hurley, 1996; Kharel et al., 2012). They contain a 4*H*-anthra (1,2-*b*)pyran-4,7,12-trione substructure with branched side chain at C2, as well as C-glycoside moieties (Figure 1). These compounds intercalate with the minor groove of DNA via sugar-mediated interactions and direct alkylation of guanine in the major groove in a sequence-specific manner (Hansen et al., 1995).

The pluramycin family compounds are divided into two subfamilies that represent different C-glycosylated patterns at C10 and C5 or C8 positions. The classical pluramycins are represented by hedamycin, which contains two amino-deoxyhexoses, an anglosamine sugar attached to the C8 position and a N,N-dimethylvancosamine group attached to the C10 position; and the altromycins, containing altroses at the C5 position (Figure 1). Rubifavinone and sapurimycin are the simplest analogs of the pluramycin family, which lack sugar substitution (Hara et al., 1991; Schumacher et al., 1995; Harunari et al., 2022). Several studies have shown that the selectivity of the pluramycin family compounds correlate with the sugar substitution pattern at the C5, C8, and C10 positions (Prakash et al., 1995; Hansen and Hurley, 1996). Based on this observation, several stereo- and regio-selective glycosylated synthetic derivatives have been developed recently (Kitamura et al., 2014a; Kitamura et al., 2014b; Hartung et al., 2014).

Kidamycin is one of the earliest known members of the pluramycin family (Kanda, 1971; Furukawa et al., 1975). Kidamycin also exhibits cytotoxic activity against various tumors (Rixson et al., 2015). The photokidamycin as well as kidamycin show selective activity on MDA-MB-231, a triple-negative breast cancer cell line (Cho et al., 2019). Although kidamycin consists of planar polycyclic angucycline identical to that of hedamycin, the structure of the subgroups differs because of the presence of a 2-butenyl residue at C2. It also harbors two aminosugars, angolosamine and N,N-dimethylvancosamine, branched at C8 and C10, respectively, attached *via* C-glycosidic bonds. However, the genetic and biochemical basis of kidamycin biosynthesis is still unknown and details of the *di*-C-glycosylation biosynthetic step has not been proposed yet.

Herein we report the cloning and characterization of the biosynthetic gene cluster (BGC) responsible for the biosynthesis of kidamycin and its aglycone with different residues at C2, which are produced by *Streptomyces* sp. W2061. The BGC for kidamycin (*kid*) includes enzymes that are responsible for the biosynthesis of the core polyketide backbone (Kid12–20), nucleotidyl-activated aminosugar moieties (Kid4-9 and Kid21-

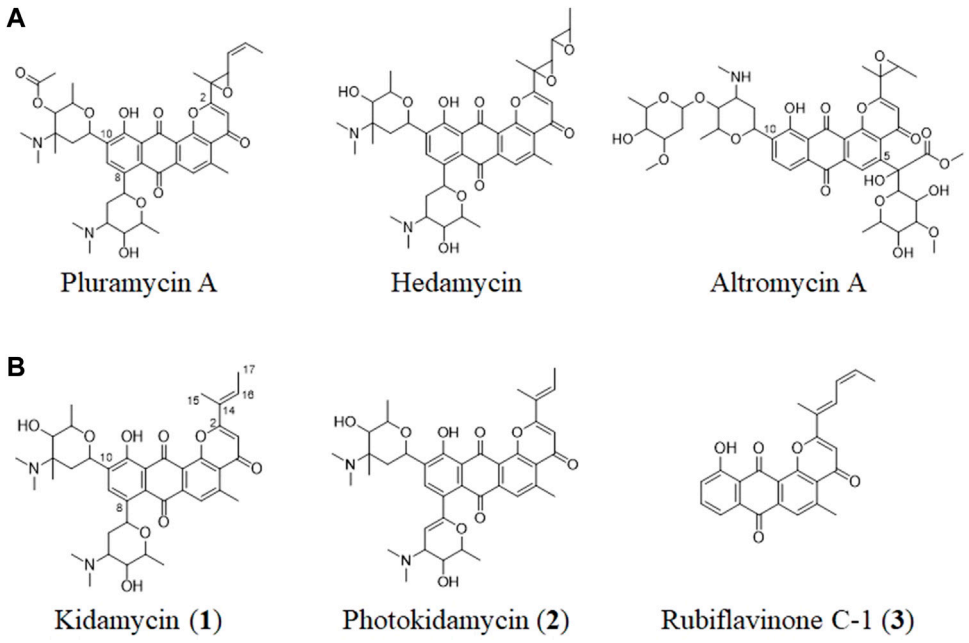


FIGURE 1
(A) Structures of representative pluramycin family angucyclines and **(B)** the kidamycin (**1**), photokidamycin (**2**), and rubiflavinone C-1 (**3**) isolated from *Streptomyces* sp. W2061.

28), and two glycosyltransferases (GTs; Kid7 and Kid21). Gene inactivation studies revealed that the subsequent glycosylation steps occurred in a sequential manner: Kid7 first attached the N,N-dimethylvancosamine moiety to C10 of the angucycline aglycone; and Kid21 then transferred an angolosamine moiety to C8 of the *mono* C-glycosylated angucycline core. In addition, we confirmed that all three methyltransferases (Kid4, Kid9, and Kid24) were involved in the biosynthesis of the aminosugar moieties.

2 Materials and methods

2.1 Plasmids, strains, culture conditions, and extraction

Antibiotics were added to the medium at the following concentrations: apramycin, 50 mg/L; kanamycin, 50 mg/L; and chloramphenicol, 25 mg/L. The restriction enzymes (NEB, United States; Takara, Japan), KOD-plus-DNA polymerase (Toyobo, Japan), PrimeSTAR® GXL DNA polymerase (Takara, Japan) and DNA ligation kit (Takara, Japan) were used according to the manufacturers' manuals. T-Blunt vector (BioFact, Deajeon, Korea) was used to clone the polymerase chain reaction (PCR) products. Gene inactivation experiments were performed using the vector pKC1139 (Kieser et al., 2000), and the kanamycin resistance gene from pFD-NEO-S (Denis and Brzezinski, 1991) was used as the selection marker. *Escherichia coli* DH5α was used for plasmid cloning and amplification, and ET12567/pUZ8002 was used for introducing the plasmid into *Streptomyces* sp. W2061 strain via conjugation. All bacterial strains and plasmids used in this work are summarized in Supplementary Table S1.

Streptomyces sp. W2061 and mutant strains were grown in ISP4 plate (10 g/L soluble starch, 1 g/L K_2HPO_4, 1 g/L $MgSO_4{\cdot}7H_2O$, 1 g/L NaCl, 2 g/L $(NH_4)_2SO_4$, 2 g/L $CaCO_3$, 0.001 g/L $FeSO_4{\cdot}7H_2O$, 0.001 g/L $MnCl_2{\cdot}4H_2O$, 0.001 g/L $ZnSO_4{\cdot}7H_2O$, and 15 g/L agar, pH 7.0–7.4) at 28°C for 4 days. Then, they were inoculated into seed culture M2 medium (2 g/L yeast extract, 5 g/L glucose, 25 ml/L glycerol, 4 g/L soytone, 0.03 g/L $CaCO_3$, and pH 7.2) and incubated for 2 days at 28°C, following which 15 ml seed culture was transferred to 1 L flask containing 300 ml of M2X medium (M2 medium + 5 g/L $MgCO_3$) and incubated for 5–7 days. For compounds isolation for NMR-accessible amounts from the present culture conditions, 30 flasks containing 300 ml of M2X medium were used. The culture broth was extracted with equal volume of ethyl acetate; next, ethyl acetate was dried and the extract was resuspended in methanol for high performance liquid chromatography (HPLC) and liquid chromatography-mass spectrometry (LC-MS) analysis.

For bioconversion, the ΔKid7 mutant was grown in ISP4 plate at 28°C for 4 days and then inoculated into seed culture M2 medium. After 2 days at 28°C, 15 ml of the seed culture was added to a 1-L flask containing 300 ml M2X medium and incubated at 28°C. Compound **7** was added after 3 days and incubated for 4 days. The extract was prepared as mentioned above and analyzed using LC-MS.

2.2 Bioinformatics analysis

The genomic DNA of *Streptomyces* W2061 was obtained and sequenced by PacBio RSII sequencer (Pacific BioSciences, Menlo Park, CA, United States). The sequence raw data were assembled using SMRT Analysis (v2.3.0 HGAP.2) and predicted using the BGC New Genome Annotation System (Newgas, Genotech, Korea) and antiSMASH (Blin et al., 2021). The amino acid sequences of Kid7, Kid21, and other GTs involved in the generation of natural products by *Streptomyces* [AknK (AAF70102), AknS (AAF73455), AraGT (ABL09968), Asm25 (AAM54103), DesVII (Q9ZGH7), ElaGT (ADP68587), EryBV (AAB84072), EryCIII (A4F7P3), GilGT (AAP69578), Gra-ORF14 (CAA09635), HedJ (AA85354), HedL (AAP85354), LanGT (AAD13562), LanGT2 (AAD13553), Med8 (BAC79040), SsfS6 (ADE34512), SunS (KIX81208), UrdGT2 (AAF00209) and VlnC (BAJ52701)] were aligned using the Clustal W method. The amino acid sequences of the kidamycin genes are described in the Supporting Information. Phylogenetic tree analysis using the minimum evolution method was based on the results of sequence alignment, and evolutionary distances were computed using the Poisson correction method in the MEGA7 software (Kumar et al., 2016).

2.3 Construction of disruption vectors and mutant strains

The gene knockout recombinants were generated using homologous recombination. The target region was replaced with a kanamycin resistance gene. The kanamycin resistance cassette was digested with KpnI and PstI or SalI from plasmid pFD-NEO-S (Denis and Brzezinski, 1991). Two homologous regions were amplified using the appropriate primers (Supplementary Table S1). For constructing all gene replacement vectors, three fragments were ligated with plasmid pKC1139 and introduced into *Streptomyces* sp. W2061 strain via conjugation from *E. coli* ET12567/pUZ8002. The exoconjugants were selected based on antibiotic resistance (kanamycin and apramycin) and PCR genotyping. The double-crossover mutants were screened based on kanamycin resistance and PCR genotyping (kid4_scF/R for ΔKid4 mutant, kid7_scF/R for ΔKid7 mutant, kid9_scF/R for ΔKid9 mutant, kid21_scF/R for ΔKid21 mutant, and kid24scF/R for ΔKid24 mutant; Supplementary Table S2).

In particular, to inactivate *kid4,* a 1 kb EcoRI/KpnI fragment generated using kid4-H1-F and kid4-H1-R, and a 1.1 kb PstI/HindIII fragment generated using kid4-H2-F and Kid4-H2-R, were ligated and cloned into the EcoRI and HindIII sites of pKC1139 to yield pKC-Kid4neo. To inactivate *kid7,* a 1.1 kb EcoR1/SalI fragment generated using kid7-H1-F and kid7-H1-R, and a 1.1 kb SalI/HindIII fragment generated using kid7-H2-F and Kid7-H2-R, were ligated and cloned in the EcoRI and HindIII sites of pKC1139 to yield pKC-Kid7neo. To inactivate *kid9,* a 1 kb EcoR1/KpnI fragment generated using kid9-H1-F and kid9-H1-R, and a 1.3 kb PstI/HindIII fragment generated using kid9-H2-F and Kid9-H2-R, were ligated and cloned in the EcoRI and HindIII sites of pKC1139 to yield pKC-Kid9neo. To inactivate the *kid19*, a 1kb EcoR1/KpnI fragment generated using kid19-H1-F and kid19-H1-R and a 1.1kb PstI/HindIII fragment generated using kid19-H2-F and Kid19-H2-R were ligated and cloned in the EcoRI and HindIII sites of pKC1139 to yield pKC-Kid19neo. To inactivate the *kid21,* a 1.5 kb EcoR1/SalI fragment generated using kid21-H1-F and kid21-H1-R, and a 1.2 kb SalI/HindIII fragment generated using kid21-H2-F and Kid21-H2-R, were ligated and cloned in the EcoRI and HindIII sites of pKC1139 to yield pKC-Kid21neo. To inactivate *kid24*, a 1.4 kb EcoR1/SalI fragment generated using kid24-H1-F and kid24-H1-R, and a 1.3 kb SalI/HindIII fragment generated using kid24-H2-F and Kid24-H2-R, were ligated and cloned in the EcoRI and HindIII sites of pKC1139 to yield pKC-Kid24neo (Supplementary Table S1 and Supplementary Figure S1).

2.4 LC-MS analysis

The samples were dissolved in methanol and analyzed using a Thermo U3000-LTQ XL ion trap mass spectrometer (Thermo Scientific, Waltham, MA, United States) equipped with an electrospray ionization (ESI) mass source. Chromatographic separation of the compounds was achieved using a Waters HSS T3 C18 column (2.1 × 150 mm; 2.5 μm) at the flow rate of 0.3 ml/min. The mobile phases, A and B, contained 0.1% formic acid along with water and acetonitrile, respectively. Gradient elution was performed as follows: 5–100% B for 0–15 min with a linear gradient, followed by 5 min of 100% B. The MS/MS system was operated in ESI mode. The operating parameters were as follows: spray needle voltage, +5 kV; ion transfer capillary temperature, 275°C; nitrogen sheath gas, 35 (arbitrary units); auxiliary gas, 5 (arbitrary units). The ion trap contained helium damping gas, which was introduced in accordance with the manufacturer's recommendations. Mass spectra were acquired in an *m/z* range of 100–2000, applying three microscans and a maximum ion injection time of 100 ms. Data-dependent mass spectrometry experiments were controlled using the menu-driven software provided in the Xcalibur system (version 4.0; Thermo Scientific).

2.5 Compound isolation

For isolation of compound **7**, the crude extract (7.3g) of ΔKid21 mutant was fractionated employing reversed-phase C18 vaccum column chromatography eluting with a stepwise MeOH:H_2O solvent system of (20: 80 to 100: 0, each × 1 L). The 70 % (1043.4 mg) fraction was further fractionated using a CombiFlash RF (Teledyne ISCO, Lincoln, NE, United States) medium-pressure chromatography system (MPLC) on a Redisep RF C18 reverse-phase column under stepwise gradient elution with MeOH-H_2O (from 20: 80, 40: 60, 60: 40, 80: 20 to 100: 0; 1 L for each step). MPLC fraction was separated using semipreparative HPLC (Waters Atlantis T3 C18 column: 10 × 250 mm, 5 μm) with an isocratic solvent system (45 % MeOH-H_2O [0.05% trifluoroacetic acid (TFA)] over 20 min, UV 254 nm detection, flow rate: 3 ml/min) to obtain fraction containing compound **7** (26.0 mg). The separated fraction (26.0 mg) was further subjected to semipreparative HPLC (Waters Atlantis T3 C18 column (10 × 250 mm, 5 μm) a gradient solvent system (30% CH_3CN-H_2O (0.05 % TFA) to 100% CH_3CN, 3 ml/min) over 20 min to yield compounds **7** (9.0 mg). The aglycone compounds **3**, **5**, and **6** are isolated by 2.1 mg, 4.2 mg, and 6.2 mg from ΔKid4 mutant, respectively. The other aglycone compound **4** (4.2 mg) is purified from ΔKid7 mutant. The structures of the purified compounds were determined based on ^{1}H and ^{13}C nuclear magnetic resonance (NMR) analysis, as well as HSQC, COSY and HMBC NMR experiments.

3 Results

3.1 Confirmation of the kidamycin BGC by disrupting the polyketide synthase gene

Among the pluramycin family angucyclines, the mechanism of hedamycin biosynthesis, which involves the assembly of the angucycline core by a rare hybrid type I/type II PKS, has been studied extensively (Bililign et al., 2004; Das and Khosla, 2009). The structural similarities between kidamycin and hedamycin was indicative of substantial shared biosynthetic logic. We identified a putative kidamycin BGC (*kid*) in *Streptomyces* sp. W2061 using detailed bioinformatics analysis from draft whole genome sequencing data (Supplementary Table S3). The *kid* cluster is similar in content and organization to the known hedamycin BGC identified in *S. griseoruber* (Bililign et al., 2004) (Figure 2). The results of antiSMASH analysis revealed only one gene cluster including the hybrid type I/type II PKS, which was 87% similar to the hedamycin BGC. Sequence analysis of the 60 kb region revealed 59 open reading frames (ORFs), which included putative genes involved in the biosynthesis of the angucycline core, construction of angolosamine and N,N-dimethylvancosamine sugars, and tailoring of the core scaffold.

The structure of the kidamycin angucycline core is identical to that of hedamycin, except that its core has a 2-butenyl residue

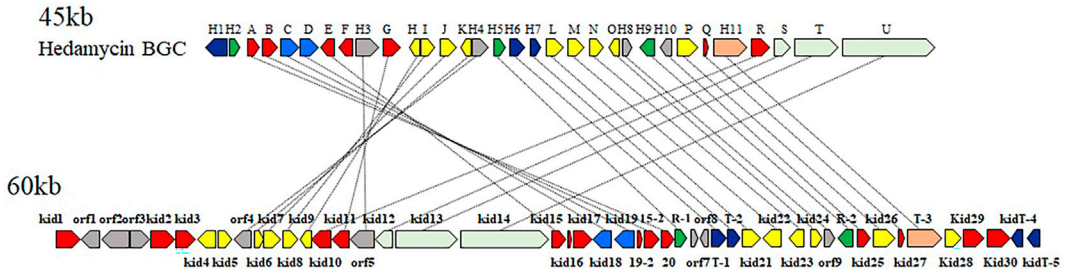


FIGURE 2
Comparison of the biosynthetic gene clusters encoding hedamycin (top) and kidamycin (bottom). Each arrow represents the direction of transcription of an open reading frame. Please see Supplementary Table S3 for complete cluster annotation.

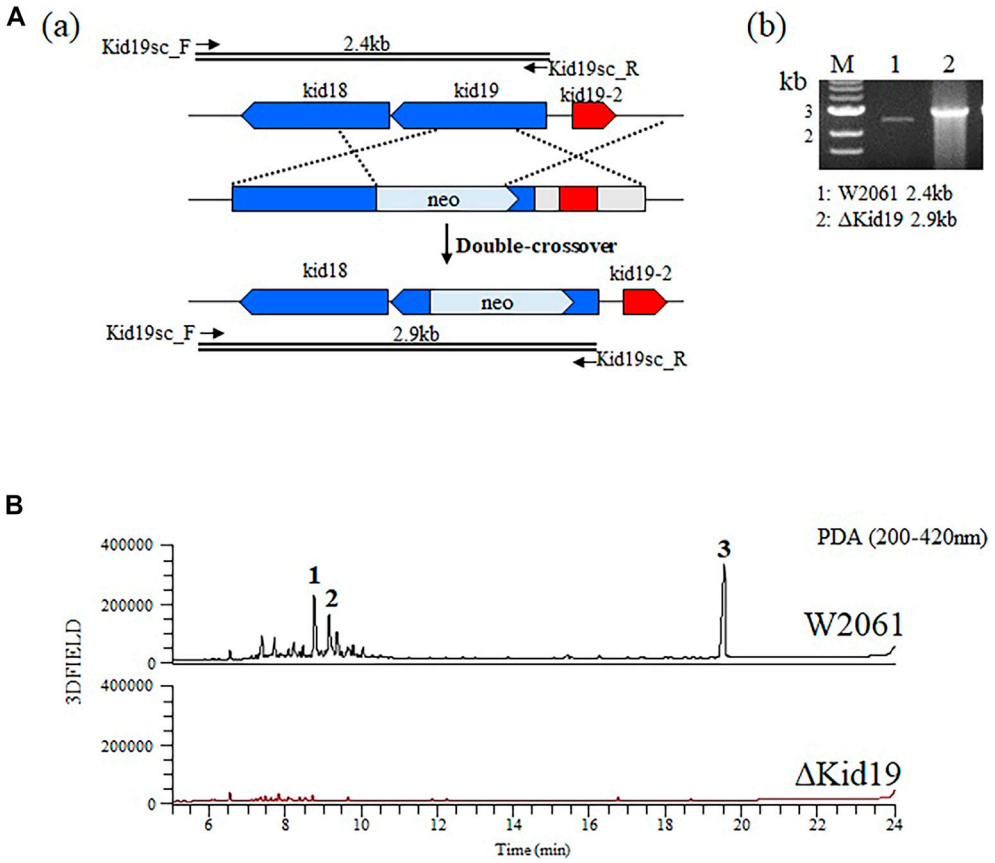


FIGURE 3
Strategies for gene disruption. **(A)** The ΔKid19 mutant resulted from a double crossover event to produce a kanamycin-resistant strain in which the *kid19* gene was disrupted (a). Confirmation of insertional *kid19* gene inactivation using PCR and the total genomic DNA of each mutant as the template (b). The primers (kid19sc_F & kid19sc_R; Supplementary Table S2) used to amplify the desired DNA fragments are indicated by solid arrows. M, 1 kb ladder; 1, wild type; 2, ΔKid19 mutant. **(B)** Comparative HPLC analysis of crude fermentation extracts showing production of kidamycins (**1** and **2**) and rubiflavinone C-1 (**3**) by strains W2061 and ΔKid19, revealing loss of kidamycins and rubiflavinone C-1 production due to the *kid19* gene deletion.

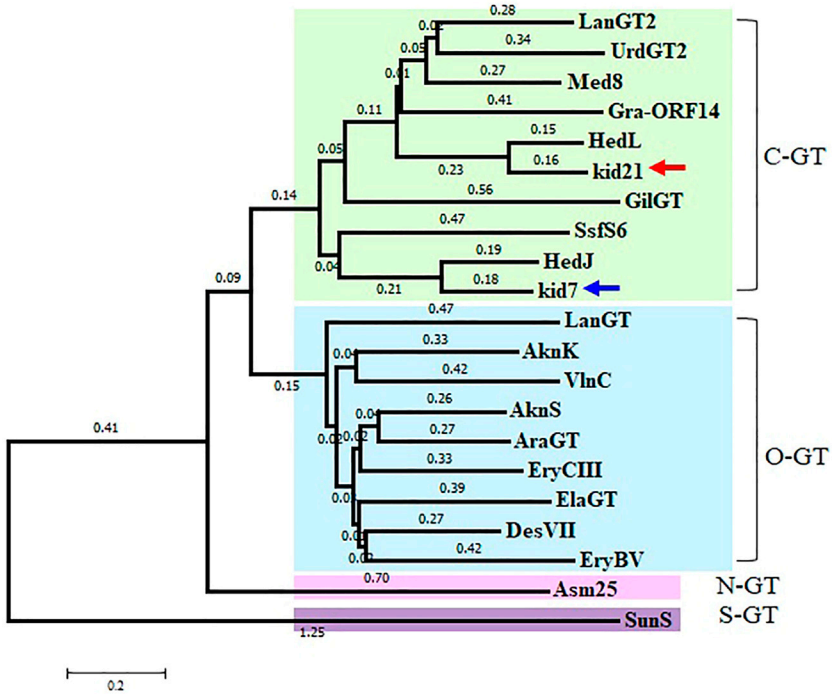


FIGURE 4
Phylogenetic analysis of 2 GTs, Kid7 and Kid21, compared to known GTs from *Streptomyces*. Arrow indicates the 2 GTs of this study. The sources and GenBank accession numbers of GTs involved in natural product biosynthesis by *Streptomyces*: AknK (AAF70102), AknS (AAF73455), AraGT (ABL09968), Asm25 (AAM54103), DesVII (Q9ZGH7), ElaGT (ADP68587), EryBV (AAB84072), EryCIII (A4F7P3), GilGT (AAP69578), Gra-ORF14 (CAA09635), HedJ (AA85354), HedL (AAP85354), LanGT (AAD13562), LanGT2 (AAD13553), Med8 (BAC79040), SsfS6 (ADE34512), SunS (KIX81208), UrdGT2 (AAF00209), and VlnC (BAJ52701). Amino acid sequences were aligned using the ClustalW method and the phylogenetic tree analysis was performed based on the sequence alignment using the Minimum Evolution method in the MEGA7 software.

attached to C2 instead of the bisepoxide group found in hedamycin. We proposed that this kidamycin core group also arises from the use of a hybrid type I/type II PKS system. The *kid* cluster contains homologs of the type II PKS core components [chain length factor (CLF) Kid18, ketosynthase (KS) α Kid19, acyl carrier protein (ACP), Kid16] and tailoring enzymes [ketoreductase (KR) Kid20, cyclases Kid15, Kid15-1, and Kid19-1, and oxidase Kid27] required to construct an angucycline core polyketide (Supplementary Figure S2).

To validate that the *kid* gene cluster was indeed responsible for kidamycin biosynthesis, an insertional inactivation was generated with a kanamycin resistance marker to disrupt the function of the putative KSα, Kid19. The mutants were confirmed using PCR. A successful double crossover mutant, ΔKid19, was confirmed using PCR and fermented parallel to the parent strain (Figure 3A). The ΔKid19 mutant did not produce any detectable kidamycins or rubiflavinone C-1, confirming the role of Kid19 and the *kid* gene cluster in the production of these products (Figure 3B). Therefore, this *kid* BGC, at least the type II PKSs, play an important role in angucycline core biosynthesis, indicating that they are used together in the production of angucycline compounds with two different side chains, kidamycin and rubiflavinone C-1.

3.2 Sequence analysis of two putative C-glycosyl transferases (GTs) in the *kid* BGC

Kidamycin is characterized by the C-glycosidic attachment of two amino sugars to the angucycline core, a process anticipated to require GTs. After formation of angucycline core polyketides, a complex series of glycosylation events is required for generation of the final di-C-glycosidic structure. We identified 12 putative genes (*kid4–kid9*, *kid21-24*, *26*, and *28*) in the *kid* BGC, consistent with the biosynthesis of two amino sugar moieties and subsequent attachment to the angucycline core (Figure 2 and Supplementary Table S3).

Identification of 2 GTs in the *kid* BGC met the general requirement of one GT-one sugar for the sugar moieties of microbial natural products. Phylogenetic tree analysis of Kid7 and Kid21 indicated their involvement in C-glycosylation (Figure 4). Kid7 and Kid21, which are

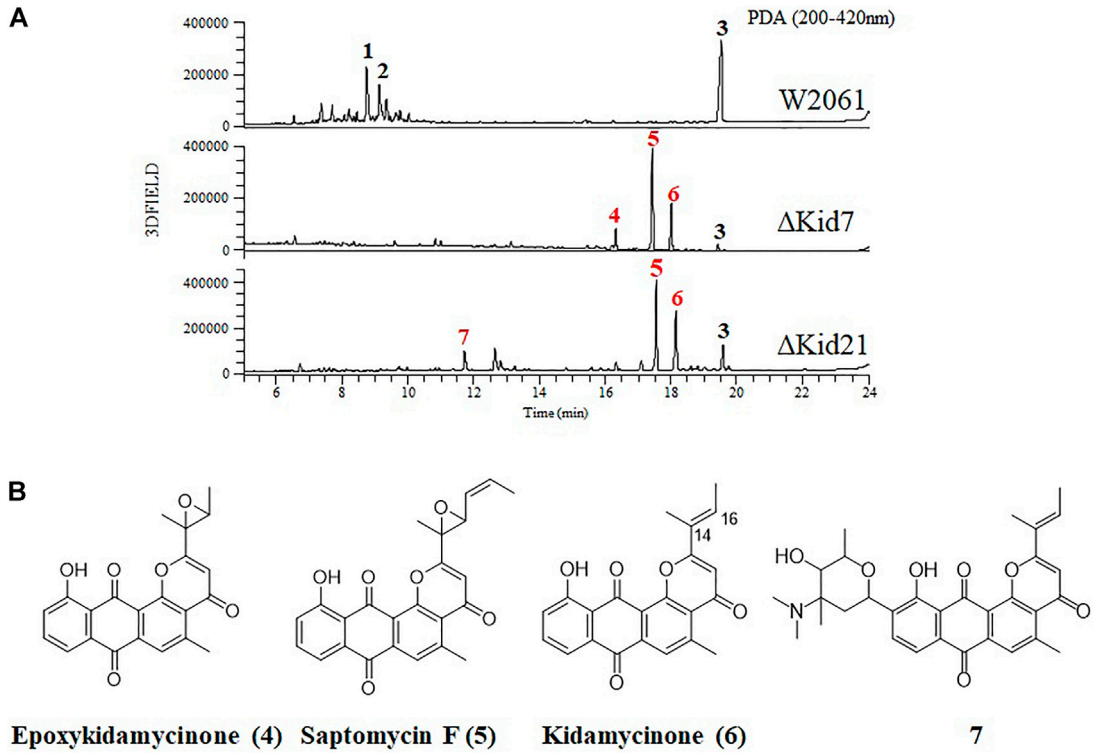


FIGURE 5
Disruption of *kid7* and *kid 21* gene loci. **(A)** HPLC analysis of the products of *Streptomyces* sp. W2061 (upper), ΔKid7 (middle), and ΔKid21 (bottom). **(B)** Structure of compounds epoxykidamycinone (**4**), saptomycin F (**5**), kidamycinone (**6**), and **7** isolated from the mutants.

located far away from the polyketide core genes, shows significant homology to the C-GT found in the hedamycin BGC (HedJ and HedL) (Figure 4). Alignment of the amino acid sequences of Kid7 and Kid21 with those of HedJ and HedL, respectively, also showed that the proteins are significantly identical (Kid7 vs. HedJ, 67.5% similarity; Kid21 vs. HedL, 71.8% similarity) (Supplementary Table S4). Additionally, both Kid7 and Kid21 resemble Med-8 (52% and 59% similarity, respectively), a C-GT responsible for angolosamine transfer in medermycin biosynthesis (Ichinose et al., 2003; Cai et al., 2021). However, Kid21 is believed to be responsible for angolosamine transfer, because of its high degree of identity with Med-8.

3.3 Identification of sequential C-glycosylation steps in kidamycin biosynthesis

The function of Kid21 was elucidated using targeted gene inactivation. The pKC-kid21neo plasmid carrying the kanamycin resistance cassette within *kid21* was used to replace the chromosomal allele of this gene in *Streptomyces* sp. W2061 via a double crossover event. The mutation was confirmed using PCR (Supplementary Figure S1).

Kidamycins were not detected in the extracts of the ΔKid21 mutant. Instead, several new peaks were detected in the extract from the mutant strain, which were absent in the extract of the wild-type strain. However, the aglycone compound, rubiflavinone C-1 (**3**), identified in the parent strain, was still produced by the ΔKid21 mutant (Figure 5). The ESI-MS spectrum of the new peak at 11.5 min (**7**) was generated at the *m/z* 532 $[M + H]^+$ (Supplementary Figure S3), which was the same as that of the N,N-dimethylvancosamine moiety attached to the aglycone of kidamycin (kidamycinone). In particular, ions with *m/z* = 361 $[M + H]^+$, indicating loss of the N,N-dimethylvancosamine moiety (171 Da), appeared at the MS/MS fragmentation profile of the ion with *m/z* = 532. This indicated that the product was N,N-dimethylvancosamine glycosylated kidamycinone. The new peaks that appeared in the ΔKid21 mutant were purified from a 9 L fermentation and the chemical structure was obtained using NMR. The structure of **7** was confirmed to be that of N,N-dimethylvancosamine C10 glycosylated kidamycinone based on a comparison of the 1D and 2D NMR spectroscopic data of kidamycin with previously reported data (Supplementary Table S5 and

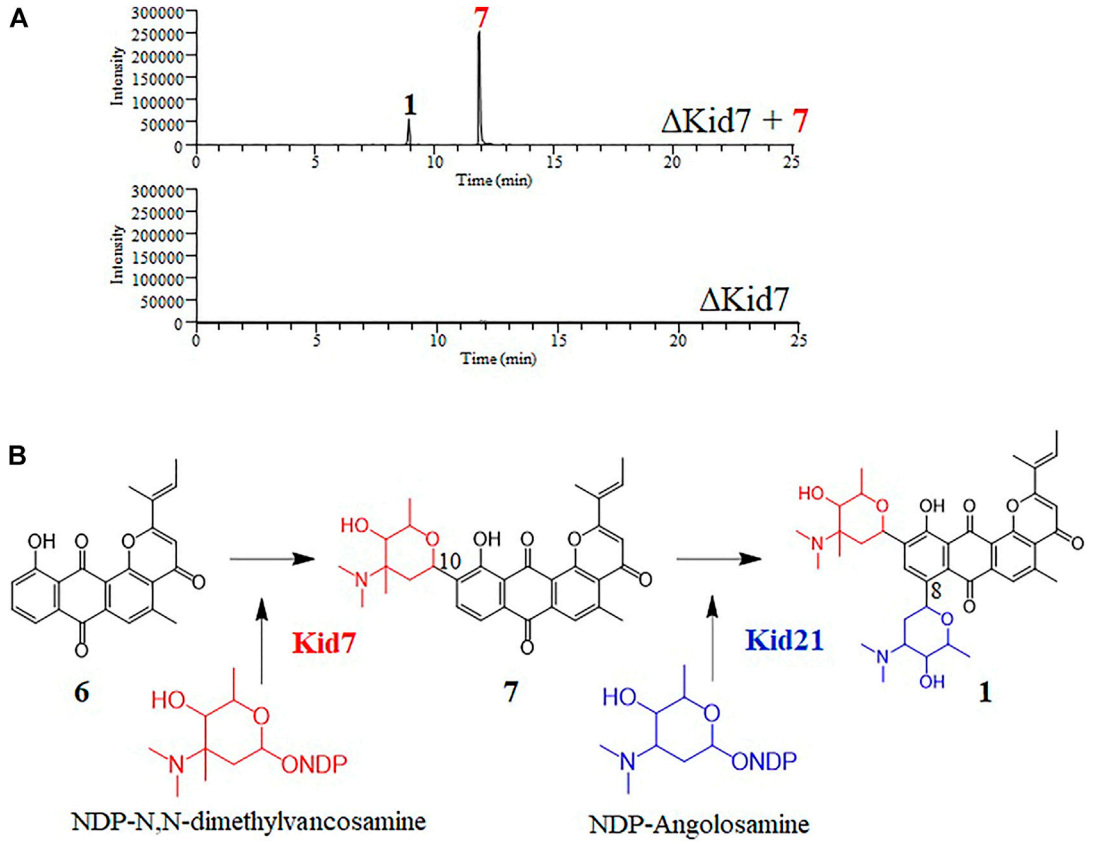


FIGURE 6
(A) Selected LC-MS analysis of the extracts of culture broth obtained from the bioconversion experiment of N,N-dimethylvancosamine glycosylated compound (**7**) by the ΔKid7 mutant. The selected ion chromatograms were pre-set at *m/z* 532 and 689 correspond to the respective molecular ion of compound **7** and kidamycin (**1**), respectively. **(B)** The proposed glycosylation pathway of kidamycin.

Supplementary Figure S4–12). The NMR spectroscopic data of **7** were similar to those of kidamycin, except that **7** did not show any additional signals corresponding to the angolosamine group connected to C8. In addition, the other new peaks at 17 min (**5**) and 18.2 min (**6**) showed the $[M + H]^+$ ion signal at *m/z* 403 and *m/z* 361, respectively. One- and two-dimensional NMR analyses identified these compounds to possess agalycone structure (Supplementary Table S6), lacking the aminosugars both at C8 and C10. Compound **5** contained an epoxide in C14, 16 of rubiflavinone C-1 (**3**) and **6** was kidamycinone, an aglycone of kidamycin. Taken together, the accumulation of N,N-dimethylvancosamine glycosylated kidamycinone (**7**) in the ΔKid21 mutant implied that Kid21 was an angolosamine GT for kidamycin. Thus, the N,N-dimethylvancosamine GT reaction takes precedence over the angolosamine GT reaction.

Next, we mutated *kid7*, another C-GT gene in *kid* BGC, using the same method as mentioned above. Results showed that the *kid7* mutant produced only aglycone compounds (**3**, **4**, **5**, and kidamycinone) (Figure 5). A new peak at 16.5 min (**4**) was generated at *m/z* = 377 $[M + H]^+$, which was identified to be a structure with an epoxide in C14,16 of kidamycinone (Figure 5). As a result, the ΔKid7 mutant produced only aglycone compounds without any attached aminosugar moieties. Therefore, the ΔKid7 mutant produced only aglycone compounds and the ΔKid21 mutant produced *mono*-glycosylated compound. *kid7* gene encodes the C10 N,N-dimethylvancosamine GT and catalyzed the first C-GT reaction in kidamycin biosynthesis. Therefore, Kid21 is expected to be responsible for the second glycosylation step in kidamycin biosynthesis. To verify this, purified N,N-dimethylvancosamine glycosylated compound **7** produced by the ΔKid21 mutant was added to the culture of the ΔKid7 mutant with activated Kid21 C-GT function and angolosamine production. The selected ion peaks were detected on the selected ion monitoring on LC/MS analysis with the pre-set at *m/z* 532 and 689 correspond to the respective molecular ion of compound **7** and kidamycin (**1**), respectively (Figure 6A). The N,N-dimethylvancosamine glycosylated compound **7** was effectively converted to kidamycin in the culture broth of the ΔKid7 mutant. This

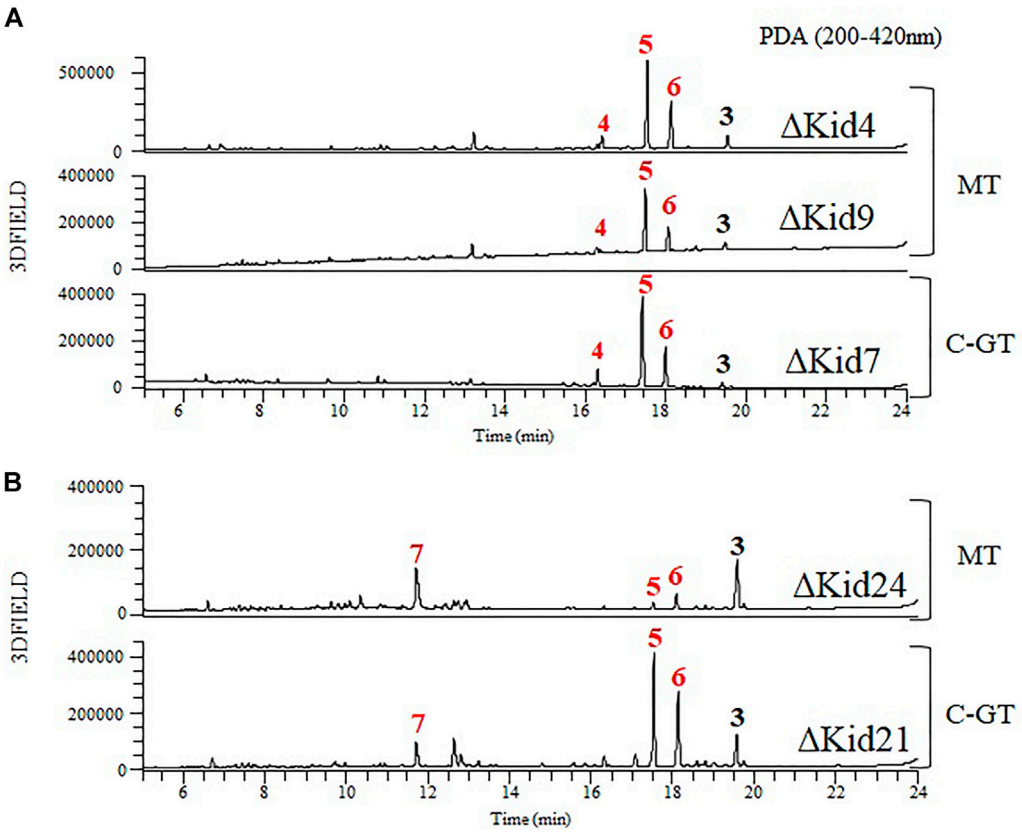


FIGURE 7
Comparative HPLC analysis of the products of the methyltransferase (MT) mutants and C-GT mutants. **(A)** ΔKid4 (upper), ΔKid9 (middle), and ΔKid7 (bottom). **(B)** ΔKid24 (upper) and ΔKid21 (bottom).

confirmed that angolosamine glycosylation at C8 is not only the second step in the kidamycin glycosylation pathway, but also that **7** is a substrate of the Kid21 angolosamine C-GT reaction (Figure 6B).

3.4 Role of three methyltransferases in aminosugar biosynthesis

The early steps are common with those of aminosugar biosynthesis, which involve NDP-glucose synthase, NDP-glucose-4,6-dehydratase, and NDP-glucose-2,3-dehydratase, all of which were reasonably assigned to the respective gene products of *kid5*, *kid8*, and *kid26*, respectively, based on significant percent identity (>65%). The Kid22 protein resembles a NDP-deoxyhexose 3-aminotransferase that is involved in the biosynthesis of *Streptomyces* antibiotics containing aminosugars. The late steps for angolosamine biosynthesis, which may be interchangeable, are catalyzed by 4′-keto-reductase and N-methyltransferase. The 4′-keto-reductase genes, med-ORF14 (Cai et al., 2021) and gra-ORF22 (Tornus and Floss, 2001), encode proteins 60% similar to the Kid23 enzyme. Kid4 is highly similar to AORI_1488 (76%), which catalyzes the C-methylation at aminohexose in the vancomycin biosynthetic pathway (van Wageningen et al., 1998; Chen et al., 2000; Xu et al., 2014). Additionally, Kid9 and Kid24 show high identity with med-15 (Kid9: 46% and Kid24: 50%), which catalyze the di-methylation at the amine moiety of angolosamine. (Cai et al., 2021) (Supplementary Table S7)

To obtain further insights regarding aminosugar biosynthesis, we inactivated three methyltransferase genes (*kid4*, *kid9*, and *kid24*) and analyzed the function of the methyltransferases. The mutant strains were selected based on kanamycin resistance and genotyping using PCR (Supplementary Figure S1). These mutant strains did not produce kidamycin but yielded new peaks. The peaks produced by the ΔKid4 and ΔKid9 mutants were the same as those of the ΔKid7 C10-GT mutant, and the peaks produced by the ΔKid24 mutant were the same as those of the ΔKid21 C8-GT mutant (Figure 7). This indicated that Kid4 and Kid9 were involved in N,N-dimethylvancosamine biosynthesis, and that

Kid24 was involved in the angolosamine biosynthetic pathway. All three methyltransferases (Kid4, Kid9, and Kid24) present in this BGC were involved in the biosynthesis of the aminosugars, N,N-dimethylvancosamine or anglosamine. Interestingly, the two methyltransferase genes (*kid4* and *kid9*) near *kid7* encoded C10 N,N-dimethylvancosamine GT, while *kid24* is a neighbor of the C8 angolosamine GT (*kid21*) gene. Therefore, it is not found in *kid* BGC that corresponds to the methyltransferase of methyl bis-epoxide residue of the hedamycin aglycone. Therefore, the gene corresponds to the methyltransferase of methyl *bis*-epoxide residue of the hedamycin aglycone is not found in *kid* BGC (Bililign et al., 2004).

4 Discussion

It is well known that the glycosylation of natural or synthetic compound-based drugs can dramatically influence the pharmacological properties of the parent molecule (Weymouth-Wilson, 1997; Elshahawi et al., 2015). Therefore, the characterization of the biosynthetic pathway of glycoside compounds and the study of the substrates specificity of GTs play an important part in the development of new drugs (Fu et al., 2003; Salas and Méndez, 2007; Singh et al., 2012). In particular, the development of new derivatives using GT enzymes with relatively broad substrate promiscuity is being actively carried out (Luzhetskyy et al., 2008; Pandey and Sohng, 2016; Xie et al., 2020; Goel et al., 2021). However, most of these approaches are used with O-GTs, only a small number of C-GTs have been found and are limited in their use.

C-GTs catalyze the transfer of activated sugar moieties to the carbon atoms in substituted aromatic rings of receptor molecules and generally show high selectivity for receptors and sugar moieties. Because of the regio- and stereospecifically attached sugar moieties, C-GTs also play important roles in biological activities and in the biosynthesis of pharmaceutically significant natural products (Kren and Martínková, 2001; Elshahawi et al., 2015; Putkaradze et al., 2021). As is evident from the recent development of various synthetic derivatives for pluramycin family compounds, modification in the sugar moieties is an important factor determining their pharmacological activities. However, compared to the studies on synthetic derivatives, those on biosynthetic pathways are rare, with the only reports being on the discovery of the 2 GTs, HedJ and HedL, in hedamycin BGC (Bililign et al., 2004; Das and Khosla, 2009).

This is the first study to show the precise glycolytic function of Kid7 and Kid21, 2 GTs involved in kidamycin biosynthesis, and the biosynthetic order of each C-GT using gene disruption and bioconversion experiments. Both Kid7 and Kid21 showed significant homology with angolosamine C-GT of medermycin (Med-8), although the identity of Kid21 was slightly more than that of Kid7. Owing to these slight differences, Kid21 was classified as angolosamine C-GT, while Kid7 was classified as N,N-dimethylvancosamine C-GT. Thus, both C-GTs showed strong selectivity for sugar moieties and regiospecifically attached sugar moieties to the different carbon atoms in the angucycline ring of kidamycin. This study provides a basis how enzymatically generates unusual sugars attached angucycline compounds and has paved the way for alteration of these glycosylated compounds via pathway or enzyme engineering (Blanchard and Thorson, 2006; Zhao and Liu, 2010; He et al., 2019). *Streptomyces* sp. W2061, a kidamycin producer, also accumulated rubiflavinone C-1, which harbors an aglycone with a hexenoate residue, unlike the 2-methylbutenoyl residue of kidamycin, under the same culture condition. The glycosylated derivatives of rubiflavinone are also detected in small amounts in the same culture by molecular networking searching (data not shown). This indicated that among the two different aglycones of varying chain lengths (kidamycinone and rubiflavinone), the 2 GTs have considerable specificity for 2-methylbutenoyl of kidamycinone, although they are produced by a shared aglycone biosynthetic system that possibly uses two different precursors. Therefore, replacement with C-GTs of hedamycin, which mainly produce aglycone with hexenoate residue, will provide new insights regarding the substrate specificity of C-GT.

Additionally, three methyltransferase genes showed homology with the genes previously reported to be involved in amniosugar biosynthesis. Kid4 and Kid9 are involved in N,N-dimethylvancosamine biosynthesis, while Kid24 is involved in the angolosamine biosynthetic pathway. Interestingly, these genes were located near the C-GT genes, and each of these methyltransferases used biosynthetic aminosugars as substrates. *kid4* and *kid9* are adjacent to the N,N-dimethylvancosamine GT (*kid7*), and *kid21* is adjacent to *kid24*. These two loci are separated by the PKS genes. However, genes involved in the early biosynthesis of aminohexose, glucose-1-phosphate thymidyl transferase (*kid5*), NDP-4-keto-6-deoxyhexose 3,5-epimerase (*kid6*), TDP-glucose 4,6-dehydratase (*kid28*), and NDP-deoxyhexose 3-aminotransferase (*kid20*) are scattered on both sides of the glycosylation gene locus.

In conclusion, we demonstrated that the two C-GTs, Kid7 and Kid21, in *Streptomyces* sp. are involved in sequential glycosylation in kidamycin biosynthesis. Kid7 first attached N,N-dimethylvancosamine to the C10 position of aglycone, following which Kid21 transferred an anglosamine moiety to C8 of the C10-glycosylated kidamycinone. This is the first report to reveal the sequential biosynthetic steps of the C-glycosylated amino-deoxyhexoses of pluramycin family angucycline compounds. Additionally, all the three methyltransferases (Kid4, Kid9, and Kid24) present in the kidamycin BGC were involved in the biosynthesis of these amino sugars, and were not related to methylation of the side chain of aglycone, as predicted in the production of methyl bis-epoxide residue of hedamycin. The nine aglycones and one mono-glycoside compound produced in

these genetically modified strains will provide new substrates for bioconversion experiments with other C-GTs or for organic synthesis, which will assist in developing pluramycin-related natural/non-natural compounds.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding authors.

Author contributions

KH and BL performed experimental work whereas J-HJ and Y-SH drafted the manuscript. All authors were involved in designing, discussing, and interpreting the results of the experiments. All authors read and approved the final manuscript.

Funding

This work was supported by the Basic Science Research Program (NRF 2020R1I1A206871313) of the Ministry of Education, the National Research Foundation of Korea (NRF) grant and the KRIBB Research Initiative Program (KGM5292221 and KGM5402221) funded by the Ministry of Science and ICT (MSIT) of the Republic of Korea.

Acknowledgments

The authors thank the National Research Foundation of Korea and KRIBB for financial support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.985696/full#supplementary-material

References

Bililign, T., Hyun, C.-G., Williams, J. S., Czisny, A. M., and Thorson, J. S. (2004). The hedamycin locus implicates a novel aromatic PKS priming mechanism. *Chem. Biol.* 11 (7), 959–969. doi:10.1016/j.chembiol.2004.04.016

Blanchard, S., and Thorson, J. S. (2006). Enzymatic tools for engineering natural product glycosylation. *Curr. Opin. Chem. Biol.* 10 (3), 263–271. doi:10.1016/j.cbpa.2006.04.001

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49 (W1), W29–W35. doi:10.1093/nar/gkab335

Cai, X., Taguchi, T., Wang, H., Yuki, M., Tanaka, M., Gong, K., et al. (2021). Identification of a C-glycosyltransferase involved in medermycin biosynthesis. *ACS Chem. Biol.* 16 (6), 1059–1069. doi:10.1021/acschembio.1c00227

Chen, H., Thomas, M. G., Hubbard, B. K., Losey, H. C., Walsh, C. T., and Burkart, M. D. (2000). Deoxysugars in glycopeptide antibiotics: enzymatic synthesis of TDP-L-epivancosamine in chloroeremomycin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 97 (22), 11942–11947. doi:10.1073/pnas.210395097

Cho, Y. Y., Hong, Y.-S., Choi, K.-I., Jang, J.-H., and Lee, C.-J. (2019). Kidamycin derivative L1-95-1, and composition for preventing or treating cancer, containing same as active ingredient. KOREA patent application WO/2019/031880.

Das, A., and Khosla, C. (2009). *In vivo* and *in vitro* analysis of the hedamycin polyketide synthase. *Chem. Biol.* 16 (11), 1197–1207. doi:10.1016/j.chembiol.2009.11.005

Denis, F. O., and Brzezinski, R. (1991). An improved aminoglycoside resistance gene cassette for use in Gram-negative bacteria and *Streptomyces*. *FEMS Microbiol. Lett.* 81, 261–264. doi:10.1111/j.1574-6968.1991.tb04769.x

Elshahawi, S. I., Shaaban, K. A., Kharel, M. K., and Thorson, J. S. (2015). A comprehensive review of glycosylated bacterial natural products. *Chem. Soc. Rev.* 44 (21), 7591–7697. doi:10.1039/c4cs00426d

Fu, X., Albermann, C., Jiang, J., Liao, J., Zhang, C., and Thorson, J. S. (2003). Antibiotic optimization via *in vitro* glycorandomization. *Nat. Biotechnol.* 21 (12), 1467–1469. doi:10.1038/nbt909

Furukawa, M., Hayakawa, I., Ohta, G., and Iitaka, Y. (1975). Structure and chemistry of kidamycin. *Tetrahedron* 31 (23), 2989–2995. doi:10.1016/0040-4020(75)80324-3

Goel, B., Tripathi, N., Mukherjee, D., and Jain, S. K. (2021). Glycorandomization: a promising diversification strategy for the drug development. *Eur. J. Med. Chem.* 213, 113156. doi:10.1016/j.ejmech.2021.113156

Hansen, M. R., and Hurley, L. H. (1996). Pluramycins. Old drugs having modern friends in structural biology. *Acc. Chem. Res.* 29 (5), 249–258. doi:10.1021/ar950167a

Hansen, M., Yun, S., and Hurley, L. (1995). Hedamycin intercalates the DNA helix and, through carbohydrate-mediated recognition in the minor groove, directs N7-alkylation of guanine in the major groove in a sequence-specific manner. *Chem. Biol.* 2 (4), 229–240. doi:10.1016/1074-5521(95)90273-2

Hara, M., Takiguchi, T., Ashizawa, T., Gomi, K., and Nakano, H. (1991). Sapurimycin, new antitumor antibiotic produced by *Streptomyces*. Producing organism, fermentation, isolation and biological properties. *J. Antibiot.* 44 (1), 33–39. doi:10.7164/antibiotics.44.33

Hartung, J., Wright, D. B. J. D., and Danishefsky, S. J. (2014). Studies toward the total synthesis of pluraflavin A. *Chem. Eur. J.* 20 (28), 8731–8736. doi:10.1002/chem.201402254

Harunari, E., Bando, M., and Igarashi, Y. (2022). Rausuquinone, a non-glycosylated pluramycin-class antibiotic from *Rhodococcus*. *J. Antibiot.* 75 (2), 86–91. doi:10.1038/s41429-021-00489-0

He, J. B., Zhao, P., Hu, Z. M., Liu, S., Kuang, Y., Zhang, M., et al. (2019). Molecular and structural characterization of a promiscuous C-glycosyltransferase from *Trollius chinensis*. *Angew. Chem. Int. Ed. Engl.* 131 (33), 11513–11520. doi:10.1002/anie.201905505

Ichinose, K., Ozawa, M., Itou, K., Kunieda, K., and Ebizuka, Y. (2003). Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: towards comparative analysis of the benzoisochromanequinone gene clusters. *Microbiology* 149 (7), 1633–1645. doi:10.1099/mic.0.26310-0

Kanda, N. (1971). A new antitumor antibiotic, kidamycin. I. Isolation, purification and properties of kidamycin. *J. Antibiot.* 24 (9), 599–606. doi:10.7164/antibiotics.24.599

Kharel, M. K., Pahari, P., Shepherd, M. D., Tibrewal, N., Nybo, S. E., Shaaban, K. A., et al. (2012). Angucyclines: biosynthesis, mode-of-action, new natural products, and synthesis. *Nat. Prod. Rep.* 29 (2), 264–325. doi:10.1039/c1np00068c

Kieser, T., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). *Practical streptomyces genetics*. Norwich: John Innes Foundation.

Kitamura, K., Ando, Y., Matsumoto, T., and Suzuki, K. (2014a). Synthesis of the pluramycins 1: two designed anthrones as enabling platforms for flexible bis-C-glycosylation. *Angew. Chem. Int. Ed.* 53 (5), 1258–1261. doi:10.1002/anie.201308016

Kitamura, K., Maezawa, Y., Ando, Y., Kusumi, T., Matsumoto, T., and Suzuki, K. (2014b). Synthesis of the pluramycins 2: total synthesis and structure assignment of saptomycin B. *Angew. Chem. Int. Ed.* 53 (5), 1262–1265. doi:10.1002/anie.201308017

Kren, V., and Martínková, L. (2001). Glycosides in medicine: "The role of glycosidic residue in biological activity". *Curr. Med. Chem.* 8 (11), 1303–1328. doi:10.2174/0929867013372193

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054

Luzhetskyy, A., Méndez, C., Salas, J. A., and Bechthold, A. (2008). Glycosyltransferases, important tools for drug design. *Curr. Top. Med. Chem.* 8 (8), 680–709. doi:10.2174/156802608784221514

Pandey, R. P., and Sohng, J. K. (2016). Glycosyltransferase-mediated exchange of rare microbial sugars with natural products. *Front. Microbiol.* 7, 1849. doi:10.3389/fmicb.2016.01849

Prakash, A. S., Moore, A. G., Murray, V., Matias, C., McFadyen, W. D., and Wickham, G. (1995). Comparison of the sequence selectivity of the DNA-alkylating pluramycin antitumour antibiotics DC92-B and hedamycin. *Chem. Biol. Interact.* 95 (1), 17–28. doi:10.1016/0009-2797(94)03341-2

Putkaradze, N., Teze, D., Fredslund, F., and Welner, D. H. (2021). Natural product C-glycosyltransferases – a scarcely characterised enzymatic activity with biotechnological potential. *Nat. Prod. Rep.* 38 (3), 432–443. doi:10.1039/D0NP00040J

Rixson, J. E., Abraham, J. R., Egoshi, Y., Skelton, B. W., Young, K., Gilbert, J., et al. (2015). The synthesis and biological activity of novel anthracenone-pyranones and anthracenone-furans. *Bioorg. Med. Chem.* 23 (13), 3552–3565. doi:10.1016/j.bmc.2015.04.032

Salas, J. A., and Méndez, C. (2007). Engineering the glycosylation of natural products in actinomycetes. *Trends Microbiol.* 15 (5), 219–232. doi:10.1016/j.tim.2007.03.004

Schumacher, R. W., Davidson, B. S., Montenegro, D. A., and Bernan, V. S. (1995). γ-Indomycinone, a new pluramycin metabolite from a deep-sea derived actinomycete. *J. Nat. Prod.* 58 (4), 613–617. doi:10.1021/np50118a024

Singh, S., Phillips, G. N., Jr., and Thorson, J. S. (2012). The structural biology of enzymes involved in natural product glycosylation. *Nat. Prod. Rep.* 29 (10), 1201–1237. doi:10.1039/c2np20039b

Tornus, D., and Floss, H. G. (2001). Identification of four genes from the granaticin biosynthetic gene cluster of Streptomyces violaceoruber Tü22 involved in the biosynthesis of L-rhodinose. *J. Antibiot.* 54 (1), 91–101. doi:10.7164/antibiotics.54.91

van Wageningen, A. A., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., et al. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.* 5 (3), 155–162. doi:10.1016/S1074-5521(98)90060-6

Weymouth-Wilson, A. C. (1997). The role of carbohydrates in biologically active natural products. *Nat. Prod. Rep.* 14 (2), 99–110. doi:10.1039/NP9971400099

Xie, K., Zhang, X., Sui, S., Ye, F., and Dai, J. (2020). Exploring and applying the substrate promiscuity of a C-glycosyltransferase in the chemo-enzymatic synthesis of bioactive C-glycosides. *Nat. Commun.* 11 (1), 5162. doi:10.1038/s41467-020-18990-9

Xu, L., Huang, H., Wei, W., Zhong, Y., Tang, B., Yuan, H., et al. (2014). Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*. *BMC Genomics* 15 (1), 363. doi:10.1186/1471-2164-15-363

Zhao, L., and Liu, H.-W. (2010). "Pathway and enzyme engineering and applications for glycodiversification," in *Enzyme technologies*. Editors W.-K. Yeh, H.-C. Yang, and J. R. McCarthy (John Wiley & Sons), 309–362.

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Mini Review
PUBLISHED 29 November 2022
DOI 10.3389/fbioe.2022.1057938

OPEN ACCESS

EDITED BY
Fengjie Cui,
Jiangsu University, China

REVIEWED BY
Guang Zhao,
Shandong University, China
Fu-Li Li,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS), China

*CORRESPONDENCE
Cheng Zhou,
zhoucheng@im.ac.cn

SPECIALTY SECTION
This article was submitted to Industrial Biotechnology,
a section of the journal
Frontiers in Bioengineering and Biotechnology

RECEIVED 30 September 2022
ACCEPTED 10 November 2022
PUBLISHED 29 November 2022

CITATION
Huang S, Xue Y, Ma Y and Zhou C (2022), Microbial (*E*)-4-hydroxy-3-methylbut-2-enyl pyrophosphate reductase (IspH) and its biotechnological potential: A mini review.
Front. Bioeng. Biotechnol. 10:1057938.
doi: 10.3389/fbioe.2022.1057938

Microbial (*E*)-4-hydroxy-3-methylbut-2-enyl pyrophosphate reductase (IspH) and its biotechnological potential: A mini review

Shiyong Huang[1,2], Yanfen Xue[1], Yanhe Ma[1] and Cheng Zhou[1]*

[1]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China

(*E*)-4-hydroxy-3-methylbut-2-enyl pyrophosphate (HMBPP) reductase (IspH) is a [4Fe-4S] cluster-containing enzyme, involved in isoprenoid biosynthesis as the final enzyme of the methylerythritol phosphate (MEP) pathway found in many bacteria and malaria parasites. In recent years, many studies have revealed that isoprenoid compounds are an alternative to petroleum-derived fuels. Thus, ecofriendly methods harnessing the methylerythritol phosphate pathway in microbes to synthesize isoprenoid compounds and IspH itself have received notable attention from researchers. In addition to its applications in the field of biosynthesis, IspH is considered to be an attractive drug target for infectious diseases such as malaria and *tuberculosis* due to its survivability in most pathogenic bacterium and its absence in humans. In this mini-review, we summarize previous reports that have systematically illuminated the fundamental and structural properties, substrate binding and catalysis, proposed catalytic mechanism, and novel catalytic activities of IspH. Potential bioengineering and biotechnological applications of IspH are also discussed.

KEYWORDS

IspH, MEP pathway, HMBPP reductase, structural properties, catalytic mechanism, biotechnological potential

Introduction

Isoprenoids, including steroids and terpenes, are one of the largest and most diverse classes of natural products. They include essential biological compounds such as vitamins, cholesterol, steroid hormones, carotenoids, and quinines (Rohmer, 1999; Eisenreich et al., 2004). In organisms they are derived from the same precursors: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Oldfield and Lin, 2012). Two distinct biosynthetic pathways are known to produce both IPP and DMAPP: the mevalonate (MVA) pathway, which is present in mammals as well as some microorganisms, and the methylerythritol phosphate (MEP) pathway, found in many pathogenic bacteria such as *Mycobacterium tuberculosi* and *Plasmodium falciparum* (Kuzuyama and Seto, 2003).

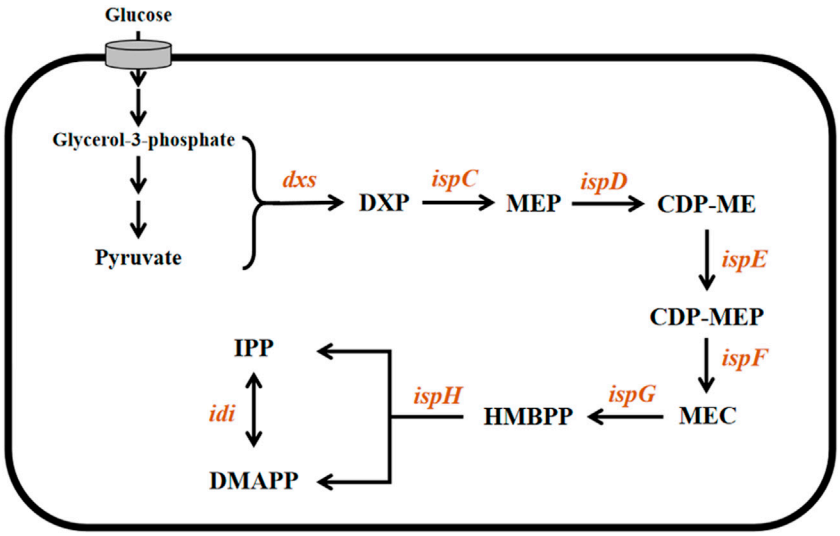


FIGURE 1
Microbial MEP pathway. The involved genes and corresponding enzymes are *dxs*: 1-deoxyxylose-5-phosphate synthase; *ispC*: 1-deoxyxylose-5-phosphate reductoisomerase; *ispD*: 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase; *ispE*: 4-(cytidine-5′-diphosphate)-2-c-methyl-d-erythritol kinase; *ispF*: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; *ispG*: 4-hydroxy-3-methyl-2-(*E*)-butenyl pyrophosphate synthase; and *ispH* (*E*)-4-hydroxy-3-methylbut-2-enyl pyrophosphate reductase. The involved intermediate metabolites are DXP: 1-Deoxy-D-xylulose-5-phosphate; MEP: 2C-methyl-D-erythritol-4-phosphate; CDP-ME: 4-(cytidine-5′-diphosphate)- 2-c-methyl-d-erythritol; CDP-MEP: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; MEC: 4-hydroxy-3-methyl-2-(E)-butenyl pyrophosphate; and HMBPP (*E*)-4-hydroxy-3-methylbut-2-enyl pyrophosphate.

The MEP pathway (Figure 1) begins with the condensation of pyruvate and glyceraldehyde 3-phosphate to form 1-deoxy-D-xylulose-5-phosphate (DXP), catalyzed by the enzyme 1-deoxyxylulose-5-phosphate synthase (DXS). DXP is then converted into 2-C-methyl-D-erythritol-4-phosphate by IspC and is also called 1-deoxyxylulose-5-phosphate reductoisomerase (DXR). A sequence of steps catalyzed by the enzymes IspD, IspE, and IspF converts 2-C-methyl-D-erythritol-4-phosphate into 2-C-methyl-D-erythritol-2,4-cyclodiphosphate *via* cytidine diphosphate intermediates (4-diphosphocytidyl-2-C-methyl-D-erythritol and 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate). IspG/GcpE catalyze the penultimate reaction in the pathway by reducing and opening the cyclic diphosphate intermediate to form (*E*)-4-hydroxy-3-methylbut-2-enyl pyrophosphate (HMBPP). The final step is the conversion of HMBPP into a mixture of IPP and DMAPP (Wang et al., 2012; Span et al., 2014). This reductive dehydroxylation (Figure 1) is catalyzed by IspH, an oxygen-sensitive monomeric protein with a [4Fe-4S] cluster at the active site. Because it plays a key role in the biosynthesis of isoprenoids and is essential for survival, IspH has attracted great interest, particularly with regard to the development of new antimicrobial drugs as well as novel biofuels as alternatives to petroleum-derived fuels.

In this mini-review, we summarize the fundamental features of IspH, its structural properties, substrate binding and catalysis, proposed catalytic mechanisms, and novel catalytic activities. Potential bioengineering and biotechnological applications of IspH are also discussed.

Fundamental features of IspH

As an enzyme in the MEP pathway, IspH occurs in most bacteria, plant chloroplasts, green algae, and apicomplexan, but it is absent in humans (Rohmer et al., 2004). The *lytB* gene, encoding LytB (EC.1.17.7.4), also called IspH, was first described in *Escherichia coli* as a gene involved in penicillin resistance (Gustafson et al., 1993) and was later reported to be present in other bacteria (Potter et al., 1998). Cunningham *et al.* discovered that the deletion of *lytB* from *Synechocystis* PCC6803 was fatal, but the strain was able to recover when supplied with an analog of either IPP or DMAPP (3-methyl-3-butene-1-alcohol or 3-methyl-2-butene-1-alcohol, respectively) in the culture medium. Moreover, *lytB* can increase the biosynthesis of carotenoids when expressed in *E. coli* (Cunningham et al., 2000). These findings confirm that *lytB* is involved in the MEP pathway. Since these studies, a series of *in vitro* experiments have shown that IspH is the last enzyme in the MEP pathway, responsible for the conversion of HMBPP to IPP and DMAPP (Altincicek et al., 2002; Petra et al., 2002).

Because IspH contains an oxygen-sensitive [4Fe-4S] cluster at its active site, it is easily oxidized and inactivated when exposed

to air. Thus, purification of this protein is usually carried out in an anaerobic chamber. Gräwert *et al.* reported that a purified (in anaerobic conditions) solution of IspH was iron-green in color. When exposed to air for 1 h, the protein gradually deactivated and turned light brown (Gräwert et al., 2010). Decomposition of the oxygen-sensitive [4Fe-4S] cluster in IspH leads to partial loss of the tertiary structure and to complete loss of function, as also observed for aconitase (Kent et al., 1982) and radical *S*-adenosyl-L-methionine (SAM) enzymes (Tamarit et al., 2000; Layer et al., 2004).

Numerous reports on the biochemical properties and enzyme kinetics of IspH also exist. Altincicek *et al.* determined that the optimum pH of *Aquifex aeolicus* IspH (AaIspH) was in the range 7.0–7.5. Activity was observed at 30°C–70°C, with maximum activity occurring at 60°C. The K_m value of AaIspH for HMBPP was 590 ± 60 μM (Altincicek et al., 2002).

Structure, substrate binding, and catalysis of IspH

Even though IspHs from different microorganisms exhibit a low sequence homology, for example, the sequence similarity of *E. coli* and *A. aeolicus* IspHs is only 43%, high structural similarity between different IspHs is observed. X-ray crystallographic structures of IspHs are available for the protein in *E. coli* (EcIspH, PDB: 3F7T) (Gräwert et al., 2010), *A. aeolicus* (AaIspH, PDB: 3DNF) (Rekittke et al., 2008), and *P. falciparum* (PfIspH, PDB: 4N7B) (Rekittke et al., 2013). These structures present a similar "trefoil" arrangement consisting of three α/β domains with the Fe-S cluster bound at the center of the structure (Figures 2A, B). The [4Fe-4S] cluster is usually present in the crystallographic structure of IspH. However, the crystal structure of IspH in *A. aeolicus* exhibited a [3Fe-4S] center; it is believed that the [3Fe-4S] center observed in the crystal structure lost an iron atom during co-crystallization with HMBPP (Rekittke et al., 2008). In functional IspHs, the oxidized cluster is in the form $[4Fe\text{-}4S]^{2+}$, which is bound to three highly conserved cysteine residues. The fourth (unique) Fe atom is involved in ligand binding and electron transport during catalysis.

The catalytic reaction of IspH is a redox reaction. The ligand can be an electron donator, such as a flavodoxin/flavodoxin reductase/NADH system (*E. coli*) (Wolff et al., 2002; Puan et al., 2005), or ferredoxin (in *P. falciparum*) (Röhrich et al., 2005), and is essential during catalysis. Chemical reductants, such as artificial electron donors including photoactivated deazaflavin (Xiao et al., 2008) and methyl dithionite (Altincicek et al., 2002), can usually be used in *in vitro* experiments (Gräwert et al., 2009).

Co-crystallization of an enzyme with a substrate is the common method to explore the substrate binding and to identify active regions of the enzyme. Although most studies have merely obtained the ligand-free IspH structure which is challenging to co-crystallize with a substrate, the crystal structure of the EcIspH-ligand complex has been obtained and analyzed. In the complex structure, the HMBPP substrate interacted with multiple amino acid residues *via* hydrogen bonding. The substrate-binding region and the key binding sites in IspH were determined (Figures 2C, D) (Rekittke et al., 2013; Xu et al., 2016). A number of amino acids residues were very highly conserved in different IspHs (Figure 2E), and these residues are mainly involved in catalysis or maintaining enzyme activity. Three cysteine residues (Cys12, Cys96, and Cys197 in *E. coli* IspH) bind to the iron-sulfur center; mutation of any one of these residues causes a complete loss of enzymatic activity (Petra et al., 2002; Gräwert et al., 2010). In *A. aeolicus* IspH, three histidine residues, His42, His74, and His124, are also involved in substrate binding (Gräwert et al., 2009). Another conserved amino acid, Glu126 in *E. coli* IspH, has been shown to play a crucial role as a proton donor during catalysis (Span et al., 2012a). The residue His41 in *E. coli* IspH (His42 in *A. aeolicus* IspH) is predicted to be involved in the binding of HMBPP, is essential for catalysis, and may be involved in delivering H^+ from Glu126 to the bound HMBPP. The residue Thr167 may act as a proton relay and Glu126 serves as the ultimate proton donor (Gräwert et al., 2009).

The mechanism of enzyme catalysis for IspH is controversial and there have been several proposals related to this (Altincicek et al., 2002; Wolff et al., 2002; Rohdich et al., 2003; Xiao et al., 2008; Xiao and Liu, 2008; Gräwert et al., 2009). Three main hypotheses for the catalytic mechanism have been proposed. In the Birch Reduction Theory (Figure 3A), the process of reduction and dehydroxylation of HMBPP is similar to a Birch reduction reaction. When the HMBPP substrate binds to the protein, an electron is first transferred from the [4Fe-4S] center to HMBPP, which then protonates and loses a H_2O molecule. After the hydroxyl is lost, HMBPP forms an allyl carbon free radical; the intermediate then accepts a second electron and is protonated to form IPP and DMAPP (Rohdich et al., 2003). The Birch Reduction Theory is regarded as the most likely mechanism occurring for IspH-based catalysis.

The second proposal is the Bioorganometallic Theory (Figure 3B), which was proposed based on the biophysical properties of the complex formed between the HMBPP substrate and the IspH Glu126Ala mutant (Wang et al., 2010). In this mechanism, the substrate hydroxyl group first binds to the IspH [4Fe-4S] cluster and then receives an electron to form an HMBPP-[4Fe-4S] complex, similar to a π or 2-alkenyl/metallacycle complex, which is then dehydrated to form a 1-allyl intermediate. This receives an electron and is reduced to form a 3-allyl intermediate complex, which is finally protonated to form IPP and DMAPP.

The third theory is the allyl cation model (Figure 3C). In this proposal, the metal center, as a Lewis acid, is more conducive to fracture of the C4-OH bond and the subsequent formation of an

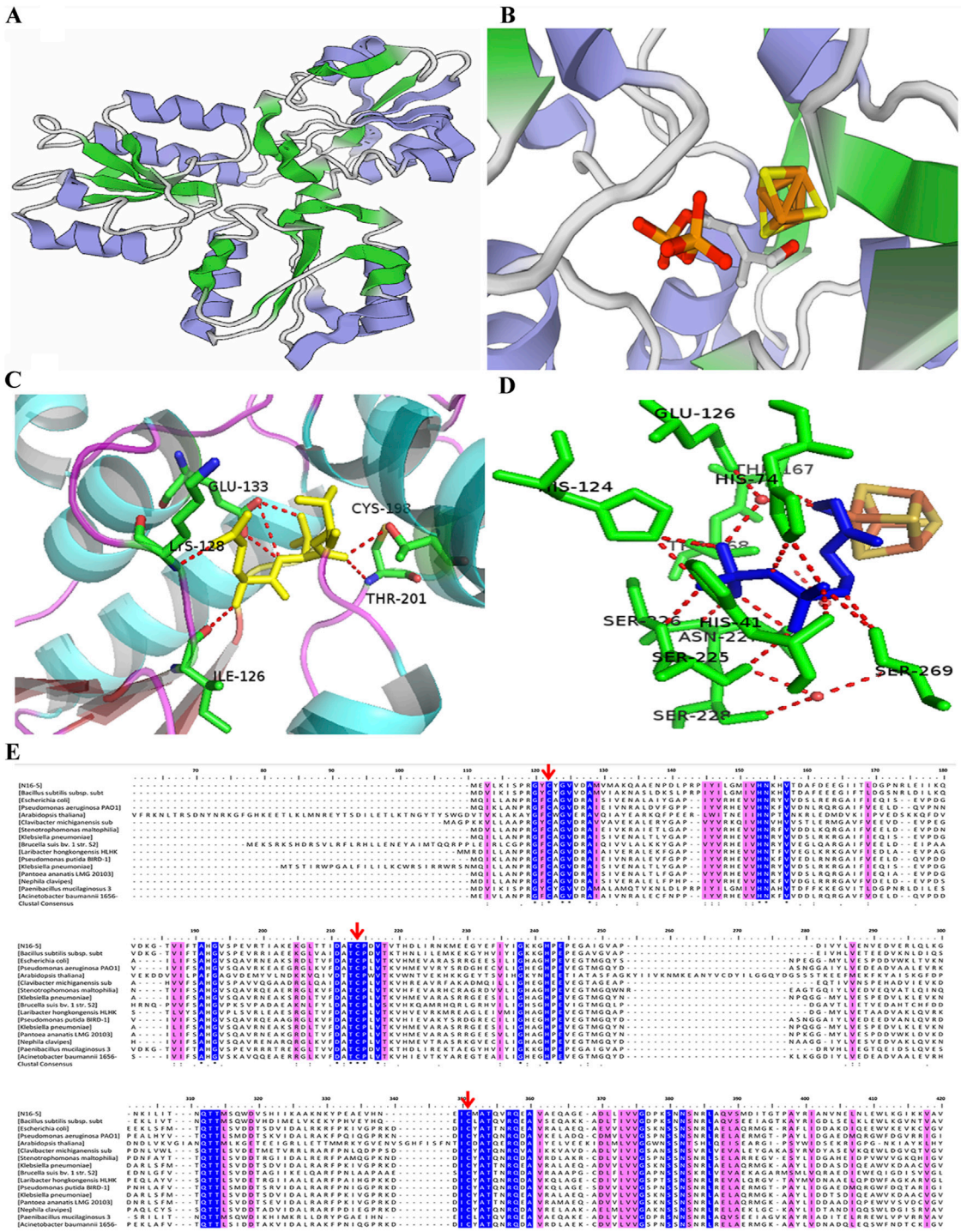


FIGURE 2
Structure, active sites, and multiple sequence alignment of IspHs from different species. **(A)** Three-dimensional structure model of monomeric IspH. **(B)** Crystal structure of the IspH-HMBPP complex; the [4Fe-4S] cluster center is shown as a ball-and-stick model with iron atoms colored in orange and sulfur atoms in gold. **(C)** Binding of HMBPP at the active site in IspH. HMBPP is shown as a ball-and-stick model in yellow. **(D)** Binding of HMBPP shown as a ball-and-stick model in blue and the hydrogen-bonding network shown by the dotted red lines. **(E)** Multiple sequence alignment of IspHs in different species. Cysteines coordinating the iron sulfur cluster are indicated by the red arrows. Stringently conserved residues are highlighted in blue.

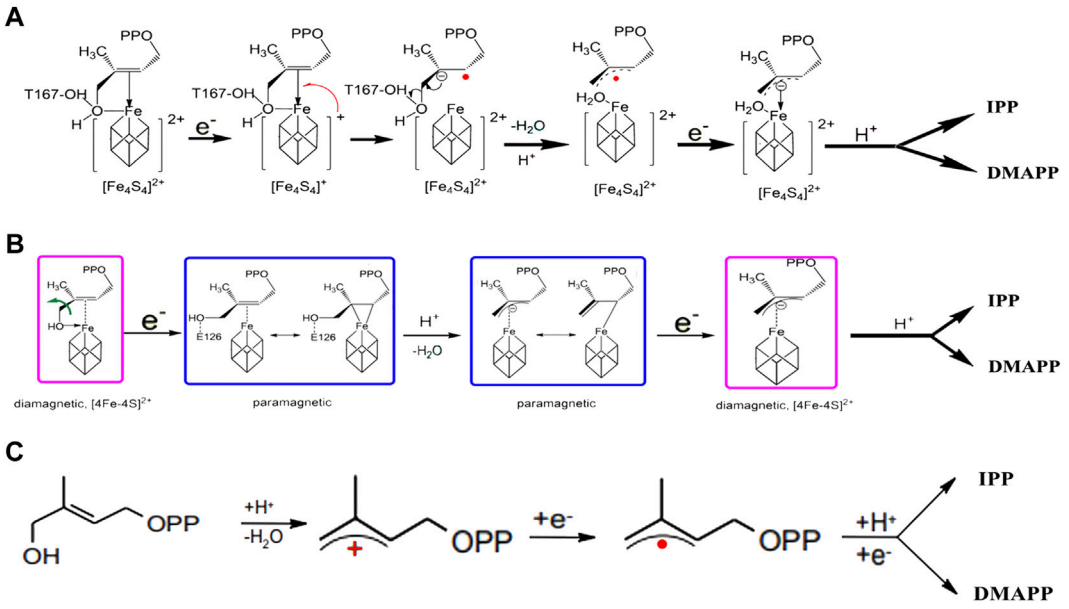


FIGURE 3
Catalytic reaction mechanism in IspH. **(A)** According to Birch reduction theory; **(B)** according to the Bioorganometallic mechanism; **(C)** according to allyl cation model theory.

allyl cation intermediate product, which receives two electrons and is protonated to eventually form the products IPP and DMAPP (Altincicek et al., 2002).

The common point in these three theories is that the initial reaction requires the C4-OH in the HMBPP substrate to be combined with the [4Fe-4S] center in the enzyme, as well as formation of the final reaction product during a prolongation step. It is believed that the proportions of IPP and DMAPP are controlled by the position of the prolongated carbon. Clearly, more studies are required to further delineate the catalytic mechanism of IspH. Additional experiments are also needed to determine how the reaction flux (IPP *vs*. DMAPP) varies in the IspH-catalyzed reaction.

Catalytic promiscuity of IspH

IspH is the final enzyme in the MEP pathway, and its natural catalytic products IPP and DMAPP are crucial precursors for downstream terpenoids compounds. As the simplest terpenoid, isoprene is an important platform compound that is widely used in the production of rubber. In 1957, Sanadze *et al.* discovered isoprene emission from woody plants (Sharkey and Yeh, 2001). Since then, it has also been found that isoprene is produced by other organisms, including animals, fungi, and bacteria (Kuzma et al., 1995; Sharkey and Yeh, 2001). Isoprene can be synthesized *via* the MEP pathway with DMAPP as the precursor, and this reaction is catalyzed by an isoprene synthase (IspS) that is only found in higher plants (Monson et al., 2013). However, in the study of *Bacillus*, IspH was found to have the same activity for catalyzing the formation of isoprene. Hess *et al.* used transcriptomic analysis to predict the possible genes involved in isoprene production in the MEP pathway of *B. subtilis*, and showed that the overexpression of IspH was negatively correlated with isoprene production (Hess et al., 2013). Julsing *et al.* analyzed all genes in the MEP pathway by means of a conditional knockout strategy, aiming to explore the influence of different genes on the isoprenoids of *B. subtilis* (Julsing et al., 2007). The results showed that the yield of isoprene from the mutant of which the IspH-coding gene *yqfP* was deleted was 9.8 times lower than that in the wild-type strain. Currently, there is no direct evidence that *B. subtilis* IspH is responsible for producing isoprene; however, Ge *et al.* confirmed that IspH derived from an alkaliphilic *Bacillus* sp. N16-5 exhibits isoprene synthase activity (it can catalyze the production of isoprene from endogenous HMBPP and isoamylene from DMAPP) (Ge et al., 2016). To date, IspS has not been found in microorganisms (Köksal et al., 2010). The discovery of an isoprene synthase in microorganisms will undoubtedly be important.

Some studies have reported other novel activities of IspH. IspH from *E. coli* could catalyze the conversion of acetylene to acetaldehyde and ketone through hydration (Span et al., 2012b). Other studies confirmed that IspH can use a number of HMBPP analogs as substrates (Xiao et al., 2011; Wang et al., 2012). Obviously, microbial IspH presents catalytic promiscuity;

however, these novel catalytic activities are poorly understood. If the mechanism of the multi-substrate, catalytic function of IspH can be further explored in future studies, it will provide a theoretical basis for engineering IspH, providing new enzyme resources for terpenoid biosynthesis.

Biotechnological potential

Isoprenoids are a large family, many of which are important industrial compounds with high added values. The simplest is isoprene, which is used as a raw material for the production of rubber. In addition, complicated-structure terpenoids such as limonene, linalool, vitamin E, vitamin K, and β-carotene are antioxidants. Many studies have already reported that isoprenoid compounds can be synthesized in microorganisms *via* either the MEP or MVA pathways (Rohmer, 1999; Kuzuyama, 2002). Using synthetic biology approaches, many studies have achieved high-level production of many isoprenoid compounds in engineered microbes, including in *E. coli*, *Saccharomyces cerevisiae*, and *B. subtilis*. It was proposed that the MEP pathway may present several potential control points, each exhibiting different degrees of control (Li et al., 2018). IspH is a crucial rate-limiting step for amplifying the isoprenoid flux. To increase β-carotene production in *E. coli,* a strong promoter was employed to replace the native promoter IspH in order to increase the isoprenoid flux (Yuan et al., 2006; Suh, 2012). *lytB* (encoding IspH) from *Thermosynechococcus elongatus* was introduced into *E. coli*, resulting in efficient isoprene production (Chotani et al., 2013).

Development of novel anti-infectives

In addition to serving as a target in MEP pathways to improve the production of terpenes *via* metabolic engineering, IspH can also be used as a target in protein inhibitors for the development of antibacterial drugs. The MEP pathway is absent in mammals but is essential for the survival of many pathogenic bacteria; thus, it provides a new route for the development of novel antibacterial drugs (Rohdich et al., 2005). Most studies have focused on the development of compounds that inhibit the activity of IspH. Reported IspH inhibitors include substrate analogs, pyridine diphosphates, alkyne derivatives, and non-diphosphate compounds.

Two HMBPP analogs were designed, wherein a thiol or amino group replaced the hydroxyl group in HMBPP. A complete kinetic investigation in anaerobic conditions revealed that these analogs were extremely potent inhibitors of *E. coli* IspH, displaying competitive modes of inhibition (Ahrens-Botzong et al., 2011; Janthawornpong et al., 2013). Wang *et al.* studied a set of pyridine derivatives substituted in the ortho, meta, and para positions as potential inhibitors of IspH in *A. aeolicus*, and found that two of them exhibited superior inhibitory potencies (Wang et al., 2010). In addition, Wang *et al.* also demonstrated that alkyne diphosphate can function as an inhibitor of *A. aeolicus* IspH (Wang et al., 2010). O'Dowd *et al.* used an *in silico* approach to screen a series of compounds from ZINC and NCI libraries; the authors identified two drug-like compounds that acted as IspH inhibitors (O'Dowd et al., 2017). However, although a number of potent IspH inhibitors have been discovered, to date, no reports exist on research using these inhibitors against pathogenic bacteria.

Conclusions and perspectives

The microbial MEP pathway has long been known; however, few studies regarding individual enzymes in the pathway have been reported, particularly the last enzyme in the MEP pathway, IspH. Because IspH contains an iron-sulfur cluster that degrades in air, the technical conditions required for the expression and purification of the protein are relatively strict. However, considering its importance in both the MEP pathway and isoprenoid compound biosynthesis, further research on this protein is required, and should include: 1) Studies for improving the enzyme's properties, such as stability, catalytic activity, substrate binding, product specificity, and cofactor-binding affinity to further enhance its application potential. 2) Solving the mismatch in the reduction potential issue and the difficulties faced in [4Fe-4S] cluster reduction in IspH. 3) Research on IspH inhibitors has been limited to *in vitro* experiments, therefore future research should focus on *in vivo* experiments to provide a theoretical basis for the development of new antibacterial drugs targeting IspH. A systematic and in-depth study of the biochemical properties and structure–function relationships in IspH will provide valuable information for engineering and using IspH in future biotechnological applications.

Author contributions

SH prepared the original manuscript. YX, YM and CZ revised the manuscript. CZ supervised the manuscript. All authors approved this version of the article.

Funding

This research was financially supported by a grant from National Natural Science Foundation of China (Grant no. 31870789).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahrens-Botzong, A., Janthawornpong, K., Wolny, J. A., Tambou, E. N., Rohmer, M., Krasutsky, S., et al. (2011). Biosynthesis of isoprene units: Mössbauer spectroscopy of substrate and inhibitor binding to the [4Fe-4S] cluster of the LytB/IspH enzyme. *Angew. Chem. Int. Ed.* 50. 11976–11979. doi:10.1002/anie.201104562

Altincicek, B., Duin, E. C., Reichenberg, A., Hedderich, R., Kollas, A.-K., Hintz, M., et al. (2002). LytB protein catalyzes the terminal step of the 2-C-methyl-D-erythritol-4-phosphate pathway of isoprenoid biosynthesis. *FEBS Lett.* 532, 437–440. doi:10.1016/s0014-5793(02)03726-2

Chotani, G. K., McAuliffe, J. C., Miller, M. C., Vaviline, D. V., and Weyler, W. (2013). *Isoprene production using the DXP and MVA pathway. US8507235B.*

Cunningham, F. X., Lafond, T. P., and Gantt, E. (2000). Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. *J. Bacteriol.* 182, 5841–5848. doi:10.1128/JB.182.20.5841-5848.2000

Eisenreich, W., Bacher, A., Arigoni, D., and Rohdich, F. (2004). Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cell. Mol. Life Sci.* 61, 1401–1426. doi:10.1007/s00018-004-3381-z

Ge, D., Xue, Y., and Ma, Y. (2016). Two unexpected promiscuous activities of the iron–sulfur protein IspH in production of isoprene and isoamylene. *Microb. Cell Fact.* 15, 79. doi:10.1186/s12934-016-0476-9

Gräwert, T., Ingrid, S., Wolfgang, E., Felix, R., Eppinger, J., Adelbert, B., et al. (2010). Probing the reaction mechanism of IspH protein by x-ray structure analysis. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1077–1081. doi:10.1073/pnas.0913045107

Gräwert, T., Rohdich, F., Span, I., Bacher, A., Eisenreich, W., Eppinger, J., et al. (2009). Structure of active IspH enzyme from *Escherichia coli* provides mechanistic insights into substrate reduction. *Angew. Chem. Int. Ed.* 48 (31), 5756–5759. doi:10.1002/anie.200900548

Gustafson, C. E., Kaul, S., and Ishiguro, E. E. (1993). Identification of the *Escherichia coli lytB* gene, which is involved in penicillin tolerance and control of the stringent response. *J. Bacteriol.* 175, 1203–1205. doi:10.1128/jb.175.4.1203-1205.1993

Hess, B. M., Xue, J., Markillie, L. M., Taylor, R. C., Wiley, H. S., Ahring, B. K., et al. (2013). Coregulation of terpenoid pathway genes and prediction of isoprene production in *Bacillus subtilis* using transcriptomics. *PLoS One* 8, e66104. doi:10.1371/journal.pone.0066104

Janthawornpong, K., Krasutsky, S., Chaignon, P., Rohmer, M., Poulter, C. D., and Seemann, M. (2013). Inhibition of IspH, a $[4Fe–4S]^{2+}$ enzyme involved in the biosynthesis of isoprenoids via the methylerythritol phosphate pathway. *J. Am. Chem. Soc.* 135, 1816–1822. doi:10.1021/ja309557s

Julsing, M. K., Rijpkema, M., Woerdenbag, H. J., Quax, W. J., and Kayser, O. (2007). Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 75 (6), 1377–1384. doi:10.1007/s00253-007-0953-5

Kent, T. A., Dreyer, J. L., Kennedy, M. C., Huynh, H. B., Emptage, M. H., Beinert, H., et al. (1982). Mossbauer studies of beef heart aconitase: Evidence for facile interconversions of iron-sulfur clusters. *Proc. Natl. Acad. Sci. U. S. A.* 79 (4), 1096–1100. doi:10.1073/pnas.79.4.1096

Köksal, M., Zimmer, I., Schnitzler, J. P., and Christianson, W. D. (2010). Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. *J. Mol. Biol.* 402 (2), 363–373. doi:10.1016/j.jmb.2010.07.009

Kuzma, J., Nemecek-Marshall, M., Pollock, W. H., and Fall, R. (1995). Bacteria produce the volatile hydrocarbon isoprene. *Curr. Microbiol.* 30 (2), 97–103. doi:10.1007/BF00294190

Kuzuyama, T. (2002). Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci. Biotechnol. Biochem.* 66 (8), 1619–1627. doi:10.1271/bbb.66.1619

Kuzuyama, T., and Seto, H. (2003). Diversity of the biosynthesis of the isoprene units. *Nat. Prod. Rep.* 20 (2), 171–183. doi:10.1039/b109860h

Layer, G., Wheinz, D., Jahn, D., and Schubert, W. D. (2004). Structure and function of radical SAM enzymes. *Curr. Opin. Chem. Biol.* 8 (5), 468–476. doi:10.1016/j.cbpa.2004.08.001

Li, M., Nian, R., Xian, M., and Zhang, H. (2018). Metabolic engineering for the production of isoprene and isopentenol by *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 102, 7725–7738. doi:10.1007/s00253-018-9200-5

Monson, R. K., Jones, R. T., Rosenstiel, T., and Schnitzler, J. P. (2013). Why only some plants emit isoprene. *Plant Cell Environ.* 36 (3), 503–516. doi:10.1111/pce.12015

O'Dowd, B., Williams, S., Wang, H., No, J. H., Rao, G., Wang, W., et al. (2017). Spectroscopic and computational investigations of ligand binding to IspH: Discovery of non-diphosphate inhibitors. *Chembiochem* 18 (10), 914–920. doi:10.1002/cbic.201700052

Oldfield, E., and Lin, F. Y. (2012). Terpene biosynthesis: Modularity rules. *Angew. Chem. Int. Ed.* 51 (5), 1124–1137. doi:10.1002/anie.201103110

Petra, A., Stefan, H., Wolfgang, E., Johannes, K., Tobias, G., Duilio, A., et al. (2002). Biosynthesis of terpenes: Studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase. *Proc. Natl. Acad. Sci. U. S. A.* 99 (9), 12108–12113. doi:10.1073/pnas.182412599

Potter, S., Yang, X., Boulanger, M. J., and Ishiguro, E. E. (1998). Occurrence of homologs of the *Escherichia coli lytB* gene in gram-negative bacterial species. *J. Bacteriol.* 180, 1959–1961. doi:10.1128/jb.180.7.1959-1961.1998

Puan, K. J., Wang, H., Dairi, T., Kuzuyama, T., and Morit, C. T. (2005). *fldA* is an essential gene required in the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *FEBS Lett.* 579 (17), 3802–3806. doi:10.1016/j.febslet.2005.05.047

Rekittke, I., Olkhova, E., Wiesner, J., Demmer, U., Warkentin, E., Jomaa, H., et al. (2013). Structure of the (E)-4-hydroxy-3-methyl-but-2-enyl-diphosphate reductase from *Plasmodium falciparum*. *FEBS Lett.* 587 (24), 3968–3972. doi:10.1016/j.febslet.2013.10.029

Rekittke, I., Wiesner, J., Röhrich, R., Demmer, U., Warkentin, E., Xu, W., et al. (2008). Structure Of (e)-4-hydroxy-3-methyl-but-2-enyl Diphosphate Reductase, the Terminal Enzyme Of The Non-mevalonate Pathway. *J. Am. Chem. Soc.* 130 (51), 17206–17207. doi:10.1021/ja806668q

Rohdich, F., Bacher, A., and Eisenreich, W. (2005). Isoprenoid biosynthetic pathways as anti-infective drug targets. *Biochem. Soc. Trans.* 33 (4), 785–791. doi:10.1042/BST0330785

Rohdich, F., Zepeck, F., Adam, P., Hecht, S., Kaiser, J., Laupitz, R., et al. (2003). The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: Studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proc. Natl. Acad. Sci. U. S. A.* 100 (4), 1586–1591. doi:10.1073/pnas.0337742100

Rohmer, M., Grosdemange-Billiard, C., Seemann, M., and Tritsch, D. (2004). Isoprenoid biosynthesis as a novel target for antibacterial and antiparasitic drugs. *Curr. Opin. Investig. Drugs* 5, 154–162.

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16 (5), 565–574. doi:10.1039/a709175c

Röhrich, R. C., Englert, N., Troschke, K., Reichenberg, A., Hintz, M., Seeber, F., et al. (2005). Reconstitution of an apicoplast-localised electron transfer pathway involved in the isoprenoid biosynthesis of*Plasmodium falciparum*. *FEBS Lett.* 579, 6433–6438. doi:10.1016/j.febslet.2005.10.037

Sharkey, T. D., and Yeh, S. (2001). Isoprene emission from plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 407–436. doi:10.1146/annurev.arplant.52.1.407

Span, I., Gräwert, T., Bacher, A., Eisenreich, W., and Groll, M. (2012a). Crystal structures of mutant IspH proteins reveal a rotation of the substrate's hydroxymethyl group during catalysis. *J. Mol. Biol.* 416 (1), 1–9. doi:10.1016/j.jmb.2011.11.+033

Span, I., Wang, K., Eisenreich, W., Bacher, A., Zhang, Y., Oldfield, E., et al. (2014). Insights into the binding of pyridines to the iron--sulfur enzyme IspH. *J. Am. Chem. Soc.* 136 (22), 7926–7932. doi:10.1021/ja501127j

Span, I., Wang, K., Wang, W., Zhang, Y., Bacher, A., Eisenreich, W., et al. (2012b). Discovery of acetylene hydratase activity of the iron–sulphur protein IspH. *Nat. Commun.* 3, 1042. doi:10.1038/ncomms2052

Suh, W. (2012). High isoprenoid flux *Escherichia coli* as a host for carotenoids production. *Methods Mol. Biol.* 834, 49–62. doi:10.1007/978-1-61779-483-4_4

Tamarit, J., Gerez, C., Meier, C., Mulliez, E., Trautwein, A., and Fontecave, M. (2000). The activating component of the anaerobic ribonucleotide reductase from *Escherichia coli*. *J. Biol. Chem.* 275 (21), 15669–15675. doi:10.1074/jbc.275.21.15669

Wang, K., Wang, W., No, J. H., Zhang, Y., Zhang, Y., and Oldfield, E. (2010b). Inhibition of the Fe(4)S(4)-cluster-containing protein IspH (LytB): Electron paramagnetic resonance, metallacycles, and mechanisms. *J. Am. Chem. Soc.* 132 (19), 6719–6727. doi:10.1021/ja909664j

Wang, W., Wang, K., Liu, Y. L., No, J. H., Li, J., Nilges, M. J., et al. (2010a). Bioorganometallic mechanism of action, and inhibition, of IspH. *Proc. Natl. Acad. Sci. U. S. A.* 107 (10), 4522–4527. doi:10.1073/pnas.0911087107

Wang, W., Wang, K., Span, I., Jauch, J., Bacher, A., Groll, M., et al. (2012). Are free radicals involved in IspH catalysis? An EPR and crystallographic investigation. *J. Am. Chem. Soc.* 134 (27), 11225–11234. doi:10.1021/ja303445z

Wolff, M., Seemann, M., Grosdemange-Billiard, C., Tritsch, D., Campos, N., Rodrı´guez-Concepción, M., et al. (2002). Isoprenoid biosynthesis via the methylerythritol phosphate pathway. (E)-4-Hydroxy-3-methylbut-2-enyl diphosphate: Chemical synthesis and formation from methylerythritol cyclodiphosphate by a cell-free system from *Escherichia coli*. *Tetrahedron Lett.* 43 (14), 2555–2559. doi:10.1016/S0040-4039(02)00293-9

Xiao, Y., Chang, W. C., Liu, H. W., and Liu, P. (2011). Study of IspH, a key enzyme in the methylerythritol phosphate pathway using fluoro-substituted substrate analogues. *Org. Lett.* 13 (21), 5912–5915. doi:10.1021/ol202559r

Xiao, Y., and Liu, P. (2008). IspH protein of the deoxyxylulose phosphate pathway: Mechanistic studies with C1-deuterium-labeled substrate and fluorinated analogue. *Angew. Chem. Int. Ed.* 47 (50), 9722–9725. doi:10.1002/anie.200803452

Xiao, Y., Zhao, Z. K., and Liu, P. (2008). Mechanistic studies of IspH in the deoxyxylulose phosphate pathway: Heterolytic C-O bond cleavage at C4 position. *J. Am. Chem. Soc.* 130 (7), 2164–2165. doi:10.1021/ja710245d

Xu, W., Lees, N. S., Hall, D., Welideniya, D. T., Hoffman, B. M., and Duin, E. C. (2016). A closer look at the spectroscopic properties of possible reaction intermediates in wild-type and mutant (*E*)-4-Hydroxy-3-methylbut-2-enyl diphosphate reductase. *Biochemistry* 51 (24), 4835–4849. doi:10.1021/bi3001215

Yuan, L. Z., Rouviere, P. E., LaRossa, R. A., and Suh, W. (2006). Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* 8 (1), 79–90. doi:10.1016/j.ymben.2005.08.005

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Review
PUBLISHED 28 April 2023
DOI 10.3389/fbioe.2023.1154779

OPEN ACCESS

EDITED BY
M. Kalim Akhtar,
United Arab Emirates University, United Arab Emirates

REVIEWED BY
Pau Loke Show,
University of Nottingham Malaysia Campus, Malaysia
Rufeng Wang,
Shanghai University of Traditional Chinese Medicine, China

*CORRESPONDENCE
Fernando Perez Rojo,
fernando.perezrojo@research.uwa.edu.au
Parwinder Kaur,
parwinder.kaur@uwa.edu.au

RECEIVED 31 January 2023
ACCEPTED 18 April 2023
PUBLISHED 28 April 2023

CITATION
Perez Rojo F, Pillow JJ and Kaur P (2023), Bioprospecting microbes and enzymes for the production of pterocarpans and coumestans.
Front. Bioeng. Biotechnol. 11:1154779.
doi: 10.3389/fbioe.2023.1154779

Bioprospecting microbes and enzymes for the production of pterocarpans and coumestans

Fernando Perez Rojo[1]*, J. Jane Pillow[2] and Parwinder Kaur[1]*

[1]UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, [2]UWA School of Human Sciences, The University of Western Australia, Perth, WA, Australia

The isoflavonoid derivatives, pterocarpans and coumestans, are explored for multiple clinical applications as osteo-regenerative, neuroprotective and anti-cancer agents. The use of plant-based systems to produce isoflavonoid derivatives is limited due to cost, scalability, and sustainability constraints. Microbial cell factories overcome these limitations in which model organisms such as *Saccharomyces cerevisiae* offer an efficient platform to produce isoflavonoids. Bioprospecting microbes and enzymes can provide an array of tools to enhance the production of these molecules. Other microbes that naturally produce isoflavonoids present a novel alternative as production chassis and as a source of novel enzymes. Enzyme bioprospecting allows the complete identification of the pterocarpans and coumestans biosynthetic pathway, and the selection of the best enzymes based on activity and docking parameters. These enzymes consolidate an improved biosynthetic pathway for microbial-based production systems. In this review, we report the state-of-the-art for the production of key pterocarpans and coumestans, describing the enzymes already identified and the current gaps. We report available databases and tools for microbial bioprospecting to select the best production chassis. We propose the use of a holistic and multidisciplinary bioprospecting approach as the first step to identify the biosynthetic gaps, select the best microbial chassis, and increase productivity. We propose the use of microalgal species as microbial cell factories to produce pterocarpans and coumestans. The application of bioprospecting tools provides an exciting field to produce plant compounds such as isoflavonoid derivatives, efficiently and sustainably.

KEYWORDS

microbial bioprospecting, enzyme bioprospecting, microbial-based production, isoflavonoids, pterocarpans and coumestans

1 Introduction

Bioprospecting enables researchers to explore biodiversity, and identify novel molecules, enzymes and microbes relevant to research and industrial applications (Beattie et al., 2011; AfifaHussain et al., 2022). Bioprospecting strategies can be sub-divided into microorganisms (metagenomics), enzymes (transcriptomics/proteomics), or individual molecules (metabolomics). Microbial bioprospecting explores vastly diverse microorganisms, which constitute more than two-third of global life forms, present in a diverse range of environments including extreme ones (Becker and Wittmann, 2020). Metagenomics tools provide information from even non-culturable microbes, being a rich source of novel metabolic pathways, enzymes and their related catalyzed products. The combination of omics-based technologies allows researchers to link genes with

enzymatic pathways and metabolite biosynthesis, fuelling the discovery of novel molecules for diverse applications (Bansal et al., 2022). To understand how these molecules are produced, the whole enzymatic pathway needs to be identified, including the genes involved in the synthesis and its regulatory framework. *In silico* enzyme bioprospecting allows the identification of novel homologous sequences, and the possibility to model and trial enzymatic activities. After this *in silico* approach, heterologous biosynthetic pathways can be incorporated into model microorganisms using a combination of the most efficient enzymes through synthetic biology tools. Optimized microbial chassis are proven to increase yield and reduce the production's environmental impact towards a circular bioeconomy model (Krüger et al., 2020). Bioprospecting strategies for enzymes and chassis offer an unprecedented number of novel biocatalysts for a more efficient and sustainable production.

The use of microbes such as *Saccharomyces cerevisiae* and *Escherichia coli* is one possible strategy to overcome the limitations of plant-based production systems. As the isoflavonoids derivatives pterocarpans and coumestans are mainly produced in plant species, microbes provide an interesting alternative for boosting production titer, sustainably. Additionally, bioprospecting microbes that naturally produce isoflavonoids allow the diversification of production chassis, moving away from model organisms and allowing the implementation of even more sustainable solutions.

Primary metabolites such as amino acids, carbohydrates and lipids are essential for the growth and development of living organism. Secondary metabolites are typically modified chemical derivatives of these metabolites; such modifications include methylation, glycosylation, and hydroxylation (Twaij and Hasan, 2022). Though not essential, secondary metabolites have significant influence in the survivability of organisms and, in the case of plants, are quite often vital for plant-biome interactions.

The secondary metabolites, pterocarpans and coumestans, have clinical potential in both plants and humans. These isoflavonoid derivatives are associated with antimicrobial and antifungal properties (considered a phytoalexin), and its action is explored to deal with plant diseases such as fungal infections (Gupta et al., 2022). Furthermore, they have shown some promise for the treatment diseases of Alzheimer's disease, osteoporosis, and cancer (Dixit et al., 2015; Li et al., 2021; Tu et al., 2021). Thus, isoflavonoid derivatives merit further exploration with respect to not only its clinical effects but also its method of production.

This review summarizes research efforts to identify the biosynthetic pathways and microbial-based production systems for these plant secondary metabolites, identifying knowledge gaps and providing insights into unexplored production systems. The scope for this review is limited to the most cited coumestans and pterocarpan molecules: medicarpin, pisatin, maackiain, glyceollins, coumestrol, wedelolactone, psoralidin, and glycyrol. The main purpose of this review is to consolidate the link between the current knowledge of pterocarpans and coumestans biosynthesis (including gaps and limitations) and bioprospecting strategies to identify enzymes (address gaps) and novel production systems. We also discuss microbial and enzyme bioprospecting solutions to enhance the sustainable production of pterocarpans and coumestans using microbial-based production systems.

2 Isoflavonoids: A key family of phenolics

Phenolics are vastly diverse molecules distributed within several plant species (Tohge et al., 2013). They contain an aromatic ring and at least one hydroxyl group, and are derivatives of the aromatic amino acids, phenylalanine and tyrosine, through the shikimate or acetate pathways (Bravo, 1998). They are related to pigmentation and astringency, but also have diverse roles in plant defense against pathogens (Durazzo et al., 2019). The phenolics biosynthetic pathway starts from the amino acids phenylalanine and tyrosine, from which the enzymes phenylalanine ammonia-lyase (PAL) and tyrosine ammonia-lyase (TAL) generate cinnamate. Subsequently, a subfamily of 4-coumarate ligases produces different CoA-esters, including p-Coumaroyl-CoA, a key member of polyphenols biosynthetic pathways (Vogt, 2010). The phenolic family comprises phenolic acids, tannins, stilbenes, lignans, and flavonoids. Multiple applications are investigated for all these groups from the pharmaceutical, food, and cosmetic industries (Albuquerque et al., 2021).

Flavonoids are the largest family of polyphenols, ubiquitous in land plants and some algal species, potentially sharing a pathway from a common ancestor (Yonekura-Sakakibara et al., 2019). Epidemiological and meta-analysis studies show flavonoids exert positive human health effects with antioxidant, hepatoprotective, antibacterial, anti-inflammatory, antiviral, and anti-cancer properties (Kumar and Pandey, 2013; Amawi et al., 2017; Jiang et al., 2019). The substitution of chemical groups in the flavonoid backbone is related to its biological and chemical properties (Teng and Chen, 2019). Further chemical modifications such as O- and C-methylation and O- and C-glycosylation alter pharmacokinetic variables such as solubility and chemical stability, shifting the molecule's bioavailability (Sajid et al., 2021a). Seven subclasses are categorized according to the degree of oxidation: flavanones, flavonols, flavones, flavanols, chalcones, anthocyanidins, and isoflavonoids.

Isoflavonoids main structure is composed of three rings, two benzene rings (A and B) linked by a pyran ring (C). Isoflavonoids is the only subclass of flavonoids with the B ring in position 3 instead of position 2 (Figure 1). More than 2,400 isoflavonoids are reported, of which isoflavones are the most extensively cited group, predominantly found in legume species such as soy, alfalfa and chickpea (Dixon, 1999). During biosynthesis of isoflavonoids, p-Coumaroyl-CoA is converted to a chalcone product by chalcone synthase (CHS), and is then catalyzed into the precursors naringenin and liquiritigenin by the chalcone isomerase (CHI). From these precursors, the isoflavonoids, genistein and daidzein, are generated due to the action of the isoflavone synthase (IFS) (Nabavi et al., 2020). Genistein and daidzein, along with the methylated biochanin A and formononetin, are widely described with multiple osteogenic, anticancer and antioxidant properties (Brodowska, 2017; Sajid et al., 2021b). A large range of isoflavonoid derivatives is produced from these molecules, including pterocarpans and coumestans.

3 Pterocarpans and coumestans: Multipurpose molecules with incomplete biosynthetic pathways

Pterocarpans and coumestans are derivatives from the isoflavonoid daidzein. They share the differential feature of a

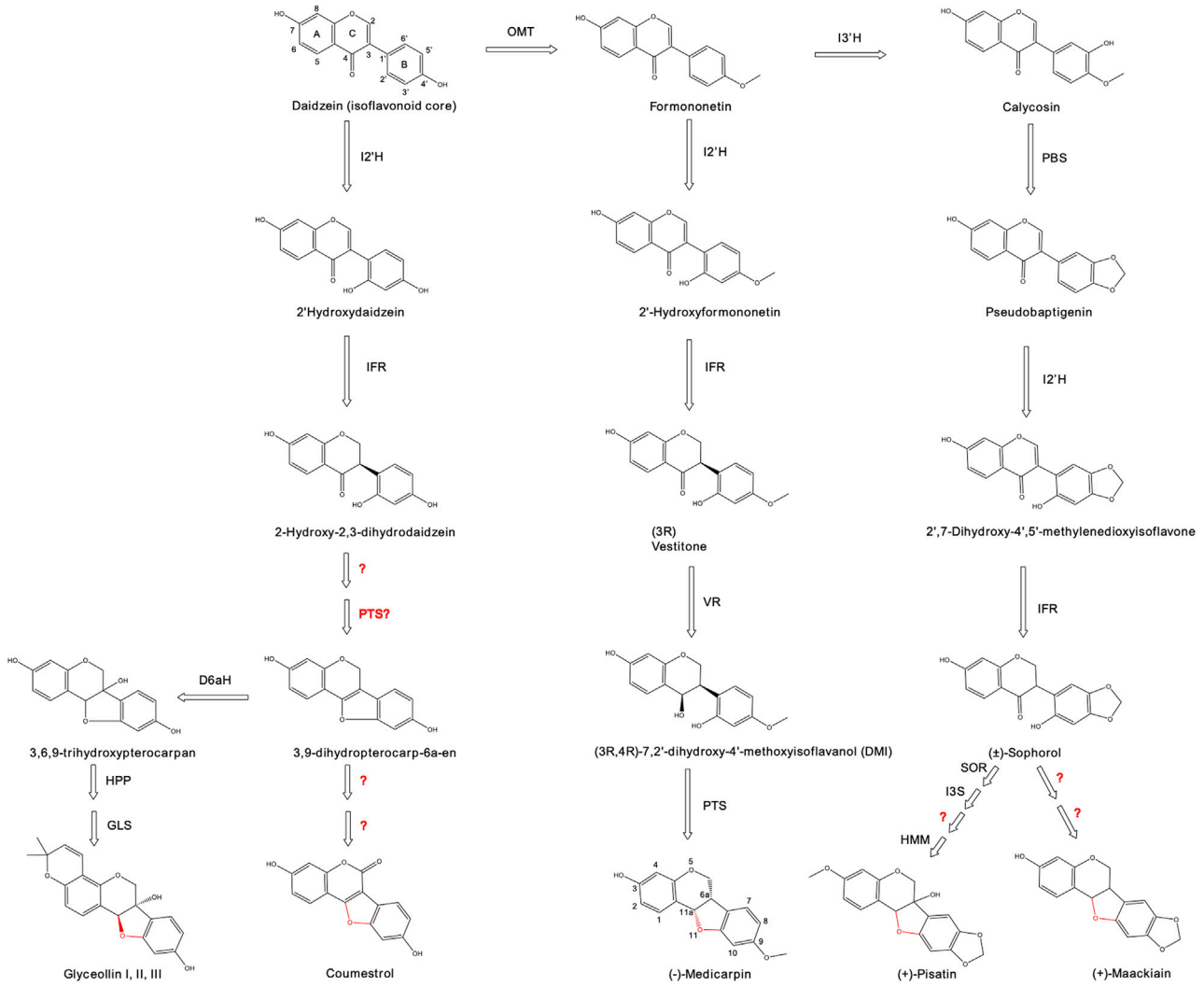


FIGURE 1
Biosynthetic pathway for the production of coumestans and pterocarpans from the precursor, daidzein. Numbers for identifying carbon positions are shown for daidzein (isoflavonoid core) and medicarpin (carbon numbering varies with the formation of the fourth ring). The formation of a fourth ring between carbons 6 and 11 is highlighted in red. Abbreviations: OMT = isoflavanone 4′-O-methyltransferase; I3′H = isoflavone 3′-hydroxylase; I2′H = isoflavone 2′-hydroxylase; IFR = isoflavone reductase; PBS = pseudobaptigenin synthase; VR = Vestitone reductase; PTS = pterocarpan synthase; SOR = sophorol reductase; I3S = isoflav-3-ene synthase; HMM = (+)-6a-hydroxymaackiain 3-O-methyltransferase; DHPH = 3,9-dihydroxypterocarpan 6a-hydroxylase; GLS = glyceollin synthase; HPP = trihydroxypterocarpan prenyltransferase; D6aH = 3,9-dihydroxypterocarpan 6a-hydroxylase. Question marks indicates unknown enzymatic reactions.

formation of a fourth ring between C and B rings, presenting two asymmetric carbons at positions C-6a and C-11a. The chemical structure of relevant pterocarpans and coumestans is reported in Figure 1.

3.1 Production of pterocarpans

Pterocarpans are an extensive group of more than 400 isoflavonoid derivatives with described beneficial human health effects (Jiménez-González et al., 2008; Goel et al., 2013). There are two possible configurations, levorotatory (−) pterocarpans (configuration 6aR, 11aR) and dextrorotatory (+) pterocarpans with the opposite configuration. Within this classification, multiple legume species produce (−) pterocarpans, while just a few plant species, such as peanut (*Arachis hypogaea*), produce (+) pterocarpans (Strange et al., 1985).

Medicarpin is a pterocarpan widely cited due to its diverse range of applications. It is primarily found in legume species, being the main isoflavone in alfalfa (*Medicago sativa L.*) (Du et al., 2010). Medicarpin chemical structure is composed of a pterocarpan core with a hydroxyl group on position C-3 and a methyl group on position C-9. Besides its function as a natural antioxidant, medicarpin is a promising neuroprotective agent for Alzheimer's disease, and also as a bone regeneration agent to treat osteoporosis (Dixit et al., 2015; Li et al., 2021; Kim et al., 2022). Medicarpin also has potential uses in the agricultural sector. As it is a phytoalexin with antimicrobial effects, medicarpin confers powdery mildew

resistance to alfalfa species (Gupta et al., 2022). The first enzyme involved in the biosynthesis of medicarpin is the isoflavanone 4′-O-methyltransferase (OMT). It is a single-step reaction that incorporates a methyl group on the C- 4′ (ring B) of the precursor daidzein, resulting in formononetin as the product. The next step is the formation of 2′-hydroxyformononetin by the enzyme isoflavone 2′ hydroxylase (I2′H), which performs an oxidation at the C-6′ position (ring B). This enzyme belongs to the cytochrome P450 superfamily (class CYP81E1/E7) and produces a set of 2′ hydroxyflavones using different acceptor molecules. It was identified in chickpea, alfalfa and licorice species (Hinderer et al., 1987; Akashi et al., 1998; Liu et al., 2003). The following step in medicarpin synthesis involves the enzyme isoflavone reductase (IFR) that reduces 2′-hydroxyformononetin to the (3R)-2′-hydroxyisoflavanone, reported also as (3R)-vestitone (Paiva et al., 1991). This enzyme is particularly interesting because it introduces a chiral centre using an achiral precursor. IFR was first reported in alfalfa, but homologous sequences were identified from several legume species (Cheng et al., 2015). Vestitone reductase (VR) is the next enzymatic step, which reduces the keto group from position C-4 (ring C) creating the alcohol derivative 7, 2′-dihydroxy4′-methoxyisoflavanol (DMI). VR was first described in *M. sativa* and lately in the pea *Pisum sativum* (Guo et al., 1994a; DiCenzo and VanEtten, 2006). The final step utilizes the enzyme pterocarpan synthase (PTS) to create the characteristic fourth ring between positions C-6a and C-11a, using the two hydroxyl residues as substrates (Fischer et al., 1990). This reaction, which produces (−)-medicarpin from DMI, was described in alfalfa (Guo et al., 1994b). PTS is a dirigent, or stereochemistry-altering, enzyme that also catalyzes the formation of (+)-medicarpin, where the configuration of the hydrogen attached to the C-3 position determines which enantiomer (+ or −) will be produced.

Pisatin and maackiain are phytoalexins from the pterocarpans family that present multiple clinical effects. Maackiain was demonstrated as a potent anti-oxidant and anti-inflammatory agent with anti-sepsis, anti-allergies and neuroprotective actions, among others (Mizuguchi et al., 2015; Tsai et al., 2020; Bai et al., 2022). Besides its antimicrobial activities, pisatin has not been extensibly explored for clinical applications. Pisatin was the first (+)-pterocarpan identified, coming from the pea *P. sativum*, and maackiain was isolated from the Chinese medicinal herb *Sophora flavescens* (Cruickshank and Perrin, 1960; He et al., 2015). The enzyme isoflavone 3′-hydroxylase (I3′H) adds a hydroxyl group at the position C-3′ of the precursor formononetin, as the first step for both pisatin and maackiain biosynthesis. The metabolite produced, named calycosin, is then converted into pseudobaptigenin by the enzyme pseudobaptigenin synthase (PBS) (Liu et al., 2003). The pathway continues with the already mentioned I2′H and IFR enzymes, which introduces chirality to obtain the precursor compounds (+)-sophorol and (−)-sophorol, of different enantiomers. From there, the biosynthetic steps for the synthesis of maackiain are not fully identified yet, but it is believed that downstream enzymes from (+)-pisatin may be involved (DiCenzo and VanEtten, 2006). The production of (+)-pisatin involves the enzyme sophorol reductase (SOR) that produces (3R and 4R)-7, 2′-dihydroxy-4′5′-methylenedioxyisoflavanol (DiCenzo and VanEtten, 2006). The next step is catalyzed by isoflav-3-ene synthase (I3S) that converts the precursor into 7, 2′-dihydroxy-4′, 5′-methylenedioxyisoflav-3-ene (DMDIF). The subsequent enzymatic reaction is still not identified, but it produces the molecule (+)-6a-hydroxymaackiain using DMDIF as a precursor. The final step requires the action of the enzyme (+)-6a-hydroxymaackiain 3-O-methyltransferase (HMM), which adds a methyl group at the C-3 position to generate the final metabolite, (+)-pisatin (Wu et al., 1997).

Glyceollins are a set of soybean-specific pterocarpans with a described set of human health benefits (Pham T. H. et al., 2019). They have an anti-estrogenic effect that competes with endogenous estrogens, and are being tested as a suppressor of breast and ovarian tumorigenesis (Yamamoto et al., 2018). Glyceollin presents the distinctive feature of a fifth prenylated ring linked to the A ring (isoflavonoid core). The first steps from daidzein involve the already discussed I2′H and IFR enzymes to produce 2-hydroxy-2, 3-dihydrodaidzein. From there the enzyme PTS generates the fourth ring to obtain the molecule 3, 9-dihydroterocarp-6a-en, also reported as glycinol (Fischer et al., 1990). Subsequently, a set of prenyltransferases (PTs) convert glycinol into a whole set of precursors, where the prenylated site determines the glyceollin type (I, II or III) (Sukumaran et al., 2018). The last step is the cyclization of the fifth ring by glyceollin synthase, creating up to six different glyceollins depending on their prenyl position (Welle and Grisebach, 1988).

3.2 Production of coumestans

Coumestans is another important subgroup of isoflavonoid derivatives. They are usually reported as a separate group, although some authors refer to them as oxidized products of the pterocarpanoid subfamily (Goel et al., 2013). They are characterized by the addition of a keto group at position C-6, and a double bond between positions 6a and 11a. To date, more than 120 coumestans have been reported with potential clinical effects (Tu et al., 2021). The most studied compound within this family is coumestrol, which is present in soy leaves. Intake of coumestrol is associated with reduced risk for breast cancer, skin photoaging protection, and neuroprotection (Hedelin et al., 2008; Castro et al., 2014; Park et al., 2015). Coumestrol biosynthesis starts from daidzein where the previously mentioned I2′H enzyme adds a hydroxyl group to produce of 2′hydroxydaidzein. The next metabolic step is the reduction of the double bond between C-2 and C-3 (ring C) by the action of the IFR enzyme. The product obtained, 2-hydroxy-2, 3-dihydrodaidzein, is then converted into 3, 9-dihydroterocarp-6a-en by at least two unknown catalytic reactions. A combination of the VR enzyme plus the effect of the isoflav-3-ene synthase (I3S) are predicted to generate the isoflav-3-ene precursor for the characteristic fourth ring formation (Uchida et al., 2020). Finally, another two to three reactions are needed to add the keto group to C-2 of the isoflavonoid core, the distinctive feature of the coumestans subgroup. A transcriptomic study identified that up to 14 genes are associated for coumestrol biosynthesis from the precursor daidzein (Ha et al., 2019). The full metabolic pathway for coumestrol remains incomplete despite efforts to identify the genes involved on its synthesis.

Apart from coumestrol, other relevant coumestans with potential applications for human health are wedelolactone,

psoralidin and glycyrol, although their biosynethic pathways are incomplete. Wedelolactone, first reported in the plant *Wedelia calendulacea*, is a coumestan skeleton with hydroxyl groups at positions 1, 8, and 9; and a methoxyl group at position 3. This metabolite has been studied for multiple clinical applications, such as anti-inflammation, anti-oxidative, inhibits breast cancer, and neuroprotective against Parkinsonism (Hsieh et al., 2015; Zhu et al., 2019; Sharma et al., 2021). Psoralidin, a coumestan first isolated from the legume *Psoralea corylifolia,* has the chemical characteristic of two hydroxyl groups at positions 3 and 9 and a prenyl group at position 2. Multiple pre-clinical studies have demonstrated its anticancer, antiosteoporotic, anti-inflammatory, anti-vitiligo, antibacterial, antiviral, and antidepressant-like effects (Sharifi-Rad et al., 2020). Lastly, glycyrol, a coumestan isolated from the Leguminosae species *Glycyrrhiza* sp., presents a distinctive feature of an O-methylation at position C-1 and a prenylation at position C-2. Glycyrol demonstrates multiple clinical uses as anti-cancer, anti-inflammatory, hepatoprotective, antimicrobial, and anti-viral agents (Tu et al., 2021). Overall, several pterocarpans and coumestans are described with clinical effects; their metabolic pathway is fully reported for some compounds (medicarpin, glyceollin), and is incompletely understood for others (coumestrol, maackiain). The identification of the missing enzymes, as well as the identification of the best production systems, are the first steps for boosting the biosynthesis of pterocarpans and coumestans.

4 Bioprospecting strategies

4.1 Exploring biodiversity

Microbial bioprospecting is a new terminology, but humans have used microbes for their benefit for many years, from the implementation of yeast for bread-making to the discovery of new molecules to treat diseases. The recent emergence of metagenomics analysis has unleashed the potential of microbial bioprospecting. It is estimated that there are more than 1×10^{16} microbes in just 1 ton of soil, of which 85%–99% are unculturable (Curtis and Sloan, 2005; Lok, 2015). Analysis of this massive data set led to the discovery of several new strains for both human (therapeutic, food production) and environmental (sustainable industries, bioremediation) applications.

Different approaches for microbial *in silico* bioprospecting are being already applied using bioinformatic resources. The two main approaches for the identification of microbial diversity are mining data from publicly available databases, or using raw data from different sampling sites following metagenomic pipelines (Vuong et al., 2022b). Within the available datasets, online trustworthy repositories are associated with the International Nucleotide Sequence Database: the DNA Databank of Japan (DDBJ; http://www.ddbj.nig.ac.jp/), the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI; http://www.ebi.ac.uk/ena/) the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/), and the Joint Genome Institute (JGI; https://img.jgi.doe.gov). For environmental metagenomic data, global sequencing initiatives such as Tara Oceans (http://oceanmicrobiome.embl.de/), the Earth Microbiome Project (https://earthmicrobiome.org/), and the Malaspina Gene Database are excellent sources of publicly available information (Acinas et al., 2021). Multiple tools are available for exploring metagenomic data to identify protein sequences and rebuild metabolic pathways. Bioinformatic tools such as QIIME (http://qiime.org/) allow the analysis of metagenomic raw data, and once individual genomes are segregated, tools such as Prodigal (https://github.com/hyattpd/Prodigal) predict genes and link them with metabolic pathways.

Different environments have been explored for novel microbes and metabolic pathways following a metagenomic approach (Quince et al., 2017). Marine samples are a rich source of new microbes, that are bio-prospected as a source of novel genes (Paoli et al., 2022). Extreme environments from polar to volcanic regions, provide a source of extremophiles that are investigated for human health applications using a multi-omics approach (Hedlund et al., 2014). Additionally, not only natural environments are considered a rich source of microorganisms. For instance, industrial effluents are targeted for microbes that degrade lipids, a key activity for biotechnological applications (Peil et al., 2016). The enzyme lipase is then recovered using liquid biphasic flotation, efficiently and sustainably (Sankaran et al., 2018). Overall, microbial bioprospecting generates a rich source of microorganisms for a diverse range of applications, including isoflavonoids production. The following section gives an example of how microalgal species can be targeted to produce isoflavonoid derivatives.

4.2 Employing microalgae

Microbes are an abundant source of new nutraceutical and pharmaceutical products. As phototrophic microbes (prokaryotic cyanobacteria and eukaryotic microalgae) share biosynthetic pathways with plants, scientists are exploring these organisms as chassis to produce plant metabolites. Microalgae are photosynthetic unicellular or colonial microorganisms with the ability to grow in different soil and underwater environments. Phototrophic microbes are divided into two prokaryotic divisions: Cyanophyta and Prochlorophyta, and nine eukaryotic divisions: Glaucophyta, Rhodophyta, Heterokontophyta, Haptophyta, Cryptophyta, Dinophyta, Euglenophyta, Chlorarachniophyta and Chlorophyta (Hemaiswarya et al., 2013). Decades ago, the species *Dunaeliella salina* was successfully implemented as a novel sustainable production system for β-carotene as it naturally produces up to 10% of its dry weight as β-carotene (Harvey and Ben-Amotz, 2020). Since then, microalgal species are bio-prospected for farming (aquaculture), biomanufacturing (nutraceutical, pharmaceutical, functional foods, biofuels), and environmental (bioremediation) applications (Choong et al., 2016; Mobin and Alam, 2017).

Using microalgal species as a sustainable production chassis to produce plant metabolites present several benefits. As a 3rd generation biorefinery, microalgae utilize renewable energy (sunlight) and CO_2 and hence microalgae are a key global contributor to CO_2 sequestration (Liu Z. et al., 2020; Prasad et al., 2021). Some microalgal species also have the potential to use wastewater as a source of nutrients, adding another positive impact on the environment as a remediation tool (Goswami et al., 2022). Additionally, similar to the yeast *S. cerevisiae,* some microalgal species are Generally Recognized As Safe (GRAS), so

TABLE 1 Microalgal species that naturally produce isoflavonoid derivatives.

Flavonoids	Microalgal species	Environment	Concentration	References
Total flavonoid content	Cyanobacteria *Chroococcidiopsis thermalis*	Extreme environments: deserts and hot springs	2.44 mg/g dry biomass	Ijaz and Hasnain (2016)
	Cyanobacteria *Leptolyngbya* sp.	Tunisian hot springs	34.9 mg/g dry biomass	Trabelsi et al. (2016)
	Chlorophyta *Desmodesmus* sp.	Wastewater treatment system	4.03 mg/g dry biomass	Safafar et al. (2015)
	Chlorophyta *Dunaliella salina*	Hypersaline environments (salt lakes)	3.61 mg/g dry biomass	Safafar et al. (2015)
	Euglenoida *Euglena tuba*	Indian aquatic bodies	1.01 mg/g dry biomass	Chaudhuri et al. (2014)
	Chlorophyta *Chlorella sorokiniana*	Freshwater (able to grow in wastewater)	2.41 mg/g dry biomass	Safafar et al. (2015)
	Chlorophyta *Chlorella vulgaris*	Freshwater (able to grow in wastewater)	118 µg/mL culture	Yadavalli et al. (2022)
Isoflavonoid precursors				
p-Coumaric acid	Diatom *Phaeodactylum tricornutum*	Marine artic polar	750 ng/g dry biomass	Goiris et al. (2014)
	Rhodophyta *Porphyridium purpureum*	Marine environment	770 ng/g dry biomass	Goiris et al. (2014)
	Chlorophyta *Haematococcus pluvialis*	Freshwater	640 ng/g dry biomass	Goiris et al. (2014)
Naringenin	Chlorophyta *Haematococcus pluvialis*	Freshwater	0.6 ng/g dry biomass	Goiris et al. (2014)
	Haptophyta *Diacronema lutheri*	Marine environment	0.6 ng/g dry biomass	Goiris et al. (2014)
	Cyanobacteria *Leptolyngbya* sp.	Tunisian hot springs	4.1 ng/g dry biomass	Trabelsi et al. (2016)
Isoflavonoids				
Daidzein	Diatom *Phaeodactylum tricornutum*	Marine artic polar	5.9 ng/g dry biomass	Goiris et al. (2014)
	Rhodophyta *Porphyridium purpureum*	Marine environment	1.27 ng/g dry biomass	Goiris et al. (2014)
	Chlorophyta *Haematococcus pluvialis*	Freshwater	0.6 ng/g dry biomass	Goiris et al. (2014)
	Cyanobacteria *Nostoc 17*	Marine, freshwater and terrestrial	7.05 ng/g dry biomass	Klejdus et al. (2010)
	Chlorophyta *Scenedesmus* sp.	Marine, freshwater and terrestrial	10.59 ng/g dry biomass	Klejdus et al. (2010)
Genistein	Diatom *Phaeodactylum tricornutum*	Marine artic polar	1.42 ng/g dry biomass	Goiris et al. (2014)
	Rhodophyta *Porphyridium purpureum*	Marine environment	0.63 ng/g dry biomass	Goiris et al. (2014)
	Chlorophyta *Haematococcus pluvialis*	Freshwater	0.4 ng/g dry biomass	Goiris et al. (2014)
	Cyanobacteria *Nostoc 17*	Marine, freshwater and terrestrial	5.91 ng/g dry biomass	Klejdus et al. (2010)
	Chlorophyta *Spongiochloris spongiosa*	Freshwater and terrestrial	4.27 ng/g dry biomass	Klejdus et al. (2010)
	Chlorophyta *Scenedesmus* sp.	Marine, freshwater and terrestrial	6.11 ng/g dry biomass	Klejdus et al. (2010)
Formononetin	Cyanobacteria *Nostoc 17*	Marine, freshwater and terrestrial	33.14 ng/g dry biomass	Klejdus et al. (2010)
	Chlorophyta *Spongiochloris spongiosa*	Freshwater and terrestrial	4.29 ng/g dry biomass	Klejdus et al. (2010)
	Chlorophyta *Scenedesmus* sp.	Marine, freshwater and terrestrial	5.92 ng/g dry biomass	Klejdus et al. (2010)
Coumestrol	Cyanobacterial multispecies blooms	Czech lakes and ponds	0.1 ng/L water sampled	Procházková et al. (2017)
IFS Enzyme	45 different algae taxa	Multiple environments	NA	Del Mondo et al. (2022)

they can be consumed as food (Villarruel-Lopez et al., 2017; Wells et al., 2017; Caporgno and Mathys, 2018; Khemiri et al., 2020). Even though, conducting a food safety risk assessment is imperative to ensure a high-quality product for human consumption (Wu et al., 2022).

Isoflavonoids derivatives and their associated enzymes were recently identified in algal species (Del Mondo et al., 2022). A summary of these findings is represented in Table 1. Microalgal species share similar metabolic pathways with plants probably due to a primordial ancestor, creating an exciting avenue as isoflavonoid sustainable producers (Klejdus et al., 2010; Del Mondo et al., 2021). The isoflavonoids genistein and daidzein were found in the prokaryotic Cyanobacteria *Nostoc,* as well as in eukaryotes divisions such as Chlorophyta and Rhodophyta (Goiris et al., 2014; Del Mondo et al., 2021; Ferdous and Balia Yusof, 2021). Massive microalgal blooms may have a negative impact on the environment and human health, but those species can be explored in confined laboratory conditions to source valuable genes, enzymes,

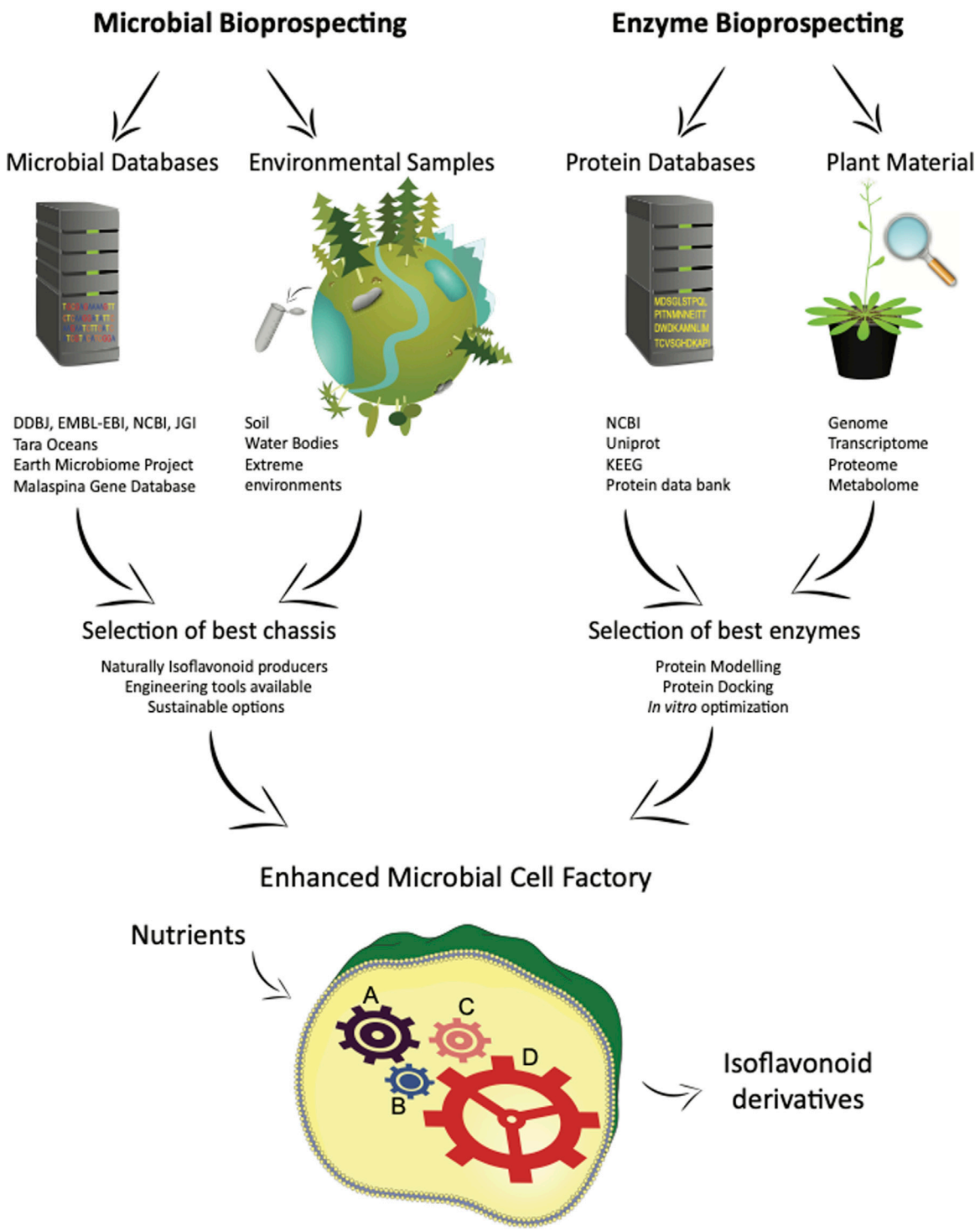


FIGURE 2
Bioprospecting strategies for microbial and enzyme bioprospecting. Microbial bioprospecting allows the identification of best chassis, whereas enzyme bioprospecting identifies the best enzymes for the production of isoflavonoid derivatives using microbial cell factories.

and molecules. The IFS, a key protein for the biosynthesis of isoflavonoids have been found in 45 of the 47 algal taxa analyzed according to sequence alignments (BLASTp) (Del Mondo et al., 2022). The IFS plus other 28 phenylpropanoid core enzymes were screened using a multistep *in silico* analysis, that demonstrates the presence of isoflavonoid-related genes across microalgal databases (Del Mondo et al., 2022).

Besides the identification of microalgal species that can naturally produce isoflavonoids, strategies should be followed to boost metabolite levels to generate a commercially viable solution. In

TABLE 2 Online sources and tools for enzyme bioprospecting.

Enzyme bioprospecting sources and tools	Application	References
Protein databases (source of information)		
Joint genome institute	Multispecies DNA, RNA and protein database (annotations and distribution)	Grigoriev et al. (2012)
NCBI	Multispecies DNA, RNA and protein database (annotations, distribution, blast)	Jenuth (1999)
Uniprot	Find, align and blast proteins	The UniProt Consortium (2019)
KEGG	Exploration of metabolic pathways and enzymes involved	Kanehisa and Goto (2000)
Protein data bank (PDB)	Experimentally determined 3D structures and computed protein models	Dutta and Berman (2005)
Pathway and enzyme discovery		
Retro-analysis	Discover biosynthetic pathways using metabolite's information	Delépine et al. (2018)
Metabolic modelling	Discover biosynthetic pathways using information from multiple omics	Zampieri et al. (2019)
Enzyme characteristics (Protein Modelling)		
ProtParam (ExPaSy resource portal)	Calculate protein main parameters	Artimo et al. (2012)
SWISS-MODEL	Protein modelling	Waterhouse et al. (2018)
ALPHAFOLD	Protein Structure Database and modelling	Jumper et al. (2021)
ColabFold	Protein modelling combining homology search technologies	Mirdita et al. (2022)
MEME Suite Tools	Protein motif discovery and analysis	Bailey et al. (2015)
Enzyme-ligand interaction (protein docking)		
AutoDock and Vina	Protein-ligand affinity interaction (PyMOL visualization)	Seeliger and de Groot (2010)
CB-Dock	User friendly web-based for protein-ligand affinity	Liu et al. (2019)
HADDOCK	Improved protein-ligand docking considering homology conformations	Koukos et al. (2021)
DEELIG	Deep-learning-based method for protein-ligand determination	Ahmed et al. (2021)
Protein Docking of Interfacial Enzymes	Virtual screening that considers both catalytic and membrane-interaction domains	Schaller et al. (2022)
In vitro techniques for further optimization		
directed evolution, random mutagenesis, error-prone amplification	Generation of unnatural homology variants	Kamble et al. (2019)
Enzyme scaffold substrate channeling	Improve biosynthetic cascades linking enzymes together	Ellis et al. (2019)

recent years, engineering tools have been developed for some phototrophic microorganisms such as cyanobacteria, chlorophytes, diatoms, and eustigmatophytes. For these organisms, culture conditions and genetic edition tools were optimized (Vavitsas et al., 2021). Phototrophic unicellular microorganisms have already shown immense potential as sustainable production platforms, and the yet unexplored field of microalgal isoflavonoid production may soon become a reality.

4.3 Optimizing metabolic pathways

Enzymes with enhanced activity is a particularly effective strategy for boosting metabolite production. The approach is slightly different to microbial bioprospecting, where instead of using metagenomic data from the environment, an organism's protein database (proteomics) is combined with its metabolic expression for discovering and optimizing enzymatic reactions. A diagram representing both microbial and enzyme bioprospecting strategies is shown in Figure 2.

The identification of the state-of-the-art is the first step for the discovery of enzymes and whole metabolic pathways. As multiple steps are required for the discovery of the best enzymes, a summary of enzyme bioprospecting tools divided by stages is detailed in Table 2. To identify biosynthetic gaps, the presence of metabolites and their precursors can provide information about enzymatic reactions. Metabolite detection using High-Performance Liquid Chromatography (HPLC) or Mass Spectrometry (MS) analysis evaluates chemo-structural diversity to build enzymatic pathways (Santana et al., 2021). Nuclear magnetic resonance

(NMR) spectroscopy is another approach that allows the identification of potentially bioactive compounds using the whole plant metabolomic set (Augustijn et al., 2021). The biosynthetic pathway of the enzymes involved in the generation of those metabolites can be predicted using retro-analysis tools, linking the precursors with the final product (Delépine et al., 2018). A machine learning approach allows not only to predict one metabolic pathway but the whole organism's metabolomics, using genomic, transcriptomic and proteomic information altogether (Zampieri et al., 2019). A different approach is followed when the metabolic information is already known. The database KEGG is the most cited source of already fully described metabolic pathways and compound relationships. Multiple public protein databases provide information about proteins, such as Joint Genome Institute (JGI; https://jgi.doe.gov/), Uniprot (https://www.uniprot.org) and Protein Data Bank (PDB; https://www.rcsb.org/). The NCBI database, as well as Uniprot, allow the short-listing of homologous proteins as potential candidates for biosynthetic pathways.

After identifying the target enzyme within a specific biosynthetic pathway, the first step for boosting its activity is to model its 3D structure. Protein modelling is performed using tools such as ProtParam software (https://web.expasy.org/protparam/), SWISS-MODEL (https://swissmodel.expasy.org/) and Alphafold (https://alphafold.ebi.ac.uk/) to predict the protein conformation, kinetic parameters and physicochemical properties (extinction coefficient, estimated half-life, instability index, and aliphatic index). The search is usually based on full protein homology and domain structures, and motif analysis tools such as the MEME Suite Tool can provide an extra layer of enzymatic information (Bailey et al., 2015). Molecular docking is the next step that allows an *in silico* test of protein-ligand interactions. This virtual screening estimates the ligand position within the modeled protein cavity and the potential amino acids involved in the enzymatic reaction. Several tools are described in the literature for applying this technology to discover novel enzymes with improved activity, considering rigid and flexible dockings (Fan et al., 2019). Within these tools, the popular molecule-viewing software PyMOL is widely used with the docking suite AutoDock and Vina (Seeliger and de Groot, 2010). A rectangular box defines the binding ligand-protein site, from where multiple binding runs are performed between the ligand and the specific amino acids, to determine the binding energies. Cutting-edge techniques like High Ambiguity Driven DOCKing (HADDOCK) incorporate the factor of ligand conformation, adding an additional layer of information in the dynamic landscape of this field (Koukos et al., 2021). There are also user-friendly and web-based tools, such as CB-Dock, that retain the power of protein docking (Autodck Vina) without compromising the quality of the results (Liu Y. et al., 2020). In the last few years, new approaches have been developed using the power of deep learning and convolutional neural networks to "train" the model for the characterization of multiple protein-ligand affinity interactions. As an example, the deep-learning-based approach (DEELIG) uses this method (Ahmed et al., 2021). For interfacial enzymes, the protein motifs (catalytic and membrane binding) are analyzed separately, to build homology models and docking activity data (Schaller et al., 2022). The implementation of all these strategies allows researchers to predict protein-ligand binding affinity and overall enzymatic activity to select the more suitable enzymes for a target reaction.

Presently, *in silico* enzyme bioprospecting is positively impacting the way where novel enzymes are screened, optimized, and tested, in a cost-effective manner. Nevertheless, an *in vitro* approach is necessary to confirm the enzyme's pharmacokinetic and predicted activity. Conventional enzyme engineering tools such as directed evolution, random mutagenesis, and error-prone amplification can add variability to the discovery of novel enzymes (Kamble et al., 2019). Once every enzymatic step is optimized for the best activity, multienzyme scaffolds can also contribute to improving biosynthetic cascades through substrate channeling, linking all the enzymes involved in a concise subcellular space (Ellis et al., 2019). Altogether, *in silico* and *in vitro* strategies enhance the biosynthetic pathway efficiency and positively contribute to obtain higher production titers.

5 Approaches for the production of pterocarpans and coumestans

Plant-based systems (homologous biosynthesis) are the main source of coumestans and pterocarpans, but have several limitations. First, plant-based systems are unsustainable due to the consumption of environmental resources (land, water, fertilizers). Second, the target plant biomass usually requires long growing periods to reach the harvest stage. Third, the concentration of coumestans and pterocarpans is low when compared with other isoflavones, and difficult to isolate and purify from the plant tissue. As an example, coumestrol from *Glycine max* cultivar "Santa rosa" (high coumestrol concentration) showed 1.85 μg/g dry material, while the same cultivar contains 560 μg/g dry material of the precursor daidzein (Mazur et al., 1998). Some of the plant-based production limitations can be solved by plant-tissue culturing. Coumestrol biosynthesis was obtained through adventitious soybean root cultivation, but it is still an unsustainable and cost-ineffective solution (Lee et al., 2022). Another production strategy is chemical synthesis, where the metabolite can be solely synthesized using chemical approaches. This process usually uses hazardous and expensive reagents, as is the case for the full chemical production of medicarpin (Yang et al., 2017). To some extent, a chemical reaction can be used to further modify plant metabolites, adding methyl or glycosyl groups to enhance bioavailability. This has been achieved for flavonoid production, where bioavailability, anti-inflammatory, and anticancer activities were enhanced by chemically adding methyl groups (Wen et al., 2017).

5.1 Evaluating microbial hosts

The application of microbial-based production systems (heterologous biosynthesis), empowered by bioprospecting, may provide an alternative solution for overcoming the limitations previously described by plant-based and chemical approaches. This lab-based production method does not require arable land, is scalable, sustainable, and cost-effective depending on the chassis (microbe) utilized. Granted, there are challenges that the utilization of microbial-based systems for plant natural products still need to overcome. The access distribution to this technology is mainly dominated by the United States, and the patenting of plant biosynthetic pathways may threaten small-acre producers from developing countries (French, 2019). A global regulatory

TABLE 3 Successful examples of heterologous biosynthesis of isoflavonoid precursors using different cell factories. Abbreviations: ARO10, phenylpyruvate decarboxylase; PDC5, pyruvate decarboxylase; TAL, tyrosine ammonia-lyase; ARO4, 3-deoxy-D-arabino-heptulosonic acid synthase; ARO7, chorismate mutase; AROL, shikimate kinase II; 4CL, 4-coumaroylCoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CHR, chalcone reductase; IFS, isoflavone synthase; HIS, 2-hydroxyisoflavanone synthase; HID, 2-hydroxyisoflavanone dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; ACC, acetyl-CoA carboxylase; CCL, cinnamate/coumarate-CoA ligase; CPR, cytochrome P450 reductase.

Precursors	Microbial cell factory	Substrate	Titer	Heterologous enzymes and other strain modifications	Strategy	References
Pterocarpans and coumestans precursors						
p-Coumaric acid	*E. coli*	Tyrosine	2.5 gr/L	TAL	Screening and overexpression of multiple TAL variants	Jones et al. (2017)
	S. cerevisiae	Tyrosine	1.9 gr/L	TAL, ARO10 and PDC5 (knock-out), ARO4, ARO7 and AROL (overexpression)	Reduction of by-products and overexpression of precursors	Rodriguez et al. (2015)
Liquiritigenin	*E. coli*	p-Coumaric acid	7.6 mg/L	4CL2, CHR, CHS, CHI	Different copy number plasmids to control enzyme expression	Yan et al. (2007)
	S. cerevisiae	Tyrosine	5.3 mg/L	TAL, 4CL, CHR, CHS, CHI	p-Coumaric acid overexpression yeast plus liquiritigenin metabolic pathway	Rodriguez et al. (2017)
	Yarrowia lipolytica	Phenylalanine	62.4 mg/L	PAL, 4CL, CHS-CHR (fused), CHI	Strain engineering adding 4 genes plus 1 fusion enzyme	Akram et al. (2021)
Daidzein (isoflavonoid)	*E. coli*	Liquiritigenin	18 mg/g dry weight	IFS	IFS engineering for using P450 enzymes in bacteria	Leonard and Koffas (2007)
	S. cerevisiae	Glucose	85.4 mg/L	PAL, C4H-4CL (fused), CHS, CHR-CHI (fused), HIS, HID	Selection of best homologous candidates, control expression with promoters, precursors overexpression and protein fusion	Liu et al. (2021)
Other Isoflavonoids precursors						
Naringenin						
	E. coli	Tyrosine	57 mg/L	PAL (TAL activity), C4H, 4CL, CHS, CHI, ACC (overproduce manolyl-CoA), ScCCL (overproduce 4-coumaroyl-CoA)	Metabolic strain engineering for overexpressing precursors	Miyahisa et al. (2005)
	S. cerevisiae/ E. coli	D-xilose	21.1 mg/L	TAL, CHS, 4CL, CHS, CHI (*S. cerevisiae*); pyk and pheA gene knock-out (*E. coli*) to overexpress acetate and tyrosine	Co-culture with multiple copies of 4CL into an extra plasmid	Zhang et al. (2017)
	S. cerevisiae	Tyrosine	90 mg/L	TAL, CHS, 4CL, CHS, CHI	Engineered strains to overexpress precursors tyrosine and malonyl-CoA	Lyu et al. (2017)
Genistein (isovlavonoid)	*E. coli*	p-Coumaric acid	18.6 mg/L	4CL, CHS, IFS, CPR	Inclusion of P450 machinery in bacteria and co-culture for split the pathway	Kim (2020)
	S. cerevisiae/ E. coli	Tyrosine	6 mg/L	PAL, 4CL, CHS, CHI (*E. coli*), IFS (*S. cerevisiae*)	Co-culture for achieving higher titer expressing P450 enzyme IFS in yeast	Katsuyama et al. (2007)
	S. cerevisiae	Phenylalanine	0.1 mg/L	PAL, C4H, 4CL, IFS, CPR, CHS, CHI	Introduction of a *de novo* biosynthetic pathway in one strain	Trantas et al. (2009)

framework is essential to guarantee fair access to this technology, and to protect the original sources of natural products.

The chassis selection plays a central role in the production of plant metabolites. The commonly used model species, *Escherichia coli* and *S. cerevisiae,* are largely employed as there are numerous of engineering tools available (Cravens et al., 2019). Between these two organisms, *S. cerevisiae* is more suitable for plant metabolites as it naturally contains the plant machinery (Cytochrome P450 enzymes and others) required for plant metabolite production. *S. cerevisiae* is also Generally Recognized As Safe (GRAS), being an excellent candidate to produce food supplements and other nutraceutical derivatives (Singh and Gaur, 2021). Other GRAS yeast species such as *Aspergillus sp.* and *Hansenula polymorpha* have been used for the expression of plant products (Pham J. V. et al., 2019). Bioprospecting strategies have already been implemented to identify bacterial host systems compatible for large-scale cultivation. Poorly investigated microorganisms such as the Gram-positive *Streptomyces sp.*, *Bacillus sp.*, and *Lactococcus lactis* are also promising cell factories for the production of produce plant chemicals. Different microbial production

strategies have been explored for the production of isoflavonoids (Chouhan et al., 2017; Sajid et al., 2021b). Table 3 summarizes the current reports for coumestans, pterocaptans precursors and other isoflavonoid derivatives. The microbial-based production of the precursors p-coumaric acid, liquiritigenin and daidzein have already been achieved using both bacterial and yeast species, but the biosynthesis of coumestans and pterocarpans is incompletely described in the literature. In the case of coumestans (coumestrol), a heterologous production is still not feasible as its biosynthetic pathway is not fully described. Similarly, the *de novo* biosynthesis of medicarpin is complex and difficult to optimize as it requires several enzymatic reactions.

A "divide and conquer" strategy may ease the metabolic burden of incorporating multiple genes in one strain. This co-culture strategy also allows the exploration of branched molecules combining different enzyme cascades. The selected metabolic pathway is separated and integrated into different strains, which can be the same or different species (Chen et al., 2019). As an example, an *E. coli* bi-culture was used to produce the flavonoid catechin using one strain to obtain naringenin (upstream precursor) while the second strain uses the precursor to generate the target metabolite (Jones et al., 2016). For isoflavonoids, up to 6 mg/L of the metabolite genistein was produced by dividing the pathway between bacterial and yeast species (Katsuyama et al., 2007). The use of polycultures is an efficient strategy for the biosynthesis of complex molecules. For instance, multiple enzymatic reactions were divided into several strains of *E. coli* to produce anthocyanins from simple sugars (Jones et al., 2017). Also, a symbiotic relationship can be established by forming a consortium where one species provides substrates to the other one and *vice versa*. This is the case of co-culturing *S. cerevisiae* with microalgal species, in which the algae produce essential nutrients and the yeast provides extra CO_2 (Nguyen et al., 2020; Alam et al., 2022). Overall, the use of bioprospecting strategies for sourcing better enzymes and microbial chassis offers a promising strategy for boosting plant metabolite production.

Apart from the model species, *E. coli* and *S. Cerevisiae*, other microbes need be explored as production chassis for the production of pterocarpans and coumastans. The presence of pterocarpans and coumastans in cyanobacterial blooms indicates that the full metabolic pathway is already present in phototrophic bacteria (Procházková et al., 2017). Daidzein, a precursor of pterocarpans and coumastans, was found both in prokaryotic and eukaryotic microalgal species (Klejdus et al., 2010; Goiris et al., 2014). The addition of three to four enzymatic steps will allow the production of pterocarpans and coumastans from those microalgal species. Furthermore, the activation of promoters through nuclease-dead Cas9 systems can help to overexpress key enzymes of the pathway, as was demonstrated in rice plants (Gong et al., 2020). Lastly, the use of co-culture between yeast and microalgal species presents a strategy for enhancing secondary metabolite production, not yet been explored for pterocarpans and coumestans. Scientific advances in recent years provide an excellent set of tools to unleash the production of pterocarpans and coumastans. Filling the enzyme characterization gaps, and selecting the best production chassis would help to meet the escalating demand for these isoflavonoid derivatives.

6 Evaluating enzyme orthologs

Key enzymes for pterocarpans and coumenstans production such as OMT, I2H, IFR, VR, and PTS were originally identified from enzyme characterization studies. The OMT enzyme, key for the conversion of daidzein to formononetin, was first characterized in the legumes *Medicago truncatula* and *Glycyrrhiza echinate* (Akashi et al., 2003; Liu et al., 2006; Li et al., 2016). More recently, another enzyme from the species *Pueraria lobata* (PIOMT9) was identified as an alternative route for formononetin biosynthesis (Li et al., 2016). The membrane-bound CYP450 protein I2′H, key to hydroxylate isoflavones, was cloned and expressed in *E. coli* to test its ability to act on formononetin. The subtype sequence from *Lotus japonicus* generated 8.4 mg/L of hydroxyformononetin, while the *G. echinate* variety failed to produce that compound (Uchida et al., 2015). An ortholog from *Astragalus membranaceus* was also cloned and purified, but its functional activity was not fully reported (Chen et al., 2015). The IFR enzyme from *G. max* was purified and characterized to detect its activity *in vitro,* and overexpressed in soybean seeds to confirm the overproduction of glyceollins (Cheng et al., 2015). The IFR subtype from *Medicago sativa* was successfully implemented to produce vestitone (Uchida et al., 2017). The VR orthologous sequence from *M. sativa* was tested using an *E. coli* chassis, and its activity was tested to produce DMI (Guo and Paiva, 1995). The activity of PTS, a key enzyme for pterocarpan biosynthesis, was characterized for *Glycyrrhiza echinata*, *L. japonicus*, and *G. max* species. From that study, the *G. echinate* ortholog was identified as the most efficient for medicarpin production (Uchida et al., 2017).

Enzyme characterization using *in silico* tools is a useful approach for enhancing the heterologous production of pterocarpans and coumenstans. Flavones and flavonol synthases have been bio-prospected using some of the tools previously described. In an interesting approach, researchers curated 44 enzymes using PyMOL and AutoDock Vina for protein docking and based on the instability index and conserved domain data selected the best candidates (Wang et al., 2021). In a recent publication, the IFS, a critical enzyme for the biosynthesis of isoflavonoids, was analyzed by considering its interaction with the molecules, liquiritigenin and naringenin. The use of Alphafold and Swiss-Model for the modelling, and AutoDock tools for the docking, allowed the authors to determine *in silico* the IFS from *Trifolium pretense* as the best candidate for liquiritigenin conversion (Sajid et al., 2022). For pterocarpans and coumestans production, the enzyme VR has previously been crystallized, modelled and docked *in silico* to test its ability to produce DMI (Shao et al., 2007). The enzyme PTS, key in the cycling stage for producing the fourth ring structure, was recently analyzed using protein docking with the ligand DMI. The PTS enzyme from licorice (*G. echinata*) was determined as the best dirigent protein for the formation of both (−) and (+)-medicarpins (Meng et al., 2020). Besides these analyses, no other efforts have been reported so far for modelling and docking enzymes related to pterocarpans and coumestans biosynthesis.

7 Challenges and prospects

The main challenge that the scientific community will face shortly is how to apply sustainable production systems to cope with the increasing global demand for isoflavonoids. The discovery

of a novel production chassis through microbial bioprospecting may help to identify more efficient and sustainable solutions. Bioprospecting allows the discovery of novel microorganisms that produce isoflavonoid derivatives, besides plant species. Phototropic microbes such as chlorophyta and cyanobacteria may become an alternative chassis, boosting metabolite production, and conserving the environment at the same time. Notably, the use of these species are potentially beneficial to the environment on account of their capacity to act as carbon-sinks and wastewater remediators, in line with the United Nations Sustainable Developments Goals (Vuong et al., 2022a). Efforts should be implemented to develop more engineering tools for algal species. A genetic toolkit is available for diatoms and cyanobacteria species, but gene delivery, transformation, and selection of recombinant microalgal strains require significant improvement to replace model species *S. saccharomyces* and *E. Coli* (Lee et al., 2023).

A full comprehension of the metabolic pathway is essential to determine the enzymes involved in pterocarpans and coumestans biosynthesis. Enzyme bioprospecting allows the identification and selection of the best enzyme homologs, using *in silico* bioinformatic tools for modelling and docking simulations. Although some pterocarpans and coumestans metabolic enzymes were analyzed, such as the IFS and PTS, further contributions should deliver more information about the best homologous enzymes for the whole pathway. Once the best production chassis and enzymes are identified, microbial-based systems should be implemented to boost pterocarpans and coumestans expression. The employment of a symbiotic poly-culture strategy may positively contribute to the target molecule production, and ease the optimization of fermentation conditions. Other tools that can help to boost production titer besides the incorporation of the metabolic pathway are: a) determining the perfect expression ratio of all the enzymes involved through techniques such as multiplex automatic genome engineering (MAGE) (Wang et al., 2009), b) implementation of strains with an increased expression of precursors and cofactors (Akhtar and Jones, 2014), c) applying substrate channeling through enzyme scaffolds and protein fusions (Kang et al., 2018; Choi et al., 2019). Overall, a link between researchers with metabolic pathway/enzymes background, and engineering specialists who work on microbial sustainable production systems may overcome the isoflavonoid production bottleneck.

8 Conclusion

Isoflavonoids derivatives such as pterocarpans and coumestans have applications in multiple fields. Diseases such as breast cancer, osteoporosis and neurodegenerative syndromes can be targeted using these molecules to improve human health and wellbeing. Nevertheless, further experimental contributions are needed to enhance their production, purification and overall implementation. Though efforts have been implemented to identify the biosynthetic pathways of some pterocarpans and coumestans, others remain unclear. A holistic bioprospecting strategy that considers both pathway identification (enzymes) and microbial discovery (production systems) is very much required. We describe the "state of the art" microbial-based systems for pterocarpans and coumestans production, and propose the implementation of microalgae species as sustainable production platforms. This strategy is not limited to isoflavonoid derivatives, and can be applied to other plant natural product such as terpenes, steroids, other phenolics. The shared knowledge contributed by the scientific community regarding the bioprospecting of enzymes and microbes will be invaluable for the development of production platforms for the isoflavonoid derivatives, pterocarpans and coumestans.

Author contributions

FP conceived, designed and wrote the initial draft of this review article. JP and PK reviewed and edited the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Funding

PK is supported by the University of Western Australia with additional research funding from ExPlanta Pty Ltd., towards Biosynthesis of Isoflavones—Formononetin (FMN) Proof of Concept project. JP is supported by an NHMRC Investigator Grant (GNT1196188).

Acknowledgments

FPR acknowledges the University of Western Australia for providing international fee-offset scholarship and the University Postgraduate Award.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Acinas, S. G., Sánchez, P., Salazar, G., Cornejo-Castillo, F. M., Sebastián, M., Logares, R., et al. (2021). Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun. Biol.* 4, 604–615. doi:10.1038/s42003-021-02112-2

AfifaHussain, N., Baqar, Z., Mumtaz, M., El-Sappah, A. H., Show, P. L., Iqbal, H. M. N., et al. (2022). Bioprospecting fungal-derived value-added bioproducts for sustainable pharmaceutical applications. *Sustain. Chem. Pharm.* 29, 100755. doi:10.1016/j.scp.2022.100755

Ahmed, A., Mam, B., and Sowdhamini, R. (2021). Deelig: A deep learning approach to predict protein-ligand binding affinity. *Bioinforma. Biol. Insights* 15, 117793222110303. doi:10.1177/11779322211030364

Akashi, T., Aoki, T., and Ayabe, S. (1998). CYP81E1, a cytochrome P450 cDNA of licorice (Glycyrrhiza echinataL.), encodes isoflavone 2′-hydroxylase. *Biochem. Biophys. Res. Commun.* 251, 67–70. doi:10.1006/bbrc.1998.9414

Akashi, T., Sawada, Y., Shimada, N., Sakurai, N., Aoki, T., and Ayabe, S. (2003). cDNA cloning and biochemical characterization of S-Adenosyl-l-Methionine: 2,7,4′-Trihydroxyisoflavanone 4′-O-methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway. *Plant Cell Physiol.* 44, 103–112. doi:10.1093/pcp/pcg034

Akhtar, M. K., and Jones, P. R. (2014). Cofactor engineering for enhancing the flux of metabolic pathways. *Front. Bioeng. Biotechnol.* 2, 30. doi:10.3389/fbioe.2014.00030

Akram, M., Rasool, A., An, T., Feng, X., and Li, C. (2021). Metabolic engineering of Yarrowia lipolytica for liquiritigenin production. *Chem. Eng. Sci.* 230, 116177. doi:10.1016/j.ces.2020.116177

Alam, Md. A., Wan, C., Tran, D. T., Mofijur, M., Ahmed, S. F., Mehmood, M. A., et al. (2022). Microalgae binary culture for higher biomass production, nutrients recycling, and efficient harvesting: A review. *Environ. Chem. Lett.* 20, 1153–1168. doi:10.1007/s10311-021-01363-z

Albuquerque, B. R., Heleno, S. A., Oliveira, M. B. P. P., Barros, L., and Ferreira, I. C. F. R. (2021). Phenolic compounds: Current industrial applications, limitations and future challenges. *Food Funct.* 12, 14–29. doi:10.1039/d0fo02324h

Amawi, H., Ashby, C. R., and Tiwari, A. K. (2017). Cancer chemoprevention through dietary flavonoids: what's limiting? *Chin. J. Cancer* 36, 50. doi:10.1186/s40880-017-0217-4

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi:10.1093/nar/gks400

Augustijn, D., de Groot, H. J. M., and Alia, A. (2021). HR-MAS NMR applications in plant metabolomics. *Molecules* 26, 931. doi:10.3390/molecules26040931

Bai, X., Zhu, Y., Jie, J., Li, D., Song, L., and Luo, J. (2022). Maackiain protects against sepsis via activating AMPK/Nrf2/HO-1 pathway. *Int. Immunopharmacol.* 108, 108710. doi:10.1016/j.intimp.2022.108710

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi:10.1093/nar/gkv416

Bansal, M., Tiwari, N., and Sharma, J. G. (2022). "Chapter 3 - revolution in microbial bioprospecting via the development of omics-based technologies," in *Bioprospecting of microbial diversity*. Editors P. Verma and M. P. Shah (Netherlands: Elsevier).

Beattie, A. J., Hay, M., Magnusson, B., de Nys, R., Smeathers, J., and Vincent, J. F. V. (2011). Ecology and bioprospecting. *Austral Ecol.* 36, 341–356. doi:10.1111/j.1442-9993.2010.02170.x

Becker, J., and Wittmann, C. (2020). Microbial production of extremolytes — High-value active ingredients for nutrition, health care, and well-being. *Curr. Opin. Biotechnol.* 65, 118–128. doi:10.1016/j.copbio.2020.02.010

Bravo, L. (1998). Polyphenols: Chemistry, dietary sources, metabolism, and nutritional significance. *Nutr. Rev.* 56, 317–333. doi:10.1111/j.1753-4887.1998.tb01670.x

Brodowska, K. M. (2017). Natural flavonoids: Classification, potential role, and application of flavonoid analogues. *Eur. J. Biol. Res.* 7, 108–123. doi:10.5281/zenodo.545778

Caporgno, M. P., and Mathys, A. (2018). Trends in microalgae incorporation into innovative food products with potential health benefits. *Front. Nutr.* 5, 58. doi:10.3389/fnut.2018.00058

Castro, C. C., Pagnussat, A. S., Moura, N., da Cunha, M. J., Machado, F. R., Wyse, A. T. S., et al. (2014). Coumestrol treatment prevents Na+, K+-ATPase inhibition and affords histological neuroprotection to male rats receiving cerebral global ischemia. *Neurol. Res.* 36, 198–206. doi:10.1179/1743132813Y.0000000286

Chaudhuri, D., Ghate, N. B., Deb, S., Panja, S., Sarkar, R., Rout, J., et al. (2014). Assessment of the phytochemical constituents and antioxidant activity of a bloom forming microalgae Euglena tuba. *Biol. Res.* 47, 24. doi:10.1186/0717-6287-47-24

Chen, J., Yuan, H., Zhang, L., Pan, H., Xu, R., Zhong, Y., et al. (2015). Cloning, expression and purification of isoflavone-2′-hydroxylase from Astragalus membranaceus Bge. Var. mongolicus (Bge.) Hsiao. *Protein Expr. Purif.* 107, 83–89. doi:10.1016/j.pep.2014.11.010

Chen, T., Zhou, Y., Lu, Y., and Zhang, H. (2019). Advances in heterologous biosynthesis of plant and fungal natural products by modular co-culture engineering. *Biotechnol. Lett.* 41, 27–34. doi:10.1007/s10529-018-2619-z

Cheng, Q., Li, N., Dong, L., Zhang, D., Fan, S., Jiang, L., et al. (2015). Overexpression of soybean isoflavone reductase (GmIFR) enhances resistance to phytophthora sojae in soybean. *Front. Plant Sci.* 6, 1024. doi:10.3389/fpls.2015.01024

Choi, K. R., Jang, W. D., Yang, D., Cho, J. S., Park, D., and Lee, S. Y. (2019). Systems metabolic engineering strategies: Integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol.* 37, 817–837. doi:10.1016/j.tibtech.2019.01.003

Choong, W. P., Tan, C. H., Show, P. L., Lam, H. L., Mohamad Annuar, M. S. B., Juan, J. C., et al. (2016). Efficient enzyme-catalysed transesterification of microalgal biomass from Chlamydomonas sp. *Energy* 116, 1370–1373. doi:10.1016/j.energy.2016.06.032

Chouhan, S., Sharma, K., Zha, J., Guleria, S., and Koffas, M. A. G. (2017). Recent advances in the recombinant biosynthesis of polyphenols. *Front. Microbiol.* 8, 2259. doi:10.3389/fmicb.2017.02259

Cravens, A., Payne, J., and Smolke, C. D. (2019). Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.* 10, 2142. doi:10.1038/s41467-019-09848-w

Cruickshank, I. a. M., and Perrin, D. R. (1960). Isolation of a phytoalexin from Pisum sativum L. *Nature* 187, 799–800. doi:10.1038/187799b0

Curtis, T. P., and Sloan, W. T. (2005). Exploring microbial diversity--A vast below. *Science* 309, 1331–1333. doi:10.1126/science.1118176

Del Mondo, A., Sansone, C., and Brunet, C. (2022). Insights into the biosynthesis pathway of phenolic compounds in microalgae. *Comput. Struct. Biotechnol. J.* 20, 1901–1913. doi:10.1016/j.csbj.2022.04.019

Del Mondo, A., Smerilli, A., Ambrosino, L., Albini, A., Noonan, D. M., Sansone, C., et al. (2021). Insights into phenolic compounds from microalgae: Structural variety and complex beneficial activities from health to nutraceutics. *Crit. Rev. Biotechnol.* 41, 155–171. doi:10.1080/07388551.2021.1874284

Delépine, B., Duigou, T., Carbonell, P., and Faulon, J.-L. (2018). RetroPath2.0: A retrosynthesis workflow for metabolic engineers. *Metab. Eng.* 45, 158–170. doi:10.1016/j.ymben.2017.12.002

DiCenzo, G. L., and VanEtten, H. D. (2006). Studies on the late steps of (+) pisatin biosynthesis: Evidence for (-) enantiomeric intermediates. *Phytochemistry* 67, 675–683. doi:10.1016/j.phytochem.2005.12.027

Dixit, M., Raghuvanshi, A., Gupta, C. P., Kureel, J., Mansoori, M. N., Shukla, P., et al. (2015). Medicarpin, a natural pterocarpan, heals cortical bone defect by activation of notch and wnt canonical signaling pathways. *PLOS ONE* 10, e0144541. doi:10.1371/journal.pone.0144541

Dixon, R. A. (1999). "1.28 - isoflavonoids: Biochemistry, molecular biology, and biological functions," in *Comprehensive natural products chemistry*. Editors S. D. Barton, K. Nakanishi, and O. Meth-Cohn (Oxford: Pergamon), 773–823. doi:10.1016/B978-0-08-091283-7.00030-8

Du, H., Huang, Y., and Tang, Y. (2010). Genetic and metabolic engineering of isoflavonoid biosynthesis. *Appl. Microbiol. Biotechnol.* 86, 1293–1312. doi:10.1007/s00253-010-2512-8

Durazzo, A., Lucarini, M., Souto, E. B., Cicala, C., Caiazzo, E., Izzo, A. A., et al. (2019). Polyphenols: A concise overview on the chemistry, occurrence, and human health. *Phytother. Res. PTR* 33, 2221–2243. doi:10.1002/ptr.6419

Dutta, S., and Berman, H. M. (2005). Large macromolecular complexes in the protein Data Bank: A status report. *Structure* 13, 381–388. doi:10.1016/j.str.2005.01.008

Ellis, G. A., Klein, W. P., Lasarte-Aragonés, G., Thakur, M., Walper, S. A., and Medintz, I. L. (2019). Artificial multienzyme scaffolds: Pursuing *in vitro* substrate channeling with an overview of current progress. *ACS Catal.* 9, 10812–10869. doi:10.1021/acscatal.9b02413

Fan, J., Fu, A., and Zhang, L. (2019). Progress in molecular docking. *Quant. Biol.* 7, 83–89. doi:10.1007/s40484-019-0172-y

Ferdous, U. T., and Balia Yusof, Z. N. (2021). Insight into potential anticancer activity of algal flavonoids: Current status and challenges. *Molecules* 26, 6844. doi:10.3390/molecules26226844

Fischer, D., Ebenau-Jehle, C., and Grisebach, H. (1990). Purification and characterization of pterocarpan synthase from elicitor-challenged soybean cell cultures. *Phytochemistry* 29, 2879–2882. doi:10.1016/0031-9422(90)87096-D

French, K. E. (2019). Harnessing synthetic biology for sustainable development. *Nat. Sustain.* 2, 250–252. doi:10.1038/s41893-019-0270-x

Goel, A., Kumar, A., and Raghuvanshi, A. (2013). Synthesis, stereochemistry, structural classification, and chemical reactivity of natural pterocarpans. *Chem. Rev.* 113, 1614–1640. doi:10.1021/cr300219y

Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paepe, D., E Baart, G. J., et al. (2014). Detection of flavonoids in microalgae from different evolutionary lineages. *J. Phycol.* 50, 483–492. doi:10.1111/jpy.12180

Gong, X., Zhang, T., Xing, J., Wang, R., and Zhao, Y. (2020). Positional effects on efficiency of CRISPR/Cas9-based transcriptional activation in rice plants. *aBIOTECH* 1, 1–5. doi:10.1007/s42994-019-00007-9

Goswami, R. K., Mehariya, S., Karthikeyan, O. P., Gupta, V. K., and Verma, P. (2022). Multifaceted application of microalgal biomass integrated with carbon dioxide reduction and wastewater remediation: A flexible concept for sustainable environment. *J. Clean. Prod.* 339, 130654. doi:10.1016/j.jclepro.2022.130654

Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., et al. (2012). The genome portal of the department of energy Joint genome Institute. *Nucleic Acids Res.* 40, D26–D32. doi:10.1093/nar/gkr947

Guo, L., Dixon, R. A., and Paiva, N. L. (1994a). Conversion of vestitone to medicarpin in alfalfa (Medicago sativa L.) is catalyzed by two independent enzymes. Identification, purification, and characterization of vestitone reductase and 7,2'-dihydroxy-4'-methoxyisoflavanol dehydratase. *J. Biol. Chem.* 269, 22372–22378. doi:10.1016/s0021-9258(17)31799-4

Guo, L., Dixon, R. A., and Paiva, N. L. (1994b). The 'pterocarpan synthase" of alfalfa: Association and co-induction of vestitone reductase and 7,2′-dihydroxy-4′-methoxyisoflavanol (DMI) dehydratase, the two final enzymes in medicarpin biosynthesis. *FEBS Lett.* 356, 221–225. doi:10.1016/0014-5793(94)01267-9

Guo, L., and Paiva, N. L. (1995). Molecular cloning and expression of alfalfa (Medicago sativa L.) vestitone reductase, the penultimate enzyme in medicarpin biosynthesis. *Arch. Biochem. Biophys.* 320, 353–360. doi:10.1016/0003-9861(95)90019-5

Gupta, A., Awasthi, P., Sharma, N., Parveen, S., Vats, R. P., Singh, N., et al. (2022). Medicarpin confers powdery mildew resistance in Medicago truncatula and activates the salicylic acid signalling pathway. *Mol. Plant Pathol.* 23, 966–983. doi:10.1111/mpp.13202

Ha, J., Kang, Y.-G., Lee, T., Kim, M., Yoon, M. Y., Lee, E., et al. (2019). Comprehensive RNA sequencing and co-expression network analysis to complete the biosynthetic pathway of coumestrol, a phytoestrogen. *Sci. Rep.* 9, 1934. doi:10.1038/s41598-018-38219-6

Harvey, P. J., and Ben-Amotz, A. (2020). Towards a sustainable Dunaliella salina microalgal biorefinery for 9-cis β-carotene production. *Algal Res.* 50, 102002. doi:10.1016/j.algal.2020.102002

He, X., Fang, J., Huang, L., Wang, J., and Huang, X. (2015). Sophora flavescens Ait. Traditional usage, phytochemistry and pharmacology of an important traditional Chinese medicine. *J. Ethnopharmacol.* 172, 10–29. doi:10.1016/j.jep.2015.06.010

Hedelin, M., Löf, M., Olsson, M., Adlercreutz, H., Sandin, S., and Weiderpass, E. (2008). Dietary phytoestrogens are not associated with risk of overall breast cancer but diets rich in coumestrol are inversely associated with risk of estrogen receptor and progesterone receptor negative breast tumors in Swedish women. *J. Nutr.* 138, 938–945. doi:10.1093/jn/138.5.938

Hedlund, B. P., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter. *Extremophiles* 18, 865–875. doi:10.1007/s00792-014-0664-7

Hemaiswarya, S., Raja, R., Ravikumar, R., and Carvalho, I. (2013). Microalgae taxonomy and breeding. *Biofuel Crops Prod. Physiol. Genet.*, 44–53.

Hinderer, W., Flentje, U., and Barz, W. (1987). Microsomal isoflavone 2′- and 3′-hydroxylases from chickpea (Cicer arietinum L.) cell suspensions induced for pterocarpan phytoalexin formation. *FEBS Lett.* 214, 101–106. doi:10.1016/0014-5793(87)80021-2

Hsieh, C.-J., Kuo, P.-L., Hou, M.-F., Hung, J.-Y., Chang, F.-R., Hsu, Y.-C., et al. (2015). Wedelolactone inhibits breast cancer-induced osteoclastogenesis by decreasing Akt/mTOR signaling. *Int. J. Oncol.* 46, 555–562. doi:10.3892/ijo.2014.2769

Ijaz, S., and Hasnain, S. (2016). Antioxidant potential of indigenous cyanobacterial strains in relation with their phenolic and flavonoid contents. *Nat. Prod. Res.* 30, 1297–1300. doi:10.1080/14786419.2015.1053088

Jenuth, J. P. (1999). "The NCBI," in Bioinformatics Methods and protocols *methods in molecular Biology™*. Editors S. Misener and S. A. Krawetz (Totowa, NJ: Humana Press), 301–312. doi:10.1385/1-59259-192-2:301

Jiang, D., Rasul, A., Batool, R., Sarfraz, I., Hussain, G., Mateen Tahir, M., et al. (2019). Potential anticancer properties and mechanisms of action of formononetin. *Biomed. Res. Int.* 2019, 1–11. doi:10.1155/2019/5854315

Jiménez-González, L., Álvarez-Corral, M., Muñoz-Dorado, M., and Rodríguez-García, I. (2008). Pterocarpans: Interesting natural products with antifungal activity and other biological properties. *Phytochem. Rev.* 7, 125–154. doi:10.1007/s11101-007-9059-z

Jones, J. A., Vernacchio, V. R., Collins, S. M., Shirke, A. N., Xiu, Y., Englaender, J. A., et al. (2017). Complete biosynthesis of anthocyanins using *E. coli* polycultures. *mBio* 8, e00621–17. doi:10.1128/mBio.00621-17

Jones, J. A., Vernacchio, V. R., Sinkoe, A. L., Collins, S. M., Ibrahim, M. H. A., Lachance, D. M., et al. (2016). Experimental and computational optimization of an *Escherichia coli* co-culture for the efficient production of flavonoids. *Metab. Eng.* 35, 55–63. doi:10.1016/j.ymben.2016.01.006

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Kamble, A., Srinivasan, S., and Singh, H. (2019). *In-silico* bioprospecting: Finding better enzymes. *Mol. Biotechnol.* 61, 53–59. doi:10.1007/s12033-018-0132-1

Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kang, S.-Y., Heo, K. T., and Hong, Y.-S. (2018). Optimization of artificial curcumin biosynthesis in *E. coli* by randomized 5′-UTR sequences to control the multienzyme pathway. *ACS Synth. Biol.* 7, 2054–2062. doi:10.1021/acssynbio.8b00198

Katsuyama, Y., Miyahisa, I., Funa, N., and Horinouchi, S. (2007). One-pot synthesis of genistein from tyrosine by coincubation of genetically engineered *Escherichia coli* and *Saccharomyces cerevisiae* cells. *Appl. Microbiol. Biotechnol.* 73, 1143–1149. doi:10.1007/s00253-006-0568-2

Khemiri, S., Khelifi, N., Nunes, M. C., Ferreira, A., Gouveia, L., Smaali, I., et al. (2020). Microalgae biomass as an additional ingredient of gluten-free bread: Dough rheology, texture quality and nutritional properties. *Algal Res.* 50, 101998. doi:10.1016/j.algal.2020.101998

Kim, B.-G. (2020). Biological synthesis of genistein in *Escherichia coli*. *J. Microbiol. Biotechnol.* 30, 770–776. doi:10.4014/jmb.1911.11009

Kim, J.-H., Kang, D.-M., Cho, Y.-J., Hyun, J.-W., and Ahn, M.-J. (2022). Medicarpin increases antioxidant genes by inducing NRF2 transcriptional level in HeLa cells. *Antioxidants* 11, 421. doi:10.3390/antiox11020421

Klejdus, B., Lojková, L., Plaza, M., Snóblová, M., and Stěrbová, D. (2010). Hyphenated technique for the extraction and determination of isoflavones in algae: Ultrasound-assisted supercritical fluid extraction followed by fast chromatography with tandem mass spectrometry. *J. Chromatogr. A* 1217, 7956–7965. doi:10.1016/j.chroma.2010.07.020

Koukos, P. I., Réau, M., and Bonvin, A. M. J. J. (2021). Shape-restrained modeling of protein–small-molecule complexes with high ambiguity driven DOCKing. *J. Chem. Inf. Model.* 61, 4807–4818. doi:10.1021/acs.jcim.1c00796

Krüger, A., Schäfers, C., Busch, P., and Antranikian, G. (2020). Digitalization in microbiology - paving the path to sustainable circular bioeconomy. *New Biotechnol.* 59, 88–96. doi:10.1016/j.nbt.2020.06.004

Kumar, S., and Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: An overview. *Sci. World J.* 2013, 1–16. doi:10.1155/2013/162750

Lee, E. J., Song, M. C., and Rha, C.-S. (2022). Mass biosynthesis of coumestrol derivatives and their isomers via soybean adventitious root cultivation in bioreactors. *Front. Plant Sci.* 13, 923163. doi:10.3389/fpls.2022.923163

Lee, T.-M., Lin, J.-Y., Tsai, T.-H., Yang, R.-Y., and Ng, I.-S. (2023). Clustered regularly interspaced short palindromic repeats (crispr) technology and genetic engineering strategies for microalgae towards carbon neutrality: A critical review. *Bioresour. Technol.* 368, 128350. doi:10.1016/j.biortech.2022.128350

Leonard, E., and Koffas, M. A. G. (2007). Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Appl. Environ. Microbiol.* 73, 7246–7251. doi:10.1128/AEM.01411-07

Li, D., Cai, C., Liao, Y., Wu, Q., Ke, H., Guo, P., et al. (2021). Systems pharmacology approach uncovers the therapeutic mechanism of medicarpin against scopolamine-induced memory loss. *Phytomedicine* 91, 153662. doi:10.1016/j.phymed.2021.153662

Li, J., Li, C., Gou, J., Wang, X., Fan, R., and Zhang, Y. (2016). An alternative pathway for formononetin biosynthesis in Pueraria lobata. *Front. Plant Sci.* 7, 861. doi:10.3389/fpls.2016.00861

Liu, C.-J., Deavours, B. E., Richard, S. B., Ferrer, J.-L., Blount, J. W., Huhman, D., et al. (2006). Structural basis for dual functionality of isoflavonoid O-methyltransferases in the evolution of plant defense responses. *Plant Cell* 18, 3656–3669. doi:10.1105/tpc.106.041376

Liu, C.-J., Huhman, D., Sumner, L. W., and Dixon, R. A. (2003). Regiospecific hydroxylation of isoflavones by cytochrome P450 81E enzymes from Medicago truncatula. *Plant J.* 36, 471–484. doi:10.1046/j.1365-313X.2003.01893.x

Liu, L., Liu, H., Zhang, W., Yao, M., Li, B., Liu, D., et al. (2019). Engineering the biosynthesis of caffeic acid in *Saccharomyces cerevisiae* with heterologous enzyme combinations. *Engineering* 5, 287–295. doi:10.1016/j.eng.2018.11.029

Liu, Q., Liu, Y., Li, G., Savolainen, O., Chen, Y., and Nielsen, J. (2021). De novo biosynthesis of bioactive isoflavonoids by engineered yeast cell factories. *Nat. Commun.* 12, 6085. doi:10.1038/s41467-021-26361-1

Liu, Y., Grimm, M., Dai, W., Hou, M., Xiao, Z.-X., and Cao, Y. (2020a). CB-dock: A web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacol. Sin.* 41, 138–144. doi:10.1038/s41401-019-0228-6

Liu, Z., Wang, K., Chen, Y., Tan, T., and Nielsen, J. (2020b). Third-generation biorefineries as the means to produce fuels and chemicals from CO2. *Nat. Catal.* 3, 274–288. doi:10.1038/s41929-019-0421-5

Lok, C. (2015). Mining the microbial dark matter. *Nature* 522, 270–273. doi:10.1038/522270a

Lyu, X., Ng, K. R., Lee, J. L., Mark, R., and Chen, W. N. (2017). Enhancement of naringenin biosynthesis from tyrosine by metabolic engineering of *Saccharomyces cerevisiae*. *J. Agric. Food Chem.* 65, 6638–6646. doi:10.1021/acs.jafc.7b02507

Mazur, W. M., Duke, J. A., Wähälä, K., Rasku, S., and Adlercreutz, H. (1998). Isoflavonoids and Lignans in Legumes: Nutritional and Health Aspects in Humans 11The method development and synthesis of the standards and deuterium-labelled compounds was supported by National Institutes of Health Grants No. 1 R01 CA56289-01 and No. 2 R01 CA56289-04, and analytical work by the EU research contract FAIR-CT95-0894. *J. Nutr. Biochem.* 9, 193–200. doi:10.1016/S0955-2863(97)00184-8

Meng, Q., Moinuddin, S. G. A., Kim, S.-J., Bedgar, D. L., Costa, M. A., Thomas, D. G., et al. (2020). Pterocarpan synthase (PTS) structures suggest a common quinone methide–stabilizing function in dirigent proteins and proteins with dirigent-like domains. *J. Biol. Chem.* 295, 11584–11601. doi:10.1074/jbc.RA120.012444

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi:10.1038/s41592-022-01488-1

Miyahisa, I., Kaneko, M., Funa, N., Kawasaki, H., Kojima, H., Ohnishi, Y., et al. (2005). Efficient production of (2S)-flavanones by *Escherichia coli* containing an artificial biosynthetic gene cluster. *Appl. Microbiol. Biotechnol.* 68, 498–504. doi:10.1007/s00253-005-1916-3

Mizuguchi, H., Nariai, Y., Kato, S., Nakano, T., Kanayama, T., Kashiwada, Y., et al. (2015). Maackiain is a novel antiallergic compound that suppresses transcriptional upregulation of the histamine H1 receptor and interleukin-4 genes. *Pharmacol. Res. Perspect.* 3, e00166. doi:10.1002/prp2.166

Mobin, S., and Alam, F. (2017). Some promising microalgal species for commercial applications: A review. *Energy Procedia* 110, 510–517. doi:10.1016/j.egypro.2017.03.177

Nabavi, S. M., Šamec, D., Tomczyk, M., Milella, L., Russo, D., Habtemariam, S., et al. (2020). Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. *Biotechnol. Adv.* 38, 107316. doi:10.1016/j.biotechadv.2018.11.005

Nguyen, T. M. N., Nguyen, T. K. A., and Vu, T. K. L. (2020). Symbiotic photobioreactor using immobilized microalgae-yeast consortium for saccharomyces cerevisiae and chlorella vulgaris biomass production. *J. Tech. Educ. Sci.*, 17–29.

Paiva, N. L., Edwards, R., Sun, Y. J., Hrazdina, G., and Dixon, R. A. (1991). Stress responses in alfalfa (Medicago sativa L.) 11. Molecular cloning and expression of alfalfa isoflavone reductase, a key enzyme of isoflavonoid phytoalexin biosynthesis. *Plant Mol. Biol.* 17, 653–667. doi:10.1007/BF00037051

Paoli, L., Ruscheweyh, H.-J., Forneris, C. C., Hubrich, F., Kautsar, S., Bhushan, A., et al. (2022). Biosynthetic potential of the global ocean microbiome. *Nature* 607, 111–118. doi:10.1038/s41586-022-04862-3

Park, G., Baek, S., Kim, J.-E., Lim, T., Lee, C. C., Yang, H., et al. (2015). Flt3 is a target of coumestrol in protecting against UVB-induced skin photoaging. *Biochem. Pharmacol.* 98, 473–483. doi:10.1016/j.bcp.2015.08.104

Peil, G. H. S., Kuss, A. V., Rave, A. F. G., Villarreal, J. P. V., Hernandes, Y. M. L., and Nascente, P. S. (2016). Bioprospecting of lipolytic microorganisms obtained from industrial effluents. *An. Acad. Bras. Ciênc.* 88, 1769–1779. doi:10.1590/0001-3765201620150550

Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., et al. (2019a). A review of the microbial production of bioactive natural products and biologics. *Front. Microbiol.* 10, 1404. doi:10.3389/fmicb.2019.01404

Pham, T. H., Lecomte, S., Efstathiou, T., Ferriere, F., and Pakdel, F. (2019b). An update on the effects of glyceollins on human health: Possible anticancer effects and underlying mechanisms. *Nutrients* 11, 79. doi:10.3390/nu11010079

Prasad, R., Gupta, S. K., Shabnam, N., Oliveira, C. Y. B., Nema, A. K., Ansari, F. A., et al. (2021). Role of microalgae in global CO2 sequestration: Physiological mechanism, recent development, challenges, and future prospective. *Sustainability* 13, 13061. doi:10.3390/su132313061

Procházková, T., Sychrová, E., Javůrková, B., Večerková, J., Kohoutek, J., Lepšová-Skácelová, O., et al. (2017). Phytoestrogens and sterols in waters with cyanobacterial blooms - analytical methods and estrogenic potencies. *Chemosphere* 170, 104–112. doi:10.1016/j.chemosphere.2016.12.006

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. doi:10.1038/nbt.3935

Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., and Nielsen, J. (2015). Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. *Metab. Eng.* 31, 181–188. doi:10.1016/j.ymben.2015.08.003

Rodriguez, A., Strucko, T., Stahlhut, S. G., Kristensen, M., Svenssen, D. K., Forster, J., et al. (2017). Metabolic engineering of yeast for fermentative production of flavonoids. *Bioresour. Technol.* 245, 1645–1654. doi:10.1016/j.biortech.2017.06.043

Safafar, H., van Wagenen, J., Møller, P., and Jacobsen, C. (2015). Carotenoids, phenolic compounds and tocopherols contribute to the antioxidative properties of some microalgae species grown on industrial wastewater. *Mar. Drugs* 13, 7339–7356. doi:10.3390/md13127069

Sajid, M., Channakesavula, C., Stone, S., and Kaur, P. (2021a). Synthetic biology towards improved flavonoid pharmacokinetics. *Biomolecules* 11, 754. doi:10.3390/biom11050754

Sajid, M., Stone, S., and Kaur, P. (2021b). Recent advances in heterologous synthesis paving way for future green-modular bioindustries: A review with special reference to isoflavonoids. *Front. Bioeng. Biotechnol.* 9, 673270. doi:10.3389/fbioe.2021.673270

Sajid, M., Stone, S. R., and Kaur, P. (2022). Phylogenetic analysis and protein modelling of isoflavonoid synthase highlights key catalytic sites towards realising new bioengineering endeavours. *Bioengineering* 9, 609. doi:10.3390/bioengineering9110609

Sankaran, R., Show, P. L., Yap, Y. J., Tao, Y., Ling, T. C., and Tomohisa, K. (2018). Green technology of liquid biphasic flotation for enzyme recovery utilizing recycling surfactant and sorbitol. *Clean. Technol. Environ. Policy* 20, 2001–2012. doi:10.1007/s10098-018-1523-5

Santana, K., do Nascimento, L. D., Lima e Lima, A., Damasceno, V., Nahum, C., Braga, R. C., et al. (2021). Applications of virtual screening in bioprospecting: Facts, shifts, and perspectives to explore the chemo-structural diversity of natural products. *Front. Chem.* 9, 662688. doi:10.3389/fchem.2021.662688

Schaller, K. S., Molina, G. A., Kari, J., Schiano-di-Cola, C., Sørensen, T. H., Borch, K., et al. (2022). Virtual bioprospecting of interfacial enzymes: Relating sequence and kinetics. *ACS Catal.* 12, 7427–7435. doi:10.1021/acscatal.2c02305

Seeliger, D., and de Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.* 24, 417–422. doi:10.1007/s10822-010-9352-6

Shao, H., Dixon, R. A., and Wang, X. (2007). Crystal structure of vestitone reductase from alfalfa (Medicago sativa L.) *J. Mol. Biol.* 369, 265–276. doi:10.1016/j.jmb.2007.03.040

Sharifi-Rad, J., Kamiloglu, S., Yeskaliyeva, B., Beyatli, A., Alfred, M. A., Salehi, B., et al. (2020). Pharmacological activities of psoralidin: A comprehensive review of the molecular mechanisms of action. *Front. Pharmacol.* 11, 571459. doi:10.3389/fphar.2020.571459

Sharma, S., Trivedi, S., Pandey, T., Ranjan, S., Trivedi, M., and Pandey, R. (2021). Wedelolactone mitigates parkinsonism via alleviating oxidative stress and mitochondrial dysfunction through NRF2/SKN-1. *Mol. Neurobiol.* 58, 65–77. doi:10.1007/s12035-020-02080-4

Singh, N., and Gaur, S. (2021). "GRAS fungi: A new horizon in safer food product," in *Fungi in sustainable food production fungal biology*. Editors X. Dai, M. Sharma, and J. Chen (Cham: Springer International Publishing), 27–37. doi:10.1007/978-3-030-64406-2_3

Strange, R. N., Ingham, J. L., Cole, D. L., Cavill, M. E., Edwards, C., Cooksey, C. J., et al. (1985). Isolation of the phytoalexin medicarpin from leaflets of *Arachis hypogaea* and related species of the tribe aeschynomeneae. *Z. Für Naturforsch. C* 40, 313–316. doi:10.1515/znc-1985-5-605

Sukumaran, A., McDowell, T., Chen, L., Renaud, J., and Dhaubhadel, S. (2018). Isoflavonoid-specific prenyltransferase gene family in soybean: GmPT01, a pterocarpan 2-dimethylallyltransferase involved in glyceollin biosynthesis. *Plant J.* 96, 966–981. doi:10.1111/tpj.14083

Teng, H., and Chen, L. (2019). Polyphenols and bioavailability: An update. *Crit. Rev. Food Sci. Nutr.* 59, 2040–2051. doi:10.1080/10408398.2018.1437023

The UniProt Consortium (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi:10.1093/nar/gky1049

Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.* 48, 123–152. doi:10.3109/10409238.2012.758083

Trabelsi, L., Mnari, A., Abdel-Daim, M. M., Abid-Essafi, S., and Aleya, L. (2016). Therapeutic properties in Tunisian hot springs: First evidence of phenolic compounds in the cyanobacterium leptolyngbya sp. biomass, capsular polysaccharides and releasing polysaccharides. *BMC Complement. Altern. Med.* 16, 515. doi:10.1186/s12906-016-1492-3

Trantas, E., Panopoulos, N., and Ververidis, F. (2009). Metabolic engineering of the complete pathway leading to heterologous biosynthesis of various flavonoids and stilbenoids in *Saccharomyces cerevisiae*. *Metab. Eng.* 11, 355–366. doi:10.1016/j.ymben.2009.07.004

Tsai, R.-T., Tsai, C.-W., Liu, S.-P., Gao, J.-X., Kuo, Y.-H., Chao, P.-M., et al. (2020). Maackiain ameliorates 6-hydroxydopamine and SNCA pathologies by modulating the PINK1/parkin pathway in models of Parkinson's disease in *Caenorhabditis elegans* and the SH-SY5Y cell line. *Int. J. Mol. Sci.* 21, 4455. doi:10.3390/ijms21124455

Tu, Y., Yang, Y., Li, Y., and He, C. (2021). Naturally occurring coumestans from plants, their biological activities and therapeutic effects on human diseases. *Pharmacol. Res.* 169, 105615. doi:10.1016/j.phrs.2021.105615

Twaij, B. M., and Hasan, M. N. (2022). Bioactive secondary metabolites from plant sources: Types, synthesis, and their therapeutic uses. *Int. J. Plant Biol.* 13, 4–14. doi:10.3390/ijpb13010003

Uchida, K., Akashi, T., and Aoki, T. (2015). Functional expression of cytochrome P450 in *Escherichia coli*: An approach to functional analysis of uncharacterized enzymes for flavonoid biosynthesis. *Plant Biotechnol.* 32, 205–213. doi:10.5511/plantbiotechnology.15.0605a

Uchida, K., Akashi, T., and Aoki, T. (2017). The missing link in leguminous pterocarpan biosynthesis is a dirigent domain-containing protein with isoflavanol dehydratase activity. *Plant Cell Physiol.* 58, 398–408. doi:10.1093/pcp/pcw213

Uchida, K., Aoki, T., Suzuki, H., and Akashi, T. (2020). Molecular cloning and biochemical characterization of isoflav-3-ene synthase, a key enzyme of the biosyntheses of (+)-pisatin and coumestrol. Plant Biotechnol. *advpub* 20, 301–310. doi:10.5511/plantbiotechnology.20.0421a

Vavitsas, K., Kugler, A., Satta, A., Hatzinikolaou, D. G., Lindblad, P., Fewer, D. P., et al. (2021). Doing synthetic biology with photosynthetic microorganisms. *Physiol. Plant.* 173, 624–638. doi:10.1111/ppl.13455

Villarruel-Lopez, A., Ascencio, F., and Nuño, K. (2017). Microalgae, a potential natural functional food source- A review. *Pol. J. Food Nutr. Sci.* 67, 251–263. doi:10.1515/pjfns-2017-0017

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Phenylpropanoid Biosynth.* Mol. Plant 3, 2–20. doi:10.1093/mp/ssp106

Vuong, P., Chong, S., and Kaur, P. (2022a). The little things that matter: How bioprospecting microbial biodiversity can build towards the realization of united Nations sustainable development Goals. *Npj Biodivers.* 1, 4–5. doi:10.1038/s44185-022-00006-y

Vuong, P., Wise, M. J., Whiteley, A. S., and Kaur, P. (2022b). Small investments with big returns: Environmental genomic bioprospecting of microbial life. *Crit. Rev. Microbiol.* 48, 641–655. doi:10.1080/1040841X.2021.2011833

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898. doi:10.1038/nature08187

Wang, Z., Huang, X., Liu, J., Xiao, F., Tian, M., Ding, S., et al. (2021). Screening and heterologous expression of flavone synthase and flavonol synthase to catalyze hesperetin to diosmetin. *Biotechnol. Lett.* 43, 2161–2183. doi:10.1007/s10529-021-03184-0

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427

Welle, R., and Grisebach, H. (1988). Induction of phytoalexin synthesis in soybean: Enzymatic cyclization of prenylated pterocarpans to glyceollin isomers. *Arch. Biochem. Biophys.* 263, 191–198. doi:10.1016/0003-9861(88)90627-3

Wells, M. L., Potin, P., Craigie, J. S., Raven, J. A., Merchant, S. S., Helliwell, K. E., et al. (2017). Algae as nutritional and functional food sources: Revisiting our understanding. *J. Appl. Phycol.* 29, 949–982. doi:10.1007/s10811-016-0974-5

Wen, L., Jiang, Y., Yang, J., Zhao, Y., Tian, M., and Yang, B. (2017). Structure, bioactivity, and synthesis of methylated flavonoids. *Ann. N. Y. Acad. Sci.* 1398, 120–129. doi:10.1111/nyas.13350

Wu, G., Zhuang, D., Chew, K. W., Ling, T. C., Khoo, K. S., Van Quyen, D., et al. (2022). Current status and future trends in removal, control, and mitigation of algae food safety risks for human consumption. *Mol. Basel Switz.* 27, 6633. doi:10.3390/molecules27196633

Wu, Q., Preisig, C. L., and VanEtten, H. D. (1997). Isolation of the cDNAs encoding (+)6a-hydroxymaackiain 3-O-methyltransferase, the terminal step for the synthesis of the phytoalexin pisatin in Pisum sativum. *Plant Mol. Biol.* 35, 551–560. doi:10.1023/a:1005836508844

Yadavalli, R., Ratnapuram, H., Motamarry, S., Reddy, C. N., Ashokkumar, V., and Kuppam, C. (2022). Simultaneous production of flavonoids and lipids from Chlorella vulgaris and Chlorella pyrenoidosa. *Biomass Convers. Biorefinery* 12, 683–691. doi:10.1007/s13399-020-01044-x

Yamamoto, T., Sakamoto, C., Tachiwana, H., Kumabe, M., Matsui, T., Yamashita, T., et al. (2018). Endocrine therapy-resistant breast cancer model cells are inhibited by soybean glyceollin I through Eleanor non-coding RNA. *Sci. Rep.* 8, 15202. doi:10.1038/s41598-018-33227-y

Yan, Y., Huang, L., and Koffas, M. A. G. (2007). Biosynthesis of 5-deoxyflavanones in microorganisms. *Biotechnol. J.* 2, 1250–1262. doi:10.1002/biot.200700119

Yang, X., Zhao, Y., Hsieh, M.-T., Xin, G., Wu, R.-T., Hsu, P.-L., et al. (2017). Total synthesis of (+)-Medicarpin. *J. Nat. Prod.* 80, 3284–3288. doi:10.1021/acs.jnatprod.7b00741

Yonekura-Sakakibara, K., Higashi, Y., and Nakabayashi, R. (2019). The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* 10, 943. doi:10.3389/fpls.2019.00943

Zampieri, G., Vijayakumar, S., Yaneske, E., and Angione, C. (2019). Machine and deep learning meet genome-scale metabolic modeling. *PLOS Comput. Biol.* 15, e1007084. doi:10.1371/journal.pcbi.1007084

Zhang, W., Liu, H., Li, X., Liu, D., Dong, X.-T., Li, F.-F., et al. (2017). Production of naringenin from D-xylose with co-culture of *E. coli* and *S. cerevisiae*. *Eng. Life Sci.* 17, 1021–1029. doi:10.1002/elsc.201700039

Zhu, M., Wang, L., Yang, D., Li, C., Pang, S., Li, X., et al. (2019). Wedelolactone alleviates doxorubicin-induced inflammation and oxidative stress damage of podocytes by IκK/IκB/NF-κB pathway. *Biomed. Pharmacother.* 117, 109088. doi:10.1016/j.biopha.2019.109088

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 19 May 2023
DOI 10.3389/fbioe.2023.1085797

OPEN ACCESS

EDITED BY
Luan Luong Chu,
Phenikaa University, Vietnam

REVIEWED BY
Weiwen Zhang,
Tianjin University, China
Ben Jacob Novak,
Revive and Restore, United States

*CORRESPONDENCE
Dong Wang,
wang0817@mail.ustc.edu.cn
Fahad Asmi,
fasmie@ustc.edu.cn

RECEIVED 31 October 2022
ACCEPTED 02 May 2023
PUBLISHED 19 May 2023

CITATION
Wang G, Kong Q, Wang D and Asmi F (2023), Ethical and social insights into synthetic biology: predicting research fronts in the post-COVID-19 era. *Front. Bioeng. Biotechnol.* 11:1085797. doi: 10.3389/fbioe.2023.1085797

Ethical and social insights into synthetic biology: predicting research fronts in the post-COVID-19 era

Gaofeng Wang, Qingqing Kong, Dong Wang* and Fahad Asmi*

School of Humanities and Social Sciences, University of Science and Technology of China, Hefei, Anhui Province, China

As a revolutionary biological science and technology, synthetic biology has already spread its influence from natural sciences to humanities and social sciences by introducing biosafety, biosecurity, and ethical issues to society. The current study aims to elaborate the intellectual bases and research front of the synthetic biology field in the sphere of philosophy, ethics, and social sciences, with knowledge mapping and bibliometric methods. The literature records from the Social Sciences Citation Index and Arts & Humanities Citation Index in the Web of Science Core Collection from 1982 to 2021 were collected and analyzed to illustrate the intellectual structure of philosophical, ethical, and social research of synthetic biology. This study profiled the hotspots of research focus on its governance, philosophical and ethical concerns, and relevant technologies. This study offers clues and enlightenment for the stakeholders and researchers to follow the progress of this emerging discipline and technology and to understand the cutting-edge ideas and future form of this field, which takes on greater significance in the post-COVID-19 era.

KEYWORDS

synthetic biology, social sciences, bioethical concerns, playing God, biosecurity, biosafety

1 Introduction

Life sciences experienced three major revolutions since the 20th century (Sharp et al., 2011), the first revolution featuring cellular and molecular biology began with the discovery of the DNA double-helix structure in the 1950s, the second is the genomic revolution, beginning with the exploration of the entire genome of organisms after 2000, and the third is convergence revolution marked by the merging of life science, physical sciences, and engineering (National Research Council, 2014). Synthetic biology (SynBio) serves as one of the most representative convergent disciplines following the idea aforementioned. It has evolved over a huge spectrum of the interdisciplinary continuum which includes life-science branches, such as biochemistry, microbiology, molecular biology, systematic biology, and non-life-science branches like computer sciences and engineering sciences.

Due to its abundant upstream knowledge sources from these disciplines, SynBio has been understood and practiced divergently from the perspective of the respective discipline. For scientists in the bioengineering field, SynBio is defined as "the engineering-driven building of increasingly complex biological entities for novel applications" (Heinemann and Panke, 2006). Luis Serrano (2007) put forward that a group of European experts defined SynBio together as "the engineering of biology: the synthesis of complex, biologically

based (or inspired) systems, which display functions that do not exist in nature," and distinguished SynBio from systems biology by "engineering" and "synthesis of novel functions." Meanwhile, in the conceptual framework of Gilbert et al. (2010), there are two approaches in SynBio research, the top–bottom approach stressing the function modification of existing cells and the bottom–up approach, which is "interested in the construction of artificial systems" like protocells from a chemical perspective. SynBio in the chemical context is usually emphasized as a science providing microbial chassis and biological devices to explore the interaction mechanism of natural products and yield them which sometimes serves as part of the chemical synthesis, i.e., synthetic chemistry (Keasling, 2008; Goss et al., 2012). Gómez-Tatay and Hernández-Andreu (2019) also stressed the constructive and non-naturally existing feature of SynBio products with the goal that "aims toward the creation of something fundamentally new, biological parts or systems not otherwise found in nature," which is maintained with a bottom–up strategy.

Above all, most definitions emphasized both the construction of novel biological entities and the modification of existing ones, and a group of leading scientists summarized it as "the design and construction of novel biological parts, devices, and systems, as well as the redesign of existing natural biological systems, for useful purposes" (Calvert, 2010). This definition distinguishes SynBio from other biological disciplines that obtained tremendous scientific gains by exploring the structure and function of naturally existing organisms and entities with a "top–down" strategy locally and deconstructively with a disparate moral status from SynBio, like biochemistry, systematic biology, genetic engineering, and its genetic modification technologies. SynBio emerged as a new concern by its novelty of productions which does not exist naturally or previously, causing uncertainties in ethical issues and biosafety and biosecurity concerns. Compared to the enormous social and ethical discussions on the "top–down" life-science sub-issues like genetically modified organisms (GMOs) (Dong et al., 2019; Nawaz et al., 2019; Zhou et al., 2019), SynBio serves as a novel and probably crucial issue for sociologists, ethicists, and stakeholders. The world has witnessed many breakthroughs in various fields, such as biomedicine and the energy industry with SynBio's support (Martin et al., 2003; Lindahl et al., 2006). SynBio is used to design DNA-based information storage and communication and processing systems for genetic coding (Cello et al., 2002; Weber and Fussenegger, 2012). In industry, engineered microbial cells were used for biofuel production (Radakovits et al., 2010). In addition to all the advances in the laboratory, additional applications and commercially available products have been made including the chemicals produced by engineered cells and the engineered cells themselves (Voigt, 2020) and more possibilities await. Generally, SynBio appears to be a game-changer for its prospective tremendous potential.

However, despite all of the findings and commercial products SynBio brings to human beings, it also comes with risks and concerns as dual-use research. SynBio's products are non-natural artificial life with an endogenous uncertainty. Before its emergence, human beings have no experience or knowledge of its existence and nature, and this product itself may have the possibility of self-reproduction and genetic evolution, which could magnify the uncertainty dramatically. This uncertainty itself implies biosafety and biosecurity risks that could cause harm to human beings, organisms, or living environment and leads to psychological and cognitive risks at the individual and social levels, as well as other significant impacts on human society. Therefore, SynBio has also become the research object in humanities and social sciences, attracting wide attention from researchers and stakeholders in philosophy and ethics, policy and laws, science communication and public understanding, and intellectual property. As ethical concerns, SynBio prompts us to think about the intrinsic value of life (Link, 2013), the dignity of life, the integrity of nature, and the relationship between God and His creation (Heavey, 2013). Meanwhile, the novelty of SynBio undermined the feasibility of previous policies or laws for biological or bioengineering practices and appealed for new governance methods and ethical regulations. Some researchers are concerned about SynBio from the public point of view and tested the relationship with more information and deliberation and public support (Kronberger et al., 2012). Moreover, the intellectual property and commercialization of SynBio have also become important research areas and issues for sociologists (Calvert, 2008). All these demonstrate that SynBio has had a significant impact on humanities, featuring philosophy and ethics regarding SynBio and social science research, and would further spread to human society globally.

Meanwhile, in the context of the significant influence of global pandemics such as COVID-19, H1N1, SARS, and MERS on people (Lakoff and Collier, 2008), the public's concern about the

TABLE 1 Search strategy.

Step	Operation	Result
1	Search in the SSCI and AHCI based on the search strategy offered by Shapira et al.'s (2017) and check the search results	Found many irrelevant records due to improper searching terms such as "artificial * cell phone"
2	Search the terms on SynBio provided by Shapira one by one in the SSCI and AHCI and check the results; remove the terms with irrelevant records and maintain the effective ones	"Synthetic biolog *" "synthetic DNA" "synthetic genom*" "synthetic gene"
3	Add relevant terms from the retrieved literature and other experts' definitions of SynBio and test the terms in turn	New keywords: "synthetic cell" "artificial cell" "do-it-yourself biology *" "DIYbio" "DIY biology *" "biobrick" "protocell virus" and "protocell"
4	The formula for the search strategy with the terms mentioned above and Boolean operators, and remove the irrelevant ones with the Boolean operator "NOT"	The final search strategy is as follows: TS=("synthetic biology*" or "synthetic DNA" or "synthetic genom*" or "synthetic gene" or "synthetic cell" or "artificial cell" or "do-it-yourself biology*" or "DIYbio" or "DIY biology*" or "biobrick") or (TS= ("artificial life" or "artificial lives") and TS=(protocell or virus) or [TS=(protocell) NOT TS=("dichroic reflector" or architecture)]
	Combine the keywords obtained by the two methods and conduct the fourth screening to obtain the final search formula	

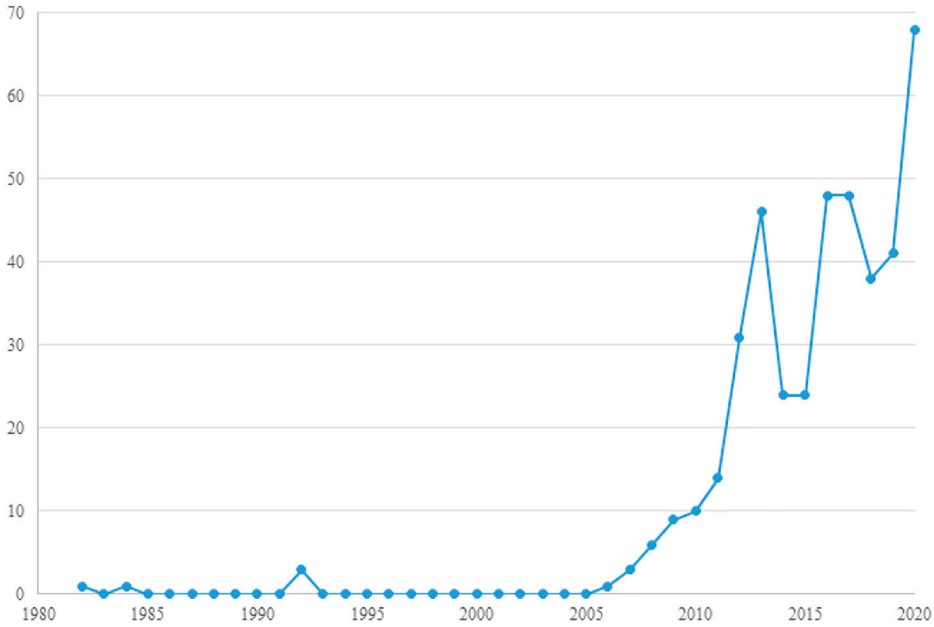


FIGURE 1
Yearly publication records of the SynBio literature in social sciences.

development of novel biotechnology will be magnified significantly. Thus, in the post-COVID-19 era, while SynBio, as a dual-use discipline (Gronvall, 2014), can provide the world with more possibilities and latent enormous benefits, it may encounter new possible risks and obstacles that are concomitant with its emergence and progression as it challenges the long-standing norms and culture, along with the nature, safety, and even existence of human beings. These challenges are mainly revealed, examined, and analyzed by sociologists, ethicists, and philosophers, and are probably overcome as a result of their significant contributions. Therefore, it is necessary to systematically review and summarize the research in humanities, especially philosophical and ethical domains, and social sciences concerning SynBio and elaborate the research fronts ("an emergent and transient grouping of concepts and underlying research issues" (Chen, 2006), which can be conceptualized as the research themes of the collected literature and can be operationally extracted from the keywords and titles of the literature) and intellectual bases (conceptualized as the scientific publications cited by research-front concepts according to Chen (2006)) of this field to provide some references for future research on public understanding, ethics, and governance of SynBio. Some studies have attempted to review the natural science progress of SynBio or argued on partial aspects of its social sciences so far. For instance, Shapira et al.'s (2017) reviewed the SynBio research in the sphere of natural sciences and provided the practical definition and conceptual boundary of SynBio in the study, which significantly helped in collecting the literature. Hayry (2017) elaborated on ethical concerns related to SynBio and its relatedness to GMOs. Similarly, Patrick Heavey (2017) underlined the moral perspective, integrity of nature, and essence of spirituality (Heavey, 2013).

However, there is little research that has hitherto reviewed the entire pool of SynBio social science literature, instead of the natural science ones. In consideration of the huge amounts of research articles for review, traditional reviews are either unable to avoid bias, subjectivity, and incompleteness or time consuming and lack the diversity of analysis methods (Daim et al., 2006). Therefore, in the current research, we introduce bibliometric methods, co-word analysis, co-citation analysis, and some other algorithms to analyze the literature systematically and holistically and use knowledge mapping and visualization technology to provide intuitive and comprehensible insights and conclusions. Specifically, a set of favorable bibliometric software such as CiteSpace and VOSviewer is used for analysis.

Above all, the current study aims to solve the following questions:

RQ1: (Research Question 1): Generally, what does the literature distribution look like? Its aim is to illustrate the yearly growth status, the distribution of the literature at country and institution levels, and the dominant research areas of all the regarding literature.

RQ2: What are the intellectual bases of this area?

RQ3: What are the research hotspots in this research area?

RQ4: What are the potential research trends of SynBio social science studies?

2 Methodology

2.1 Bibliometrics, co-occurrence analysis, and visualization

In the current research, bibliometric analysis methods were used to illustrate and explicitly describe the evolution path and knowledge structure of SynBio literature in the social, philosophical, and ethical

TABLE 2 Publication records of top prolific countries or regions.

Rank	Country/region	Publication	Proportion (%)
1	United States	126	29.439
2	England	83	19.393
3	Germany	36	8.411
4	The Netherlands	33	7.71
5	France	30	7.009
6	Spain	30	7.009
7	Scotland	27	6.308
8	Switzerland	24	5.607
9	Austria	20	4.673
10	Canada	18	4.206

spheres. Bibliometric approaches, referring to the measurement of text and information, help understand the pattern of literature growth, highlight the remarkable research categories, productive research units such as researchers, institutions, and countries, and impactful literature and references. Instead of the partial understanding of certain areas from individual views, bibliometric methods provide a bigger and neutral image of the whole area, specifically the pool of the literature herein, which was later upgraded by Moss's Database Information Visualization and Analysis (DIVA) system (Morris et al., 2002), which combines bibliometric data with visualization techniques to assist researchers in understanding the scientific literature systematically and visually. Moreover, co-occurrence analysis, including co-word analysis, co-citation analysis, and collaboration analysis, works as a powerful and mature method for detecting the knowledge structure in recent years.

The co-word analysis of keywords is to generate the networks of keywords based on their co-occurrence relationship in the same literature, which implies their similarity and relevance at the level of meaning and helps explore the thematic evolution, research fronts, and trends of a given research field (Cobo et al., 2011). Co-citation analysis has been proven as a well-known approach for intellectual structure detection (Chen, 2006), functions better when the co-citation networks are well clustered (Chen et al., 2010), and can be used to predict the future/emerging trends (Van Eck and Waltman. 2010). Given that SynBio is highly interdisciplinary and dominated by highly specialized research themes, it is much more convenient to find out research gaps among its divergent intellectual knowledge bases with a co-citation approach. In recent years, visualization techniques have been used in co-occurrence analysis, the target units in co-occurrence networks such as cited references, keywords exist as nodes, and the co-occurrence relationship between the units is exhibited as edges that connected the nodes; then, all the target units can be shown together in an informative visual network.

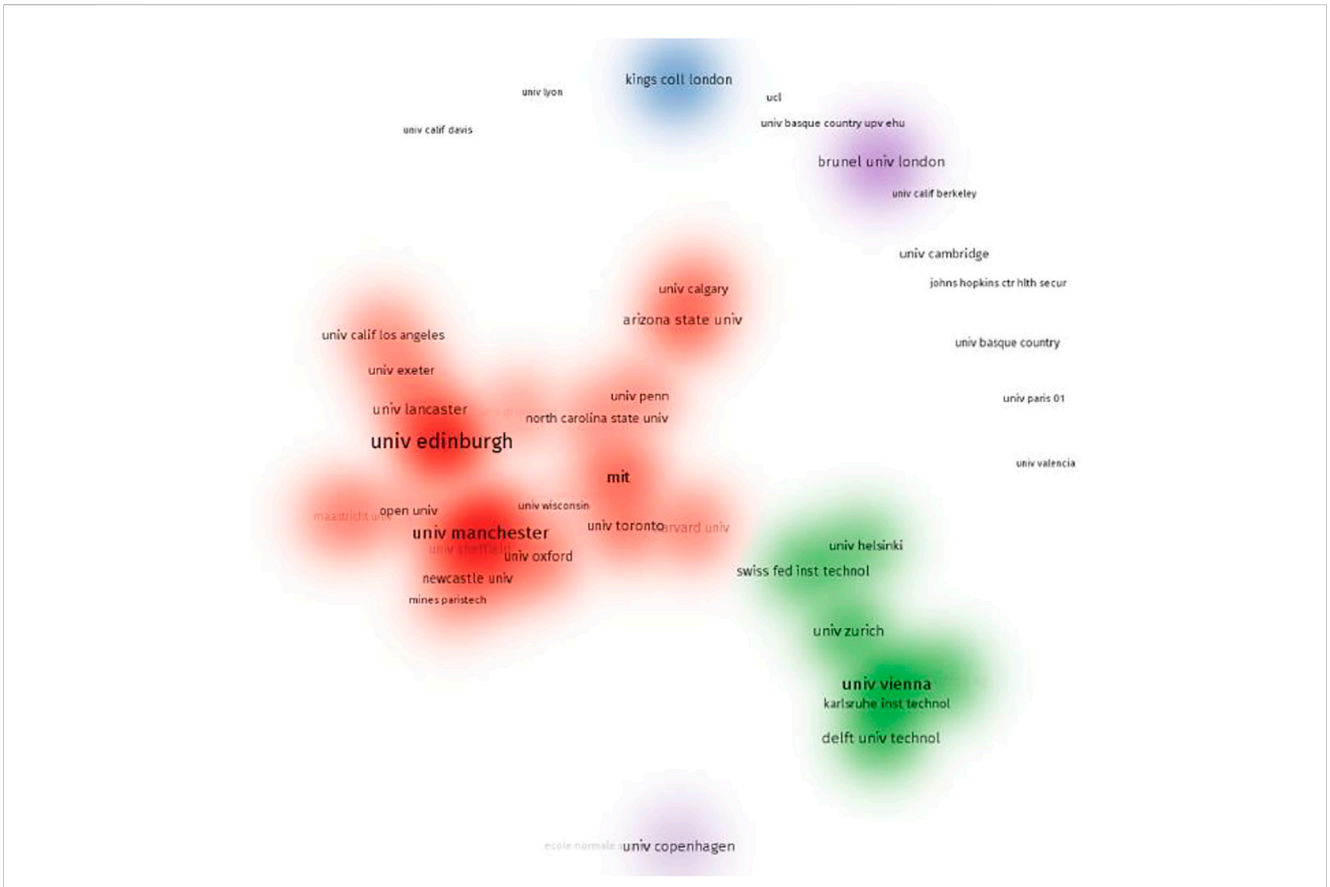


FIGURE 2
Institution collaboration networks in this field.

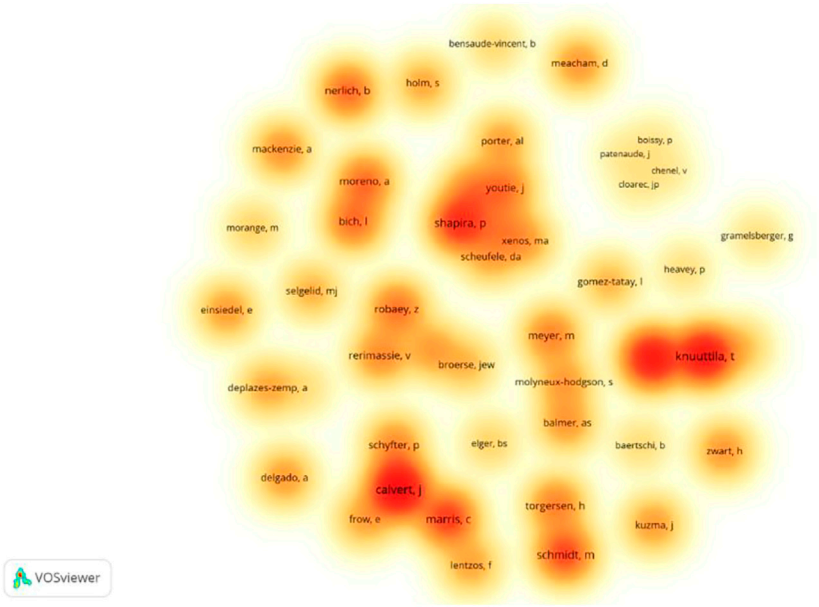


FIGURE 3
Author collaboration networks in this field.

Several tools and software have been developed to analyze and visualize bibliometric data in the last few years, such as CiteSpace and VOSviewer. CiteSpace is a Java software designed for analyzing and visualizing co-occurrence data (Chen, 2017) that is especially good at co-citation analysis, which facilitates the analysis of emerging trends and transient patterns in the scientific literature (Chen, 2006). VOSviewer was developed for creating, visualizing, and exploring bibliometric maps of science and good at text mining and visualization (Van Eck and Waltman. 2011). This study uses VOSviewer and CiteSpace, two mature visualization tools with respective advantages, to analyze the SynBio literature in the sphere of social sciences.

2.2 Data collection

This study focuses on the original and impactful research of SynBio in social, philosophical, and ethical domains, so we collected the article and reviewed records from the Web of Science Core Collection, including SSCI and AHCI, as a data source. To eliminate the risk of skewness from collecting records, the conference proceedings, letters, retracted manuscripts, and book chapters were excluded.

Owing to its rich connotation and without a well-accepted definition of SynBio, the relevant terms for retrieving need to be organized systematically. To confront the challenges of dealing with interdisciplinary definitions (Kuzhabekova and Kuzma, 2014), many scholars adopted expert-defined keyword methods, such as Porter et al. (2008); Small et al. (2014), or other semi-automatic retrieving methods. To cope with the validity- and reliability-related challenges, we cited the concept system of SynBio from the work of Shapira et al.'s (2017) and upgraded and created a novel definition system of SynBio according to the relevant elements or definitions of SynBio in the social science context. For instance, we removed the terms which refer to cell phones instead of artificial cells for disambiguation and added more terms about SynBio like "protocell" and "DIY biology." DIY biology refers to do-it-yourself biology (DIYbio), and several DIYbio groups have formed from the encounter of amateur sciences with synthetic biology (Seyfried et al., 2014). As involved with the high level of publicity and engagement of public by its very definition, DIYbio has been catching the attention of ethical and social researchers from the outset with plenty of social science articles published and has been held accountable for the rise of biohacking (Meyer and Vergnaud, 2020). Moreover, since we need to examine the research regarding SynBio rather than the long-standing debate on GMO, which is distinguished from SynBio by the definition mentioned above, the concepts closely related to GMO were excluded like "transgen*," "genetic engineering," and "genetically modified," so as to the general technology used in GMO and SynBio like "gene editing" and "CRISPR." Eventually, 428 article records and 2,474 references were collected on 24 April 2021. The process of creating the definition system of SynBio is shown in Table 1.

3 Empirical results

3.1 RQ1: literature distribution

The number of the literature grew in succession from 2006 to 2020 despite the slight vibration, and the peak appears in 2020, which implies the increasing and consistent attention from social and ethical researchers, so there may be more research oncoming. As the data of total publications of 2021 was not available when it was

TABLE 3 Top 10 research areas of the SynBio-related social science literature.

Rank	Research area	Number of publications	Proportion (%)
1	History and Philosophy of Science	124	28.972
2	Ethics	91	21.262
3	Social Sciences Biomedical	55	12.85
4	Philosophy	51	11.916
5	Social Issues	43	10.047
6	Multidisciplinary Sciences	34	7.944
7	Environmental Studies	29	6.776
8	Medical Ethics	27	6.308
9	Engineering Multidisciplinary	24	5.607
10	Communication	22	5.14

TABLE 4 Top 10 highly co-cited references.

No.	Frequency	Title	Author	Year	DOI
1	46	Creation of a bacterial cell controlled by a chemically synthesized genome	Daniel G. Gibson et al.	2010	10.1126/science.1190719
2	32	Knowledge-making distinctions in synthetic biology	Maureen A. O'Malley et al.	2008	10.1002/bies.20664
3	26	Foundations for engineering biology	Drew Endy	2005	10.1038/nature04342
4	25	Developing a framework for responsible innovation	Jack Stilgoe, et al., 2020	2013	10.1016/j.respol. 2013.05.008
5	19	Synthetic biology and the ethics of knowledge	Thomas Douglas and Julian Savulescu	2010	10.1136/jme. 2010.038232
6	18	That was the synthetic biology that was	Luis Campos	2009	10.1007/978-90-481-2,678-1_2
7	17	Tales of emergence: Synthetic biology as a scientific community in the making	Susan Molyneux-Hodgson and Morgan Meyer	2009	10.1017/S1745855209990019
8	16	Five hard truths for synthetic biology	Roberta Kwok	2010	10.1038/463288a
9	16	Responsible research and innovation: From science in society to science for society, with society	Richard Owen et al.	2012	10.1093/scipol/scs093
10	15	Diffusion of synthetic biology: A challenge to biosafety	Markus Schmidt et al.	2008	10.1007/s11693-008-9018-z

collected and the incomplete data might impede the understanding of the publishing trend, therefore the data for 2021 is excluded here. Figure 1 shows the yearly growth of publication records. The earliest record in this collection is "Social Responsibility in an Age of Synthetic Biology" by Sheldon Krimsky (1982), which highlighted the risk of rDNA as a bioweapon and appealed to harness rDNA technology and the social responsibility of SynBio research to society earlier then the emergence of the first artificial cell "Synthia" (Gibson et al., 2010), which represented the early concerns on SynBio. Since 2006, consecutive articles are published owing to the bioweapon concerns like synthetic virus genome and other ethical issues on SynBio, and it manifests the continuous attention on SynBio from social and philosophical spheres. The first surge occurred in 2010–2013, right after the growth of the first synthetic cell (Gibson et al., 2010) which made a huge stir globally. It has such a profound impact that we believe this event directly led to this sharp spike. Hereafter, the publication number went up and down as the novel development of SynBio emerged worldwide like the invention of CRISPR technology in 2013 (Cong et al., 2013), the first synthetic minimal cell (Hutchison et al., 2016), the first synthetic yeast genome (Richardson et al., 2017), and so on.

3.1.1 Most productive countries and institutions

The list of the top productive countries is dominated by developed countries (Table 2). Specifically, two English-speaking countries, i.e., the US and UK, took the top two positions and account for over 50% of the total publications, which are followed by Germany, the Netherlands, France, and other European countries. While this result is in line with the advances in the lab of SynBio research and development happening in the developed countries like the United States and European countries, the enormous endeavor like the reflection of social and philosophical impact of SynBio and

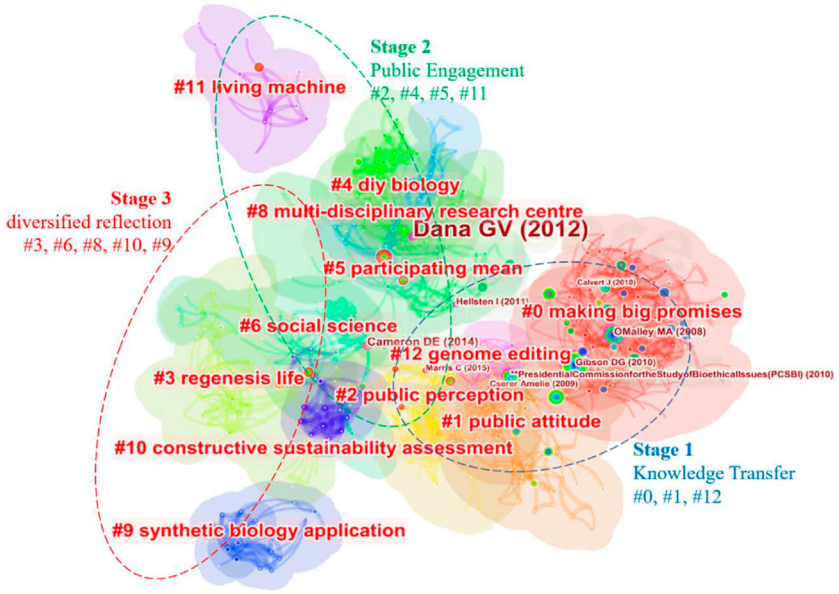


FIGURE 4
Clusters in the co-cited reference network. Knowledge transfer stage is the earliest one and contains three clusters, including the biggest two clusters (#0 and #1), well connected and located on the right part. In this stage, the concepts and distinctions from relative disciplines (O'Malley et al., 2007), relative technologies, and potential applications of SynBio were widely introduced into humanities and social sciences, as well as philosophical studies, during which SynBio served as a novel issue, and people crave for its bright prospect, commercial value, intellectual property (Calvert, 2008; Saukshmya and Chugh, 2010), and technology potentials. The huge commercial value of SynBio intellectual properties and patents are not completely consistent with the open-source philosophy of SynBio like what happens in International Genetically Engineered Machine (iGEM) competitions, sparking debates on legal issues (Contreras et al., 2015). In addition, some studies focus on the ethical issues of SynBio as well.

deliberations, conferences, and policies regarding this novel biotechnology account for the prosperity and productivity in publications of these countries in a way. Some of the policies have been put into effect such as *New Directions: The Ethics of Synthetic Biology and Emerging Technologies* released in 2010 by the Presidential Commission for the Study of Bioethical Issues in the United States and *Final Opinion on Synthetic Biology III: Risks to the environment and biodiversity related to synthetic biology and research priorities in the field of synthetic biology* released in 2015 by European Commission—Scientific Committees. Additionally, it should be noted that the current research mainly follows the opinions published in articles by social and philosophical researchers rather than public opinion in massive or social media which can be probed better in the perspective of communications. Furthermore, we chose only the SSCI and AHCI as data sources, which may not draw on some articles outside the index, especially the ones from non-English-speaking countries. Therefore, the developed countries might become overrepresentative when looking into the SynBio social research.

The most prolific institutes include the University of Edinburgh (25 records), the University of Manchester (18), the Georgia Institute of Technology (10), the University of Copenhagen (9), and the University of Zurich (9). British universities are distinguished from all the contributors. Professor Shapira of the University of Manchester is also affiliated with the Georgia Institute of Technology, and these two institutions share the publication records. To understand the collaboration among institutions, VOSviewer was used to illustrate their co-occurrence networks here (Figure 2). We use the density visualization function in VOSviewer to mark the most productive institutions and their collaborations, in which two major clusters are generated in red and green color. Two research teams in the UK are closely internally connected in their network, led by the University of Edinburgh, with the University of Lancaster and the University of Exeter, and led by the University of Manchester, with the University of Sheffield and the University of Oxford. Moreover, in the red cluster, several universities from the United States are also well connected, including the Massachusetts Institute of Technology, the University of California at Los Angeles, the University of Pennsylvania, and Harvard University. Institutions from continental Europe have also formed networks in the green cluster, including the University of Helsinki, the University of Vienna, the University of Zurich, and some other institutions. With the same approach, we illustrated the collaborative network of researchers in this field (Figure 3), three teams standout, including Calvert's team at the University of Edinburgh, Shapira's team at the University of Manchester, and Knuuttila's team at the University of Vienna.

3.1.2 Research areas

In this field, the research area with the most publications is the history and philosophy of science (Table 3), followed by ethics, social sciences biomedical, and philosophy. Evidently, amid the top 10 research areas, the philosophical and ethical associated issues (including #1, #2, #4, and #7 areas) serve as the dominant theme of this field and account for 68.46% of the publication records. In

TABLE 5 Size, silhouette, and labels of each cluster in the co-cited reference network.

Stage	Cluster-ID	Size	Silhouette	Mean (year)	Label (LLR)
1 knowledge transfer	0	97	0.852	2007	Making big promises
	1	57	0.746	2011	Public attitude
	12	10	0.959	2012	Genome editing
2 public engagement	2	43	0.831	2013	Public perception
	4	41	0.949	2013	DIY biology
	5	35	0.725	2013	Participating mean
	11	13	0.973	2013	Living machine
3 diversified reflection	3	43	0.916	2014	Regenesis life
	6	34	0.944	2014	Social science
	8	24	0.951	2015	Multidisciplinary research center
	10	16	0.934	2015	Constructive sustainability assessment
	9	18	0.962	2016	Synthetic biology application

addition to the long-standing debates in ethics, philosophy, and social science domains, other areas like environmental studies and communication emerge as novel perspectives to explore the social impact of SynBio.

3.2 RQ2: intellectual bases of the current field

Research fronts and intellectual bases are fundamental concepts in information science, the former focus on the grouping of concepts and underlying research issues, while the latter refers to the citation and co-citation footprint in the scientific literature (Chen, 2006), which helps detect the evolvement and knowledge origin of certain research fields. In this section, we used CiteSpace software to analyze the references of the collected records, especially the co-citation relationship of these references and their source titles.

We found the top highly co-cited references, which are visualized as nodes by the software in the co-cited reference network, and the co-citation relationship of each pair of references is visualized as the edge between the nodes. Then, we conducted a cluster analysis of the co-cited references and labeled the clusters with the embedded algorithm log-likelihood ratio to elaborate the themes of the clustered references within the network. The highly co-cited journals are listed at the end of this section to facilitate the analysis of the intellectual bases. CiteSpace software has several advantages when dealing with co-citation data. First, it can detect highly co-cited references (Table 4) and the nodes with high co-citation frequency in the network. Meanwhile, the nodes in the network can also be clustered according to their spatial distance, so the nodes with details can be classified and labeled as clusters to extract knowledge from detailed information by CiteSpace's built-in text mining algorithms, such as the log-likelihood ratio whose effectiveness was introduced by Chen et al. (2010).

3.2.1 Highly co-cited references related to SynBio in social sciences

We listed the top 10 references (in Table 4) with a high frequency of co-citation, which means they were highly co-cited by the citing articles we collected from the Web of Science, which indicated their high impact as the intellectual base for this field. Only two among the top 10 co-cited references came from the natural science fields, including a review article about the brief history and research progress of SynBio (#3) written by Endy (2005). The other references from the natural science realm with the highest frequency (#1) is the first research of self-replicating bacterial cells using synthetic DNA in 2010, achieved by a team with Venter's leading (Gibson et al., 2010). This research was a blast appealing to global attention and triggered great concerns throughout natural and non-natural science fields and the public ever since.

Beyond the natural science perspectives, the #2 highly co-cited reference (A. O'Malley et al., 2008) helped explicate the rationale and relationship between synthesis and analysis, which was followed by the discussion of bottom-up and top-down strategies in SynBio research and innovation (Gilbert et al., 2010); distinct SynBio from disciplines like biological engineering and systems biology, meanwhile contributed to locating and delimiting research objects and laying a foundation for further SynBio social and humanities research in respect to the fact that it was published relatively early, and calling for attention on knowledge itself. Subsequently, the # 5 reference (Douglas and Savulescu, 2010) brought about the rethinking of ethics of knowledge in SynBio. It introduced ethics of knowledge into the research on SynBio based on the "misuse of knowledge," initiated a new research domain, and provoked widespread debate. For instance, Pierce (2012) "challenges(d) an ethics of knowledge to respond to concerns of procedural and substantive justice," and invited more consideration on the decision by whom, whose interest and effectiveness of ethics of knowledge regarding SynBio.

As early as in 2008 in the #10 reference, Schmidt (2008) highlighted the unprecedented biosafety challenges caused by

TABLE 6 Top 20 highly co-cited journals of the SynBio-related social science literature.

Frequency	Centrality	Cited journal	Mean year	Half-life
258	0.1	Nature	2007	8.5
229	0.05	Science	2007	8.5
149	0.06	Nature Biotechnology	2007	8.5
120	0.07	EMBO Reports	2007	8.5
101	0.1	Proceedings of the National Academy of Sciences	2009	7.5
80	0.24	Biosocieties	2009	5.5
73	0.07	Nature Reviews Genetics	2009	6.5
70	0.08	Molecular Systems Biology	2010	4.5
66	0.11	BioEssays	2009	5.5
66	0.1	Social Studies of Science	2012	4.5
64	0.08	PLOS One	2012	5.5
63	0.02	Science, Technology, & Human Value	2009	7.5
62	0.13	Public Understanding of Science	2012	4.5
59	0.09	Nanoethics	2011	5.5
59	0.03	Science and Engineering Ethics	2012	4.5
57	0.12	Research Policy	2014	3.5
52	0.11	PLOS Biology	2012	3.5
50	0.06	Trends in Biotechnology	2012	6.5
49	0.07	Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences	2012	4.5
43	0.09	Science and Public Policy	2014	3.5

engineered de-skilling SynBio practices (like garage biology or do-it-yourself biology), especially by newcomers without formal biosafety training including researchers from other disciplines and public, as it may lead to biohackery, illicit bioeconomy, and appealed for appropriate safety standards. In 2009, the #6 reference, a chapter written by Campos (2009), examined the coinage of "synthetic biology" back to 1912 by Stéphane Leduc in his *La Biologie Synthétique* and elucidated the nature and features of SynBio as technology and engineering from a historical perspective. In addition, Molyneux-Hodgson and Meyer (2009) viewed SynBio from a perspective of the scientific community, as the community of SynBio contains not only researchers but also interested citizens distinctively, probed the emergence and four types of formation of the SynBio discipline community, and constructed a framework of "movements" and "stickiness" to understand this community.

Regarding the prevailing framework for the governance, administration, and normalization of SynBio, Owen, Macnaghten, and Stilgoe (in #9 reference) provided a historical overview of the concept and three features of responsible innovation (RI) after the *Horizon 2020 Strategy* was put forward by the EU (Owen et al., 2012). Furthermore, they (in #4 reference) systematically constructed a framework to integrate SynBio and RI, illuminated four integrated dimensions of RI in SynBio research, including anticipation, reflexivity, inclusion, and responsiveness, and discussed the universality of RI in SynBio beyond the use in UK Research Councils and the scientific communities (Stilgoe and Macnaghten, 2013).

For #8, Kwok (2010) introduced five challenges that SynBio had to confront back in 2010 in a Nature news feature and poured cold water on the hypes regarding SynBio's overrated prospects. This news feature was later criticized by Voigt (2020) as an "infamous article" based on the evidence of commercially available products from SynBio research. The top 10 highly co-cited references not only highlighted diverse dimensions regarding the research and innovation of SynBio but also harnessed most of the substantial issues, especially on the social, philosophical, and ethical fronts.

3.2.2 Intellectual bases: cluster analysis of co-cited references

All the co-cited references in the network are clustered, labeled, and visualized with CiteSpace, and the distribution of each cluster is shown in Figure 4. The clusters are identified by their size (Cluster #0 has the most nodes) and ranked by their mean year (Table 5). The silhouette value indexed the similarity of nodes in the same cluster, and each of them is over 0.7, which indicates the homogeneity of the nodes within the same cluster for analysis. The clusters can be

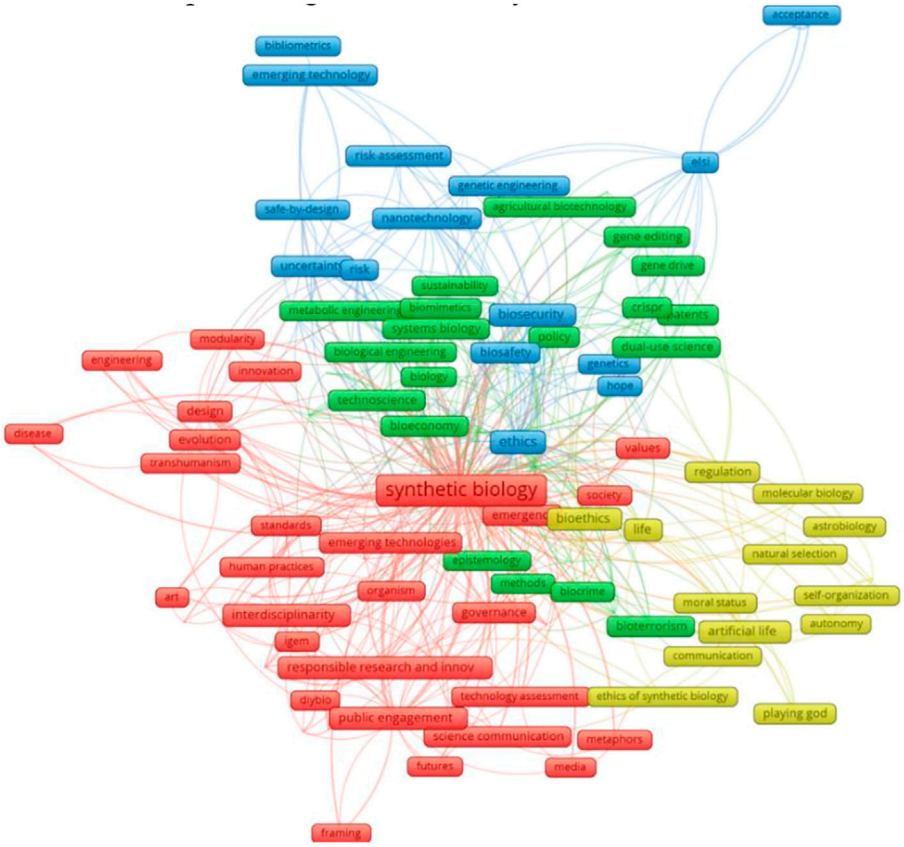


FIGURE 5
Co-word analysis of the keyword network. "ELSI" refers to Ethical, Legal, and Social Implication; "CRISPR" refers to clustered regularly interspaced short palindromic repeats; "iGEM" refers to International Genetically Engineered Machine; and "DIYbio" refers to do-it-yourself biology.

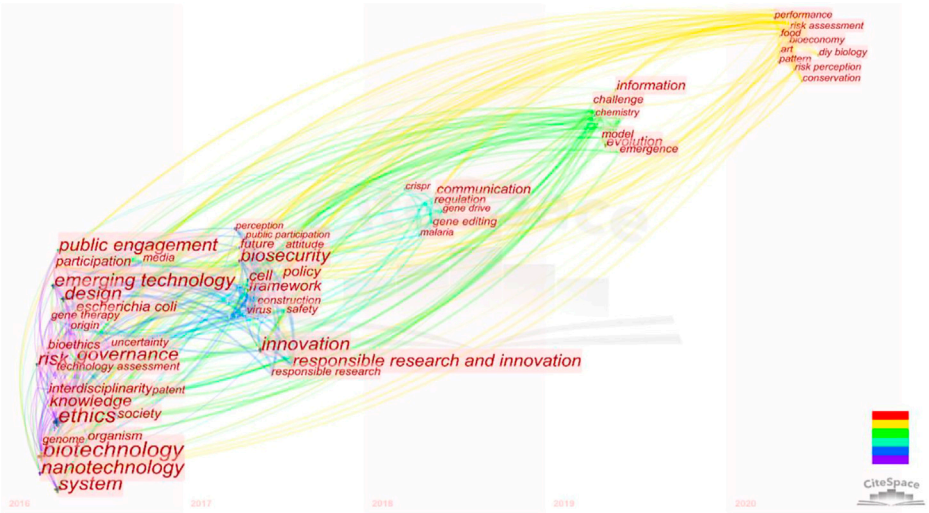


FIGURE 6
Keyword-based trend analysis to highlight emerging trends and possible future.

divided into three stages generally based on the mean year of references in each cluster and extracted labels, i.e., the "knowledge transfer" stage (Cluster #0, #1, and #12) during which the concepts, ethical concerns, and knowledge produced in the natural science research flooded into the social and humanities sphere, triggered the knowledge transfer from natural science to

social and ethical domains; the "public engagement" stage (Cluster #2, #4, #5, and #11) featuring public engagement such as DIYbio and public participation as major topics; and the "diversified reflection" stage (Cluster #3, #6, #8, #10, and #9) which contains not only the application and impact of SynBio but also the assessment and reflection on SynBio itself.

In the public engagement stage, public perception and public participation became the major concern in this field. No matter the experiments and practices of the public on biological materials with SynBio technology, like DIYbio, or the perception, governance, monitoring, and policymaking on SynBio, public engagement appears as the non-negligible and inevitable issue since the emergence of SynBio itself. The clusters about public participation connected closely in this stage, while the Cluster #11 living machine, located separately from them in the same period, indicates the concerns on the ethical and philosophical discussion about SynBio.

The latest stage "diversified reflection" turned out to be multiple regarding themes, and the researchers reevaluated and reflected on SynBio from divergent perspectives. In addition to the previous issues in the last two stages, novel technologies used in SynBio, innovative administration and management, sustainability assessment, and other various themes supplied various knowledge bases for this field in the latest period. Novel and continuously improving technologies like CRISPR (Heidari et al., 2017) and gene editing are hot topics of controversy in recent years, especially after the "gene-edited babies" event that two baby girls were born, whose genomes were edited using CRISPR/Cas9, a gene-editing tech, during the embryo stage to confer the babies innate resistance to HIV as one of their parent was HIV-positive. The genome-editing research on human embryos aiming to "help people with HIV-related fertility problems" turned out to be associated with unpredictable risks (Wang and Yang, 2019), provoked a fierce global backlash (Cyranoski et al., 2018), and prompted people's attention to biotechnological ethics. Meanwhile, the social scientists also have conducted research on genome editing for decades (Thompson, 2012) and published piles of articles which provide plenty of important theories and experiences for the ethical reflection on SynBio for reference so that "genome editing" becomes the representative label of references in Cluster #12. These studies reconsidered the general problems that emerged in the development of new technologies, such as the lack of transparent regulatory standards and outdated management regulations, and provide references to avoid similar problems which harass SynBio. The multi-disciplinary research center (Vermeulen, 2018) and its innovative management have also caught much attention in recent years. Especially, "responsible innovation" stands out as one of the promising solutions for the bioethical issues (Vermeulen et al., 2017) of emerging biosciences and biotechnologies, which impacts the daily practices and behaviors of scientists and outcomes of their studies (Pansera et al., 2020).

3.2.3 Highly co-cited journals

Using the same research method as in Section 3.2.2, we obtained the top 20 highly co-cited journal's table (Table 6), including the journal titles, co-cited frequency of journals, centrality in the network, the mean year of being cited, and the half-life of each journal. It indicated that the early source titles are mainly from the field of natural sciences, while in recent years, the journals from the field of social sciences take the dominant position, and there is a shift in the sources of knowledge in this field. Among all the high-centrality nodes which have structural importance, *Biosocieties*, *BioEssays*, *Social Studies of Science*, *Public Understanding of Science,* and *Research Policy* are the prominent knowledge sources for the development of this field in the social science sphere.

3.3 RQ3: hotspots in the SynBio-related literature in social sciences

In this section, VOSviewer software was utilized to conduct a co-word analysis on the keywords of all the literature collected in this field and visualize the network of co-word relationships, as shown in Figure 5, with the built-in algorithms. In this science mapping practice, keywords in the collected literature are linked according to their co-occurrence relationship, and the keywords with a larger node size (and proportionally larger fonts) have a higher degree which refers to the number of edges and indicates its impact. With the built-in clustering algorithm of the software, four clusters emerged and were marked as red, green, blue, and yellow, as shown in Figure 6, and each cluster represents a sub-theme of the current field.

Cluster 1 (red): This cluster represents the themes mainly about the governance of SynBio, containing three aspects of keywords such as assessments and standards, Responsible Research and Innovation (RRI), and public engagement. In addition to the searched terms like "synthetic biology," other high-impact keywords in this cluster are governance, standards, values, and technology assessment; Responsible Research and Innovation; and public engagement, science communication, media, DIYbio, and iGEM. Technology assessment (TA), a set of wildly used methods for assessment and governance (Trump et al., 2019) of emerging technologies during the 1970s–1990s, aimed to provide an early warning by evaluating emerging technologies so that potential and unexpected negative impacts can be corrected and remedied promptly, while they were more concerned on risk evaluation rather than ethical issues. After that, the Ethics, Legal, and Social Issues (ELSI) program was put forward alongside the Human Genome Project by the National Institute of Health and focused more on the interdisciplinary research and the interaction between different stakeholders in society and refers to the ethics and legitimation of the research studies (Gregorowius and Deplazes-Zemp, 2016). However, the ELSI practice enacted the epistemological gap as a division of labor "where scientists perform science and leave social, moral, and ethical questions to experts–ethicists, theologians, lawyers, and social scientists" and are often based on a simplified linear model of innovation pathways and outcomes (Marris et al., 2015). Then, RI (or RRI) was promoted and supplied an interdisciplinary frame for the governance of SynBio and emphasized public participation in the whole process of innovation, including the assessment of technology itself, offering the idea of social needs and wishes, and decision on the final plan to meet them. Gregorowius and Deplazes-Zemp (2016) pointed out that TA and ELSI research can be regarded as the early development stage and predecessor of RRI. So, in this cluster, all three aspects of keywords refer to the responsible governance of SynBio and can be covered in the frame of RRI.

Cluster 2 (yellow): Keywords in this cluster represent the studies mainly related to philosophical issues of synthetic biology, including "moral status," "playing God," "artificial life," "self-organization,"

"natural selection," and bioethics. These studies highlighted the philosophical issues of SynBio and its product, i.e., artificial life. Some religious strong believers oppose SynBio because the creation of artificial organisms by human beings encroaches on a domain of activity that has been considered to be God's divine prerogative (Dragojlovic and Einsiedel, 2013). Moreover, how to position the moral status of artificial life, and human beings, as peer life and creator, may confront the challenges of moral statuses from artificial life. Moreover, these creation practices substitute for natural selection in the evolution process and lead to new philosophical concerns.

Cluster 3 (blue): Keywords in this cluster primarily related to the ethical concerns of SynBio, especially the potential risk and danger that SynBio may bring about. The high-frequency keywords in this cluster are "ethics," "risk assessment," "uncertainty," "ELSI," "biosafety," and "biosecurity". The uncertainty of the process of research and practice on SynBio and its products and the risks they may bring about are related to biosafety and biosecurity at the material level. Gómez-Tatay and Hernández-Andreu (2019) tried to distinguish biosafety from biosecurity, and "biosafety refers to the prevention of the risks to public health and the environment that could be produced by accidental interactions between dangerous biological agents and other organisms or the environment" and biosecurity can be defined as "the protection, control, and accountability for valuable biological materials [...] within laboratories, to prevent their unauthorized access, loss, theft, misuse, diversion, or intentional release" (WHO, 2006). Biosafety and biosecurity identified the risks that can be posed in active or passive situations by intentional or accidental mistakes of certain experiment practitioners in labs or garages (for garage biology or DIYbio cases). Both the laboratory safety and the garage biology safety constitute the major parts of biosafety regarding SynBio, especially the former one as laboratories are still the major sources of knowledge and material production at present (Douglas and Savulescu, 2010; Thompson, 2012).

Cluster 4 (green): This cluster is related to the technologies and upstream science disciplines of SynBio. The high-frequency keywords here are "methods," "gene drive," "gene editing," "Clustered Regularly Interspaced Short Palindromic Repeats," "systems biology," "biological engineering," "metabolic engineering," and "agricultural biotechnology," and the issues about technology such as "patents," "policy," "bioeconomy," "biocrime," and "bioterrorism." Systems biology, bioengineering, metabolic engineering, biochemistry, and other disciplines and technologies constitute the cornerstone of SynBio. In recent years, accurate, economical, and accessible powerful gene-editing technologies like CRISPR and its upgraded versions have provided significant impetus for the leap-forward development of SynBio. In addition, the research about technology-related intellectual property and patents (Calvert, 2012), as well as policy and regulation issues (Wiek et al., 2012), has also evolved with SynBio.

3.4 RQ4: potential research trends of the SynBio social research

A time zone view of the distribution of the literature topics can be used to illustrate the research focus of each period and show the evolution of research topics and then supply some clues to analyze the potential research trends. In this section, we redistributed and re-visualized the keywords of the collected literature published in the last 5 years, as shown in Figure 6, and switch our focal spots to the latest and cutting-edge topics of the current field. Keywords like "public engagement," "RRI," "ethics," "risk," "biotechnology," and "governance" still have a high frequency in the last 5 years, and these issues will not fade out soon as they deeply involve human beings. However, instead of the four cluster themes analyzed in RQ3 which would dominate the field in the long term and serve as the mainstream themes for the following years; here, we focus on the three emerging themes that might share some novel perspectives for further research. Although these literature that indicate research trends might not be highly cited yet, the extended framework and concerning topics put forward by them can also be valuable as compared with the long-standing themes mentioned previously.

(1) New possibilities and challenges brought by SynBio for the transformation of bioeconomy

The influence of SynBio no longer stays in the fields of basic scientific research or technological development but has penetrated the social and economic fields and demonstrated its considerable economic value and commercial potential. SynBio has been offering possibilities and opportunities for global governments to formulate visions of a transition toward a bioeconomy and has played the enabling and transformational role of entrepreneurship (Kuckertz et al., 2020) and enlightened research and development with its exemplary convergent nature like carefully designed, stakeholder-inclusive, and community-directed evolution in bioeconomy (Bueso and Tangney, 2017). Moreover, the public engagement access offered by SynBio helps recruit and train the future workforce for bioeconomy, such as the competition iGEM and citizen science practices (Warmbrod et al., 2020). However, researchers have also noticed the necessary considerations and points of focus in this endeavor owing to the risks and uncertainties of utility in upcoming deployments outside the labs in bioeconomy practices (Parker and Kunjapur, 2020). Particularly, the self-driven commercialization process would propel the influence of SynBio broadly and rapidly, highlighting the lag of policy and entangling more opposing forces against governance.

(2) The remodeling of SynBio's multi-dimensional impacts by art

Not many articles discussed the relationship between art and SynBio, but several feasible perspectives have been explored according to the current publications, such as the fusion of art and SynBio in material practices, and the reflection on SynBio with artworks. Merritt et al. (2020) used the novel living media interfaces (LMIs) to show the case of the interaction between biological materials and digital systems as responsive living media and SynBio, as engineered biology, could offer plenty of options for the feasible fusion practice. Vaage (2020) appealed to expand the utility of artworks to provide counter-images to challenge mechanistic assumptions, i.e., the living machines metaphor and the perception of life as controllable in the context of SynBio. Meanwhile, Calvert and Schyfter (2017) argued that "engaging more closely with art and design can enrich STS work by enabling an emergent form of critique" and "open up the science

by exploring implicit assumptions and interrogating dominant research agendas." Above all, art offers an interdisciplinary perspective to examine the impacts of SynBio practically and theoretically, and more interaction approaches and science communication patterns regarding SynBio need to be explored.

(3) The role of knowledge of SynBio

The role of knowledge of SynBio has changed since its inception. Its initial form appeared as the knowledge of a certain scientific discipline, which was mainly disseminated within the scientific community and then gradually diffused to the public and exerted social influences. In this process, not only the products but also the knowledge itself became the object of ethical research. Therefore, its role has evolved from "knowledge for use" to "knowledge for public communication" and "knowledge as an ethical object." Nordmann (2015) appealed to examine SynBio on "epistemic values, the ethos and authority of science, and the relation of knowledge and power." Douglas and Savulescu (2010) put forward the concept of "ethics of knowledge" about SynBio based on the "misuse of knowledge" and appealed not to regard the creation and dissemination of knowledge of SynBio as granted or beyond doubt. Instead, it called for attention to the risk of misuse of knowledge and advocated for reflection on it as the object of ethical governance, during which its role had been radically changed.

4 Discussion

Based on Professor Shapira et al.'s (2017) research, this paper reconstructed the conceptual system and search strategies of SynBio to retrieve the literature related to it in the sphere of social sciences and humanities and formulated a set of effective retrieval strategies in the social context based on the search result of each related terms. After visualizing the publishing trends in this field, we found that the publications in this field had been growing rapidly since 2006, which encouraged the researchers greatly. In addition, we investigated the top productive countries and regions in this field, which mainly consisted of the developed world. The United States and the United Kingdom had the top two highest numbers of publications, which contained several prolific research teams from universities in regions such as Edinburgh and Manchester. However, this kind of cooperation was basically within the institutions and the inter-institution cooperation was less significant here.

The research areas of the collected literature involved philosophy and ethics mostly, such as the history and philosophy of science, ethics, philosophy, medical ethics, *etc.* In addition, communication, environmental studies, and other social sciences also took a certain proportion, which implied the diversity of research themes about SynBio in a social context. After that, we investigated the intellectual base of the current field and the high-frequency co-cited references and found that knowledge sources of natural sciences and social sciences took the dominant places in the earlier and later stages separately. With cluster analysis and visualization technology, a co-cited reference network was generated and visualized to find the three representative stages of knowledge background and the research topics of each stage, namely, knowledge transfer, public engagement, and diversified reflection. The shift of the stages manifests how scientists, social scientists, and public interact with each other with knowledge creation, transaction, diffusion, and reinvention occurring within respective communities. In public practice, SynBio experiences the diffusion of innovations when spreading among citizens in breadth and the individual engagement from the perception and attitude to behavioral practices in depth. Owing to the engagement of public in SynBio via DIY biology practice, the diffusion of SynBio knowledge *per se* leads to the ethical issues of potential knowledge misuse and simultaneously push forward the epistemological thinking of philosophical researchers rather than the previous insights into the nature, feature, benefits, and risks of SynBio as a discipline and knowledge system. This is where the intellectual base "public engagement" and the potential research trend "the role of knowledge of SynBio" converge and shows the paradigm shift from value and essence to knowledge in the philosophical perspective. In addition, this trend enlightens us and offers theoretical frameworks to review other disciplines and dual-use technologies which invite close public engagement and participation like artificial intelligence, cybersecurity, and autonomous weapons.

Source titles of the collected literature were also listed to help researchers follow the progress and development of this field. We tried to identify hotspots in this area with co-word methodology, managed to cluster the keywords into four domains, and introduced the theme of each domain. These hotspots could help locate the research fronts in the current field, to access novel ideas, and systematically understand the development of the field. Finally, we extracted and visualized the network of keywords in the last 5 years, tried to find out the potential research trend, and identified the three novel research domains, which might inspire the researchers and stakeholders. Apart from the epistemological trend regarding SynBio expatiated previously, the practical trends of SynBio like bioeconomy and art, on the one hand, echo to the third stage of intellectual base "diversified reflections;" on the other hand, it indicates that researchers stride into further exploration of emergent application and corresponding social impacts of SynBio commercialization and of potentials to boost social good.

These relatively positive and optimistic trends of waning ethical vigilance is due to the fact that despite the rapid development of SynBio, no relevant biosafety or biosecurity incident has occurred so far, owing to the persistent normative efforts of research management and policies on emerging technologies. Moreover, it is because of the promising latent and applicational prospect of SynBio in recent years, such as the synthesis of starch (Cai et al., 2021), glucose, and fatty acids (Zheng et al., 2022) from carbon dioxide. Therefore, it might be safe to boldly set about investigating the issues of diversified commercial possibilities, the impact of novel forms of public involvement, and other feasible perspectives within and beyond SynBio field.

5 Conclusion

In this study, we constructed a novel conceptual framework and search strategy about SynBio in the context of social sciences and systematically reviewed the relevant literature in this field with mature and well-proven methods such as co-occurrence analysis and knowledge mapping. We also

illustrated and clarified the literature from three aspects: hotspots, potential research trends, and intellectual base.

During the COVID-19 pandemic, considerable conspiracies and rumors regarding bioweapon and biohazard leakage have hindered the control of pandemic spread (Romer and Jamieson, 2020; Zhang et al., 2021) and shattered people's confidence over biotechnologies and institutions. This negative attitude toward biotechnologies aggravated as conspiracies about COVID-19 (accounted as bioweapon) entered collective sense-making (Nadesan, 2022), and a substantial portion of people even endorse contradictory conspiracy theories (Petrović and Žeželj, 2022) However, in these unexpected emerging situations, SynBio can benefit the treatment of infectious diseases and the development of vaccines. SynBio and its methodologies, such as RNA delivery using lipid nanoparticles, have already functioned in the production of COVID-19 RNA vaccines and medicines (Liu et al., 2020) and become one of the fundamental methods for the rapid development of fully synthetic RNA vaccines (Rappuoli et al., 2021) in the post-COVID-19 era. Therefore, in such a complex situation, how to release the positive potentials of SynBio, avoid the risks and dangers it might bring to the world by institutionalized means, and explore appropriate ethical rules and policies to accelerate the benign development of SynBio and benefit human beings has become an inescapable issue. COVID-19 only magnifies the urgency. Thus, in addition to the natural science research, non-natural sciences like social sciences, philosophy, and ethics also need to pay more attention to conduct research in advance before the possible events and the cooperation of natural science scientists, government officers, and stakeholders in society are required (Komen et al., 2020). Moreover, the concrete and practical RRI framework, public engagement through democratic deliberation subject to the post-COVID-19 working manner, and ethics of knowledge in production and diffusion should be taken into the research agenda.

However, this study also has its limitations. For example, regarding data sources, only the records from the SSCI and AHCI were included in this study, even to ensure the impact and authority of the literature. Literature like monographs, book chapters, dissertations, academic reports, non-English literature, and public opinion on media and government, white papers are excluded from the current research in consideration of the consistency of data format and data source for bibliometric practices, but the importance of these literature should not be neglected and deserve to be studied exclusively. Owing to the delays in publishing and the agenda gap between social or ethical research and natural science research, academia and industry, public and researchers, the cutting-edge scientific discoveries, industrial success, and public hotspots do take time to get into social researchers' vision and gain attention. While the social and ethical research articles cannot cover all even the latest progress, particularly the breakthroughs of leading research teams in China and the US, looking into the long-standing social and ethical issues and the history of the philosophical thinking in research articles regarding SynBio can be an effective trajectory to rethink SynBio. For further research, on the one hand, the sub-areas such as ethics or philosophy can be worthy of further study to explore enlightening ideas. On the other hand, the potential trends proposed in this paper have pointed out possible directions for further research.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization: QK; formal analysis: DW; funding acquisition: GW; investigation: QK; methodology: FA; software: DW; supervision: GW; validation: GW; visualization: FA; writing: QK; and writing—review and editing: QK and DW. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Fundamental Research Funds for the Central Universities (YD2110002004 and YD2110002014).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bueso, Y. F., and Tangney, M. (2017). Synthetic biology in the driving seat of the bioeconomy. *Trends Biotechnol.* 35 (5), 373–378. doi:10.1016/j.tibtech.2017.02.002

Cai, T., Sun, H., Qiao, J., Zhu, L., Zhang, F., Zhang, J., et al. (2021). Cell-free chemoenzymatic starch synthesis from carbon dioxide. *Science* 373 (6562), 1523–1527. doi:10.1126/science.abh4049

Calvert, J. (2012). Ownership and sharing in synthetic biology: A 'diverse ecology'of the open and the proprietary? *BioSocieties* 7 (2), 169–187. doi:10.1057/biosoc.2012.3

Calvert, J., and Schyfter, P. (2017). What can science and technology studies learn from art and design? Reflections on 'synthetic aesthetics. *Soc. Stud. Sci.* 47 (2), 195–215. doi:10.1177/0306312716678488

Calvert, J. (2010). Synthetic biology: Constructing nature? *Sociol. Rev.* 58, 95–112. doi:10.1111/j.1467-954X.2010.01913Calvert

Calvert, J. (2008). The commodification of emergence: Systems biology, synthetic biology and intellectual property. *BioSocieties* 3 (4), 383–398. doi:10.1017/s1745855208006303

Campos, L. (2009). "That was the synthetic biology that was," in *Synthetic biology* (Dordrecht: Springer), 5–21. doi:10.1007/978-90-481-2678-1_2

Cello, J., Paul, A. V., and Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science* 297 (5583), 1016–1018. doi:10.1126/science.1072266

Chen, C. (2017). *CiteSpace: A practical guide for mapping scientific literature*. New York: Nova Science Pub Inc. Available at: http://cluster.cis.drexel.edu/~cchen/citespace/books/(Accessed May 10, 2017).

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.* 57 (3), 359–377. doi:10.1002/asi.20317

Chen, C., Ibekwe-SanJuan, F., and Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *J. Am. Soc. Inf. Sci. Technol.* 61 (7), 1386–1409. doi:10.1002/asi.21309

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., and Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *J. Inf.* 5 (1), 146–166. doi:10.1016/j.joi.2010.10.002

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339 (6121), 819–823. doi:10.1126/science.1231143

Contreras, J. L., Rai, A. K., and Torrance, A. W. (2015). Intellectual property issues and synthetic biology standards. *Nat. Biotechnol.* 33 (1), 24–25. doi:10.1038/nbt.3107

Cyranoski, D., and Ledford, H. (2018). Genome-edited baby claim provokes international outcry. *Nature* 563 (7731), 607–609. doi:10.1038/d41586-018-07545-0

Daim, T. U., Rueda, G., Martin, H., and Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technol. Forecast. Soc. Change* 73 (8), 981–1012. doi:10.1016/j.techfore.2006.04.004

Dong, W., Anwar, M. A., Nawaz, M. A., Rongting, Z., Asmi, F., Sajjad, A., et al. (2019). Knowledge mapping of GMO/GMF research in social sphere. *Cogent Food and Agric.* 5 (1), 1602103. doi:10.1080/23311932.2019.1602103

Douglas, T., and Savulescu, J. (2010). Synthetic biology and the ethics of knowledge. *J. Med. Ethics* 36 (11), 687–693. doi:10.1136/jme.2010.038232

Dragojlovic, N., and Einsiedel, E. (2013). Playing god or just unnatural? Religious beliefs and approval of synthetic biology. *Public Underst. Sci.* 22 (7), 869–885. doi:10.1177/0963662512445011

Endy, D. (2005). Foundations for engineering biology. *Nature* 438 (7067), 449–453. doi:10.1038/nature04342

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329 (5987), 52–56. doi:10.1126/science.1190719

Gilbert, D., Jaramillo, A., Krasnogor, N., and de Lorenzo, V. (2010). Synthetic biology gains momentum in Europe. *Syst. Synthetic Biol.* 4 (3), 145–147. doi:10.1007/s11693-010-9065-0

Gómez-Tatay, L., and Hernández-Andreu, J. M. (2019). Biosafety and biosecurity in synthetic biology: A review. *Crit. Rev. Environ. Sci. Technol.* 49 (17), 1587–1621. doi:10.1080/10643389.2019.1579628

Goss, R. J., Shankar, S., and Abou Fayad, A. (2012). The generation of "unnatural" products: Synthetic biology meets synthetic chemistry. *Nat. Product. Rep.* 29 (8), 870–889. doi:10.1039/C2NP00001F

Gregorowius, D., and Deplazes-Zemp, A. (2016). Societal impact of synthetic biology: Responsible research and innovation (RRI). *Essays Biochem.* 60 (4), 371–379. doi:10.1042/EBC20160039

Gronvall, G. K. (2014). The existing guidance for "dual-use" research. *Hastings Cent. Rep.* 44 (S5), S34–S35. doi:10.1002/hast.396

Häyry, M. (2017). Synthetic biology and ethics: Past, present, and future. *Camb. Q. Healthc. Ethics* 26 (2), 186–205. doi:10.1017/S0963180116000803

Heavey, P. (2017). Consequentialism and the synthetic biology problem. *Camb. Q. Healthc. Ethics* 26 (2), 206–229. doi:10.1017/S0963180116000815

Heavey, P. (2013). Synthetic biology ethics: A deontological assessment. *Bioethics* 27 (8), 442–452. doi:10.1111/bioe.12052

Heidari, R., Shaw, D. M., and Elger, B. S. (2017). CRISPR and the rebirth of synthetic biology. *Sci. Eng. Ethics* 23 (2), 351–363. doi:10.1007/s11948-016-9768-z

Heinemann, M., and Panke, S. (2006). Synthetic biology - putting engineering into biology. *Bioinformatics* 22 (22), 2790–2799. doi:10.1093/bioinformatics/btl469

Hutchison, C. A., III, Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351 (6280), aad6253. doi:10.1126/science.aad6253

Keasling, J. D. (2008). Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* 3 (1), 64–76. doi:10.1021/cb7002434

Komen, J., Tripathi, L., Mkoko, B., Ofosu, D. O., Oloka, H., and Wangari, D. (2020). Biosafety regulatory reviews and leeway to operate: Case studies from sub-sahara africa. *Front. Plant Sci.* 11, 130–213. doi:10.3389/fpls.2020.00130

Krimsky, S. (1982). Social responsibility in an Age of synthetic biology. *Sci. Policy Sustain. Dev.* 24 (6), 2–5. doi:10.1080/00139157.1982.9929782

Kronberger, N., Holtz, P., and Wagner, W. (2012). Consequences of media information uptake and deliberation: Focus groups' symbolic coping with synthetic biology. *Public Underst. Sci.* 21 (2), 174–187. doi:10.1177/0963662511400331

Kuckertz, A., Berger, E. S., and Brändle, L. (2020). Entrepreneurship and the sustainable bioeconomy transformation. *Environ. Innovation Soc. Transitions* 37, 332–344. doi:10.1016/j.eist.2020.10.003

Kuzhabekova, A., and Kuzma, J. (2014). Mapping the emerging field of genome editing. *Technol. Analysis Strategic Manag.* 26 (3), 321–352. doi:10.1080/09537325.2013.850657

Kwok, R. (2010). Five hard truths for synthetic biology. *Nature* 463, 288–290. doi:10.1038/463288a

Lakoff, A., and Collier, S. J. (2008). *Biosecurity interventions: Global health and security in question*. New York: Columbia University Press.

Lindahl, A. L., Olsson, M. E., Mercke, P., Tollbom, Ö., Schelin, J., Brodelius, M., et al. (2006). Production of the artemisinin precursor amorpha-4,11-diene by engineered *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 28 (8), 571–580. doi:10.1007/s10529-006-0015-6

Link, H. J. (2013). Playing god and the intrinsic value of life: Moral problems for synthetic biology? *Sci. Eng. Ethics* 19 (2), 435–448. doi:10.1007/s11948-012-9353-z

Liu, J., Li, J., Arnold, K., Pawlinski, R., and Key, N. S. (2020). Using heparin molecules to manage COVID-2019. *Res. Pract. thrombosis haemostasis* 4 (4), 518–523. doi:10.1002/rth2.12353

Marris, C., Balmert, A., Calvert, J., Molyneux-Hodgson, S., Frow, E., Kearnes, M., et al. (2015). Taking roles in interdisciplinary collaborations: Reflections on working in post-ELSI spaces in the UK synthetic biology community. *Sci. Technol. Stud.* 28 (3), 3–25. doi:10.23987/sts.55340

Martin, V. J. J., Piteral, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21 (7), 796–802. doi:10.1038/nbt833

Merritt, T., Hamidi, F., Alistar, M., and DeMenezes, M. (2020). Living media interfaces: A multi-perspective analysis of biological materials for interaction. *Digit. Creat.* 31 (1), 1–21. doi:10.1080/14626268.2019.1707231

Meyer, M., and Vergnaud, F. (2020). The rise of biohacking: Tracing the emergence and evolution of DIY biology through online discussions. *Technol. Forecast. Soc. Change* 160, 120206. doi:10.1016/j.techfore.2020.120206

Molyneux-Hodgson, S., and Meyer, M. (2009). Tales of emergence—Synthetic biology as a scientific community in the making. *BioSocieties* 4 (2-3), 129–145. doi:10.1017/S1745855209990019

Morris, S., DeYong, C., Wu, Z., Salman, S., and Yemenu, D. (2002). Diva: A visualization system for exploring document databases for technology forecasting. *Comput. Industrial Eng.* 43 (4), 841–862. doi:10.1016/S0360-8352(02)00143-2

Nadesan, M. (2022). Crises narratives defining the COVID-19 pandemic: Expert uncertainties and conspiratorial sensemaking. *Am. Behav. Sci.* 2022, 1085893. doi:10.1177/00027642221085893

National Research Council (2014). *Convergence: Facilitating transdisciplinary integration of life sciences, physical sciences, engineering, and beyond*. Washington: National Academies Press, 18–20.

Nawaz, M. A., Anwar, M. A., Zhou, R., Nawaz, A., Dong, W., and Asmi, F. (2019). Factors influencing willingness to consume GMF in Chinese population: The moderating role of'Information literacy'. JAPS. *J. Animal Plant Sci.* 29 (4), 1.

Nordmann, A. (2015). "Synthetic biology at the limits of science," in *Synthetic biology* (Cham: Springer), 31–58. doi:10.1007/978-3-319-02783-8_2

O'Malley, A. M., Powell, A., Davies, J. F., and Calvert, J. (2008). Knowledge-making distinctions in synthetic biology. *BioEssays* 30 (1), 57–65. doi:10.1002/bies.20664

O'Malley, M. A., Calvert, J., and Dupré, J. (2007). The study of socioethical issues in systems biology. *Am. J. Bioeth.* 7 (4), 67–78. doi:10.1080/15265160701221285

Owen, R., Macnaghten, P., and Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Sci. Public Policy* 39 (6), 751–760. doi:10.1093/scipol/scs093

Pansera, M., Owen, R., Meacham, D., and Kuh, V. (2020). Embedding responsible innovation within synthetic biology research and innovation: Insights from a UK multi-disciplinary research centre. *J. Responsible Innovation* 7 (3), 384–409. doi:10.1080/23299460.2020.1785678

Parker, M. T., and Kunjapur, A. M. (2020). Deployment of engineered microbes: Contributions to the bioeconomy and considerations for biosecurity. *Health Secur.* 18 (4), 278–296. doi:10.1089/hs.2020.0010

Petrović, M., and Žeželj, I. (2022). Both a bioweapon and a hoax: The curious case of contradictory conspiracy theories about COVID-19. *Think. Reason.*, 1–32. doi:10.1080/13546783.2022.2088618

Pierce, R. L. (2012). Whose ethics of knowledge? Taking the next step in evaluating knowledge in synthetic biology: A response to douglas and Savulescu. *J. Med. Ethics* 38 (10), 636–638. doi:10.1136/medethics-2012-100563

Porter, A. L., Youtie, J., Shapira, P., and Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *J. Nanoparticle Res.* 10 (5), 715–728. doi:10.1007/s11051-007-9266-y

Radakovits, R., Jinkerson, R. E., Darzins, A., and Posewitz, M. C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* 9 (4), 486–501. doi:10.1128/EC.00364-09

Rappuoli, R., De Gregorio, E., Del Giudice, G., Phogat, S., Pecetta, S., Pizza, M., et al. (2021). Vaccinology in the post– COVID-19 era. *Proc. Natl. Acad. Sci.* 118 (3), e2020368118. doi:10.1073/pnas.2020368118

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., et al. (2017). Design of a synthetic yeast genome. *Science* 355 (6329), 1040–1044. doi:10.1126/science.aaf4557

Romer, D., and Jamieson, K. H. (2020). Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. *Soc. Sci. Med.* 263, 113356. doi:10.1016/j.socscimed.2020.113356

Saukshmya, T., and Chugh, A. (2010). Commercializing synthetic biology: Socio-ethical concerns and challenges under intellectual property regime. *J. Commer. Biotechnol.* 16, 135–158. doi:10.1057/jcb.2009.28

Schmidt, M. (2008). Diffusion of synthetic biology: A challenge to biosafety. *Syst. Synthetic Biol.* 2 (1), 1–6. doi:10.1007/s11693-008-9018-z

Serrano, L. (2007). Synthetic biology: Promises and challenges. *Mol. Syst. Biol.* 3 (1), 158. doi:10.1038/msb4100202

Seyfried, G., Pei, L., and Schmidt, M. (2014). European do-it-yourself (DIY) biology: Beyond the hope, hype and horror. *Bioessays* 36 (6), 548–551. doi:10.1002/bies.201300149

Shapira, P., Kwon, S., and Youtie, J. (2017). Tracking the emergence of synthetic biology. *Scientometrics* 112 (3), 1439–1469. doi:10.1007/s11192-017-2452-5

Sharp, P., Wong, S. T. .C., Newcomb, R., and Hafliger, P. (2011). The third revolution: The convergence of the life sciences, physical sciences, and engineering. *MIT whitepaper* 9, 4. doi:10.1109/MCAS.2012.2206989

Small, H., Boyack, K. W., and Klavans, R. (2014). Identifying emerging topics in science and technology. *Res. Policy* 43 (8), 1450–1467. doi:10.1016/j.respol.2014.02.005

Stilgoe, J., Owen, R., and Macnaghten, P. (2020). "Developing a framework for responsible innovation," in *The ethics of nanotechnology, geoengineering and clean energy* (Milton Park: Routledge), 347–359.

Stilgoe, J. R. O., Macnaghten, P., and Macnaghten, P. (2013). Developing a framework for responsible innovation. *Res. Policy* 42, 1568–1580. doi:10.1016/j.respol.2013.05.008

Thompson, P. B. (2012). Synthetic biology needs a synthetic bioethics. *Ethics, Policy and Environ.* 15 (1), 1–20. doi:10.1080/21550085.2012.672676

B. D. Trump, C. L. Cummings, J. Kuzma, and I. Linkov (2019). *Synthetic biology 2020: Frontiers in risk analysis and governance* (Germany: Springer Nature).

Vaage, N. S. (2020). Living machines: Metaphors we live by. *NanoEthics* 14 (1), 57–70. doi:10.1007/s11569-019-00355-2

van Eck, N. J., and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84 (2), 523–538. doi:10.1007/s11192-009-0146-3

Van Eck, N. J., and Waltman, L. (2011). Text mining and visualization using VOSviewer. arXiv preprint arXiv:1109.2058.

Vermeulen, N., Haddow, G., Seymour, T., Faulkner-Jones, A., and Shu, W. (2017). 3D bioprint me: A socioethical view of bioprinting human organs and tissues. *J. Med. Ethics* 43 (9), 618–624. doi:10.1136/medethics-2015-103347

Vermeulen, N. (2018). The choreography of a new research field: Aggregation, circulation and oscillation. *Environ. Plan. A Econ. Space* 50 (8), 1764–1784. doi:10.1177/0308518X17725317

Voigt, C. A. (2020). Synthetic biology 2020–2030: Six commercially-available products that are changing our world. *Nat. Commun.* 11 (1), 6379–6386. doi:10.1038/s41467-020-20122-2

Wang, H., and Yang, H. (2019). Gene-edited babies: What went wrong and what could go wrong. *PLoS Biol.* 17 (4), e3000224. doi:10.1371/journal.pbio.3000224

Warmbrod, K. L., Trotochaud, M., and Gronvall, G. K. (2020). iGEM and the biotechnology workforce of the future. *Health Secur.* 18 (4), 303–309. doi:10.1089/hs.2020.0017

Weber, W., and Fussenegger, M. (2012). Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.* 13, 21–35. doi:10.1038/nrg3094

WHO (2006). Biorisk management: Laboratory biosecurity guidance. Available at: http://www.who.int/ihr/publications/WHO_CDS_EPR_2006_6.

Wiek, A., Guston, D., Frow, E., and Calvert, J. (2012). Sustainability and anticipatory governance in synthetic biology. *Int. J. Soc. Ecol. Sustain. Dev. (IJSESD)* 3 (2), 25–38. doi:10.4018/jsesd.2012040103

Zhang, X., Guo, Y., Zhou, Q., Tan, Z., and Cao, J. (2021). The mediating roles of medical mistrust, knowledge, confidence and complacency of vaccines in the pathways from conspiracy beliefs to vaccine hesitancy. *Vaccines* 9 (11), 1342. doi:10.3390/vaccines9111342

Zheng, T., Zhang, M., Wu, L., Guo, S., Liu, X., Zhao, J., et al. (2022). Upcycling CO2 into energy-rich long-chain compounds via electrochemical and metabolic engineering. *Nat. Catal.* 5 (5), 4838–5396. doi:10.1038/s41929-022-00775-6

Zhou, R., Wang, D., Siddiquei, A. N., Anwar, M. A., Hammad, A., Asmi, F., et al. (2019). GMO/GMF on social media in China: Jagged landscape of information seeking and sharing behavior through a valence view. *Int. J. Environ. Res. Public Health* 16 (23), 4838. doi:10.3390/ijerph16234838

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 01 June 2023
DOI 10.3389/fbioe.2023.1161911

OPEN ACCESS

EDITED BY
Hyun-Dong Shin,
Bereum Co., Ltd., Republic of Korea

REVIEWED BY
Mostafa Elshobary,
Tanta University, Egypt
Ranjith Kumar Manoharan,
Yeungnam University, Republic of Korea

*CORRESPONDENCE
Wael A. Fathy,
wael.ahmed@science.bsu.edu.eg

RECEIVED 09 February 2023
ACCEPTED 23 May 2023
PUBLISHED 01 June 2023

CITATION
Fathy WA, AbdElgawad H, Essawy EA, Tawfik E, Abdelhameed MS, Hammouda O, Korany SM and Elsayed KNM (2023), Glycine differentially improved the growth and biochemical composition of *Synechocystis* sp. PAK13 and *Chlorella variabilis* DT025. *Front. Bioeng. Biotechnol.* 11:1161911. doi: 10.3389/fbioe.2023.1161911

Glycine differentially improved the growth and biochemical composition of *Synechocystis* sp. PAK13 and *Chlorella variabilis* DT025

Wael A. Fathy[1]*, Hamada AbdElgawad[1,2], Ehab A. Essawy[3], Eman Tawfik[4], Mohamed S. Abdelhameed[1], Ola Hammouda[1], Shereen Magdy Korany[5] and Khaled N. M. Elsayed[1]

[1]Botany and Microbiology Department, Faculty of Science, Beni-Suef University, Beni-Suef, Egypt, [2]Integrated Molecular Plant Physiology Research, Department of Biology, University of Antwerp, Antwerpen, Belgium, [3]Biochemistry Division, Chemistry Department, Faculty of Science, Helwan University, Helwan, Egypt, [4]Botany and Microbiology Department, Faculty of Science, Helwan University, Helwan, Egypt, [5]Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

The potential of microalgae to produce valuable compounds has garnered considerable attention. However, there are various challenges that hinder their large-scale industrial utilization, such as high production costs and the complexities associated with achieving optimal growth conditions. Therefore, we investigated the effects of glycine at different concentrations on the growth and bioactive compounds production of *Synechocystis* sp. PAK13 and *Chlorella variabilis* cultivated under nitrogen availability. Glycine supplementation resulted in increased biomass and bioactive primary metabolites accumulation in both species. Sugar production, particularly glucose content, significantly improved in *Synechocystis* at 3.33 mM glycine (1.4 mg/g). This led to enhanced organic acid, particularly malic acid, and amino acids production. Glycine stress also influenced the concentration of indole-3-acetic acid, which was significantly higher in both species compared to the control. Furthermore, fatty acids content increased by 2.5-fold in *Synechocystis* and by 1.36-fold in *Chlorella*. Overall, the exogenous application of glycine is a cheap, safe, and effective approach to enhancing sustainable microalgal biomass and bioproducts production.

KEYWORDS

microalgae, mixotrophic medium, glycine, biochemical composition, fatty acid profile

Introduction

Microalgae have garnered much attention due to their ability to produce a diverse array of compounds with significant economic, medicinal, and industrial importance (Guedes et al., 2011; Santhosh et al., 2016). In this regard, microalgae have emerged as one of the most attractive and promising feedstocks for these industrial products due to their rapid growth rate, self-renewability, fast generation time, and high metabolites content (Liu et al., 2017; Fathy et al., 2021). For instance, microalgae produce astaxanthin, beta-carotene, phycocyanin, and omega-3 fatty acids (Yaakob et al., 2014; Rahman, 2020). These compounds have been shown to possess powerful

antioxidant properties (Dantas et al., 2019). Recently, microalgae have been increasingly utilized in the production of cosmetic products, such as anti-aging creams and sunscreens (Yarkent et al., 2020). They also contain organic matter, which stores solar energy as biochemical energy in the form of carbon, hydrogen, oxygen, and nitrogen components (Cuellar-Bermudez et al., 2015). Despite their immense potential for a wide range of industrial products, the use of microalgae at an industrial scale is impeded by several obstacles. The high cost of production is a significant challenge, as microalgae require a controlled environment to get high biomass accumulation. Thus achieving optimal growth conditions for microalgae outdoors is a complex process (Yin et al., 2020). Consequently, the approach of combining microalgae cultivation with media containing cheap sources of nutrients (e.g., nitrogen) provided a cost-effective and eco-friendly perspective in microalgae-based bioproducts production (Nateghpour et al., 2021; Sarma et al., 2023). In this regard, enhancing the microalgal growth and bioproducts accumulation with chemicals or environmental factors had been recently used (Maity et al., 2014; Gao et al., 2023).

Nitrogen is a constituent in all structural and functional proteins such as peptides and enzymes, as well as chlorophyll, energy transfer molecules, and genetic elements in algal cells (Cai et al., 2013; Kim et al., 2016). This indeed makes nitrogen one of the most important dietary factors for algal growth. Providing enough nitrogen through the culture medium significantly speeds up cell development and enriches the biochemical content of microalgae (Wang et al., 2013; Song et al., 2023). On the other hand, nitrogen deficiency changes the organism's metabolic route. For example, it shifted lipid metabolism away from membrane lipid production toward neutral lipid storage, resulting in the total lipid content being raised (De Bhowmick et al., 2015; Kozan et al., 2023). In this regard, most microalgae may use a range of nitrogen sources, and each source is first converted to ammonium and then metabolized into amino acids via different key metabolic enzymes (e.g., glutamine synthetase, glutamate synthase, or NADP glutamate dehydrogenase) (Cai et al., 2013; Salbitani and Carfagna, 2021). In a related study, ammonium induced a reciprocal increase in the amino acids level in *Synechocystis* sp. (Mérida et al. (1991). Through the synthesis of cyanophycin, *Synechocystis* has the remarkable capacity to internally store ammonium nitrogen (Yu et al., 2013). This nitrogen storage capacity is a promising potential for removing nitrogen from wastewater, and can be utilized as a valuable biofertilizer (Chittora et al., 2020). Increased Amino acids levels, consequently improved organic acids, and fatty acids production, where nitrogen supplementation as ammonium increased the protein content in *Dunaliella salina* (Norici et al., 2002), and lipid content in *Chlorella sorokiniana* (Wan et al., 2012). Microalgae fatty acids, for instance, are of great commercial importance in several fields, including bioenergy (Tabatabaei et al., 2011; Almomani et al., 2023). Therefore, more research is needed to increase the productivity of commercially valuable fatty acids and reduce the cost of their production. Particularly, saturated fatty acids are preferred in biodiesel production, while polyunsaturated fatty acids may have an unfavorable effect on biodiesel attributes such as ignition quality and oxidative stability (Doan et al., 2011). To this end, efficient technologies and highly productive strains were used (Maltsev and Maltseva, 2021).

Glycine, being an amino acid, contains both carbon and nitrogen that can be utilized by microalgae for many metabolic processes such as protein synthesis (De la Hoz Siegler et al., 2011; Matantseva et al., 2018). The bioavailability of glycine as a nitrogen source for microalgae is high due to its solubility and stability in the aquatic environment (Bowden et al., 2018). Glycine is readily assimilated by microalgae; thus, it can support the high growth rates and biomass production in certain microalgal species (Kim et al., 2016; Wang et al., 2016). Furthermore, glycine is a cheap and commercially available amino acid that can be easily obtained in massive quantities. Its application does not require complex and costly equipment, making it a simple and cost-effective alternative to traditional nitrogen sources such as nitrate and ammonium. This makes glycine as a potentially attractive way for large-scale microalgal cultivation.

Consequently, the cultivation of microalgae using glycine holds the potential to increase biomass production and this would also result in biomolecules accumulation including lipids, and essential amino acids. This bioproduct rich biomass has a potential as a feedstock for producing biofuel, bio fertilizers, pharmaceuticals, nutraceuticals and other bio-based products. Enhanced yield of valuable bioproducts per same microalgae biomass makes industrial production processes more efficient and profitable (Harun et al., 2010). Moreover, scaling up microalgae production on an industrial level can have cost-saving benefits, as it requires a smaller amount of cultivation space, energy, time and resources to generate the targeted level of biomolecules. These reduced production costs contribute to the increased competitiveness of microalgae-based technologies. Additionally, there is a growing market demand for products derived from microalgae. Bioactive substances such as omega-3 fatty acids, antioxidants, pigments, and biofuels are becoming more accessible and affordable (Michalak and Chojnacka, 2014; Khan et al., 2018). This accessibility has the potential to drive innovation in diverse industries including pharmaceuticals, cosmetics, and bioenergy.

The present investigation explores the impact of exogenous glycine as a nitrogenous source on the growth and bioproduct accumulation induction in the cyanobacterium *Synechocystis* sp. PAK13 and eukaryotic microalga *Chlorella variabilis*. To this end, this comprehensive study sheds light on the role of glycine in increasing the production of primary metabolites (carbohydrates, amino acids, organic acids, and lipids) and secondary metabolites such as IAA.

Materials and methods

Strains and cultural conditions

Microalgae strains *Synechocystis* sp. PAK13 and *Chlorella variabilis* DT025 were kindly provided from Algal Biotechnology Lab, Faculty of Science, Beni-Suef University, Egypt which were isolated from the marine habitat of the Red Sea and identified by using 16S and 18S rRNA gene (Supplementary Figures S1, S2). Then they were cultivated on Wuxal medium (WM), which is a universal dünger liquid plant fertilizer that has 8% N, 8% P_2O_5, 6% K_2O,

0.01% B, 0.004% Cu, 0.02% Fe, 0.012% Mn and 0.004% Zn (Wilhelm Haug GmbH and Co., KG, Germany). Our strains have exhibited favorable growth in the synthetic Wuxal medium, indicating their adaptability to the employed salt concentration. To prepare the culture medium, 800 µL of WM per liter of tap water was used. To imitate the salinity of the Red Sea habitat, 1 g/L NaCl was added to the medium. This NaCl concentration was selected based on the salinity of the natural environment from which our strains were isolated (Winckelmann et al., 2015; Fathy et al., 2020). In addition to the selected NaCl concentration, we applied the optimal neutral pH 7.5 for the microalgae culture to achieve the highest yield of growth and byproducts (Wang et al., 2010). Microalgae strains were allowed to grow in the appropriate conditions, which included a temperature of 28°C ± 2°C, and light intensity was 30.4 µmol$m^{-2}s^{-1}$, a white, fluorescent light source used with a specific wavelength of 450–650 nm, and a photoperiod of 14 h light/10 h dark cycle. The temperature of microalgae cultivation was selected according to Ras et al. (2013), where the optimal temperature for our specific strains falls within the range of 28°C–30°C and they are very close to outdoors temperatures in Egypt. The photon flux density was measured using a digital lux meter (LX1330B, China). In addition, different amounts of glycine (1.66, 3.33, 6.66, 13.33, and 26.66 mM) were added to the culture mediums. These glycine concentrations were selected based on a thorough review of the literature, e.g., (Wang et al., 2016; Yue et al., 2021), and preliminary experiments conducted in our laboratory. This was also a logarithmic scale that covers a wide range of concentrations to investigate potential dose-dependent effects while maintaining a constant ratio between each concentration.

Growth curves

In this study, a spectrophotometer (NanBei Instrument®, China) was used to measure the optical density of the microalgae samples at 700 nm. This allowed us to construct growth curves that provide insights into the sample's growth dynamics as reported by Elsayed et al. (2017). To decide the wet-weight biomass of microalgae culture, we utilized a centrifugation-based approach. Specifically, we centrifuged 5 mL of our culture, then we drained the remained liquid in an inverted position on a paper tissue for 5 min at room temperature. The weight of the wet biomass was then measured. Furthermore, the duplication time, the number of generations, and the specific growth rate of our sample were measured by the established protocols by Fabregas et al. (1986) and utilizing the methodologies described by Jin et al. (2011) and Krzemińska et al. (2014). These metrics are crucial in understanding the growth characteristics of microalgae samples and were calculated to ensure the accuracy and reproducibility of the results.

Biochemical composition investigation

Estimation of pigment content

To quantify the photosynthetic pigments in microalgae samples, the protocol of Moran and Porath (1980) was followed. Whereas, at the exponential phase of growth, 2 mL were collected from strains culture by centrifugation at 13,000 rpm. The resulting pellets were weighed and placed in a covered glass tube, which was mixed with 5 mL of 80% ice acetone at 4°C for 24 h. After centrifugation at 13,000 rpm for 3 min, the supernatants were collected for analysis. The estimation of pigments was performed following the established protocols by Metzner et al. (1965) and Pflanz and Zude (2008), while spectrophotometric measurements were measured at 480, 645, and 663 nm.

Estimation of soluble carbohydrates content

The quantitative estimation of soluble carbohydrates was performed by Yemm and Willis (1954). Whereas 2 mL of the exponential phase culture were centrifuged at 13.000 rpm and changed into glass-capped tubes soaked in 5 mL of absolute ethanol before boiling for 10 min in a water bath. In a test tube, 1 mL of the ethanolic extract was mixed with 4 mL of ice-concentrated anthrone reagent freshly made by dissolving 200 mg anthrone in 100 mL Conc. H_2SO_4 and vortexed. The tubes were then placed in a boiling water bath for 10 min before cooling in an ice bath for 5 min. Furthermore, di-saccharides were identified following van Handel (1968), whereas, to decompose reactive sucrose, 1 mL of the extract was added to 0.1 mL of 5.4 N KOH at 97°C for 10 min. The reaction mixture was then treated with 3 mL of anthrone reagent and placed in a boiling water bath for 10 min before cooling in an ice bath for 5 min. Finally, spectrophotometry was performed at 620 nm against a water reagent blank. The content of mono and di-saccharides was quantified using a standard curve of glucose and sucrose. Another standard method depends on the devices performed according to Al Jaouni et al. (2018), 2 mL microalgae was extracted in ethanol (80% v/v) by boiling for 30 s three times and once at room temperature. Derivation and resuspension of samples in dH_2O, the supernatants were kept at 20°C for more investigation. The concentration of soluble sugars was figured out (CE in a Coulter PACE system 5500) and detected using a diode array detector. Using corresponding standards (glucose, fructose, and sucrose), concentrations were calculated.

Estimation of organic acids content

Organic acids from microalgae strains in a mixture of 0.3% (w/v) butylated hydroxy anisole and 0.1% phosphoric acid were extracted. HPLC with a SUPELCOGEL C-610H column equipped with a UV detection system operating at 210 nm was used to estimate the concentrations of citric, succinic, fumaric, and malic acids (LaChromL-7455 diode array, LaChrom, Tokyo, Japan). Phosphoric acid (0.1% v/v) was eluted at a rate of 0.45 mL/min as the mobile phase. Oxalic, malic, succinic, citric, isobutyric, and fumaric acids were used as standards.

Estimation of amino acids content

Amino acids were extracted from 2 mL of microalgae cultures in an aqueous ethanol concentration of 80% (v/v). The extraction buffer contained an internal standard (norvaline) to compensate for the loss of

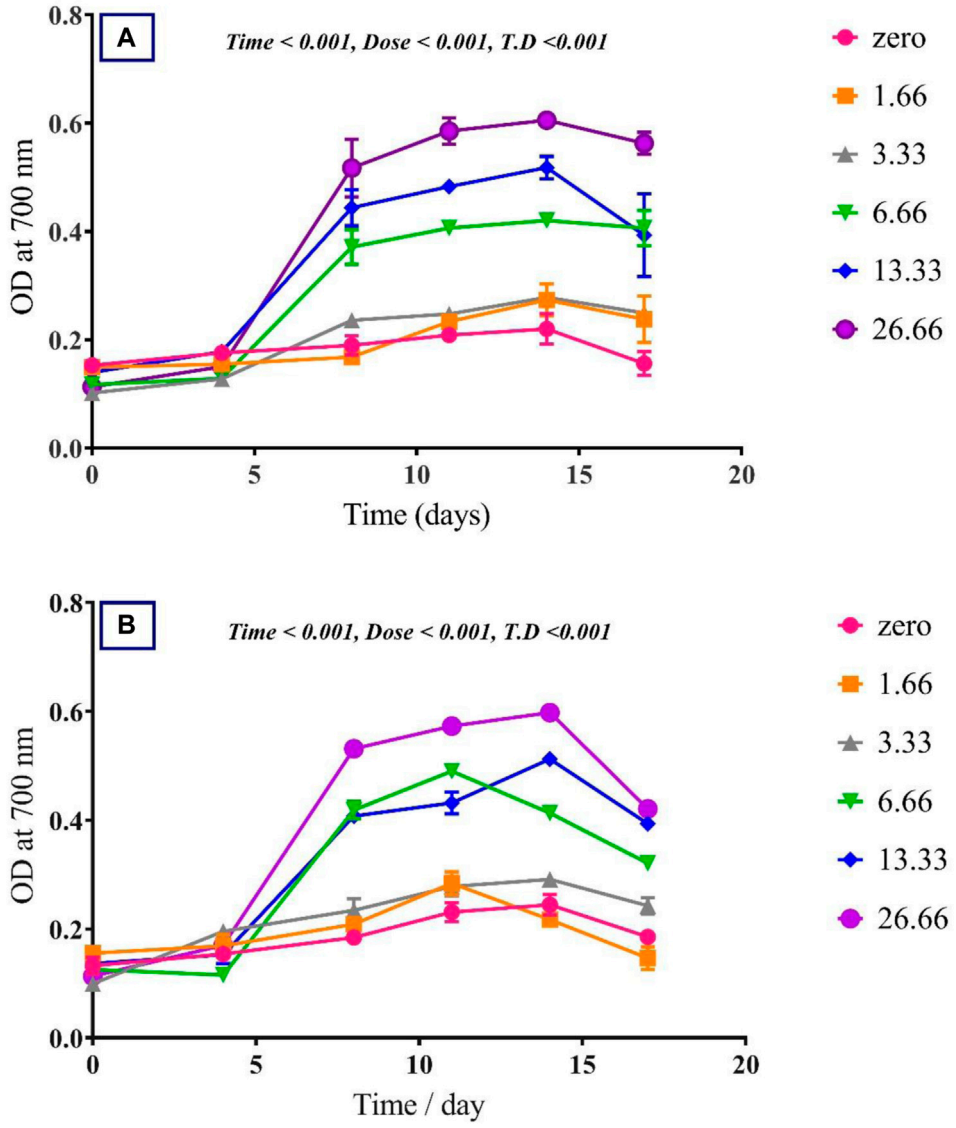


FIGURE 1
The effect of different glycine concentrations on the growth curve of **(A)** *Synechocystis* sp. and **(B)** *Chlorella* sp. Plotted points show mean daily averages ± SE, $n = 3$. The p-value is considered significant < 0.05 using t. test analysis.

amino acids during ethanol extraction and centrifugation at 20,000 rpm for 20 min. The pellet was resuspended in chloroform after the supernatant had been evaporated. The microalgae residue was re-extracted in water (HPLC grade), and the supernatant was mixed with the chloroform-suspended pellet after centrifugation. The extracts were then centrifuged and filtered through Millipore microfilters with 0.2-µm pore size. Amino acids were separated using a BEH amide (2.1 mm × 50 mm) column and quantified using a Waters Acquity UPLC-tqd mass spectrometer (Sinha et al., 2013). A list of amino acids standards (glycine, alanine, isoleucine, leucine, methionine, valine, phenylalanine, glutamine, asparagine, threonine, serine, cystine, tyrosine, lysine, histidine, arginine, glutamic acid, and aspartate) was used as standard.

Estimation of fatty acids

Fatty acids were extracted from 100 mg of microalgae biomass in 50% aqueous methanol at 25°C. For identification, a GC/MS analysis (Hewlett Packard 6890, MSD 5975 mass spectrometer, United States) with an HP-5 MS column (30 µm × 0.25 µm x 0.25 µm) was used, followed by fatty acid identification using the NIST 05 database and the Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de) (AbdElgawad et al., 2020). A list of fatty acids standards (myristic, palmitic, heptadecanoic, stearic, arachidic, docosanoic, tricosanoic, pentacosanoic, palmitoleic, heptadecenoic, oleic, linolenic, linoleic, and eicosenoic) was used as standard.

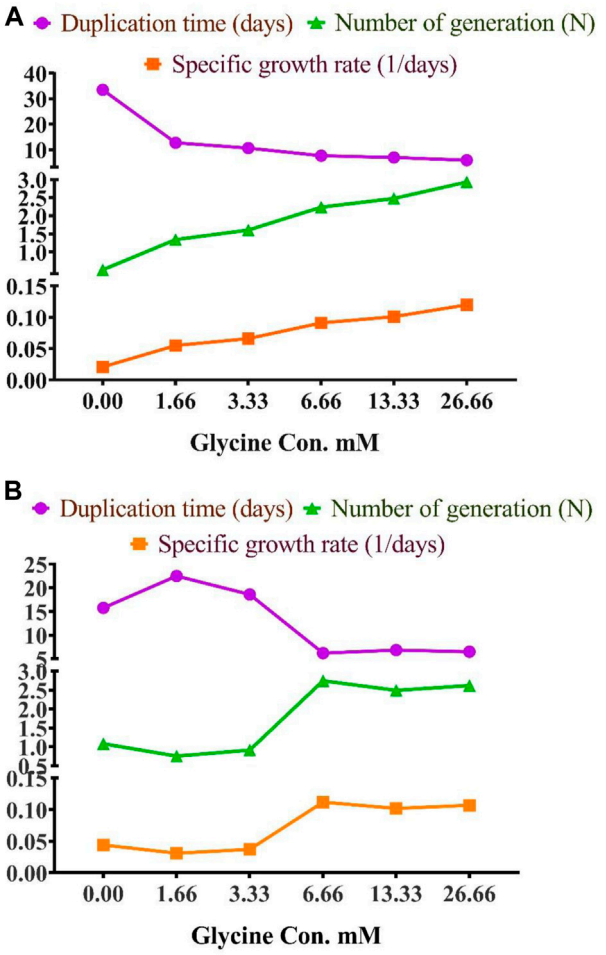


FIGURE 2
The effect of different glycine concentrations on duplication time, number of generations, and specific growth rate of **(A)** *Synechocystis* sp. and **(B)** *Chlorella* sp.

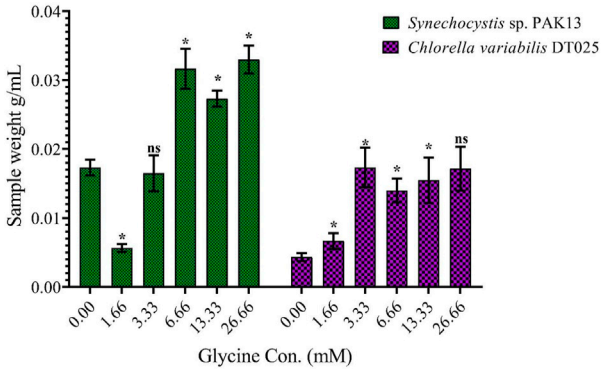


FIGURE 3
Explaining the effect of different glycine concentrations for 14 days on the wet-weight biomass of *Synechocystis* sp. and *Chlorella* sp. Data are presented as the average of three independent replicates ± SE. The statistical significances $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$ were marked by the symbols ns, *, **, ***, and ****, respectively.

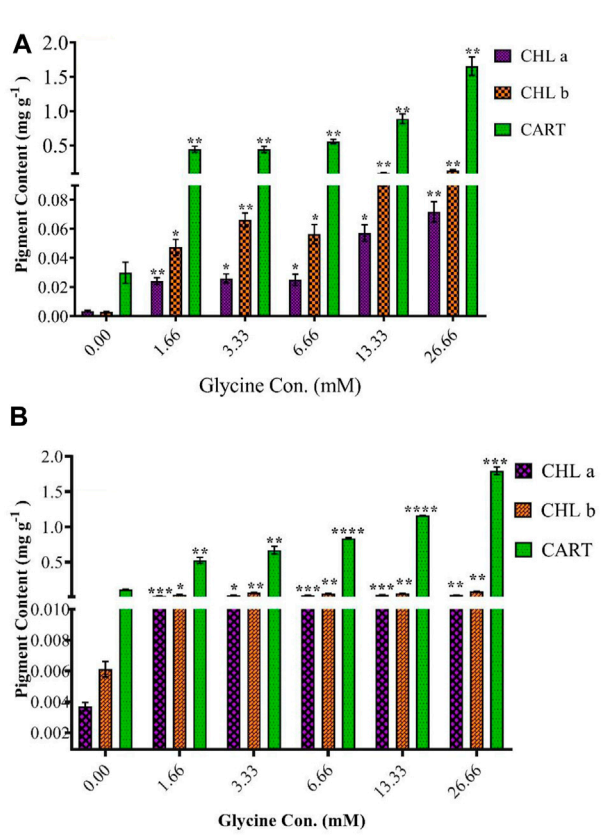


FIGURE 4
The effect of different glycine concentrations for 14 days on chlorophyll a (CHLa), chlorophyll b (CHLb), and carotenoids (CART) content in **(A)** *Synechocystis* sp. and **(B)** *Chlorella* sp. Data are presented as an average of three independent replicates ± SE. The statistical significances $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$ were marked by the symbols ns, *, **, ***, and ****, respectively.

Statistical analysis

All trials were set up in triplicate using a completely randomized design. The data are reported as (means ± standard error) and visualized using GraphPad Prism 8.4.2 software. The Student's t-test ($p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$) were used to compare each experimental value was compared to the corresponding control value. A two-way ANOVA was performed (Tukey test, $p \leq 0.05$) using the statistical package SPSS (version 20; IBM).

Results

Growth

To assess the effect of glycine on *Synechocystis* and *Chlorella*, the growth curves of the two microalgae strains were investigated (Figure 1). The growth pattern can be divided into three distinct phases based on the growth rate, i.e., lag phase, exponential phase, and stationary phase. During the lag phase, which was seen during

TABLE 1 IAA content in *Synechocystis* sp. and *Chlorella* sp. have grown under different concentrations of glycine for 14 days.

Strain	Glycine dose	IAA mean (mg/g) ± SE
Synechocystis sp. PAK13	0	0.103 ± 0.002
	1.66	0.119 ± 0.003 ^{ns}
	3.33	0.138 ± 0.004 ^{ns}
	6.66	0.175 ± 0.013*
	13.33	0.217 ± 0.006**
	26.66	0.278 ± 0.013 ^{ns}
Chlorella variabilis DT025	0	1.724 ± 0.116
	1.66	2.298 ± 0.136*
	3.33	3.958 ± 0.098****
	6.66	4.339 ± 0.123****
	13.33	5.250 ± 0.173****
	26.66	6.253 ± 0.036****

The statistical significances $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$ were marked by the symbols ns, *, **, ***, and ****, respectively.

the initial few days, *Synechocystis* and *Chlorella* acclimatized to the new environment and the growth is slow. This phase was reflected in the data as a plateau in the OD values at time points 0–4 days. The exponential phase was characterized by a rapid increase in growth rate as the microalgae population adapts to the environment and reaches its maximum growth rate. This phase was reflected in the data as an increase in OD values at 5–14 days. The stationary phase was the period in which the microalgae population reaches it carrying capacity and the growth rate becomes stable. This phase was reflected in the data as a plateau in the OD values at 14 and 17 days. Glycine at the concentrations of 6.6, 13.3, and 26.6 mM significantly increased the growth rate of both strains. On the other hand, low doses had a slight effect on growth compared to control samples. These findings were supported (Figure 2), which showed that the duplication time decreased with increasing glycine concentrations. This increased the number of generations and specific growth rate for both strains. Furthermore, estimating biomass analysis (Figure 3) revealed that glycine concentration at 26.66 mM increased *Synechocystis* and *Chlorella* biomass production by 194% and 395% respectively, compared to the control sample. Similarly, the glycine concentrations at 6.6- and 13.3-mM glycine increased the growth by 2 folds. In contrast, 1.66 and 3.33 mM of glycine decreased the *Synechocystis* biomass compared to the control growth strain. Subsequently, it boosted both strains' growth rates, and duplication time decreased as glycine concentrations increased, resulting in a rise in the number of generations and specific growth rate for both strains. Interestingly, glycine boosted biomass production more in *Chlorella* than *Synechocystis*.

Photosynthetic pigments

To understand the observed increases in growth, the photosynthetic pigment was estimated (Figure 4). In line with increased growth, the highest glycine concentration (26.66 mM) increased chlorophyll in Synechocystis and Chlorella by 19.4 and 9.3 folds, respectively. Similarly, chlorophyll b was induced to 44.3 and 13.7 folds at 26.66 mM in Synechocystis and Chlorella, respectively. To protect the photosynthesis system, carotenoids were significantly increased in Synechocystis by 55.3 folds, and in Chlorella by 16.3 folds higher than non-treated strain. In a nutshell, in conjunction with enhanced growth, the greatest glycine boosted chlorophyll a, b, and carotenoids in Synechocystis by 19.4, 44.3, and 55.3 folds higher than in non-treated strain, and these levels were more pronounced in Chlorella.

Indole-3-acetic acid

The effect of glycine on microalgae cell hormone levels was demonstrated by indole-3-acetic acid (IAA) content (Table 1). IAA levels gradually increased to 2.7 folds 26.66 mM in *Synechocystis*, which was also observed in *Chlorella* increased, which showed 3.6 times increase in IAA. In summary, glycine had a positive effect on microalgal cell hormone levels.

Soluble sugar production

Increased sugar production could reflect the improvement in photosynthesis under glycine exposure. Soluble sugar profiling revealed that glucose content was dramatically increased in *Synechocystis* with increasing glycine concentrations, specifically at 3.33 mM glycine (increased by 10.5 folds). On the other hand, glucose content was reduced in *Chlorella*. Moreover, sucrose levels in *Synechocystis* and *Chlorella* gradually were declined with increasing glycine concentrations at a range from 2.4 mg/g to 1.97 mg/g and 0.9 mg/g to 0.74 mg/g, respectively (Figure 5). We

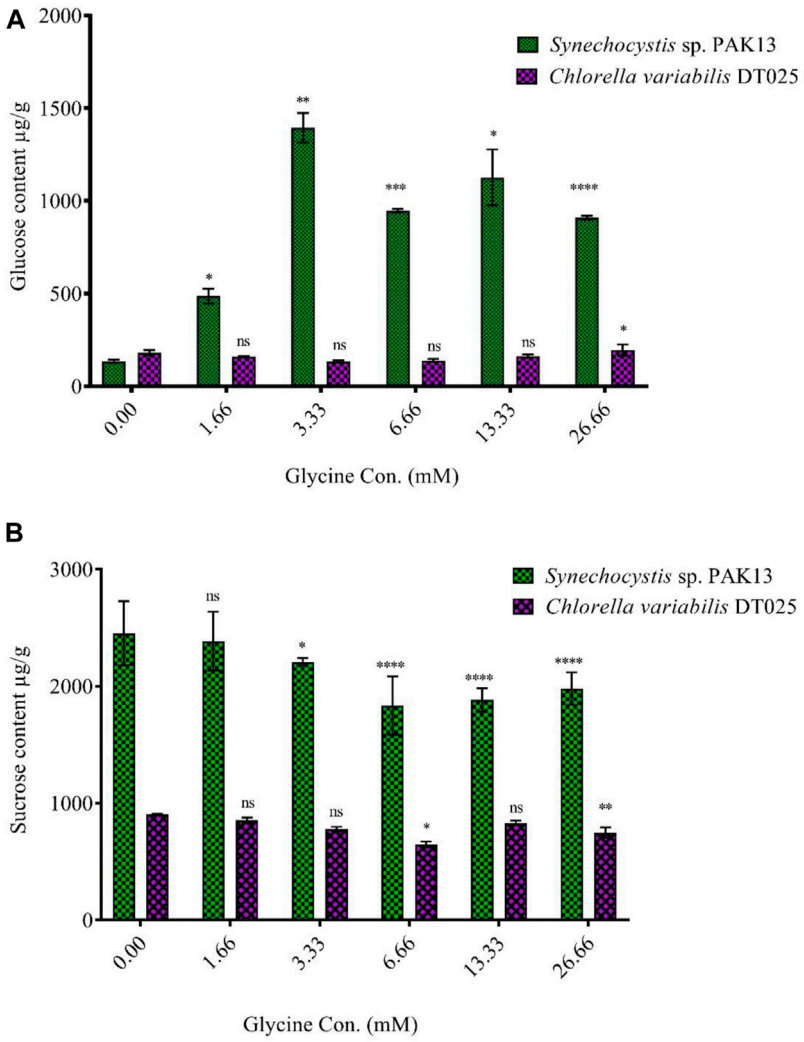


FIGURE 5
Effect of different glycine concentrations for 14 days on **(A)** glucose and **(B)** sucrose content in wet weight for *Synechocystis* sp. and *Chlorella* sp. Data are presented as the average of three independent replicates ± SE. The statistical *significances* $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and $p \leq 0.0001$ were marked by the symbols ns, *, **, ***, and ****, respectively.

further measured the fructose, total soluble sugar, and glycogen content as reported in (Table 2). For *Synechocystis,* the fructose content was highest at the 1.66 glycine dose level (1.70 mg/g) and lowest in the 26.66 glycine dose level (0.66 mg/g). The total soluble sugar content was highest at the 6.66 glycine dose level (5.44 mg/g) and lowest at the control sample (3.03 mg/g). The glycogen content was highest at the 3.33 glycine dose level (88.29 mg/g) and lowest in the 26.66 glycine dose level (29.63 mg/g). While for *Chlorella,* the fructose content was highest in the 3.33 glycine dose level (1.08 mg/g) and lowest at the 1.66 glycine dose level (0.58 mg/g). The total soluble sugars content was highest in the 3.33 glycine dose level (4.00 mg/g) and lowest at the 13.33 glycine dose level (1.38 mg/g). The glycogen content was highest in the 3.33 glycine dose level (4.08 mg/g) and lowest in the control sample (3.54 mg/g). Overall, the results suggest that glycine concentration affects the fructose, total soluble sugars, and glycogen content of these two strains in different ways.

Organic acid

The changes in the levels of sugars under glycine treatment are expected to affect the tricarboxylic acid cycle intermediates, such as organic acids. Organic acids analysis indicated that they are quite different in the two strains (Figure 6) and (Supplementary Table S1). *Synechocystis* produced much higher levels of oxalic acid, citric acid, and isobutyric acid than *Chlorella*, while *Chlorella* produced higher levels of succinic acid. In *Synechocystis*, the production of oxalic acid and citric acid was significantly decreased at higher glycine doses, while the production of isobutyric acid was increased. On the other hand, the production of malic acid and fumaric acid increases at higher glycine concentrations. For *Chlorella*, the production of oxalic acid was significantly decreased at higher glycine doses, while the production of citric acid and isobutyric acid was increased. Interestingly, total organic acids concentrations were higher in *Synechocystis*. Whereas *Synechocystis* content declined

TABLE 2 Estimating fructose, total soluble sugars, and glycogen content in *Synechocystis* sp. and *Chlorella* sp. have grown under different concentrations of glycine for 14 days.

Strain	Glycine dose	Fructose (mg/g FW)	Total S sugars (mg/g FW)	Glycogen (mg/g FW)
		Mean ± SE	Mean ± SE	Mean ± SE
Synechocystis sp. PAK13	0	1.01 ± 0.03	3.03 ± 0.02	48.44 ± 0.36
	1.66	1.70 ± 0.03***	4.22 ± 0.18 ^{ns}	71.10 ± 2.52*
	3.33	1.15 ± 0.04*	6.21 ± 0.37**	88.29 ± 4.88^{ns}
	6.66	1.49 ± 0.01*	5.44 ± 0.15*	83.21 ± 1.69**
	13.33	0.99 ± 0.03^{ns}	2.90 ± 0.05^{ns}	46.72 ± 0.37***
	26.66	0.66 ± 0.03***	1.80 ± 0.08**	29.63 ± 0.99*
Chlorella variabilis DT025	0	0.93 ± 0.03	3.48 ± 0.18	3.54 ± 0.16
	1.66	0.58 ± 0.02**	3.63 ± 0.15*	3.37 ± 0.13*
	3.33	1.08 ± 0.03***	4.00 ± 0.29*	4.08 ± 0.24^{ns}
	6.66	0.71 ± 0.03*	2.69 ± 0.17**	2.73 ± 0.15**
	13.33	0.80 ± 0.02**	1.38 ± 0.06*	1.75 ± 0.07*
	26.66	0.75 ± 0.02***	1.21 ± 0.09*	1.58 ± 0.08**

Data are presented as the average of three independent replicates ± SE. The statistical significances $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$ were marked by the symbols ns, *, ** and ***, respectively.

in the initial dosages, after that it was gradually increased to hit 153.3% more than the control strain at a glycine dose of 26.66 mM. However, organic acid content was dramatically increased in *Chlorella* with rising glycine levels, reaching 152.7% higher than the control sample at a glycine dose of 3.33 mM. These results can be mainly attributable to malic acid. Under the most severe glycine stress, organic acid concentration was equally improved in *Synechocystis* and *Chlorella*. This is due to malic acid, which was greatly influenced by exogenous glycine. This may suggest that *Synechocystis* was more effective in organic acid metabolism and production. In conclusion, our data showed that the effect of glycine was species-specific.

Amino acids

The impact of glycine provision on the amino acid profiles of the two strains was assessed. Glycine altered the levels of both nonpolar and polar amino acids. The concentrations of the nonpolar amino acidswere gradually increased with rising glycine concentration (Figure 7; Supplemenatry Table S2). Under 26.6 mM, nonpolar amino acids were increased to hit 4.8 folds in *Synechocystis* glycine compared to those in the non-treated strain. While *Chlorella* showed an increase of 2.2 folds higher than the control strain under 26.6 mM glycine. Consequently, it was obvious from the table interpretation that glycine was the main nonpolar amino acid affected by glycine exposure. Where 70% of the nonpolar amino acids content, specifically in *Synechocystis* was increased by 5.93 folds higher than the control one at the dose of 26.66 mM glycine. Overall, it can be observed that both *Synechocystis* and *Chlorella* had an increased nonpolar amino acids content with increasing glycine concentration. This trend is observed for most tested amino acids, including alanine, isoleucine, leucine, methionine and valine. Interestingly, *Synechocystis* had a much larger increase in nonpolar amino acid content with increasing glycine concentration. For example, at a glycine concentration of 26.66 mM, *Synechocystis* had a total nonpolar amino acid content of 144.47 mg/g, while *Chlorella* had a total nonpolar amino acid content of only 81.20 mg/g. Phenylalanine content also showed a different pattern, *Synechocystis* had a slight increase in phenylalanine content, while *Chlorella* showed a more significant increase in phenylalanine content at 13.33 mM and 26.66 mM. On the other hand, polar amino acids in *Synechocystis* (Figure 7) and (Supplementary Table S3) were significantly declined under all glycine dosages. For instance, at 26.66 mM, glycine was reduced by 3.5 folds compared to the control strain. Moreover, polar amino acids were decreased in *Chlorella*, with the most glycine doses around half the control one, except for 1.6- and 26.6-mM glycine. This suggested that the polar amino acids content in both species varies depending on the concentration of glycine. For example, in *Synechocystis*, the total polar amino acids content was decreased from 9.43 at the control to 7.79 at 13.33 mM glycine and it was decreased to 2.751 at 26.66 mM glycine. While, in *Chlorella*, the total polar amino acids content was steadily increased from 4.89 at control to 8.97 at 26.66 mM glycine. The differences in the specific polar amino acid content between the two species. For example, at control conditions, *Synechocystis* had a higher content of glutamine, asparagine, and tyrosine, while *Chlorella* had a higher content of threonine, serine and cysteine. However, as the glycine concentration was increased, the differences between the two species became less pronounced.

Furthermore, basic and acidic amino acid analyses revealed a steady state under all glycine treatments, while the greatest basic amino acid concentration in *Synechocystis* was 1.3 folds greater at

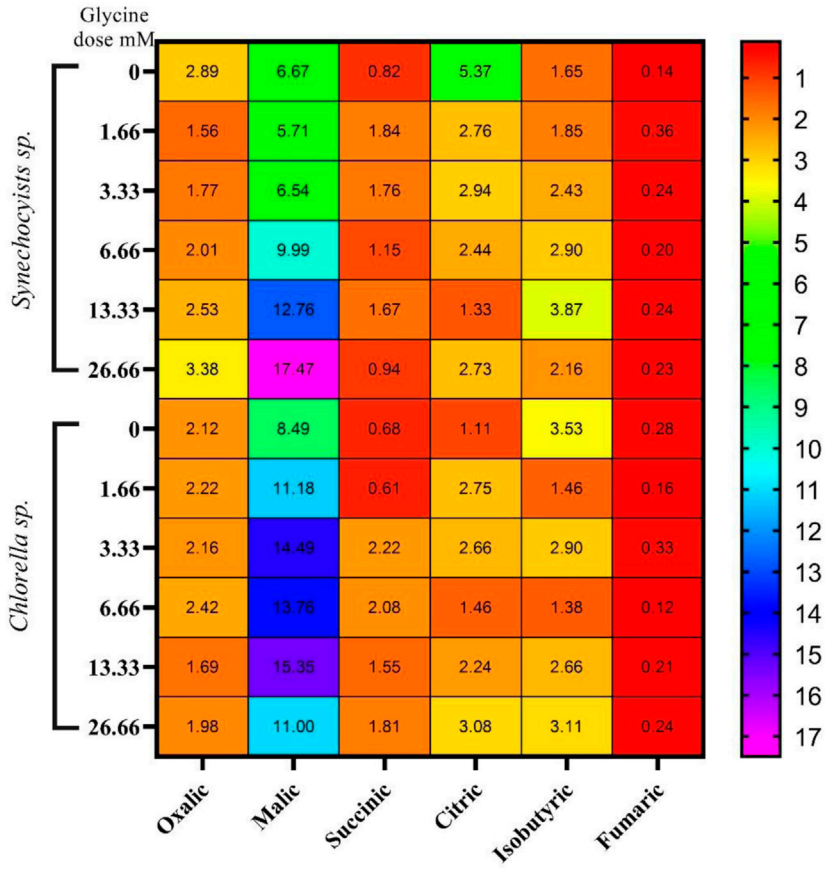


FIGURE 6
The heat map represents the effect of different glycine concentrations for 14 days on organic acids content (mg/g) wet-weight biomass *Synechocystis* sp. and *Chlorella* sp. Data are presented as the average of three independent replicates.

6.6 mM glycine (Figure 7; Supplementary Table 4). However, this was slightly different, but it was 1.5 folds higher at 3.3 mM glycine in *Chlorella*. The basic amino acids content was generally increased, except for lysine at the highest dose of 13.33, where it was significantly decreases. The acidic amino acids content, on the other hand, showed no consistent trend across the different glycine doses. For *Chlorella*, the results showed that the increased glycine dose, generally increased the basic amino acids content, except for histidine at 26.66 mM, where it was decreased significantly. The acidic amino acid content also showed no consistent trend across the different glycine doses.

Fatty acids

The induced effects of glycine could also be extended to include the tricarboxylic acid cycle intermediates, such as fatty acids. Thus, we measured the individual and total fatty acids (saturated and unsaturated), to get an overview of their potential for biodiesel production. To demonstrate these changes, we measured saturated fatty acids levels (Table 3). SFA levels were steadily increased, whereas SFA levels in *Synechocystis* reached 247% more than the control sample at 26.66 mM, and this was also observed in *Chlorella* (242% increase compared to control).

The results showed that the total SFA content increased with increasing glycine concentration in both microalgal species. For instance, the total SFA in *Synechocystis* was increased from 18.71 to 46.20 mg/g, at 26.66 mM. While in *Chlorella* increased from 19.62 to 33.73 mg/g, over the same range of glycine concentrations. This indicated that glycine positively influenced the SFA synthesis in these microalgae. Regarding individual SFA, the results showed that the two microalgal species showed different types and amounts of FA produced in response to glycine. In *Synechocystis*, the major produced SFAs were palmitic (C16:0) and myristic (C14:0) acids. However, the content of heptadecanoic (C17:0) and stearic (C18:0) acids did not show significant changes, while arachidic (C20:0), docosanoic (C22:0), tricosanoic (C23:0), and pentacosanoic (C25:0) acids were significantly increased only at 6.66 and 13.33 mM. In contrast, in *Chlorella*, the major produced SFAs were palmitic (C16:0) and stearic (C18:0) acids. The content of myristic (C14:0) acid did not show significant changes, but heptadecanoic (C17:0), arachidic (C20:0), docosanoic (C22:0), tricosanoic (C23:0), and pentacosanoic (C25:0) acids were significantly increased only 13.33 mM. Overall, the results suggested that glycine supplementation differentially enhanced the production of SFA in both species.

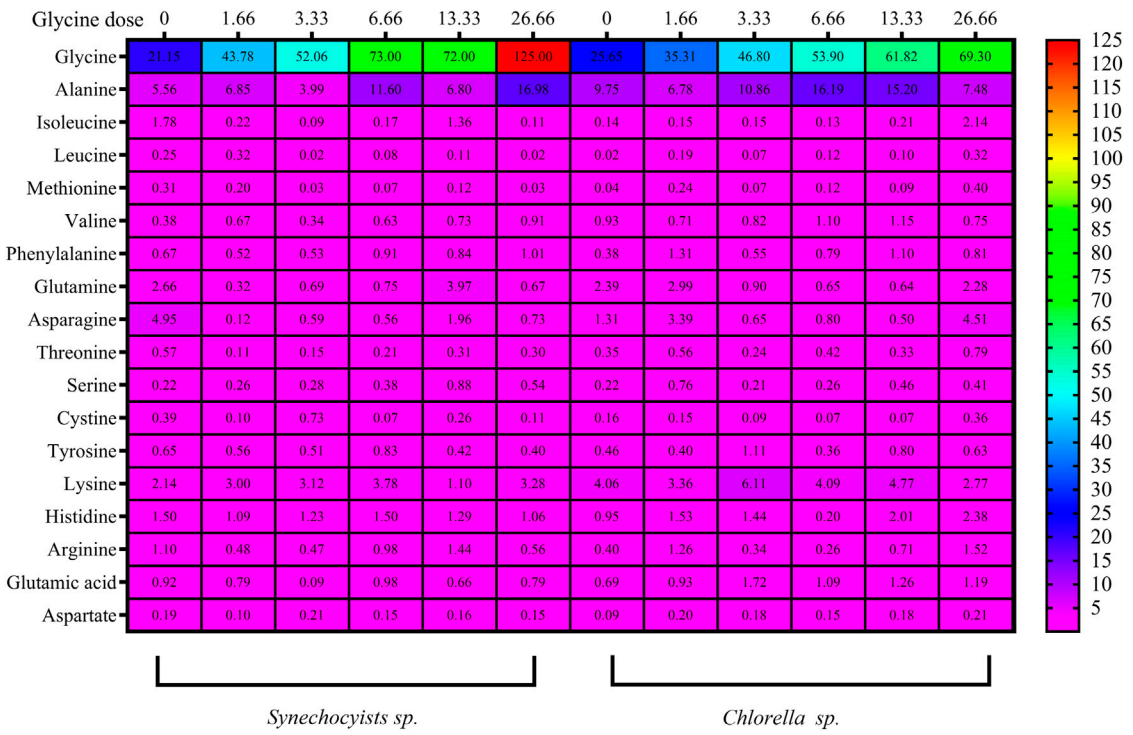


FIGURE 7
The heat map represents the effect of different glycine concentrations for 14 days on polar and nonpolar amino acids, besides acidic and basic amino acids content (mg/g) in *Synechocystis* sp. and *Chlorella* sp. Data are presented as the average of three independent replicates.

Furthermore, unsaturated fatty acids (UFA) demonstrate that at a dose of 26.66 mM glycine, UFA in *Synechocystis* was increased to 250% greater than the control strain (Table 4). At 26.66 Mm, the UFA in *Chlorella* increased by 136%. In *Synechocystis* 26.66 mM of glycine led to the highest total fatty acid content (174.14 mg/g). The content of UFA at the highest dose led to the highest content of UFAs (127.943 mg/g). Among the UFAs, oleic acid (C18:1) showed the highest content in all samples, ranging from 27.631 to 72.671 mg/g. The other UFAs showed a slight increase, with linolenic acid (C18:3) of the most significant increase. While in *Chlorella* the total fatty acids content was also increased. At 26.66 mM lead to the highest total fatty acids content (118.54 mg/g). At UFAs levels, the content of palmitoleic acid (C16:1) was decreased, while the content of heptadecenoic acid (C17:1) showed a slight increase. Oleic acid (C18:1) showed a slight decrease, while linolenic acid (C18:3) and linoleic acid (C18:2) showed an increase. Overall, our findings suggested that glycine-treated *Chlorella* or *Synechocystis* cultivation can be used to produce biofuels and other bioproducts.

Discussion

Glycine improved biomass accumulation

The biomass results revealed that glycine at 6.6- and 13.3 and/or 26.66 mM increased biomass production in *Synechocystis* and *Chlorella*, compared to the control sample. In contrast, 1.66 and 3.33 mM of glycine decreased *Synechocystis* biomass compared to the control growth strain. To understand this increase in the biomass accumulation, the growth curves of both *Synechocystis* and *Chlorella* strains were analyzed. During the lag phase, the microalgae acclimatized to the new environment, resulting in slow growth. The exponential phase was characterized by a rapid increase in growth rate. The results of the present study showed that glycine significantly increased the growth rate of both strains, particularly at 6.6, 13.3, and 26.6 mM. The low concentration of glycine had a slight effect on growth compared to control samples. These findings were supported by the results of glycine concentration, which showed that the duplication time decreased with increasing glycine concentrations, increasing the number of generations and specific growth rate for both strains.

Furthermore, at high concentrations glycine showed the fastest cell growth rate. These findings were previously validated by Kumar et al. (2014) who reported that glycine treatment increased *Chlorella* cell production, furthermore, both biomass and growth rate were much improved. The highest concentration was higher (18.95%) than the control value (Garcıa et al., 2005). Similar to our results, Wang et al. (2016) found that *Chlorella sorokiniana* dry weight increased in a dose-dependent manner with biomass to 121% compared to the control. We hypothesized, that glycine boosts biomass and biomolecule productivity by stabilizing and improving the C4 system (Sun et al. (2016). Thus, enhancing photosynthetic

TABLE 3 Estimating saturated fatty acids content (mg/g) in *Synechocystis* sp. and *Chlorella* sp. have grown under different concentrations of glycine for 14 days.

Strain	Glycine dose	Myristic (C14:0)	Palmitic (C16:0)	Heptadecanoic (C17:0)	Stearic (C18:0)	Arachidic (C20:0)	Docosanoic (C22:0)	Tricosanoic (C23:0)	Pentacosanoic (C25:0)	Sum of saturated FA ± SE
		Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	
Synechocystis sp. PAK13	0	0.332 ± 0.005	15.193 ± 0.333	0.025 ± 0.001	1.393 ± 0.092	1.136 ± 0.013	0.609 ± 0.020	0.020 ± 0.002	0.002 ± 0.0002	18.71 ± 0.47
	1.66	0.366 ± 0.002 ^{ns}	17.590 ± 0.507 ^{ns}	0.024 ± 0.001^{ns}	1.421 ± 0.079 ^{ns}	1.041 ± 0.020 ^{ns}	0.486 ± 0.022 ^{ns}	0.020 ± 0.002 ^{ns}	0.002 ± 0.0002 ^{ns}	20.95 ± 0.63
	3.33	0.564 ± 0.001**	20.286 ± 0.656 ^{ns}	0.032 ± 0.001^{ns}	1.778 ± 0.095 ^{ns}	0.725 ± 0.016**	0.404 ± 0.083 ^{ns}	0.026 ± 0.003 ^{ns}	0.003 ± 0.0003 ^{ns}	23.82 ± 0.86
	6.66	0.672 ± 0.012**	25.701 ± 1.960 ^{ns}	0.047 ± 0.002^{ns}	2.591 ± 0.092*	1.613 ± 0.150 ^{ns}	0.531 ± 0.056 ^{ns}	0.037 ± 0.003 ^{ns}	0.002 ± 0.0003 ^{ns}	31.19 ± 2.28
	13.33	0.772 ± 0.004****	32.081 ± 0.975*	0.067 ± 0.002*	2.014 ± 0.112 ^{ns}	1.799 ± 0.015***	0.406 ± 0.019*	0.047 ± 0.005 ^{ns}	0.005 ± 0.0005 ^{ns}	37.19 ± 1.13
	26.66	0.870 ± 0.019**	41.207 ± 2.038*	0.088 ± 0.002**	1.426 ± 0.140 ^{ns}	2.124 ± 0.183 ^{ns}	0.433 ± 0.012 ^{ns}	0.045 ± 0.013 ^{ns}	0.007 ± 0.0008 ^{ns}	46.20 ± 2.41
Chlorella variabilis DT025	0	0.313 ± 0.003	16.254 ± 0.597	0.061 ± 0.003	1.262 ± 0.054	1.225 ± 0.043	0.462 ± 0.077	0.040 ± 0.005	0.004 ± 0.0005	19.62 ± 0.78
	1.66	0.307 ± 0.005 ^{ns}	17.541 ± 0.733 ^{ns}	0.028 ± 0.001*	1.099 ± 0.056 ^{ns}	1.033 ± 0.029 ^{ns}	0.455 ± 0.024 ^{ns}	0.021 ± 0.002 ^{ns}	0.002 ± 0.0002 ^{ns}	20.49 ± 0.85
	3.33	0.572 ± 0.024 ^{ns}	24.133 ± 1.625 ^{ns}	0.084 ± 0.006 ^{ns}	1.094 ± 0.027 ^{ns}	1.370 ± 0.102 ^{ns}	0.680 ± 0.071 ^{ns}	0.036 ± 0.005 ^{ns}	0.004 ± 0.0005 ^{ns}	27.97 ± 1.86
	6.66	0.912 ± 0.008***	28.813 ± 0.806**	0.098 ± 0.004 ^{ns}	2.301 ± 0.098*	0.924 ± 0.032 ^{ns}	0.610 ± 0.039 ^{ns}	0.066 ± 0.008 ^{ns}	0.007 ± 0.0008 ^{ns}	33.73 ± 1.00
	13.33	0.733 ± 0.004****	31.220 ± 1.012*	0.072 ± 0.002 ^{ns}	2.219 ± 0.114 ^{ns}	1.597 ± 0.179 ^{ns}	1.018 ± 0.041 ^{ns}	0.037 ± 0.002 ^{ns}	0.004 ± 0.0004 ^{ns}	36.90 ± 1.35
	26.66	0.554 ± 0.014*	40.354 ± 0.904**	0.045 ± 0.001 ^{ns}	2.822 ± 0.070**	2.152 ± 0.390 ^{ns}	1.424 ± 0.048*	0.023 ± 0.002 ^{ns}	0.004 ± 0.0009 ^{ns}	47.38 ± 1.43

Data are presented as the average of three independent replicates ± SE. The statistical significances $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$ were marked by the symbols ^{ns}, *, ** and ***, ****, respectively.

10.3389/fbioe.2023.1161911

TABLE 4 Estimating unsaturated fatty acids and the total amount of fatty acids content (mg/g) in *Synechocystis* sp. and *Chlorella* sp. have grown under different concentrations of glycine for 14 days.

Strain	Dose	Palmitoleic (C16:1)	Heptadecenoic (C17:1)	Oleic (C18:1)	Linolenic (C18:3)	Linoleic (C18:2)	Eicosenoic (C20:1)	Sum of unsaturated FA ± SE	Total FA
		Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE		
Synechocystis sp. PAK13	0	0.06 ± 0.008	0.138 ± 0.013	27.631 ± 1.424	4.719 ± 0.156	17.780 ± 0.879	0.876 ± 0.038	51.208 ± 2.517	69.92 ± 2.98
	1.66	0.06 ± 0.007 ^{ns}	0.136 ± 0.013 ^{ns}	30.253 ± 1.039 ^{ns}	6.184 ± 0.303 ^{ns}	19.703 ± 1.438 ^{ns}	0.772 ± 0.056 ^{ns}	57.110 ± 2.857	78.06 ± 3.49
	3.33	0.06 ± 0.011 ^{ns}	0.123 ± 0.012 ^{ns}	41.040 ± 1.004*	7.325 ± 0.201*	21.082 ± 1.032 ^{ns}	0.841 ± 0.038 ^{ns}	70.478 ± 2.298	94.30 ± 3.15
	6.66	0.082 ± 0.007*	0.183 ± 0.019 ^{ns}	47.114 ± 0.387*	8.381 ± 0.525 ^{ns}	25.904 ± 2.137 ^{ns}	1.013 ± 0.084 ^{ns}	82.678 ± 3.159	113.87 ± 5.44
	13.33	0.124 ± 0.014**	0.167 ± 0.016 ^{ns}	49.564 ± 2.531 ^{ns}	8.410 ± 0.399 ^{ns}	32.773 ± 2.442 ^{ns}	1.323 ± 0.097 ^{ns}	92.362 ± 5.499	129.55 ± 6.63
	26.66	0.166 ± 0.021 ^{ns}	0.150 ± 0.014 ^{ns}	72.671 ± 1.675**	14.651 ± 0.834*	38.714 ± 4.579 ^{ns}	1.592 ± 0.191 ^{ns}	127.943 ± 7.313	174.14 ± 9.72
Chlorella variabilis DT025	0	0.093 ± 0.004	0.148 ± 0.015	33.155 ± 1.589	7.239 ± 0.285	25.670 ± 0.615	1.023 ± 0.020	67.328 ± 2.527	86.95 ± 3.31
	1.66	0.070 ± 0.006*	0.134 ± 0.008 ^{ns}	32.987 ± 0.711 ^{ns}	6.042 ± 0.349 ^{ns}	21.021 ± 1.208*	0.832 ± 0.048 ^{ns}	61.086 ± 2.331	81.57 ± 3.18
	3.33	0.111 ± 0.006 ^{ns}	0.150 ± 0.019 ^{ns}	46.128 ± 1.446***	6.968 ± 0.510 ^{ns}	24.407 ± 1.403*	0.968 ± 0.050 ^{ns}	78.732 ± 3.434	106.70 ± 5.29
	6.66	0.206 ± 0.020 ^{ns}	0.128 ± 0.013 ^{ns}	50.098 ± 3.236*	6.422 ± 0.466*	23.040 ± 1.526 ^{ns}	0.923 ± 0.059 ^{ns}	80.818 ± 5.319	114.55 ± 6.32
	13.33	0.109 ± 0.003***	0.197 ± 0.017 ^{ns}	35.187 ± 1.463**	9.243 ± 0.666 ^{ns}	29.286 ± 1.998*	1.292 ± 0.089 ^{ns}	75.315 ± 4.236	112.21 ± 5.59
	26.66	0.040 ± 0.009**	0.267 ± 0.021 ^{ns}	20.011 ± 0.481 ^{ns}	12.060 ± 0.878**	37.122 ± 1.469**	1.660 ± 0.119 ^{ns}	71.159 ± 2.978	118.54 ± 4.41

Data are presented as the average of three independent replicates ± SE. The statistical significances $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$ were marked by the symbols ^{ns}, *, ** and ***, ****, respectively.

efficiency allowed strains to use more solar energy when treated with glycine, this added energy was subsequently converted into more biomass and lipids. Similarly, mixotrophic cultures experience a rise in pigments that was proportional to the increase in biomass concentration (Camacho et al., 1999). Briefly, the increase in antioxidant carotenoids played a crucial role in the regulation of oxidative stress in microalgae, preventing oxidative damage to algal cells. Moreover, one of the most common antioxidants was glycine, which had been used to boost microalgae lipid production (Song et al., 2022).

Moreover, this study proved that glycine treatment increased the concentration of IAA, a plant hormone that promotes growth and development. The concentration of IAA was significantly higher in both strains at the highest glycine concentration. Glycine had a favorable effect on microalgae cell hormone (IAA) levels in *Chlorella*, which was higher than *Synechocystis*. Low dosages of IAA enhanced the cell proliferation by triggering genes that promote cell division (González-Garcinuño et al., 2016). It was reported before that, hormones such as IAA, gibberellic acid, and cytokinin-kinetin have a role in increasing the biomass percentages of *Chlorella sorokiniana* (Guldhe et al., 2019).

Glycine treatment directed algal metabolism toward fatty acid production

In this regard, microalgae prefer glycine as an organic nitrogen source because the metabolic cost is lower than the cost of other nitrogen forms (Mandal et al., 2018). Most microalgae may employ a variety of nitrogen sources, and each source is first converted to ammonium and then metabolized into amino acids via various mechanisms such as glutamine synthetase, glutamate synthase, or NADP glutamate dehydrogenase (Lasa et al., 2002). Moreover, glycine induced an increase in photosynthesis that induces sugar production. Sugars could supply building blocks for the biosynthesis of fatty acids, which are a part of the tricarboxylic acid cycle intermediates. Consistently, the total soluble sugar level increased at low glycine concentrations, but it decreased at the highest concentration, particularly in *Chlorella*. This decrease may indicate the shift of metabolism toward other primary or secondary metabolites. The results showed that organic acid concentration improved equally in *Synechocystis* and *Chlorella*. As a result, this sequence enhanced the generation of organic acids and fatty acids (Cai et al., 2013; Salbitani and Carfagna, 2021). Total fatty acids during our study in *Synechocystis* and

Chlorella significantly boosted compared to the control sample. In agreement, Wang et al. (2016) indicated that *Chlorella* lipid yield was also upregulated by exogenous glycine treatment at a concentration of 500 mg/L increasing lipid production to 138%. In a previous study by Zhao et al. (2016), photo-chemical modulation enhanced lipid content in *Monoraphidium* sp. The study reported the largest lipid content of 48.5%, which represented a significant improvement in lipid accumulation compared to baseline levels. One of these enhancements, membrane lipids contain the majority of unsaturated fatty acids, and their primary role was to maintain membrane fluidity under various situations (Zhila et al., 2011).

Glycine-treated *Synechocystis* and *Chlorella* are promising sources of bioactive compounds

Microalgae have emerged as a potential feedstock for bioactive compound production due to their high content of essential and unsaturated fatty acids. Chemical environmental modulations can alter microalgae metabolic pathways to accumulate high amounts of neutral lipids from 20–50 percent dry cell weight, primarily in the form of TAGs, in addition to other compounds such as carbohydrates and secondary metabolites (Elsayed et al., 2017). Various methods, such as chemical treatments, growth environment modifications, and genetic engineering, have been employed to induce bioactive primary metabolite production in microalgae (Goh et al., 2019). Whereas researchers aimed to identify microalgae strains capable of high lipid synthesis under stressful conditions. They specifically investigated the lipid content of *Chlorella* and *Synechocystis* under temperature stress. Their findings revealed that *Chlorella* had a lipid content of 6.81E-13 g/cell, while *Synechocystis* had a slightly higher lipid content of 8.19E-13 g/cell. These studies further supported and confirmed our own results, indicating that cyanobacterium cells accumulated more lipids than chlorophyta cells under glycine stress conditions (Elsayed et al., 2017). Whereas, our findings showed that under glycine stress conditions, cyanobacterium cells (*Synechocystis*) accumulated more lipids than chlorophyta cells (*Chlorella*). Glycine-treated *Synechocystis* and *Chlorella* are therefore promising feedstocks for biofuel production due to their high lipid accumulation potential.

On utilizing the microalgae on a large industrial scale, we cannot deny the role of pH and temperature on microalgae growth and metabolism, whereas pH strongly affects microalgae enzymatic activities and metabolic pathways. In this context, pH variations had significant implications for biomolecule production as they can either enhance or hinder microbial metabolism (Sassenhagen et al., 2015). While acidic pH conditions can have a detrimental effect on microbial metabolism. Enzyme activity can be inhibited, cellular processes disrupted and metabolic pathways impaired under low pH values (Lund et al., 2014). When NH_4Cl was added to *Synechocystis* cells that were utilizing nitrate as their nitrogen source, there was a rapid decline in growth and cell activity. However, by adjusting the pH and transferring the cells to a medium without ammonium, the initial activity was fully restored after washing the cells (Kolodny et al., 2006). On the other hand, high pH levels can disrupt enzyme stability and function, leading to reduced metabolic activity (Frankenberger Jr and Johanson, 1982). Thus, here to achieve the highest yield of growth and byproducts, we employed the optimal neutral pH for the microalgae culture as reported before by Wang et al. (2010).

Similar to pH, environmental conditions, such as temperature showed a notable impact on the biochemicals production of microalgae. Elevated temperatures during growth had been associated with a significant decrease in protein content, accompanied by increases in lipids and carbohydrates (Renaud et al., 2002). Moreover, different studies have shown that the response of microalgal chemical composition to high and low growth temperatures varies across species. The ideal growth temperatures are those that enable the cells to undergo photosynthesis without altering their inherent biochemical or physiological characteristics (Ras et al., 2013). In the context of Egypt, where microalgae are cultivated outdoors, they must thrive within a wide diurnal temperature range of 25°C–35°C consistently throughout the year. According to Ras et al. (2013) and our preliminary results, the optimal temperature for our specific strains falls within the range of 28°C–30°C.

Overall, our study contributes to the growing body of research on microalgae as a potential feedstock for biofuel production. Additionally, glycine can serve as both a carbon and nitrogen source, making it useful for promoting the heterotrophic growth of microalgae under stressful conditions and increasing biomass production (Yu et al., 2021). Glycine can then be utilized in various metabolic pathways within the cell, such as protein synthesis and photorespiration (Wingler et al., 2000; Ren et al., 2022). In protein synthesis, glycine can serve as one of the building blocks of proteins and can be directly incorporated into the growing polypeptide chain during translation. In photorespiration, glycine is converted into serine with the release of carbon dioxide and ammonia (Keys, 2006).

Glycine uptake can explain the differential response of *Synechocystis* and *Chlorella* responded to glycine treatment

Glycine treatment positively changed the growth and biomass besides IAA, organic acid, and amino acids and decreased total soluble sugars in *Synechocystis*. However, glycine treatment did not significantly affect any notable change in total FA and total soluble sugar levels. Overall, *Synechocystis* shows a more pronounced effect on various metabolites. While *Chlorella* shows a more specific effect on amino acids and IAA production. Specific metabolic pathways and physiological characteristics of these two microalgal strains could also play a role in differential response to glycine. The microalgae can uptake the glycine using a transport system that enables direct transportation into the cell, where it can serve as a nitrogen source for the cell (Yu et al., 2021). Here the difference in the rate of glycine uptake between prokaryotic *Synechocystis* and eukaryotic *Chlorella* cells may contribute to their different responses to glycine (Gao et al., 2009; Kanwal and De-Eknamkul, 2023). In this regard, some microalgae species have been shown to preferentially utilize glycine over other nitrogen sources such as ammonium and nitrate (Flynn and Butler, 1986; Tyler et al., 2005).

Our growth WM medium contains nitrogen in forms that could potentially affect glycine uptake. In this regard, the study of Gopalakrishnan et al. (2015) that used glycine as a nitrogen source revealed that the majority of the glycine and serine

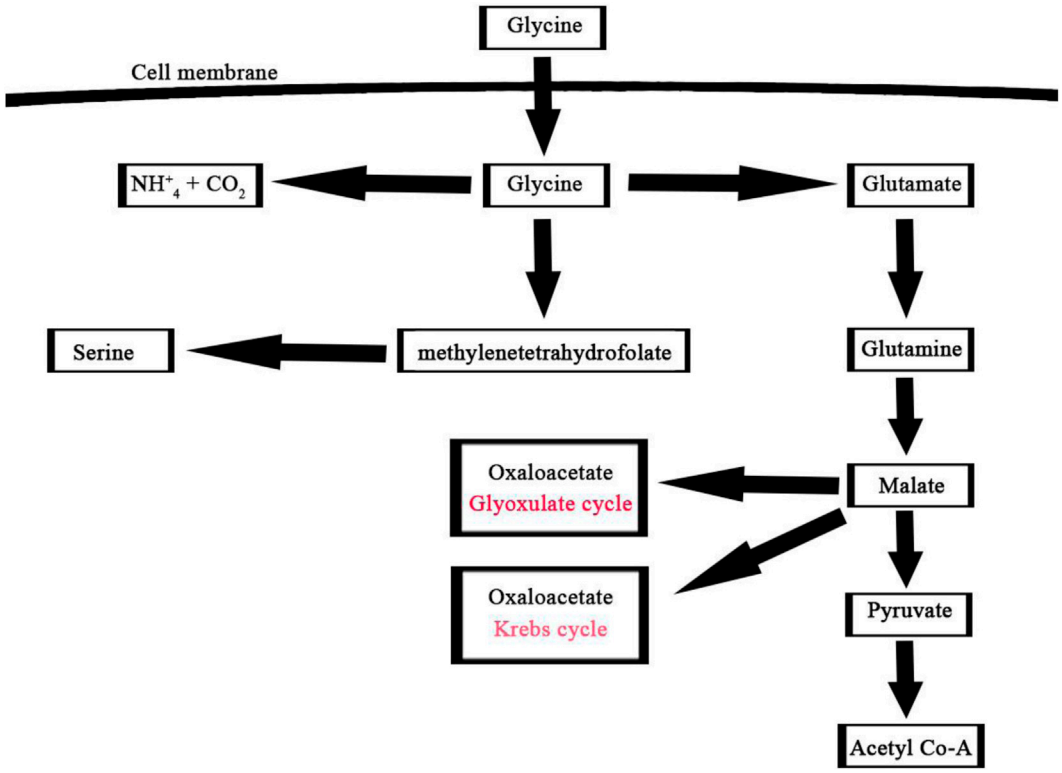


FIGURE 8
Glycine assimilation pathway in microalgae, this pathway illustrates the main three steps involved in the utilization of glycine by microalgae for metabolic processes such as glutamine, methylenetetrahydrofolate, and ammonium pathways.

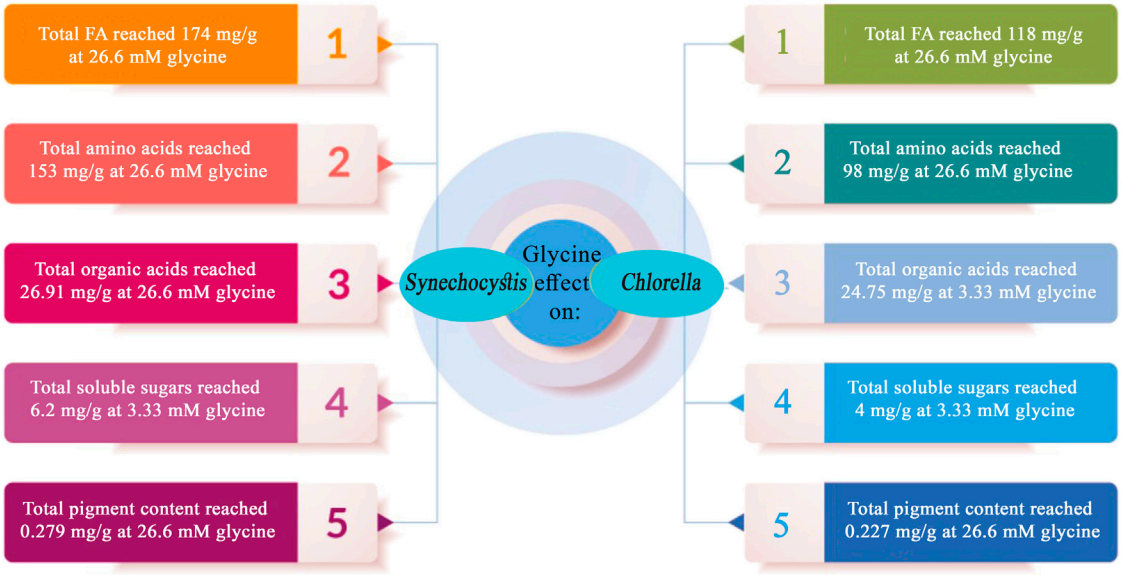


FIGURE 9
Summary of the key findings obtained through the application of external glycine on *Synechocystis* sp. and *Chlorella* sp.

detected in the *Chlorella protothecoides* intracellular were unlabeled, indicating that these amino acids were primarily taken up from the extracellular medium. Interestingly, nitrogen-limited growth conditions decreased glycine uptake, indicating that nitrogen availability in the growth media improved glycine uptake (Gopalakrishnan et al., 2015).

A noteworthy consequence of using glycine as the nitrogen source was the partial activation of the glyoxylate shunt. Therefore, glycine provided in the extracellular growth medium as a nitrogen source can either be catabolized through glycine dehydrogenase for amino acids production or directed into the glyoxylate shunt for TCA cycle intermediates as shown in (Figure 8) (Gopalakrishnan et al., 2015). Overall, algal metabolism utilizing glycine highlights the significant impact that the nature of the nitrogen source can have on the labeling patterns of intracellular metabolites and metabolite flows.

Conclusion

In a nutshell, the present study explored the effect of glycine on two microalgal strains, *Synechocystis* and *Chlorella*, and found that glycine had a positive impact on their growth, photosynthetic pigments, cell hormone levels, organic acid production, amino acid and fatty acid synthesis in microalgae as presented in (Figure 9). Glycine also had a positive effect on the levels of the microalgae cell hormone indole-3-acetic acid. Total soluble sugar was lowered in both strains, but to a greater extent in *Chlorella*. This study demonstrated the impact of glycine supplementation on the amino acid and fatty acid profiles of *Synechocystis* and *Chlorella*. The response to glycine concentration differs between the two strains, with *Chlorella* showing a significant increase in phenylalanine content at higher glycine concentrations. The findings suggest that glycine supplementation can enhance the production of nonpolar amino acids and potentially increase the potential for bioactive compounds and biodiesel production by altering the fatty acid profiles of these strains. The findings presented in this article have wide-ranging ramifications for the industry. Whereas this information is valuable for microalgae-cultivating industries because it provides insight into the potential use of glycine as a growth stimulant. In addition, the results suggest that glycine supplementation can modify the fatty acid profiles of microalgal strains, thereby potentially boosting the production of bioactive compounds and biodiesel. This has implications for industries focused on biofuel production and those investigating microalgae as a sustainable dietary source. Overall, the study emphasizes the promising potential of glycine supplementation in large-scale microalgae cultivation as a possible alternative to conventional crops for food and biofuel production.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

WF: Conceptualization, methodology, investigation, data curation, formal analysis, writing an original draft, visualization, validation. HA: Conceptualization, methodology, investigation, data curation, formal analysis, visualization, validation. EE: Investigation, data curation, formal analysis. SK and ET: Investigation, data curation, formal analysis. MA and OH: Conceptualization, methodology, investigation, data curation, formal analysis, writing-review and editing, supervision, visualization. KE: Conceptualization, methodology, investigation, data curation, formal analysis, writing-original draft, visualization, validation. All authors contributed to the article and approved the submitted version.

Funding

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP 2023R214), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Acknowledgments

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R214), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1161911/full#supplementary-material

References

AbdElgawad, H., Abuelsoud, W., Madany, M. M. Y., Selim, S., Zinta, G., Mousa, A. S. M., et al. (2020). Actinomycetes enrich soil rhizosphere and improve seed quality as well as productivity of legumes by boosting nitrogen availability and metabolism. *Biomolecules* 10, 1675. doi:10.3390/biom10121675

Al Jaouni, S., Saleh, A. M., Wadaan, M. A., Hozzein, W. N., Selim, S., and AbdElgawad, H. (2018). Elevated CO2 induces a global metabolic change in basil (Ocimum basilicum L) and peppermint (Mentha piperita L) and improves their biological activity. *J. plant physiology* 224, 121–131. doi:10.1016/j.jplph.2018.03.016

Almomani, F., Hosseinzadeh-Bandbafha, H., Aghbashlo, M., Omar, A., Joo, S.-W., Vasseghian, Y., et al. (2023). Comprehensive insights into conversion of microalgae to feed, food, and biofuels: Current status and key challenges towards implementation of sustainable biorefineries. *Chem. Eng. J.* 455, 140588. doi:10.1016/j.cej.2022.140588

Bowden, N. A., Sanders, J. P. M., and Bruins, M. E. (2018). Solubility of the proteinogenic α-amino acids in water, ethanol, and ethanol–water mixtures. *J. Chem. Eng. Data* 63, 488–497. doi:10.1021/acs.jced.7b00486

Cai, T., Park, S. Y., and Li, Y. (2013). Nutrient recovery from wastewater streams by microalgae: Status and prospects. *Renew. Sustain. Energy Rev.* 19, 360–369. doi:10.1016/j.rser.2012.11.030

Camacho, F. G., Gómez, A. C., Fernández, F. A., Sevilla, J. F., and Grima, E. M. (1999). Use of concentric-tube airlift photobioreactors for microalgal outdoor mass cultures. *Enzyme Microb. Technol.* 24, 164–172. doi:10.1016/s0141-0229(98)00103-3

Chittora, D., Meena, M., Barupal, T., Swapnil, P., and Sharma, K. (2020). Cyanobacteria as a source of biofertilizers for sustainable agriculture. *Biochem. biophysics Rep.* 22, 100737. doi:10.1016/j.bbrep.2020.100737

Cuellar-Bermudez, S. P., Garcia-Perez, J. S., Rittmann, B. E., and Parra-Saldivar, R. (2015). Photosynthetic bioenergy utilizing CO2: An approach on flue gases utilization for third generation biofuels. *J. Clean. Prod.* 98, 53–65. doi:10.1016/j.jclepro.2014.03.034

Dantas, D. M., Oliveira, C. Y. B., Costa, R. M. P. B., Carneiro-da-Cunha, M., Gálvez, A. O., and Bezerra, R. D. S. (2019). Evaluation of antioxidant and antibacterial capacity of green microalgae Scenedesmus subspicatus. *Food Sci. Technol. Int.* 25, 318–326. doi:10.1177/1082013218825024

De Bhowmick, G., Koduru, L., and Sen, R. (2015). Metabolic pathway engineering towards enhancing microalgal lipid biosynthesis for biofuel application—A review. *Renew. Sustain. Energy Rev.* 50, 1239–1253. doi:10.1016/j.rser.2015.04.131

De la Hoz Siegler, H., Ben-Zvi, A., Burrell, R. E., and McCaffrey, W. C. (2011). The dynamics of heterotrophic algal cultures. *Bioresour. Technol.* 102, 5764–5774. doi:10.1016/j.biortech.2011.01.081

Doan, T. T. Y., Sivaloganathan, B., and Obbard, J. P. (2011). Screening of marine microalgae for biodiesel feedstock. *Biomass Bioenergy* 35, 2534–2544. doi:10.1016/j.biombioe.2011.02.021

Elsayed, K. N. M., Kolesnikova, T. A., Noke, A., and Klöck, G. (2017). Imaging the accumulated intracellular microalgal lipids as a response to temperature stress. *3 Biotech.* 7, 41. doi:10.1007/s13205-017-0677-x

Fabregas, J., Herrero, C., Cabezas, B., and Abalde, J. (1986). Biomass production and biochemical composition in mass cultures of the marine microalga Isochrysis galbana Parke at varying nutrient concentrations. *Aquaculture* 53, 101–113. doi:10.1016/0044-8486(86)90280-2

Fathy, W., Elsayed, K., Essawy, E., Tawfik, E., Zaki, A., Abdelhameed, M., et al. (2020). Biosynthesis of silver nanoparticles from *Synechocystis* sp. to be used as a flocculant agent with different microalgae strains. *Curr. Nanomater.* 5, 175–187. doi:10.2174/2468187310999200605161200

Fathy, W., Essawy, E., Tawfik, E., Khedr, M., Abdelhameed, M. S., Hammouda, O., et al. (2021). Recombinant overexpression of the *Escherichia coli* acetyl-CoA carboxylase gene in Synechocystis sp. boosts lipid production. *J. Basic Microbiol.* 61, 330–338. doi:10.1002/jobm.202000656

Flynn, K. J., and Butler, I. (1986). Nitrogen sources for the growth of marine microalgae: Role of dissolved free amino acids. *Mar. Ecol. Prog. Ser.* 34, 281–304. doi:10.3354/meps034281

Frankenberger, W., Jr, and Johanson, J. (1982). Effect of pH on enzyme stability in soils. *Soil Biol. Biochem.* 14, 433–437. doi:10.1016/0038-0717(82)90101-8

Gao, Y., Cui, Y., Xiong, W., Li, X., and Wu, Q. (2009). Effect of UV-C on algal evolution and differences in growth rate, pigmentation and photosynthesis between prokaryotic and eukaryotic algae. *Photochem. Photobiol.* 85, 774–782. doi:10.1111/j.1751-1097.2008.00493.x

Gao, L., Ding, W., Xi, J., Gao, S., Zhou, X., Chen, Y., et al. (2023). Effects of different nitrogen/phosphorus ratios on the growth and metabolism of microalgae Scenedesmus obliquus cultured in the mixed wastewater from primary settling tank and sludge thickener. *Process Saf. Environ. Prot.* 170, 824–833. doi:10.1016/j.psep.2022.12.059

Garcıa, M. C., Mirón, A. S., Sevilla, J. F., Grima, E. M., and Camacho, F. G. (2005). Mixotrophic growth of the microalga phaeodactylum tricornutum: Influence of different nitrogen and organic carbon sources on productivity and biomass composition. *Process Biochem.* 40, 297–305. doi:10.1016/j.procbio.2004.01.016

Goh, B. H. H., Ong, H. C., Cheah, M. Y., Chen, W.-H., Yu, K. L., and Mahlia, T. M. I. (2019). Sustainability of direct biodiesel synthesis from microalgae biomass: A critical review. *Renew. Sustain. Energy Rev.* 107, 59–74. doi:10.1016/j.rser.2019.02.012

González-Garcinuño, Á., Sánchez-Álvarez, J. M., Galán, M. A., and Martin del Valle, E. M. (2016). Understanding and optimizing the addition of phytohormones in the culture of microalgae for lipid production. *Biotechnol. Prog.* 32, 1203–1211. doi:10.1002/btpr.2312

Gopalakrishnan, S., Baker, J., Kristoffersen, L., and Betenbaugh, M. J. (2015). Redistribution of metabolic fluxes in Chlorella protothecoides by variation of media nitrogen concentration. *Metab. Eng. Commun.* 2, 124–131. doi:10.1016/j.meteno.2015.09.004

Guedes, A. C., Amaro, H. M., and Malcata, F. X. (2011). Microalgae as sources of high added-value compounds—A brief review of recent work. *Biotechnol. Prog.* 27, 597–613. doi:10.1002/btpr.575

Guldhe, A., Renuka, N., Singh, P., and Bux, F. (2019). Effect of phytohormones from different classes on gene expression of Chlorella sorokiniana under nitrogen limitation for enhanced biomass and lipid production. *Algal Res.* 40, 101518. doi:10.1016/j.algal.2019.101518

Harun, R., Singh, M., Forde, G. M., and Danquah, M. K. (2010). Bioprocess engineering of microalgae to produce a variety of consumer products. *Renew. Sustain. energy Rev.* 14, 1037–1047. doi:10.1016/j.rser.2009.11.004

Jin, L., Junchao, H., and Feng, C. (2011). "Microalgae as feedstocks for biodiesel production," in *Biodiesel. IntechOpen, rijeka*. Editors S. Margarita and M. Gisela, 7.

Kanwal, S., and De-Eknamkul, W. (2023). A non-functional γ-aminobutyric acid shunt pathway in cyanobacterium Synechocystis sp. PCC 6803 enhances δ-aminolevulinic acid accumulation under modified nutrient conditions. *Int. J. Mol. Sci.* 24, 1213. doi:10.3390/ijms24021213

Keys, A. J. (2006). The re-assimilation of ammonia produced by photorespirationand the nitrogen economy of C3 higher plants. *Photosynth. Res.* 87, 165–175. doi:10.1007/s11120-005-9024-x

Khan, M. I., Shin, J. H., and Kim, J. D. (2018). The promising future of microalgae: Current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microb. Cell. factories* 17, 36–21. doi:10.1186/s12934-018-0879-x

Kim, G., Mujtaba, G., and Lee, K. (2016). Effects of nitrogen sources on cell growth and biochemical composition of marine chlorophyte Tetraselmis sp. for lipid production. *Algae* 31, 257–266. doi:10.4490/algae.2016.31.8.18

Kolodny, N. H., Bauer, D., Bryce, K., Klucevsek, K., Lane, A., Medeiros, L., et al. (2006). Effect of nitrogen source on cyanophycin synthesis in Synechocystis sp. strain PCC 6308. *J. Bacteriol.* 188, 934–940. doi:10.1128/jb.188.3.934-940.2006

Kozan, D. W., Derrick, J. T., Ludington, W. B., and Farber, S. A. (2023). From worms to humans: Understanding intestinal lipid metabolism via model organisms. *Biochimica Biophysica Acta (BBA)-Molecular Cell. Biol. Lipids* 1868, 159290. doi:10.1016/j.bbalip.2023.159290

Krzemińska, I., Pawlik-Skowrońska, B., Trzcińska, M., and Tys, J. (2014). Influence of photoperiods on the growth rate and biomass productivity of green microalgae. *Bioprocess Biosyst. Eng.* 37, 735–741. doi:10.1007/s00449-013-1044-x

Kumar, V., Muthuraj, M., Palabhanvi, B., Ghoshal, A. K., and Das, D. (2014). High cell density lipid rich cultivation of a novel microalgal isolate Chlorella sorokiniana FC6 IITG in a single-stage fed-batch mode under mixotrophic condition. *Bioresour. Technol.* 170, 115–124. doi:10.1016/j.biortech.2014.07.066

Lasa, B., Frechilla, S., Aparicio-Tejo, P. M., and Lamsfus, C. (2002). Role of glutamate dehydrogenase and phosphoenolpyruvate carboxylase activity in ammonium nutrition tolerance in roots. *Plant Physiology Biochem.* 40, 969–976. doi:10.1016/s0981-9428(02)01451-1

Liu, Y., Zhan, J.-j., and Hong, Y. (2017). Effects of metal ions on the cultivation of an oleaginous microalga Chlorella sp. *Environ. Sci. Pollut. Res.* 24, 26594–26604. doi:10.1007/s11356-017-0258-x

Lund, P., Tramonti, A., and De Biase, D. (2014). Coping with low pH: Molecular strategies in neutralophilic bacteria. *FEMS Microbiol. Rev.* 38, 1091–1125. doi:10.1111/1574-6976.12076

Maity, J. P., Bundschuh, J., Chen, C.-Y., and Bhattacharya, P. (2014). Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives–A mini review. *Energy* 78, 104–113. doi:10.1016/j.energy.2014.04.003

Maltsev, Y., and Maltseva, K. (2021). Fatty acids of microalgae: Diversity and applications. *Rev. Environ. Sci. Bio/Technology* 20, 515–547. doi:10.1007/s11157-021-09571-3

Mandal, S., Shurin, J. B., Efroymson, R. A., and Mathews, T. J. (2018). Functional divergence in nitrogen uptake rates explains diversity–productivity relationship in microalgal communities. *Ecosphere* 9, e02228. doi:10.1002/ecs2.2228

Matantseva, O., Pozdnyakov, I., Voss, M., Liskow, I., and Skarlato, S. (2018). The uncoupled assimilation of carbon and nitrogen from urea and Glycine by the bloom-forming dinoflagellate prorocentrum minimum. *Protist* 169, 603–614. doi:10.1016/j.protis.2018.05.006

Mérida, A., Candau, P., and Florencio, F. J. (1991). Regulation of glutamine synthetase activity in the unicellular cyanobacterium Synechocystis sp. strain PCC 6803 by the nitrogen source: Effect of ammonium. *J. Bacteriol.* 173, 4095–4100. doi:10.1128/jb.173.13.4095-4100.1991

Metzner, H., Rau, H., and Senger, H. (1965). Untersuchungen zur synchronisierbarkeit einzelner pigmentmangel-mutanten von *Chlorella*. *Planta* 65, 186–194. doi:10.1007/bf00384998

Michalak, I., and Chojnacka, K. (2014). Algal extracts: Technology and advances. *Eng. Life Sci.* 14, 581–591. doi:10.1002/elsc.201400139

Moran, R., and Porath, D. (1980). Chlorophyll determination in intact tissues using N,N-Dimethylformamide. *Plant Physiol.* 65, 478–479. doi:10.1104/pp.65.3.478

Nateghpour, B., Kavoosi, G., and Mirakhorli, N. (2021). Amino acid profile of the peel of three citrus species and its effect on the combination of amino acids and fatty acids Chlorella vulgaris. *J. Food Compos. Analysis* 98, 103808. doi:10.1016/j.jfca.2021.103808

Norici, A., Dalsass, A., and Giordano, M. (2002). Role of phosphoenolpyruvate carboxylase in anaplerosis in the green microalga Dunaliella salina cultured under different nitrogen regimes. *Physiol. Plant.* 116, 186–191. doi:10.1034/j.1399-3054.2002.1160207.x

Pflanz, M., and Zude, M. (2008). Spectrophotometric analyses of chlorophyll and single carotenoids during fruit development of tomato (*Solanum lycopersicum* L) by means of iterative multiple linear regression analysis. *Appl. Opt.* 47, 5961–5970. doi:10.1364/ao.47.005961

Rahman, K. M. (2020). Food and high value products from microalgae: Market opportunities and challenges. *Microalgae Biotechnol. food, health high value Prod.*, 3–27. doi:10.1007/978-981-15-0169-2_1

Ras, M., Steyer, J.-P., and Bernard, O. (2013). Temperature effect on microalgae: A crucial factor for outdoor production. *Rev. Environ. Sci. Bio/Technology* 12, 153–164. doi:10.1007/s11157-013-9310-6

Ren, J., Wang, W., Nie, J., Yuan, W., and Zeng, A.-P. (2022). "Understanding and engineering Glycine cleavage system and related metabolic pathways for C1-based biosynthesis," in *One-carbon feedstocks for sustainable bioproduction*. Editors A.-P. Zeng and N. J. Claassens (Cham: Springer International Publishing), 273–298.

Renaud, S. M., Thinh, L.-V., Lambrinidis, G., and Parry, D. L. (2002). Effect of temperature on growth, chemical composition and fatty acid composition of tropical Australian microalgae grown in batch cultures. *Aquaculture* 211, 195–214. doi:10.1016/s0044-8486(01)00875-4

Salbitani, G., and Carfagna, S. (2021). Ammonium utilization in microalgae: A sustainable method for wastewater treatment. *Sustainability* 13, 956. doi:10.3390/su13020956

Santhosh, S., Dhandapani, R., and Hemalatha, N. (2016). Bioactive compounds from Microalgae and its different applications-a review. *Adv. Appl. Sci. Res.* 7, 153–158.

Sarma, S., Sharma, S., Patel, A., Upadhyay, J., Rathod, V., and Narra, M. (2023). Statistical optimization of microalgal biodiesel production and protein extraction from Chlorella sorokiniana cultivated in dairy effluent. *Mater. Today Proc.* 72, 2731–2740. doi:10.1016/j.matpr.2022.09.512

Sassenhagen, I., Wilken, S., Godhe, A., and Rengefors, K. (2015). Phenotypic plasticity and differentiation in an invasive freshwater microalga. *Harmful Algae* 41, 38–45. doi:10.1016/j.hal.2014.11.001

Sinha, A. K., Giblen, T., AbdElgawad, H., De Rop, M., Asard, H., Blust, R., et al. (2013). Regulation of amino acid metabolism as a defensive strategy in the brain of three freshwater teleosts in response to high environmental ammonia exposure. *Aquat. Toxicol.* 130-131, 86–96. doi:10.1016/j.aquatox.2013.01.003

Song, X., Liu, B.-F., Kong, F., Ren, N.-Q., and Ren, H.-Y. (2022). Overview on stress-induced strategies for enhanced microalgae lipid production: Application, mechanisms and challenges. *Resour. Conservation Recycl.* 183, 106355. doi:10.1016/j.resconrec.2022.106355

Song, X., Liu, B.-F., Kong, F., Song, Q., Ren, N.-Q., and Ren, H.-Y. (2023). Lipid accumulation by a novel microalga Parachlorella kessleri R-3 with wide pH tolerance for promising biodiesel production. *Algal Res.* 69, 102925. doi:10.1016/j.algal.2022.102925

Sun, Z., Chen, Y.-F., and Du, J. (2016). Elevated CO2 improves lipid accumulation by increasing carbon metabolism in Chlorella sorokiniana. *Plant Biotechnol. J.* 14, 557–566. doi:10.1111/pbi.12398

Tabatabaei, M., Tohidfar, M., Jouzani, G. S., Safarnejad, M., and Pazouki, M. (2011). Biodiesel production from genetically engineered microalgae: Future of bioenergy in Iran. *Renew. Sustain. Energy Rev.* 15, 1918–1927. doi:10.1016/j.rser.2010.12.004

Tyler, A. C., McGlathery, K. J., and Macko, S. A. (2005). Uptake of urea and amino acids by the macroalgae Ulva lactuca (Chlorophyta) and Gracilaria vermiculophylla (Rhodophyta). *Mar. Ecol. Prog. Ser.* 294, 161–172. doi:10.3354/meps294161

van Handel, E. (1968). Direct microdetermination of sucrose. *Anal. Biochem.* 22, 280–283. doi:10.1016/0003-2697(68)90317-5

Wan, M. X., Wang, R. M., Xia, J. L., Rosenberg, J. N., Nie, Z. Y., Kobayashi, N., et al. (2012). Physiological evaluation of a new Chlorella sorokiniana isolate for its biomass production and lipid accumulation in photoautotrophic and heterotrophic cultures. *Biotechnol. Bioeng.* 109, 1958–1964. doi:10.1002/bit.24477

Wang, C., Li, H., Wang, Q., and Wei, P. (2010). [Effect of pH on growth and lipid content of Chlorella vulgaris cultured in biogas slurry]. *Sheng Wu Gong Cheng Xue Bao* 26, 1074–1079.

Wang, J., Sommerfeld, M. R., Lu, C., and Hu, Q. (2013). Combined effect of initial biomass density and nitrogen concentration on growth and astaxanthin production of Haematococcus pluvialis (Chlorophyta) in outdoor cultivation. *Algae* 28, 193–202. doi:10.4490/algae.2013.28.2.193

Wang, Y., He, B., Sun, Z., and Chen, Y.-F. (2016). Chemically enhanced lipid production from microalgae under low sub-optimal temperature. *Algal Res.* 16, 20–27. doi:10.1016/j.algal.2016.02.022

Winckelmann, D., Bleeke, F., Bergmann, P., and Klöck, G., 2015. Growth of Cyanobacterium aponinum influenced by increasing salt concentrations and temperature. *3 Biotech.* 5, 253–260. doi:10.1007/s13205-014-0224-y

Wingler, A., Lea, P. J., Quick, W. P., and Leegood, R. C. (2000). Photorespiration: Metabolic pathways and their role in stress protection. *Philosophical Trans. R. Soc. Lond. Ser. B Biol. Sci.* 355, 1517–1529. doi:10.1098/rstb.2000.0712

Yaakob, Z., Ali, E., Zainal, A., Mohamad, M., and Takriff, M. S. (2014). An overview: Biomolecules from microalgae for animal feed and aquaculture. *J. Biol. Research-Thessaloniki* 21, 6–10. doi:10.1186/2241-5793-21-6

Yarkent, Ç., Gürlek, C., and Oncel, S. S. (2020). Potential of microalgal compounds in trending natural cosmetics: A review. *Sustain. Chem. Pharm.* 17, 100304. doi:10.1016/j.scp.2020.100304

Yemm, E., and Willis, A. (1954). The estimation of carbohydrates in plant extracts by anthrone. *Biochem. J.* 57, 508–514. doi:10.1042/bj0570508

Yin, Z., Zhu, L., Li, S., Hu, T., Chu, R., Mo, F., et al. (2020). A comprehensive review on cultivation and harvesting of microalgae for biodiesel production: Environmental pollution control and future directions. *Bioresour. Technol.* 301, 122804. doi:10.1016/j.biortech.2020.122804

Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y. J., and Zhang, F. (2013). Development of Synechocystis sp. PCC 6803 as a phototrophic cell factory. *Mar. Drugs* 11, 2894–2916. doi:10.3390/md11082894

Yu, Q., He, J., Zhao, Q., Wang, X., Zhi, Y., Li, X., et al. (2021). Regulation of nitrogen source for enhanced photobiological H2 production by co-culture of Chlamydomonas reinhardtii and Mesorhizobium sangaii. *Algal Res.* 58, 102422. doi:10.1016/j.algal.2021.102422

Yue, H., Zhao, C., Yang, S., and Jia, Y. (2021). Effects of glycine on cell growth and pigment biosynthesis in Rhodobacter azotoformans. *J. basic Microbiol.* 61, 63–73. doi:10.1002/jobm.202000503

Zhao, Y., Li, D., Ding, K., Che, R., Xu, J.-W., Zhao, P., et al. (2016). Production of biomass and lipids by the oleaginous microalgae Monoraphidium sp. QLY-1 through heterotrophic cultivation and photo-chemical modulator induction. *Bioresour. Technol.* 211, 669–676. doi:10.1016/j.biortech.2016.03.160

Zhila, N. O., Kalacheva, G. S., and Volova, T. G. (2011). Effect of salinity on the biochemical composition of the alga Botryococcus braunii Kütz IPPAS H-252. *J. Appl. Phycol.* 23, 47–52. doi:10.1007/s10811-010-9532-8

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 07 August 2023
DOI 10.3389/fbioe.2023.1222812

OPEN ACCESS

EDITED BY
Yaojun Tong,
Shanghai Jiao Tong University, China

REVIEWED BY
Mario Andrea Marchisio,
Tianjin University, China
Ahmad Bazli Ramzi,
National University of Malaysia, Malaysia

*CORRESPONDENCE
Jose F. Marcos,
jmarcos@iata.csic.es
Sandra Garrigues,
sgarrigues@iata.csic.es

RECEIVED 15 May 2023
ACCEPTED 12 July 2023
PUBLISHED 07 August 2023

CITATION
Moreno-Giménez E, Gandía M, Sáez Z, Manzanares P, Yenush L, Orzáez D, Marcos JF and Garrigues S (2023), FungalBraid 2.0: expanding the synthetic biology toolbox for the biotechnological exploitation of filamentous fungi. *Front. Bioeng. Biotechnol.* 11:1222812. doi: 10.3389/fbioe.2023.1222812

FungalBraid 2.0: expanding the synthetic biology toolbox for the biotechnological exploitation of filamentous fungi

Elena Moreno-Giménez[1,2], Mónica Gandía[3], Zara Sáez[1], Paloma Manzanares[1], Lynne Yenush[2], Diego Orzáez[2], Jose F. Marcos[1]* and Sandra Garrigues[1]*

[1]Food Biotechnology Department, Instituto de Agroquímica y Tecnología de Alimentos (IATA), Consejo Superior de Investigaciones Científicas (CSIC), Valencia, Spain, [2]Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas (CSIC)-Universitat Politècnica de València (UPV), Valencia, Spain, [3]Preventive Medicine and Public Health, Food Science, Toxicology and Forensic Medicine Department. Faculty of Pharmacy. Universitat de València. Vicente Andrés Estellés s/n, Valencia, Spain

Fungal synthetic biology is a rapidly expanding field that aims to optimize the biotechnological exploitation of fungi through the generation of standard, ready-to-use genetic elements, and universal syntax and rules for contributory use by the fungal research community. Recently, an increasing number of synthetic biology toolkits have been developed and applied to filamentous fungi, which highlights the relevance of these organisms in the biotechnology field. The FungalBraid (FB) modular cloning platform enables interchangeability of DNA parts with the GoldenBraid (GB) platform, which is designed for plants, and other systems that are compatible with the standard Golden Gate cloning and syntax, and uses binary pCAMBIA-derived vectors to allow *Agrobacterium tumefaciens*-mediated transformation of a wide range of fungal species. In this study, we have expanded the original FB catalog by adding 27 new DNA parts that were functionally validated *in vivo*. Among these are the resistance selection markers for the antibiotics phleomycin and terbinafine, as well as the uridine-auxotrophic marker *pyr4*. We also used a normalized luciferase reporter system to validate several promoters, such as P*pkiA*, P*7760*, P*ef1α*, and P*afpB* constitutive promoters, and P*glaA*, P*amyB*, and P*xlnA* inducible promoters. Additionally, the recently developed dCas9-regulated GB_SynP synthetic promoter collection for orthogonal CRISPR activation (CRISPRa) in plants has been adapted in fungi through the FB system. In general, the expansion of the FB catalog is of great interest to the scientific community since it increases the number of possible modular and interchangeable DNA assemblies, exponentially increasing the possibilities of studying, developing, and exploiting filamentous fungi.

KEYWORDS

fungal synthetic biology, GoldenBraid, promoters, selection markers, luciferase-based reporter system, CRISPR activation, filamentous fungi

1 Introduction

Filamentous fungi have acquired a great biotechnological relevance as biofactories for the sustainable production of organic acids, proteins, enzymes, and metabolites with applications in the agri-food, chemical, pharmaceutical, textile, paper, and biofuel industries (Meyer et al., 2016). Their ability to grow on many distinct and economic substrates and plant residues, and their high secretory capacity justify the biotechnological interest of these microorganisms, which have become essential contributors to the so-called circular bio-economy (Meyer et al., 2020). Enzymes produced by fungi currently make up more than half of the enzymes used in the industry (de Vries et al., 2020). Additionally, fungal genomes contain a large number of biosynthetic gene clusters encoding potentially useful biomolecules to be exploited (Robey et al., 2021), reflecting the relevance of filamentous fungi as cell factories. However, there are still aspects that need to be improved since the conditions and levels of production of different biomolecules are highly variable, and some of them are difficult to produce in a cost-efficient manner.

Synthetic biology (SynBio) is an ever-expanding scientific field that has revolutionized genetic and metabolic engineering. SynBio provides new tools for the generation of ready-to-use, standardized, modular genetic elements to obtain microbial strains with optimized properties either by the production of specific proteins or by fine-tuning the expression of specific metabolic pathway-related genes (Benner and Sismour, 2005). In this context, fungal SynBio is rapidly evolving. Our group has adapted the GoldenBraid (GB) modular cloning platform originally developed for plants (Sarrion-Perdigones et al., 2013) to filamentous fungi, a variant called FungalBraid (FB) (https://gbcloning.upv.es/fungal/). This modular cloning method is based on type IIS restriction enzymes and pCAMBIA-derived binary vectors for *Agrobacterium tumefaciens*-mediated transformation (ATMT), with the main advantages of the full reusability of its DNA parts and their interchangeability between plants and fungi as long as they are functionally compatible (Hernanz-Koers et al., 2018; Vazquez-Vilar et al., 2020). The domestication or incorporation of new DNA parts into the FB system is achieved by cloning them into level 0 pUPD2 vectors, and transcriptional units (TUs) are then formed by combining different level 0 parts in a multipartite assembly into level 1 pDGB3α vectors. The GB and FB systems allow the combination of different TUs contained in two compatible pDGB3α vectors in a bipartite assembly into level 2 pDGB3Ω vectors, which can then be combined in the same way back into pDGB3α vectors, allowing the indefinite expansion of the multigene construct and designs of increased complexity.

Since the development of the FB system, an increasing number of SynBio-based applications in fungi have been reported (Dahlmann et al., 2021; Mózsik et al., 2021; Mózsik et al., 2022), which highlights the need for a boost in the SynBio toolkit for these organisms. However, there is still a shortage of tools for orthogonal and fine-tuned expression of genes applied to filamentous fungi. In this sense, an increase in the repertoire of promoters is required. Promoters with different expression levels or which are inducible and/or cell-specific would increase the flexibility and the ability to optimize expression systems, especially for proteins which can be toxic. These promoters may come from different organisms or may be created using synthetic designs. Considering that constitutive and inducible promoters are commonly used among the scientific community, synthetic promoters have been less exploited. These promoters often comprised a core or minimal promoter and an upstream region in which cis-regulatory elements are incorporated (Martins-Santana et al., 2018). These cis-regulatory elements are typically obtained from the binding sites of transcriptional regulators which activate or inactivate gene expression. Although natural transcriptional regulators limit the freedom in the design of cis-regulatory elements, the use of CRISPR activation (CRISPRa) strategies allows the use of virtually any 20-base pair (bp) sequence as a cis-regulatory box in the development of synthetic promoters (Moreno-Giménez et al., 2022). In this regard, the collection of nuclease-deactivated Cas9 (dCas9)-regulated synthetic promoter GB_SynP that has been recently developed for plants (Moreno-Giménez et al., 2022) could easily be adapted to fungi, given the interchangeability of DNA parts between GB and FB systems. Additionally, GB/FB systems provide a standard measurement using a luciferase/renilla transient assay to estimate relative expression levels of promoters, including the synthetic promoters (Vazquez-Vilar et al., 2017; Gandía et al., 2022).

In this study, we have incorporated 27 new genetic parts into the FB system, which include native strong and inducible fungal promoters, synthetic promoters, terminators, and selection markers. All these components have been validated *in vivo* in two economically relevant fungi: the non-model postharvest pathogen of citrus *Penicillium digitatum* (Palou, 2014), and in the well-known fungus with the Generally Recognized as Safe (GRAS) status and a long record of industrial use *Penicillium chrysogenum* (Fierro et al., 2022). The strength of the constitutive promoters has been characterized and compared in a nanoluciferase-normalized luciferase-based reporter system; the induction levels of the inducible promoters have also been quantified, and the activation of the synthetic promoters has been studied using programmable transcriptional factors based on CRISPRa (Mózsik et al., 2021). Overall, the expansion of the FB toolkit will be of great interest to the scientific community to further aid the exploitation of fungal workhorses and accelerate the discovery and production of (novel) bioactive molecules for multiple biotechnological applications.

2 Materials and methods

2.1 Strains, media, and growth conditions

The fungal strains used in this study were *P. digitatum* CECT 20796 (isolate PHI26) (Marcet-Houben et al., 2012) and *P. chrysogenum* wild-type ATCC 10002 (Q176) (Hegedüs et al., 2011). Fungi were routinely cultured on potato dextrose agar (PDA, Difco-BD Diagnostics) plates for 7 days at 25°C. For transformation, the generated vectors were amplified in the bacterium *Escherichia coli* JM109 and grown in Luria–Bertani (LB) medium at 37°C with either 50 μg/mL chloramphenicol, 50 μg/mL kanamycin, 100 μg/mL spectinomycin, or 100 μg/mL ampicillin depending on the vector. *A. tumefaciens* AGL-1 was cultured in LB medium at 28°C with 20 μg/mL rifampicin and the corresponding antibiotic depending on the vector used.

For growth profiles, 5 μL of conidial suspension (5×10^4 conidia/mL) was deposited on the center of PDA plates, and colony morphology was assessed and compared daily by visual inspection.

2.2 Design, domestication, and DNA assemblies of genetic elements

All the genetic elements that have been incorporated on the FB platform are listed in Table 1. New DNA parts were domesticated according to GB rules and tools (https://gbcloning.upv.es) and ordered from an external company as synthetic genes (gBlocks™, IDT). In the case of the coding sequence (CDS) of *pyr*4, the gene from *Trichoderma reesei* was codon-optimized according to the optimal codon frequency of *Penicillium* genera prior to GB/FB domestication. Domesticated elements were ligated into the pUPD2 entry vector via the restriction–ligation protocol, as previously described (Hernanz-Koers et al., 2018; Vazquez-Vilar et al., 2020). Positive *E. coli* clones were confirmed by routine PCR amplifications and Sanger sequencing using external specific primers OJM524 and OJM525, which are designed for pUPD2 vectors (Hernanz-Koers et al., 2018) (Table 2). Multiple assemblies into pDGB3α vectors of the DNA parts contained in pUPD2 vectors were carried out to obtain different TUs, and binary assemblies were subsequently performed to combine different TUs into multigenetic constructs within these pDGB3α or pDGB3Ω vectors, as previously described (Hernanz-Koers et al., 2018; Vazquez-Vilar et al., 2020).

The sequence of the single-guide RNA 1 (gRNA1) required to activate GB_SynP promoters was checked for the absence of off-target mutations in the *P. digitatum* genome using Geneious Prime software (https://www.geneious.com/), and was cloned into the pAMA18.0 vector, as described previously (Mózsik et al., 2021). Briefly, a primer pair was designed (OJM698 and OJM699, Table 2) which contained the target sequence of gRNA1, the hammerhead ribozyme, the inverted repetition of the 5′-end of the spacer sequence, and the recognition sites of *Bsa*I. The resulting PCR product was purified (Wizard SV Gel and PCR Clean-Up System, Promega) and inserted into pAMA18.0 via the restriction–ligation reaction with *Bsa*I and T4 ligase. Correct assemblies of the resulting pAMA18.0_gRNA1 vector were confirmed by Sanger sequencing.

2.3 Fungal transformation and mutant confirmation

Transformation of *P. digitatum* CECT 20796 (PHI26) and *P. chrysogenum* ATTC 10002 (Q176) with the corresponding FB binary vectors described in Table 1 was performed through ATMT, as previously described (Khang et al., 2007), with some modifications (Harries et al., 2015; Vazquez-Vilar et al., 2020).

In the case of *P. digitatum* uridine-auxotrophic Δ*pyrG* mutants, in which the *pyrG* gene (gene ID PDIG_38390) was deleted by homologous recombination without the insertion of any positive selection marker, mutants were selected on PDA supplemented with 1.22 g/L uridine (Sigma-Aldrich) and 1.25 g/L of 5-fluoroorotic acid (5-FOA, Formedium). For the ectopically complemented *P. digitatum* Δ*pyrG:pyr4* strains, mutants were selected on PDA plates. *P. digitatum* and *P. chrysogenum* ectopic transformants containing *ble* TU (phleomycin^R) were selected on PDA plates supplemented with 35 μg/mL and 25 μg/mL phleomycin (InvivoGen), respectively. Finally, *P. chrysogenum* ectopic transformants carrying *ergA* TU (terbinafine^R) were selected on PDA plates supplemented with 0.5 μg/mL terbinafine hydrochloride (Sigma-Aldrich).

P. digitatum and *P. chrysogenum* transformants carrying different luciferase reporter constructs to test the constitutive, inducible, and GB_SynP promoters were selected on PDA plates containing 25 μg/mL geneticin (G418) (InvivoGen).

All transformants were molecularly confirmed by PCR reactions using NZYTaq II DNA polymerase (NZYTech) (Figures 1–3; Supplementary Figures S1–S3) from genomic DNA isolated with the NZY Tissue gDNA isolation kit (NZYTech) and primers purchased from IDT (Table 2).

For the validation of GB_SynP promoters, *P. digitatum* protoplasts from strains carrying luciferase reporter vectors for each of the three synthetic promoters tested (FB398, FB399, and FB400) (Table 1) were transformed with the self-replicative AMA1-based dCas9-containing plasmid pAMA18.0_gRNA1 (Mózsik et al., 2021), as previously described (Garrigues et al., 2022). Transformants were selected on PDA plates containing 0.95 M sucrose and 35 μg/mL phleomycin. Regarding the reusability of the system, the loss of the pAMA18.0_gRNA1 plasmid was confirmed after three consecutive streaks of the transformants in non-selective PDA plates, as previously described (Garrigues et al., 2022).

2.4 Luciferase/nanoluciferase assays

P. digitatum and *P. chrysogenum* strains carrying the luciferase reporter vectors for each of the tested promoters (FB433, FB434, FB435, FB436, FB438, FB439, and FB441) and *P. digitatum* strains with the luciferase reporter for the GB_SynP synthetic promoters (FB398, FB399, and FB400), carrying or not pAMA18.0_gRNA1, were grown in duplicate for 2 days in 100 mL flasks with 25 mL of liquid potato dextrose broth (PDB, Difco-BD Diagnostics) at 25°C, with shaking (150 rpm). For induction assays, transformants were grown in duplicate for 4 days in 100 mL flasks with 25 mL of either *P. digitatum* or *P. chrysogenum* minimal medium (PdMM or PcMM, respectively) (Sonderegger et al., 2016) using 2% D-glucose (PanReac), 2% maltose (Sigma-Aldrich) or 2% D-xylose (Sigma-Aldrich) as the sole carbon source. Grown mycelia were filtered, and a sample of 20 mg was collected and immediately frozen in liquid nitrogen. Luciferase and nanoluciferase measurements were performed using the Dual-Glo® Luciferase Assay System (Promega), as previously described (Gandía et al., 2022). Briefly, frozen samples were homogenized in 180 μL passive lysis buffer with a pestle and centrifuged (12,000 ×*g*, 10 min at 4°C). Then, 10 μL of the supernatant was transferred to a white 96-well plate (Thermo Fisher Scientific) and mixed with 40 μL of the luciferase reagent to measure luciferase luminescence in a CLARIOStar microplate reader (BMG LABTECH GmbH) with a measurement of 10 s and a delay of 2 s. Nanoluciferase luminescence was quantified thereafter by adding 40 μL of the Stop & Glo reagent and measured in the same way.

The luciferase/nanoluciferase ratio was determined for each sample, and normalized luminescence was calculated as the mean value of the ratios obtained from each duplicate. Statistical analyses were performed with GraphPad Prism 8.0.1 software. Differences

TABLE 1 FB parts reported in this study. DNA parts are grouped according to the purpose for which they were used.

Selection marker				
Auxotrophy				
Code	**Name**	**Plasmid**	**Description**	**Reference**
FB271*	P*pyr4*	pUPD2	Promoter of the *pyr4* gene from *T. reesei*	This study
FB272*	*pyr4*	pUPD2	Coding sequence of the *pyr4* gene from *T. reesei*	This study
FB273*	T*pyr4*	pUPD2	Terminator of the *pyr4* gene from *T. reesei*	This study
FB293*	TU_*pyr4*	pDGB3α2	Assembly of the transcriptional unit for the auxotrophy marker *pyr4* from *T. reseei*	This study
FB359*	5′ upstream Pdig *pyrG*	pUPD2	5′ upstream region of the *pyrG* gene in *P. digitatum*	This study
FB361*	3′ downstream Pdig *pyrG*	pUPD2	3′ downstream region of the *pyrG* gene in *P. digitatum*	This study
FB372*	FB359+FB361	pDGB3α1	Assembly for *pyrG* deletion in *P. digitatum*	This study
Resistance				
Code	**Name**	**Plasmid**	**Description**	**Reference**
FB413*	*ble*	pUPD2	Coding sequence for phleomycin resistance	This study
FB414*	*ergA*	pUPD2	Coding sequence for terbinafine resistance	This study
FB411*	P*pcbC*	pUPD2	Promoter of isopenicillin N synthase from *P. rubens*	This study
FB416*	T*amdS*	pUPD2	Terminator of acetamidase-encoding gene *amdS* from *A. nidulans*	This study
FB430*	P*pcbc:ble:*T*amdS*	pDGB3α2	TU for the expression of phleomycin resistance	This study
FB431*	P*gpdA:ergA:*T*amdS*	pDGB3α2	TU for the expression of terbinafine resistance	This study
Constitutive/inducible promoters				
Code	**Name**	**Plasmid**	**Description**	**Reference**
FB007	P*gpdA*	pUPD2	Promoter of glyceraldehyde-3-phosphate dehydrogenase from *A. nidulans*	Hernanz-Koers et al. (2018)
FB291*	P*xlnA*	pUPD2	Promoter of the endo-1,4-beta-xylanase A gene from *A. nidulans*, xylose-inducible	This study
FB389*	P*pkiA*	pUPD2	Promoter of the highly expressed pyruvate kinase gene from *A. niger*	This study
FB404*	P*afpB*	pUPD2	Promoter of the antifungal protein *afpB* gene from *P. digitatum*	This study
FB405*	P*glaA*	pUPD2	Promoter of the glucoamylase gene from *A. niger*, maltose/starch-inducible	This study
FB406*	P*amyB*	pUPD2	Promoter of the TAKA-amylase A gene from *A. oryzae*, maltose/starch-inducible	This study
FB407*	P*ef1α*	pUPD2	Promoter of the elongation factor 1-α gene from *P. digitatum* (PDIG_59570)	This study
FB408*	P07760	pUPD2	Promoter of the ubiquitin ligase gene from *P. digitatum* (PDIG_07760)	This study
GB0096	Luciferase (Luc)	pUPD	Coding sequence for the firefly luciferase protein	Sarrion-Perdigones et al. (2013)
FB001	P*trpC*	pUPD2	Promoter of the multifunctional tryptophan biosynthesis protein coding gene *trpC* from *A. nidulans*	Hernanz-Koers et al. (2018)
FB002	T*tub*	pUPD2	Terminator of the tubulin-encoding gene from *N. crassa*	Hernanz-Koers et al. (2018)
FB008	T*trpC*	pUPD2	Terminator of the *trpC* gene from *A. nidulans*	Hernanz-Koers et al. (2018)
FB009	P*trpC:nptII:*T*tub*	pDGB3α2	TU for the expression of geneticin resistance	Hernanz-Koers et al. (2018)

(Continued on following page)

TABLE 1 (*Continued*) FB parts reported in this study. DNA parts are grouped according to the purpose for which they were used.

Constitutive/inducible promoters				
FB312	P*gpdA*:Nluc:T*tub*	pDGB3α1R	TU for the expression of nanoluciferase (Nluc) under P*gpdA*	Gandía et al. (2022)
FB323	P*gpdA*:Nluc:T*tub*::P*trpC*:*nptII*:T*tub*:: *Ppaf*:Luc:*Ttub*	pDGB3α1	Module for the expression of geneticin resistance and nanoluciferase, and the expression of firefly luciferase under the P*paf*	Gandía et al. (2022)
FB367*	P*gpdA*:Nluc:T*tub*::P*trpC*:*nptII*:T*tub*	pDGB3α1	Module for the expression of geneticin resistance and nanoluciferase	This study
FB417	P*afpB*:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*afpB*	This study
FB418	P*glaA*:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*glaA*	This study
FB419	P*amyB*:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*amyB*	This study
FB420	P*ef1*α:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*ef1*α	This study
FB421	P07760:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P07760	This study
FB423	P*gpdA*:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*gpdA*	This study
FB424	P*xlnA*:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*xlnA*	This study
FB426	P*pkiA*:Luc:T*trpC*	pDGB3α2	TU for the expression of luciferase under P*pkiA*	This study
FB432	FB367+FB417	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P*afpB*	This study
FB433	FB367+FB418	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P*glaA*	This study
FB434	FB367+FB419	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P*amyB*	This study
FB435	FB367+FB420	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under P*gpdA* promoter, and luciferase under P*ef1*α	This study
FB436	FB367+FB421	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P07760	This study
FB438	FB367+FB423	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P*gpdA*	This study
FB439	FB367+FB424	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P*xlnA*	This study
FB441	FB367+FB426	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under P*pkiA*	This study
dCas9-activated synthetic promoters				
Code	**Name**	**Plasmid**	**Description**	**Reference**
GB2815	RandomSequence R1	pUPD2	Random sequence R1 of 1240 bp for the A1 distal promoter position	Moreno-Giménez et al. (2022)
GB2878	G1aG2b.1	pUPD2	A2 proximal promoter sequence containing the target sequence for gRNA1 flanked by random sequences	Moreno-Giménez et al. (2022)
GB2885	G1ab.1	pUPD2	A2 proximal promoter sequence consisting of two times the target sequence for gRNA1 flanked by random sequences	Moreno-Giménez et al. (2022)
GB3276	G1abc.3	pUPD2	A2 proximal promoter sequence containing three times the target sequence for gRNA1 flanked by random sequences	Moreno-Giménez et al. (2022)
GB3413	m*PAF*	pUPD2	Minimal promoter of the *paf* gene from *P. chrysogenum*, containing 62 bp upstream the transcription start site and the 5′UTR region	Moreno-Giménez et al. (2022)
FB395*	R1:G1aG2b.1:m*PAF*:Luc:T*trpc*	pDGB3α2	TU for the expression of luciferase under a synthetic promoter containing the target sequence for gRNA1 (1xLuc)	This study
FB396*	R1:G1ab.1:m*PAF*:Luc:T*trpc*	pDGB3α2	TU for the expression of luciferase under a synthetic promoter containing two times the target sequence for gRNA1 (2xLuc)	This study

(Continued on following page)

TABLE 1 (*Continued*) FB parts reported in this study. DNA parts are grouped according to the purpose for which they were used.

dCas9-activated synthetic promoters				
FB397*	R1:G1abc.3:m*PAF*:Luc:T*trpc*	pDGB3α2	TU for the expression of luciferase under a synthetic promoter containing three times the target sequence for gRNA1 (3xLuc)	This study
FB398*	FB367+FB395	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under a synthetic promoter containing the target sequence for gRNA1	This study
FB399*	FB367+FB396	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under a synthetic promoter containing two times the target sequence for gRNA1 (2xLuc)	This study
FB400*	FB367+FB397	pDGB3Ω1	Module for the expression of geneticin resistance, Nluc under the P*gpdA* promoter, and luciferase under a synthetic promoter containing three times the target sequence for gRNA1 (3xLuc)	This study
FB403	pAMA18.0_gRNA1	pAMA18.0	Expression plasmid for the dCas9 activation system and the (GB_SynP) gRNA1	This study

*DNA parts from this study which are deposited in Addgene.

between the strains were analyzed using one-way ANOVA, followed by *post hoc* multiple comparison Tukey's test ($p < 0.05$). For induction experiments, we analyzed the differences in the growth of each strain in the presence of different carbon sources using Student's *t*-test ($p < 0.05$).

2.5 Fruit infection assays

P. digitatum parental and mutant strains were inoculated on freshly harvested oranges (*Citrus sinensis* L. Osbeck cv Lane late), as previously described (González-Candelas et al., 2010). Briefly, three replicates of five orange fruits were inoculated with 5 µL of fungal conidial suspension (10^4 conidia/mL) at four equidistant wounds around the equator. Control mock inoculations were performed with 5 µL of sterile Milli-Q H_2O. Once inoculated, fruits were maintained at 20°C and 90% relative humidity for up to 6 days. Each inoculated wound was scored daily for infection symptoms on consecutive days post-inoculation (dpi). We repeated the experiments twice. Differences in the percentage of infection for each strain compared to the control CECT 20796 were analyzed using Student's *t*-test ($p < 0.05$) for each individual dpi.

3 Results

3.1 Selection markers for antibiotic resistance

The FB platform already contains some commonly used positive fungal selection markers based on antibiotic resistance, such as *hph* (hygromycin^R, FB003) or *nptII* (geneticin^R, FB009) (Hernanz-Koers et al., 2018). However, in the case of integrative approaches, multiple genetic modifications often depend on the availability of different antibiotic resistance genes for transformant selection, which can be a bottleneck for the exploitation of filamentous fungi. In this study, we expand the range of selection markers available in the FB platform by including two alternative antibiotic resistance-inducing genes, the *ble* resistance gene from the bacterial transposon Tn5 and the squalene epoxidase *ergA* gene from *P. chrysogenum*. The expression of the *ble* gene provides the selection of the antibiotic phleomycin (Austin et al., 1990), whereas the expression of the *ergA* gene provides resistance against the antibiotic terbinafine in a broad range of filamentous fungi (Austin et al., 1990; Sigl et al., 2010).

In order to include *ble* resistance in the FB platform, a functional TU was generated. For this, we assembled the *ble* coding sequence (FB413), together with the promoter of isopenicillin N synthase (P*pcbC*) from *Penicillium rubens* (FB411) (Polli et al., 2016) and the terminator from the acetamidase (T*amds*) from *Aspergillus nidulans* (FB416) (Kelly and Hynes, 1985) into the pDGB3α2 vector to obtain FB430 (Table 2; Figure 1A) via restriction–ligation reactions. To functionally validate the resulting construct, we transformed *P. chrysogenum* and *P. digitatum* wild-type strains with the same FB430 via ATMT for the ectopic integration of the *ble* TU. *P. chrysogenum* transformants grown in the presence of 25 µg/mL phleomycin were selected and analyzed by PCR for the presence of the *ble* cassette (Figure 1B). The positive transformants, PCEM43053 and PCEM43062, showed growth on phleomycin-containing plates when compared to the parental ATCC 10002 (Figure 1C), further demonstrating the functionality of FB430. In parallel, FB430 was also validated in *P. digitatum* (Figures 1D–G). *P. digitatum* transformants grown in the presence of 35 µg/mL phleomycin were selected and confirmed by PCR (Figure 1D). The positive transformants PDZS43023, PDZS43041, and PDZS43051 were able to grow on phleomycin-containing plates (Figure 1E) and showed the same pathogenicity as the parental CECT 20796 in orange fruits (Figures 1F,G).

Similarly, to incorporate the terbinafine resistance-inducing gene in the FB platform, a functional TU for *ergA* was generated and validated in *P. chrysogenum* (Figure 2). We assembled the *ergA* coding sequence (FB414), together with the P*gpdA* promoter (FB007) and the T*amds* terminator (FB416), into the pDGB3α2 vector to obtain the FB431 construct (Table 2; Figure 2A). *P. chrysogenum* transformants grown on 0.5 µg/mL terbinafine were chosen and confirmed by PCR (Figure 2B). The positive transformants PDZS43122, PDZS43131, and PDZS43142 could grow on phleomycin-containing PDA plates in contrast to the parental ATCC 10002 (Figure 2C), demonstrating the functionality of FB431.

TABLE 2 Primers used in this study.

ID	Use*	Sequence 5′-3′**	Tm (°C)	Origin	Purpose	Reference
OJM371	F	ATAGATCTAACTGATATTGAAGGAGCA	52	P*trpC*	Molecular characterization	This study
OJM509	F	GCGCCGTCTCGCTCGGGAGTGGCGCATGCGGACAGACGG	64	P*gpdA*	Molecular characterization	Hernanz-Koers et al. (2018)
OJM522	R	GCGCCGTCTCGCTCAAGCGCATGTCTCAGACG GTCGATG	62	T*trpC*	Molecular characterization	Hernanz-Koers et al. (2018)
OJM524	F	GCTTTCGCTAAGGATGATTTCTGG	70	pUPD2	Molecular characterization	Hernanz-Koers et al. (2018)
OJM525	R	CAGGGTGGTGACACCTTGCC	66	pUPD2	Molecular characterization	Hernanz-Koers et al. (2018)
OJM555	R	TCATCATGCAACATGCATGTA	58	T*tub*	Molecular characterization	Hernanz-Koers et al. (2018)
OJM655	R	CATCCATACTCCATCCTTCCC	60	pAMA18.0	Molecular characterization and sequencing	This study
OJM656	F	CATTTTTGTCGTCATGTGCTGG	55	5′ Pdig *pyrG*	Molecular characterization	This study
OJM657	R	GAAGGCTGAACTCACTGTGG	55	3′ Pdig *pyrG*	Molecular characterization	This study
OJM662	F	GCTTTTGCTAACCATTTGGGACAC	52	GB B6 code	Cloning of FB372 into pDGB3α1	This study
OJM663	R	AGCGGTGTCCCAAATGGTTAGCAA	52	GB C1 code	Cloning of FB372 into pDGB3α1	This study
OJM696	F	TTGTCTCACTCTCTCTTTTCC	51	*pyr4*	Molecular characterization	This study
OJM697	R	ATTCCATGCTTCCAGATCC	51	*pyr4*	Molecular characterization	This study
OJM698	F	ATGGTCTCACCGACCAGTCCTGATGAGTCCGTGAG GACGAAACGAG	60	pAMA18.0	Cloning of gRNA1 into pAMA18.0	This study
OJM699	R	ATGGTCTCTAAACTCTTCTCTCACCAACCAGTC GACGAGCTTACTCGTTTCGTCCTCACGGACTCA	60	pAMA18.0	Cloning of gRNA1 into pAMA18.0	This study
OJM705	F	TCCTGGAAGTGCGTTGATCA	51	P*xlnA*	Molecular characterization	This study
OJM706	F	GGAAGAGAAAACCTCCGAGTAC	54	P*pkiA*	Molecular characterization	This study
OJM707	F	ATGAATTCCACCGAATGCAC	53	P*afpB*	Molecular characterization	This study
OJM708	F	TGCCATTGGCGGAGGGGTCC	53	P*glaA*	Molecular characterization	This study
OJM709	F	TCAACTGATTAAAGGTGCCG	53	P*amyB*	Molecular characterization	This study
OJM710	F	GTGAAAAAACGGATGGGGAC	53	P*ef1α*	Molecular characterization	This study
OJM711	F	GATAATGGTGATTCGGCGCG	53	P07760	Molecular characterization	This study
OJM715	F	GTATCTGCATGTTGCATCGG	53	P*pcbC*	Molecular characterization	This study
OJM716	R	TACCGCTCGTACCATGGGTT	53	T*amds*	Molecular characterization	This study

*F, forward; R, reverse.
**gRNA1 sequence is underlined.

Overall, both resistances were transformed ectopically to avoid any bias regarding the targeting of specific loci, and these experiments validated the use of FB430 and FB431 as standardized TUs for conferring positive selection in the transformation of different fungal species, expanding the antibiotic resistance selection markers currently available in the FB system.

3.2 Selection markers based on fungal auxotrophy

To date, no auxotrophic markers have been included in the FB platform despite the fact that they are sustainable alternatives for the use of antibiotics in transformant selection. The orotidine 5′-

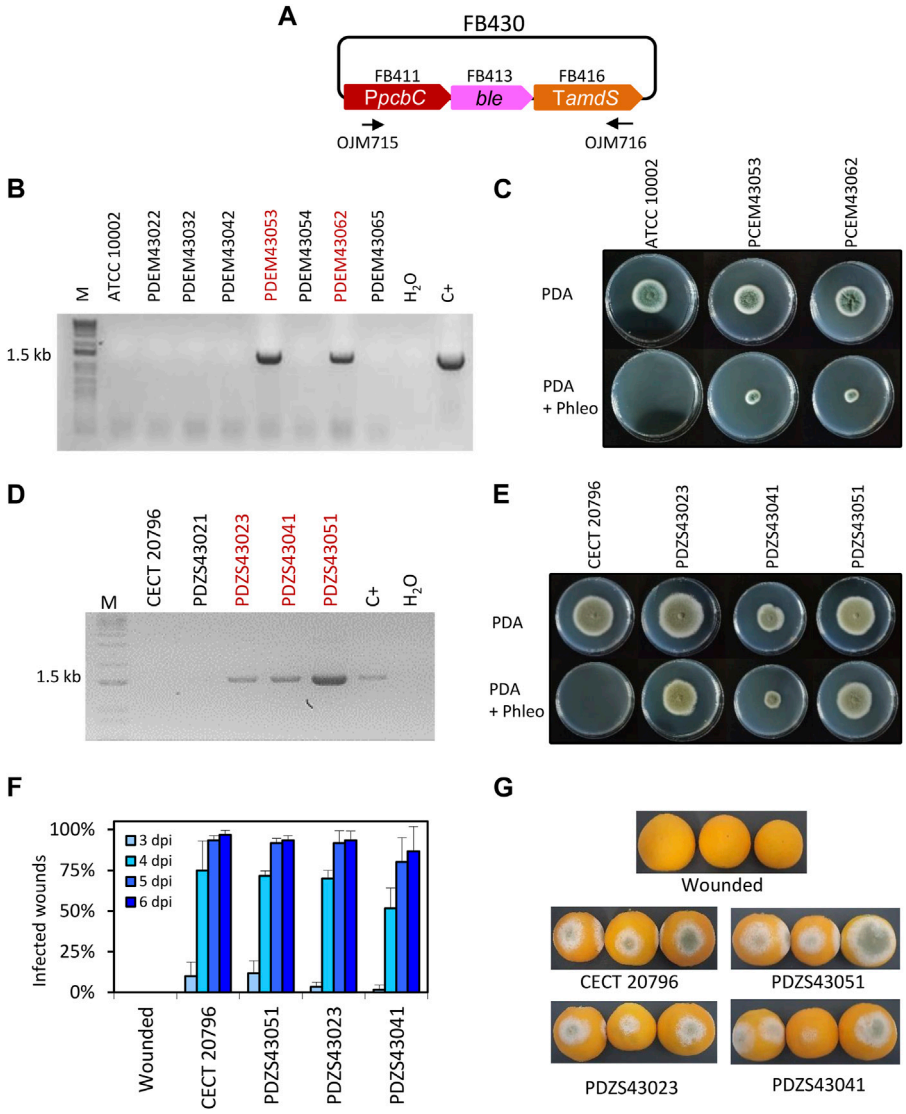


FIGURE 1
Functional validation of *ble* TU. **(A)** Plasmid pDGB3α2 FB430 for the ectopic integration of *ble* TU through ATMT to generate phleomycin resistance. Primers OJM715 and OJM716 were used for the molecular characterization of Phleo^R in *P. digitatum* and *P. chrysogenum* strains. **(B)** Molecular characterization of *P. chrysogenum* transformants. The 1.5 kb band corresponds to the complete *ble* TU. Selected strains are highlighted in red. **(C)** Growth profile of *P. chrysogenum* selected Phleo^R transformants after 7 days of growth in the presence of the antibiotic (25 μg/mL) at 25°C. **(D)** Molecular characterization of *P. digitatum* transformants. The 1.5 kb band corresponds to the complete *ble* TU as in **(B)**. Selected strains are highlighted in red. **(E)** Growth profile of *P. digitatum* Phleo^R transformants after 7 days of growth in the presence of the antibiotic (35 μg/mL) at 25°C. **(F)** Fruit infection assays of Phleo^R mutants on oranges. Data indicate the % of infected wounds (mean ± SD) at each day post-inoculation. No statistical difference was found between the parental CECT 20796 and the mutants at each dpi (*t*-test, $p < 0.05$). **(G)** Representative images of oranges infected by the indicated strains at 6 dpi.

phosphate decarboxylase *pyr4* gene from *T. reesei* is widely used as an auxotrophic selection marker that can be counter-selected using 5-FOA or fully supplemented using uridine (Díez et al., 1987; Derntl and Kiesenhofer, 2015). In this study, we set up several experiments to design, test, and validate *pyr4* as a selection marker in *pyr4/pyrG*-deficient fungal strains. As a first step, uridine-auxotrophic *P. digitatum* Δ*pyrG* mutants were generated through ATMT using FB372 as the template for homologous recombination at the *pyrG* locus (Table 2; Supplementary Figure S4). Transformants were selected on PDA plates supplemented with 1.22 g/L uridine and 1.25 g/L 5-FOA, and were molecularly and phenotypically characterized (Supplementary Figure S4). Growth profiles showed that after *pyrG* deletion, *P. digitatum* mutants could no longer grow on PDA plates unless supplemented with uridine, confirming their auxotrophic condition. Additionally, these mutants could also grow in the presence of uridine and 5-FOA in contrast to the parental CECT 20796, further confirming *pyrG* deletion. Finally, infection assays on orange fruits revealed that *P. digitatum* Δ*pyrG* mutants showed highly reduced pathogenesis compared to the control (Supplementary Figure S4). Once the *pyrG* deletion mutants were obtained, a functional TU for the *T. reesei pyr4* gene was generated. For this, we assembled the *Penicillium* codon-optimized and extensively domesticated *pyr4*

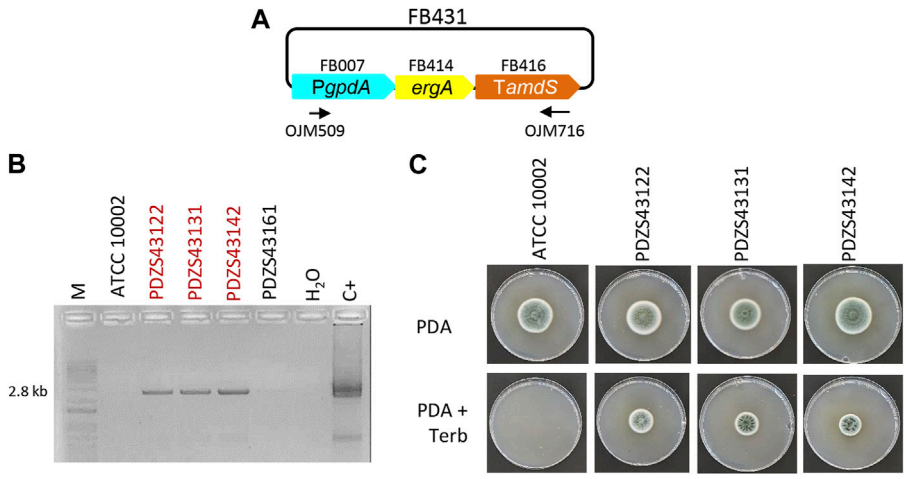


FIGURE 2
Functional validation of *ergA* TU in *P. chrysogenum*. **(A)** Plasmid pDGB3α2 FB431 for the ectopic integration of *ergA* TU through ATMT to generate terbinafine resistance. Primers OJM509 and OJM716 were used for the molecular characterization of the Terb^R strains shown in **(B)**. The 2.8 kb bands correspond to the complete *ergA* TU. Selected strains are highlighted in red. **(C)** Growth profile of selected Terb^R transformants after 7 days of growth in the presence of the antibiotic (0.5 μg/mL) at 25°C.

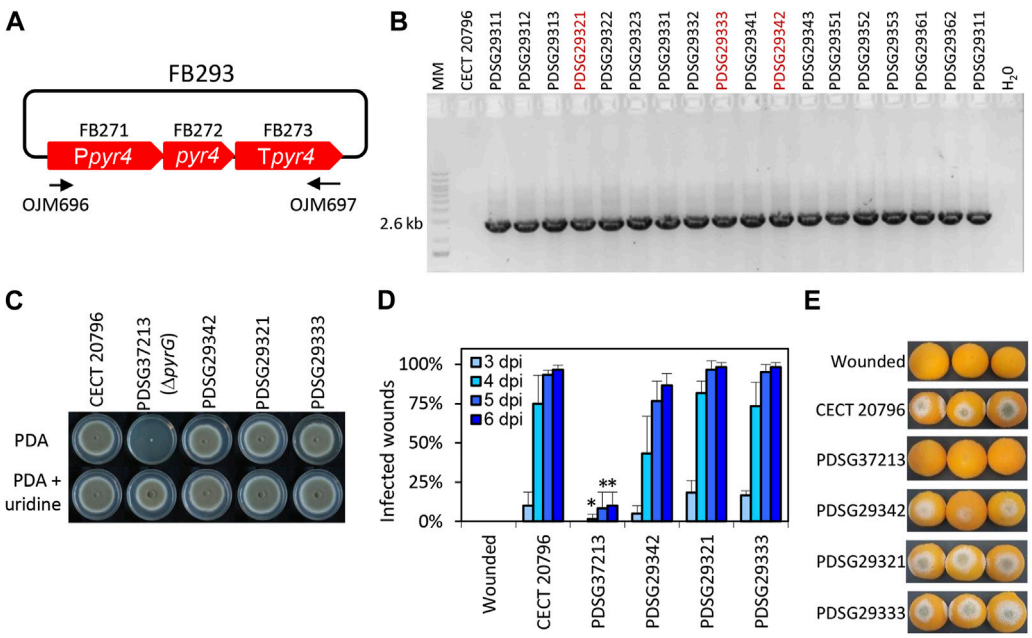


FIGURE 3
Functional validation of *T. reesei pyr4* TU in *P. digitatum*. **(A)** Plasmid pDGB3α2 FB293 for the ectopic integration of *pyr4* TU through ATMT to restore uridine auxotrophy. Primers OJM696 and OJM697 were used for the molecular characterization of the non-auxotrophic *P. digitatum* strains shown in **(B)**. The 2.6 kb bands correspond to the complete *pyr4* TU. Selected strains are highlighted in red. **(C)** Growth profile of selected Δ*pyrG:pyr4* transformants grown on PDA plates supplemented with 1.22 g/L uridine. It is to be noted that PDSG372013 was used as the parental strain for transformation with FB293. **(D)** Fruit infection assays of Δ*pyrG:pyr4* mutants on orange fruits. Data indicate the % of infected wounds (mean ± SD) at each day post-inoculation (dpi). (*) shows statistical significance between each sample compared to the control CECT 20796 at each dpi (*t*-test, $p < 0.05$). **(E)** Representative images of oranges infected by the indicated strains at 6 dpi.

coding sequence (FB272), promoter (FB271), and terminator (FB276) into the pDGB3α2 vector to obtain FB293 (Table 2; Figure 3A) via restriction–ligation reactions. To functionally validate the resulting construct FB293, we transformed the *P. digitatum* Δ*pyrG* mutant PDSG37213 with FB293 via ATMT for the ectopic integration of *pyr4* TU. Transformants grown on PDA plates were assessed by PCR (Figure 3B) and phenotypically analyzed to confirm *pyrG:pyr4* complementation and, therefore, the absence of the auxotrophy.

As shown in Figure 3C, complemented mutants PDSG29312, PDSG29321, and PDSG29333 were all able to grow on PDA plates without uridine, in contrast to the auxotrophic parental PDSG37213. Remarkably, *P. digitatum* complemented mutants that fully recovered their original pathogenicity (Figures 3D,E), which validates *pyr4* as an auxotrophic selectable marker also for (phyto)pathogenic fungi, in which the deletion of *pyrG* orthologs has been demonstrated to decrease pathogenicity and virulence in the corresponding fungi (Zameitat et al., 2007; Higashimura et al., 2022).

3.3 Constitutive and inducible promoters

In order to further expand and characterize the promoter catalog available in the FB platform, a series of promoters from different fungal species were included in the collection (Table 1) and were functionally validated using a luciferase reporter system, as previously described (Gandía et al., 2022). This reporter consists of two TUs: the nanoluciferase coding sequence (Nluc, FB310) under the regulation of the P*gpdA* promoter that serves as an internal standard for normalization, and the luciferase sequence (Luc, GB0096) under the regulation of the promoter to be tested. Promoter strength is expressed as the ratio of the Luc signal divided by the Nluc internal standard. The promoters to be evaluated included the previously characterized strong pyruvate kinase gene promoter (P*pkiA*) from *Aspergillus niger* (FB389) (de Graaff et al., 1992). Novel promoters from *P. digitatum* included the antifungal protein AfpB gene promoter P*afpB* (FB404) (Garrigues et al., 2017), and two promoters with high expression levels were reported in a previous transcriptomic study: the elongation factor 1α gene promoter (P*ef1*α) (FB407) and the ubiquitin ligase PDIG_07760 gene promoter (P07760) (FB408) (Ropero-Pérez et al., 2023). The inducible promoters included in this study are the endo-1,4-β-xylanase A gene promoter (P*xlnA*) from *A. nidulans* (FB291, xylose-responsive) (Orejas et al., 1999), the glucoamylase gene promoter (P*glaA*) from *A. niger*, and the TAKA-amylase A gene promoter (P*amyB*) from *Aspergillus oryzae* (FB405 and FB406, respectively, both maltose/starch-responsive) (Fowler et al., 1990; Tsuchiya et al., 1992b). The widely used glyceraldehyde-3-phosphate dehydrogenase promoter P*gpdA* from *A. nidulans*, which was already available in the FB collection (FB007) (Hernanz-Koers et al., 2018), was also included in the analysis, as well as the luciferase reporter construct for the *P. chrysogenum* antifungal protein PAF promoter (P*paf*) from previously published data (FB323) (Gandía et al., 2022) to serve as references. To facilitate the cloning of new luciferase reporter constructs, the nanoluciferase reference gene and the *nptII* resistance gene were cloned into a pDGB3α1 vector (FB367) to be combined in a single reaction with the luciferase TUs cloned into pDGB3α2 vectors that included the promoters to be tested (Figure 4A). Due to the requirements of the FB binary assembly, an insulator sequence (GB3458) was also included at the 3' end of the FB367 vector, which also helps prevent the interaction between Nluc and Luc TUs. The luciferase reporter constructs for each of the assayed promoters (FB432 to FB441) showed different normalized luciferase expression levels in *P. digitatum* after 2 days of growth in PDB (Figure 4B). The lowest expression levels were observed for the inducible promoters P*xlnA*, P*amyB*, and P*glaA*, in this order, due to the lack of inducers in this medium. Their expression levels, together with that driven by P*afpB,* were slightly above the basal signal observed in the control strain, but these were not statistically significant. The expression driven by P*pkiA*, P*gpdA*, and the new P*ef1*α were similar to those observed for P*paf*, while P07760 showed intermediate expression values between these expression and the inducible promoters.

The selected constructs carrying P*xlnA*, P*pkiA*, P*glaA*, P*amyB*, P*ef1*α, and P07760 were also transformed into *P. chrysogenum*, in which the luciferase expression level was about 10 times lower than the overall levels observed in *P. digitatum* (Figure 4C), except for P*ef1*α and P*amyB*, which showed a similar signal in both fungal chassis (0.27 and 0.1 for P*ef1*α in *P. digitatum* and *P. chrysogenum*, respectively, and 0.0007 and 0.0004 for P*amyB*). Unlike *P. digitatum*, the signal driven by P*amyB* was above the basal signal in *P. chrysogenum* and showed similar expression levels to that of the P7760 promoter. Relative expression levels among the other promoters were nevertheless maintained in both fungi, with P*pkiA* and P*ef1*α signals similar to that of the P*paf* reference promoter and with P*xlnA*-, P*glaA*-, and P*amyB*-driven signals similar to that of the control strains.

3.4 Induction of P*glaA*, P*amyB*, and P*xlnA* promoters

In order to further characterize the inducible promoters included in this study, we analyzed the induction of luciferase expression directed by P*glaA*, P*amyB,* and P*xlnA* promoters in *P. digitatum* (Figure 5A) and *P. chrysogenum* strains (Figure 5B) after 4 days of growth in minimal medium (PdMM for *P. digitatum* and PcMM for *P. chrysogenum*) using different inducers as the sole carbon source (2% maltose for P*glaA* and P*amyB*, and 2% xylose for P*xlnA*). When the fungi were grown in the presence of the inducer, the expression levels driven by all three promoters were significantly higher than those observed in the reference media with 2% glucose, which increase by 8x, 4x, and 10x for P*glaA*, P*amyB*, and P*xlnA*, respectively, in *P. digitatum* (Figure 5A); and by 2x, 9x, and 8x in *P. chrysogenum* (Figure 5B). Signals observed for the P*xlnA* promoter in the reference media (MM + glucose) were similar to the basal signal of the reference strains in both fungi, while P*amyB* and P*glaA* promoters showed higher basal expression in the same media, especially in *P. chrysogenum.* Remarkably, expression levels in the reference P*paf* promoter were found to increase significantly in *P. chrysogenum* when grown in the PcMM supplemented with maltose or xylose (0.22 a.u. on average) when compared to the medium supplemented with glucose (0.015 a.u. on average, 15 times lower). The same was observed, but to a lesser extent, in *P. digitatum* when P*paf* was expressed in PdMM glucose (average values of 0.6 a.u.) compared to its expression in PdMM xylose (0.15 a.u. average values, four times higher).

3.5 dCas9-activated synthetic promoters

Finally, we tested the recently developed GB_SynP (Moreno-Giménez et al., 2022) in our fungal chassis in combination with the pAMA18.0_gRNA1 plasmid, which delivers the CRISPRa system necessary to activate GB_SynP promoters in a non-integrative

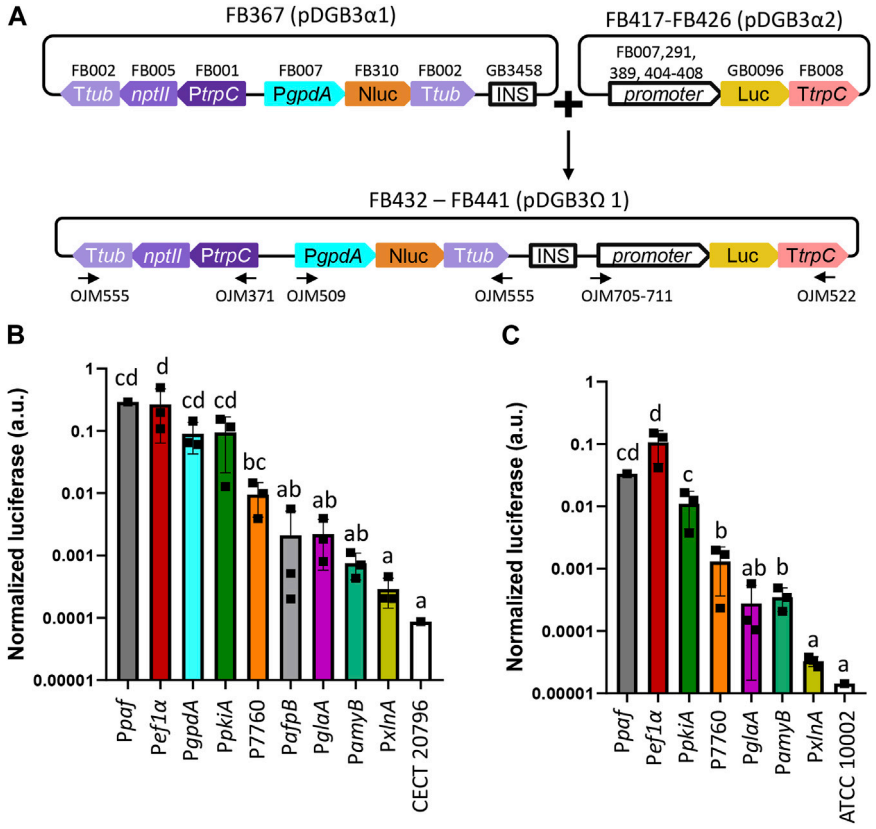


FIGURE 4
Functional promoter validation via the luciferase assay in *P. digitatum* and *P. chrysogenum*. **(A)** Scheme of the assembly architecture used to express the luciferase reporter system. Different promoters were tested using luciferase as a reporter, and the constitutive expression of nanoluciferase under the P*gpdA* promoter was used as a reference for normalization. All constructs included a geneticin resistance gene (*nptII*) for the selection of positive transformants. An insulator sequence was introduced between the nanoluciferase and luciferase genes to allow the binary assembly of plasmids. Primers used for the molecular characterization of transformants are indicated with arrows. **(B)** Normalized luciferase expression for each promoter in *P. digitatum* transformants grown in PBD for 2 days. **(C)** Normalized luciferase expression for each promoter in *P. chrysogenum* transformants grown in PBD for 2 days. Constitutive expression of luciferase under the P*paf* promoter was included as a reference. Letters denote statistical significance between values in a one-way ANOVA (Tukey's multiple comparisons test, $p \leq 0.05$). Error bars represent the average values ±SD (n = 9). Squares represent the mean value of each of the three biological replicates (transformants) measured twice. It is to be noted that the Y-axis is represented in the logarithmic scale.

manner (Mózsik et al., 2021). To this end, we developed luciferase reporter constructs following the same procedure as for natural promoters (Figure 6A). In these constructs, luciferase expression was regulated by synthetic promoters consisting of an A1 distal promoter part formed by a random sequence (GB2815), and an A2 proximal promoter part including the target sequence for gRNA1 that was repeated once (GB2878), twice (GB2885), or three times (GB3276), and the minimal promoter m*PAF* (GB3423) derived from the native P*paf* from the fungus *P. chrysogenum*, which was previously found to drive a strong induction when exposed to the dCas9 system loaded with gRNA1 in plants (Moreno-Giménez et al., 2022). The resulting constructs containing one gRNA1 target (FB395, 1xLuc), two targets (FB396, 2xLuc), or three targets (FB397, 3xLuc) were stably transformed into *P. digitatum* via ATMT. Protoplasts obtained from these strains were re-transformed with the CRISPRa expression vector pAMA18.0_gRNA1 (Figure 6B). The expression levels of 1xLuc and 2xLuc constructs were not significantly higher than the basal signal of the control strain despite the presence of pAMA18.0_gRNA1, except for one of the 1xLuc re-transformants, which showed a low, but statistically significant increase in the luciferase expression when compared to the same strain in the absence of the CRISPRa system. A higher and significant increase of approximately 16 times on average in the luciferase signal was observed for the 3xLuc construct (Figure 6C) in all tested re-transformants. Although the expression driven by this promoter was approximately 10 times lower than that observed with the reference P*paf* promoter, it is comparable to that achieved with P*glaA* and P*amyB*, showing the functionality of these dCas9-activated synthetic promoters in fungi.

4 Discussion

The FB cloning platform allows for the open exchange of standardized, ready-to-use DNA parts in the fungal research community (Hernanz-Koers et al., 2018). Moreover, if the platform is functionally compatible and validated, it also allows for the exchange of parts between plants and fungi, as occurred with

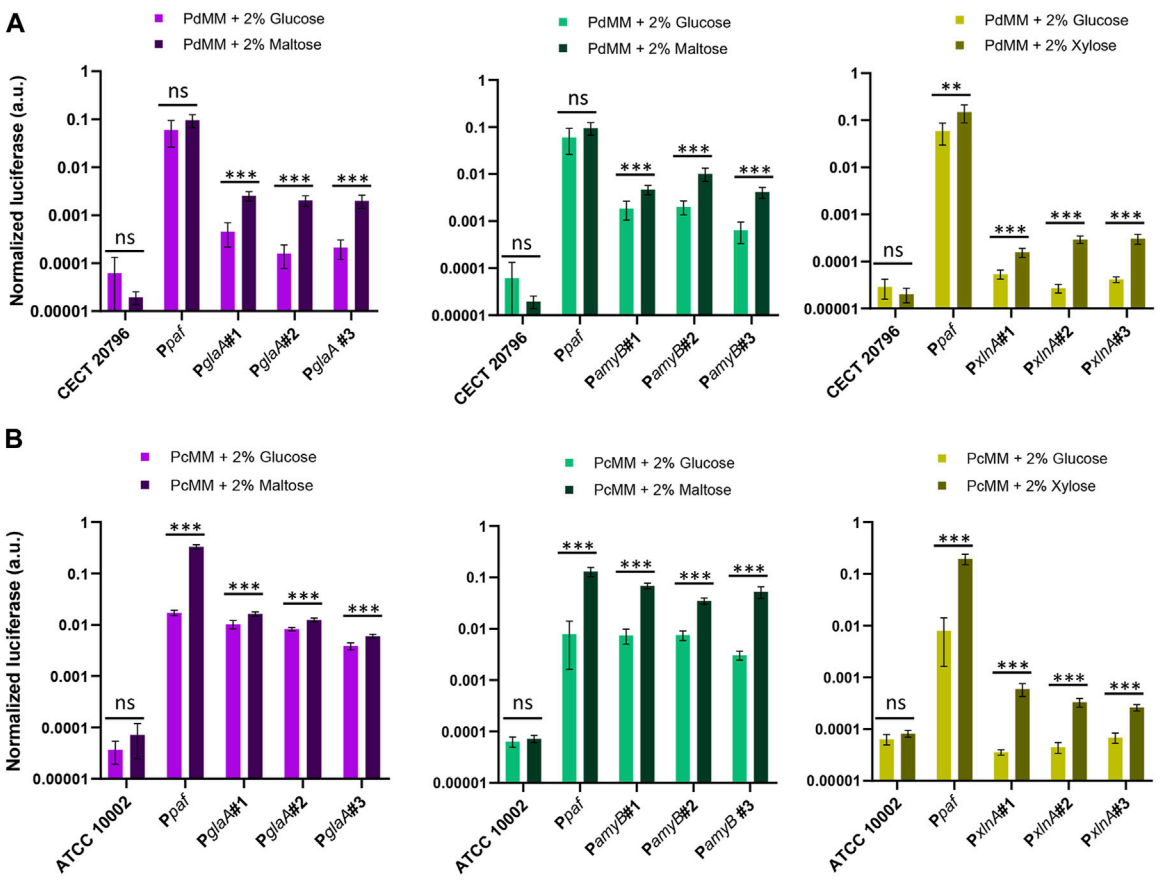


FIGURE 5
Activation of inducible promoters P*glaA*, P*amyB*, and P*xlnA* in *P. digitatum* **(A)** and *P. chrysogenum* **(B)**. Expression was measured after 4 days of growth in minimal medium, replacing glucose with maltose as the carbon source for P*amyB* and P*glaA* transformants, and with xylose for P*xlnA* transformants. The expression of luciferase under the P*paf* promoter was also included as a reference. Asterisks represent statistical significance (Student's *t*-test, ns = $p \geq 0.05$, *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$) between the expression levels of each individual transformant in MM with maltose/xylose and those observed in the reference MM with glucose. Error bars represent the average values ±SD (n = 6). It is to be noted that the Y-axis is represented in the logarithmic scale.

the fluorescent YFP protein or the hygromycin selection marker reported previously (Hernanz-Koers et al., 2018) or the synthetic promoters reported in this study. However, the number of validated genetic elements present in the FB platform was very limited to date, which hindered the biotechnological exploitation of filamentous fungi. In this study, we have expanded the available genetic elements in FB platforms by incorporating one auxotrophic selection marker (*pyr4*), two additional antibiotic resistance markers (*ergA* and *ble*), two strong promoters (P*pkiA* and P*ef1α*), two intermediate promoters (P*afpB* and P7760), three inducible promoters (P*glaA*, P*amyB*, and P*xlnA*), and three versions of the dCas9-regulated GB_SynP synthetic promoters. Even though the validation of these new parts has been performed in *Penicillium* species, the FB system has been demonstrated to mediate the expression of the same construct in different fungal genera. For instance, the FB027 construct used for the expression of YFP has been functionally validated in *P. digitatum*, *Penicillium expansum*, and *A. niger* (Hernanz-Koers et al., 2018; Vazquez-Vilar et al., 2020). Therefore, the FB system and, subsequently, the new FB parts described here, are expected to be of use in a wide range of fungal species of different genera.

Since the FB release, there have been an increasing number of SynBio-based genetic toolkits developed for filamentous fungi (Dahlmann et al., 2021; Mózsik et al., 2021; Mózsik et al., 2022). In this sense, FB, which derives from the GB cloning framework, shares most of the codes and type IIS restriction enzymes with these alternative SynBio collections (Weber et al., 2011), making it possible to combine code-compatible level 0 plasmids between these systems to assemble TUs into level 1 plasmids. However, these Golden Gate-based collections alternative to FB use plasmids derived from pAMA1 or pEHN8, which are introduced into fungal cells via protoplast transformation. In contrast, the FB collection is based on pCAMBIA-derived vectors and can be applied to a broad spectrum of fungal species that are compatible with ATMT (de Groot et al., 1998), which is considered to be a more advantageous transformation method than protoplasts as spores can be used directly for genetic transformation and transformation efficiencies are generally higher (Li et al., 2017 and references, therein). Moreover, unlike these other Golden Gate cloning systems, FB/GB systems allow for the indefinite expansion of multigenetic constructs via bipartite assemblies between pDG3α and pDGB3Ω vectors (Sarrion-Perdigones

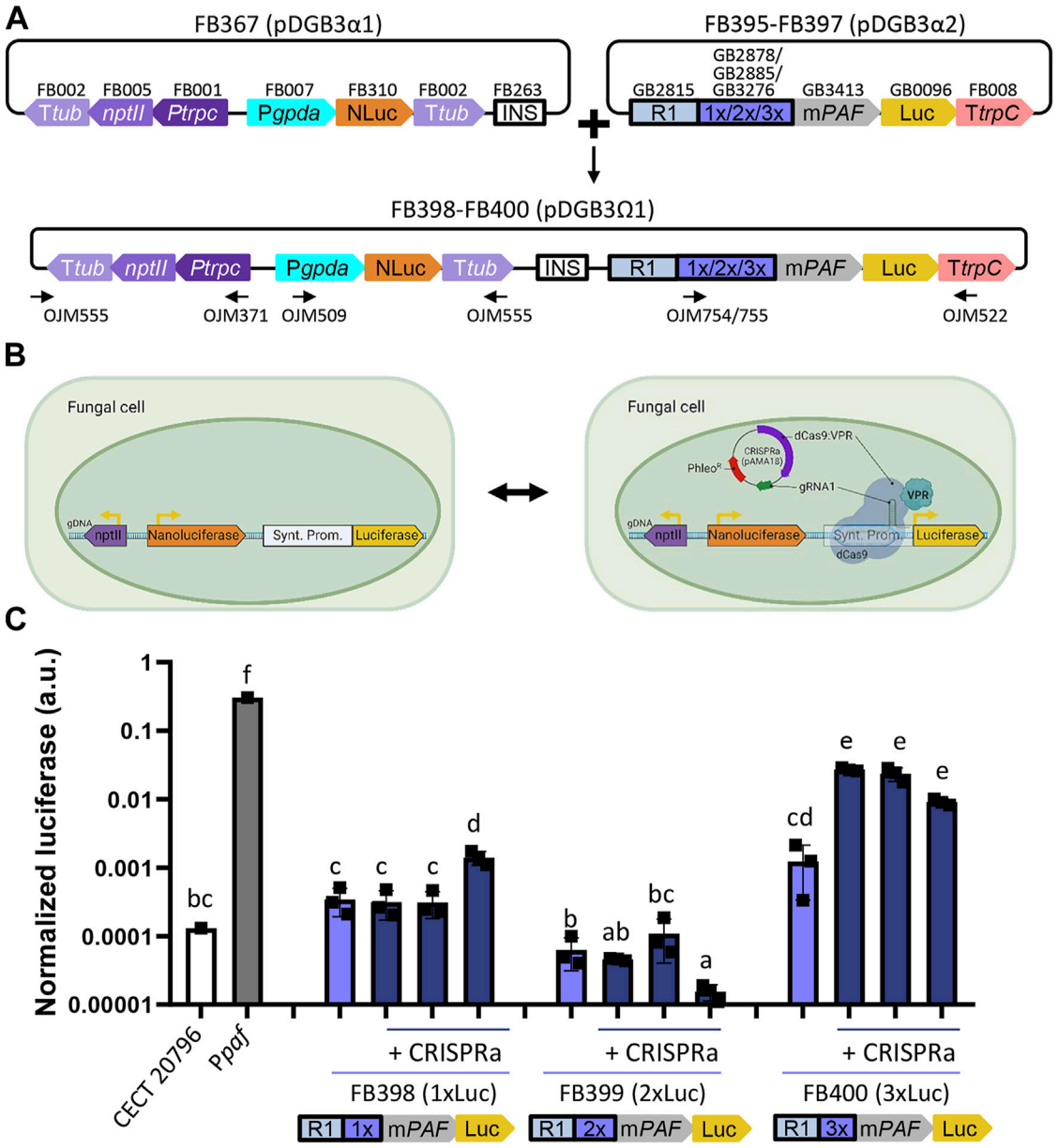


FIGURE 6
Functional validation of GB_SynP in *P. digitatum*. **(A)** Scheme of the construct architecture used to constitutively express the luciferase reporter system using the dCas9-regulated synthetic promoters with low (1xLuc, FB398), medium (2xLuc, FB399), or high (3xLuc, FB400) promoter strength. Constructs included the constitutive expression of nanoluciferase under the P*gpdA* promoter as a reference for normalization, geneticin resistance for the selection of the transformants, and an insulator sequence to allow the binary assembly of plasmids. Oligos used for the molecular characterization of transformants are indicated with arrows. **(B)** Schematic representation of the dCas9-activated luciferase reporter system. Expression of geneticin resistance and nanoluciferase is constant, while the expression of luciferase is only achieved in the presence of the dCas9-based activation system contained in the pAMA18-derived plasmid. **(C)** Expression of positive transformants for FB398 (1xLuc), FB399 (2xLuc), or FB400 (3xLuc) in the presence (+CRISPRa) or absence of the pAMA18.0_gRNA1 plasmid. Constitutive expression of luciferase under the P*paf* promoter was included as a reference. Squares represent the mean value of each of the three biological replicates (transformants) measured twice. Letters denote statistical significance between values in a one-way ANOVA (Tukey's multiple comparison test, $p \leq 0.05$). Error bars represent the average values ±SD (n = 6). It is to be noted that the y-axis is represented in the logarithmic scale. Figure includes images created with BioRender (biorender.com).

et al., 2011). To date, the GB system has permitted the assembly of up to 10 TUs (GB3243) (Selma et al., 2022) and inserted an assembly as large as 20 kb (GB4559–GB4585) (Moreno-Giménez et al., 2022), yet the transformation and propagation of larger constructs into *E. coli* might be hampered by the limitations to this host (Weber et al., 2011). In this regard, the adaptation of other ATMT-compatible vectors into FB/GB could be considered, such as the binary-BAC (BIBAC) vector reported by Hamilton (1997), which can carry >100 kb and has already been used to transform *Fusarium*, *Aspergillus*, or *Ustilago* species (Takken et al., 2004; Ali and Bakkeren, 2011).

Among the new genetic elements in the FB system, we included three commonly used fungal selection markers, two of them based

on antibiotic resistance (*ble* and *ergA*, which confer resistance to phleomycin and terbinafine, respectively) and one based on fungal auxotrophy (*pyr4*), further expanding the possibilities for fungal transformation and mutant selection within the frame of the FB platform. Although antibiotic resistance markers are among the most widely used approaches for positive transformant selection, auxotrophic markers are more sustainable alternatives for the use of antibiotics, which can have undesired side effects on the fitness of the organism under study or cause unwanted spontaneous resistance. The orotidine 5′-phosphate decarboxylase-encoding gene *pyr4* from *T. reesei*, which is an ortholog of the *Aspergillus* and *Penicillium pyrA*/*pyrG* gene, is widely applied as a strong auxotrophic selection marker that can be counter-selected using 5-FOA or fully supplemented using uracil or uridine (Díez et al., 1987; Derntl and Kiesenhofer, 2015). Interchangeability of these two orthologs has already been demonstrated between fungi from different phylogenetic classes, from Sordariomycetes to Ascomycetes and *vice versa* (Ballance and Turner, 1985; Díez et al., 1987; Gruber et al., 1990). Therefore, the FB-adapted *pyr4* TU is expected to restore uridine/uracil auxotrophy in a broad range of fungal species. In the case of fungal (plant) pathogens, for which *pyr* disruption has been reported to reduce pathogenicity and virulence (Zameitat et al., 2007; Higashimura et al., 2022), and as also demonstrated here for *P. digitatum* Δ*pyrG* for the first time (Supplementary Figure S4), *pyr4* complementation completely restored pathogenicity, thus demonstrating the suitability of this genetic element as an auxotrophic selectable marker, as well as for (phyto)pathogenic fungi.

The luciferase reporter system has allowed us to functionally validate and characterize seven distinct promoters that have been incorporated as standard DNA parts to the FB toolbox. Among these are novel promoter sequences for which their functionality had never been validated before (P*afpB*, P7760, and P*ef1*α). These promoters were selected either for their interesting behavior in a *P. digitatum*-based transcriptome analysis (Ropero-Pérez et al., 2023) or because they are well-known promoters, such as the strong P*pkiA* or the maltose-responsive P*glaA* and *PamyB*, all of which have been extensively used in *Aspergillus* species (Storms et al., 2005; Oliveira et al., 2008; Song et al., 2018). Among the inducible promoters included in this study, P*glaA* and P*xlnA* have already been implemented in alternative Golden Gate-compatible collections (Polli et al., 2016; Mózsik et al., 2021), yet the validation of either basal or induced states has not been described in *Penicillium* species. Herein, a wide expression range was found for all tested promoters, from the highly expressed P*pkiA* and P*ef1*α promoters, with levels similar to those of the well-known P*paf* and P*gpdA,* to lower- or almost no-expressed promoters, such as P07760 and P*afpB*. The expression levels of almost all these promoters were reduced in *P. chrysogenum* compared to *P. digitatum*, except for the newly characterized P*ef1*α promoter from *P digitatum*, which showed similar values in both fungal backgrounds. This likely reflects a greater orthogonality in this promoter, which may be of preferable use to ensure strong expression in other fungal chassis. Inducibility of P*glaA*, P*amyB*, and P*xlnA* was also validated using the luciferase reporter system in both *Penicillium* species, showing different expression ranges both in the presence and absence of the inducer. This allows for multiple options for the custom design of future experiments, allowing promoters with lower background expression to choose, such as P*xlnA*, when basal expression needs to be almost completely avoided or to prioritize activation over background expression with promoters such as P*amyB*. Interestingly, P*amyB* induction in *P. chrysogenum* was similar to that shown in the industrial workhorse *A. oryzae* using β-glucuronidase (GUS) as the reporter. The expression of this promoter was reported to increase 10 times in a maltose-containing medium compared to glucose (Ozeki et al., 1996). This expression was slightly higher (1250 U/mg) (Tada et al., 1991) than that reported for P*glaA* (903 mg U/mg) (Hata et al., 1992), which correlates with our results shown in Figure 5B. Both P*glaA* and P*amyB* promoters are commonly applied for the production of different proteins of interest, such as human tissue plasminogen (Wiebe et al., 2001), bovine chymosin (Ohno et al., 2011), or synthetic human lysozyme (Tsuchiya et al., 1992a), which further shows their relevance in the field of fungal biotechnology. On the other hand, the use of the P*xlnA* promoter is very limited to date, being its ortholog, P*xylP* from *P. chrysogenum* more extensively used (Yap et al., 2022). Herein, we demonstrate the possibility of implementing this promoter in the *Penicillium* genus, with more modest induction levels than those of P*glaA* and P*amyB* but with the lack of basal expression in the absence of the inducer.

Unexpectedly, the P*paf* expression was found to significantly increase in the presence of maltose in *P. chrysogenum* and xylose in both *P. chrysogenum* and *P. digitatum.* This would suggest that (i) maltose and xylose themselves or any of the maltose/xylose catabolic intermediates serve as inducers for P*paf* or (ii) P*paf* expression is partially repressed by glucose, which can be attributed to the presence of carbon catabolite repression CREA motifs in the P*paf* sequence, as previously described (Marx et al., 1995). This repression is nevertheless almost completely lost when P*paf* is expressed in a different fungal chassis, such as *P. digitatum*, suggesting different regulatory mechanisms between both fungal species despite their phylogenetic proximity.

The activation of GB_SynP promoters in *P. digitatum* was addressed using the luciferase reporter system and the CRISPRa system included in the pAMA18.0_gRNA1 vector (Mózsik et al., 2021). The non-integrative nature of this pAMA1-based plasmid makes it possible to revert promoter activation upon plasmid loss in the absence of selection pressure (Garrigues et al., 2022). Additionally, this CRISPRa system provides a method to easily assay expression variations within the same background strain, either by testing different activation domains or inducible systems, or by analyzing the induction level under different culture conditions. The activation of 1xLuc and 2xLuc constructs, however, was not achieved in *P. digitatum* using the pAMA18.0_gRNA1 vector as signals of all but one 1xLuc re-transformant were on the same range as the basal signal of the reference strain. The expression of 1xLuc was, however, not different from the expression observed in 3xLuc strains in the absence of the CRISPRa system, which could mean that this expression is within the range of basal expression of the synthetic promoters. On the other hand, in the case of the 3xLuc construct, we did observe an increase of more than one order of magnitude in the presence of the CRISPRa system compared to the non-activation control for all 3xLuc re-transformants tested. These results indicate that the activation of these promoters in fungi requires the presence of at least three repetitions of the gRNA target, which highly differ from what was observed in plants, where one repetition of the target sequence for gRNA1 was sufficient to drive a

significant increase of synthetic promoter expression (Moreno-Giménez et al., 2022). The discrepancies in the GB_SynP behavior between plants and fungi could be attributed to the differences in CRISPRa systems used in each organism. Although pAMA18.0_gRNA1 used for activation in *P. digitatum* comprised the dCas9 protein fused to the VPR activation domain, the dCasEV2.1 complex used for activation in *Nicotiana benthamiana* plants includes a dCas9 protein fused to an EDLL activation domain and an extra MS2 protein fused to the VPR domain that is able to recognize and bind the modified gRNA scaffold (Selma et al., 2019; Moreno-Giménez et al., 2022). Although VPR showed a major contribution in the activation as the expression levels dropped significantly when MS2 was fused to other activation domains (Moreno-Giménez et al., 2022), in fungi, a second activation component might be required to reach higher activation levels. Another explanation for low expression levels in fungi may reside in gRNA1 used to trigger the activation of GB_SynP promoters, which was originally designed for plants. Although no off-targets were found for gRNA1 in the *P. digitatum* CECT 20796 genome, the efficiency of this gRNA may not be optimal for this chassis, and therefore, a gRNA designed specifically for fungi might enhance the activated expression levels. Additionally, expression levels in fungi could also be enhanced by creating new A2 proximal promoter parts with more than three repetitions of the gRNA1 target sequence. Further optimization of GB_SynP promoters for filamentous fungi, following these guidelines, will be explored in the near future to better characterize this tool and its potential for wide-range expression of customizable synthetic promoters in filamentous fungi.

5 Conclusion

FungalBraid 2.0 aims to accelerate the development of fungal SynBio by the inclusion of a new repertoire of 27 domesticated DNA parts, resulting in functionally validated resistance and auxotrophic markers, as well as strong, inducible, and synthetic promoters for their contributory use by the fungal research community. With the newly expanded FB toolkit, a greater number of modular DNA assemblies are possible, exponentially increasing the possibilities for the study, development, and exploitation of filamentous fungi as cell biofactories.

Data availability statement

The datasets presented in this study can be found in the online repository DIGITAL CSIC (https://digital.csic.es/handle/10261/308066). DNA parts generated in this work can be found at: https://gbcloning.upv.es/ and https://www.addgene.org/ accession number(s) can be found in the Supplementary Material.

Author contributions

EM-G: methodology, investigation, formal analysis, visualization, and writing—original draft. MG: investigation and writing—review and editing. ZS: investigation. PM: conceptualization, supervision, funding acquisition, and writing—review and editing. LY: supervision and writing—review and editing. DO: resources, supervision and writing—review and editing. JM: conceptualization, methodology, supervision, funding acquisition, and writing—review and editing. SG: conceptualization, investigation, formal analysis, visualization, supervision, writing—original draft, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by PROMETEO/2018/066 from "Conselleria d'Educació" (Generalitat Valenciana, Comunitat Valenciana, Spain), grant PID 2021-125858OB-100, and the Severo Ochoa Excellence Program CEX 2021-001189-S funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe." EM-G was the recipient of a predoctoral grant FPU18/02019 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future." SG holds a Juan de la Cierva Incorporación grant (IJC 2020-042749-I) funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1222812/full#supplementary-material

References

Ali, S., and Bakkeren, G. (2011). Introduction of large DNA inserts into the barley pathogenic fungus, *Ustilago hordei*, via recombined binary BAC vectors and *agrobacterium*-mediated transformation. *Curr. Genet.* 57, 63–73. doi:10.1007/s00294-010-0324-0

Austin, B., Hall, R. M., and Tyler, B. M. (1990). Optimized vectors and selection for transformation of *Neurospora crassa* and *Aspergillus nidulans* to bleomycin and phleomycin resistance. *Gene* 93, 157–162. doi:10.1016/0378-1119(90)90152-h

Ballance, D. J., and Turner, G. (1985). Development of a high-frequency transforming vector for *Aspergillus nidulans*. *Gene* 36, 321–331. doi:10.1016/0378-1119(85)90187-8

Benner, S. A., and Sismour, A. M. (2005). Synthetic biology. *Nat. Rev. Genet.* 6, 533–543. doi:10.1038/nrg1637

Dahlmann, T. A., Terfehr, D., Becker, K., and Teichert, I. (2021). Golden gate vectors for efficient gene fusion and gene deletion in diverse filamentous fungi. *Curr. Genet.* 67, 317–330. doi:10.1007/s00294-020-01143-2

De Graaff, L., Van Den Broeck, H., and Visser, J. (1992). Isolation and characterization of the *Aspergillus niger* pyruvate kinase gene. *Curr. Genet.* 22, 21–27. doi:10.1007/bf00351737

De Groot, M. J. A., Bundock, P., Hooykaas, P. J. J., and Beijersbergen, A. G. M. (1998). *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. *Nat. Biotechnol.* 16, 839–842. doi:10.1038/nbt0998-839

De Vries, R. P., Patyshakuliyeva, A., Garrigues, S., and Agarwal-Jans, S. (2020). "The current biotechnological status and potential of plant and algal biomass degrading/modifying enzymes from ascomycete fungi," in *Grand challenges in fungal biotechnology*. Editor H. Nevalainen (Cham: Springer International Publishing), 81–120.

Derntl, C., and Kiesenhofer Daniel, P., Mach robert, L., and mach-aigner astrid, R. (2015). novel strategies for genomic manipulation of *Trichoderma reesei* with the purpose of strain engineering. *Appl. Environ. Microbiol.* 81, 6314–6323. doi:10.1128/aem.01545-15

Díez, B., Alvarez, E., Cantoral, J. M., Barredo, J. L., and Martín, J. F. (1987). Selection and characterization of *pyrG* mutants of *Penicillium chrysogenum* lacking orotidine-5′-phosphate decarboxylase and complementation by the *pyr4* gene of *Neurospora crassa*. *Curr. Genet.* 12, 277–282. doi:10.1007/bf00435290

Fierro, F., Vaca, I., Castillo, N. I., García-Rico, R. O., and Chávez, R. (2022). *Penicillium chrysogenum*, a vintage model with a cutting-edge profile in biotechnology. *Microorganisms* 10, 573. doi:10.3390/microorganisms10030573

Fowler, T., Berka, R. M., and Ward, M. (1990). Regulation of the *glaA* gene of *Aspergillus niger*. *Curr. Genet.* 18, 537–545. doi:10.1007/bf00327025

Gandía, M., Moreno-Giménez, E., Giner-Llorca, M., Garrigues, S., Ropero-Pérez, C., Locascio, A., et al. (2022). Development of a FungalBraid *Penicillium expansum*-based expression system for the production of antifungal proteins in fungal biofactories. *Microb. Biotechnol.* 15, 630–647. doi:10.1111/1751-7915.14006

Garrigues, S., Gandía, M., Popa, C., Borics, A., Marx, F., Coca, M., et al. (2017). Efficient production and characterization of the novel and highly active antifungal protein AfpB from *Penicillium digitatum*. *Sci. Rep.* 7, 14663. doi:10.1038/s41598-017-15277-w

Garrigues, S., Manzanares, P., and Marcos, J. F. (2022). Application of recyclable CRISPR/Cas9 tools for targeted genome editing in the postharvest pathogenic fungi *Penicillium digitatum* and *Penicillium expansum*. *Curr. Genet.* 68, 515–529. doi:10.1007/s00294-022-01236-0

González-Candelas, L., Alamar, S., Sánchez-Torres, P., Zacarías, L., and Marcos, J. F. (2010). A transcriptomic approach highlights induction of secondary metabolism in citrus fruit in response to P*enicillium digitatum* infection. *BMC Plant Biol.* 10, 194. doi:10.1186/1471-2229-10-194

Gruber, F., Visser, J., Kubicek, C. P., and De Graaff, L. H. (1990). The development of a heterologous transformation system for the cellulolytic fungus *Trichoderma reesei* based on a *pyrG*-negative mutant strain. *Curr. Genet.* 18, 71–76. doi:10.1007/bf00321118

Hamilton, C. M. (1997). A binary-BAC system for plant transformation with high-molecular-weight DNA. *Gene* 200, 107–116. doi:10.1016/s0378-1119(97)00388-0

Harries, E., Gandía, M., Carmona, L., and Marcos, J. F. (2015). The *Penicillium digitatum* protein O-mannosyltransferase Pmt2 is required for cell wall integrity, conidiogenesis, virulence and sensitivity to the antifungal peptide PAF26. *Mol. Plant Pathol.* 16, 748–761. doi:10.1111/mpp.12232

Hata, Y., Kitamoto, K., Gomi, K., Kumagai, C., and Tamura, G. (1992). Functional elements of the promoter region of the *Aspergillus oryzae glaA* gene encoding glucoamylase. *Curr. Genet.* 22, 85–91. doi:10.1007/bf00351466

Hegedüs, N., Sigl, C., Zadra, I., Pócsi, I., and Marx, F. (2011). The *paf* gene product modulates asexual development in *Penicillium chrysogenum*. *J. Basic Microbiol.* 51, 253–262. doi:10.1002/jobm.201000321

Hernanz-Koers, M., Gandía, M., Garrigues, S., Manzanares, P., Yenush, L., Orzaez, D., et al. (2018). FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. *Fungal Genet. Biol.* 116, 51–61. doi:10.1016/j.fgb.2018.04.010

Higashimura, N., Hamada, A., Ohara, T., Sakurai, S., Ito, H., and Banba, S. (2022). The target site of the novel fungicide quinofumelin, *Pyricularia oryzae* class II dihydroorotate dehydrogenase. *J. Pestic. Sci.* 47, 190–196. doi:10.1584/jpestics.d22-027

Kelly, J. M., and Hynes, M. J. (1985). Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*. *EMBO J.* 4, 475–479. doi:10.1002/j.1460-2075.1985.tb03653.x

Khang, C. H., Park, S.-Y., Rho, H.-S., Lee, Y.-H., and Kang, S. (2007). "Filamentous fungi (*Magnaporthe grisea* and *Fusarium oxysporum*)," in *Agrobacterium protocols volume 2*. Editor K. Wang (Totowa, NJ: Humana Press), 403–420.

Li, D., Tang, Y., Lin, J., and Cai, W. (2017). Methods for genetic transformation of filamentous fungi. *Microb. Cell Fact.* 16, 168. doi:10.1186/s12934-017-0785-7

Marcet-Houben, M., Ballester, A.-R., De La Fuente, B., Harries, E., Marcos, J. F., González-Candelas, L., et al. (2012). Genome sequence of the necrotrophic fungus *Penicillium digitatum*, the main postharvest pathogen of citrus. *BMC Genom* 13, 646. doi:10.1186/1471-2164-13-646

Martins-Santana, L., Nora, L. C., Sanches-Medeiros, A., Lovate, G. L., Cassiano, M. H. A., and Silva-Rocha, R. (2018). Systems and synthetic biology approaches to engineer fungi for fine chemical production. *Front. Bioeng. Biotechnol.* 6, 117. doi:10.3389/fbioe.2018.00117

Marx, F., Haas, H., Reindl, M., Stöffler, G., Lottspeich, F., and Redl, B. (1995). Cloning, structural organization and regulation of expression of the *Penicillium chrysogenum paf* gene encoding an abundantly secreted protein with antifungal activity. *Gene* 167, 167–171. doi:10.1016/0378-1119(95)00701-6

Meyer, V., Andersen, M. R., Brakhage, A. A., Braus, G. H., Caddick, M. X., Cairns, T. C., et al. (2016). Current challenges of research on filamentous fungi in relation to human welfare and a sustainable bio-economy: A white paper. *Fungal Biol. Biotechnol.* 3, 6. doi:10.1186/s40694-016-0024-8

Meyer, V., Basenko, E. Y., Benz, J. P., Braus, G. H., Caddick, M. X., Csukai, M., et al. (2020). Growing a circular economy with fungal biotechnology: A white paper. *Fungal Biol. Biotechnol.* 7, 5. doi:10.1186/s40694-020-00095-z

Moreno-Giménez, E., Selma, S., Calvache, C., and Orzáez, D. (2022). GB_SynP: A modular dCas9-regulated synthetic promoter collection for fine-tuned recombinant gene expression in plants. *ACS Synth. Biol.* 11, 3037–3048. doi:10.1021/acssynbio.2c00238

Mózsik, L., Iacovelli, R., Bovenberg, R. a. L., and Driessen, A. J. M. (2022). Transcriptional activation of biosynthetic gene clusters in filamentous fungi. *Front. Bioeng. Biotechnol.* 10, 901037. doi:10.3389/fbioe.2022.901037

Mózsik, L., Pohl, C., Meyer, V., Bovenberg, R. a. L., Nygård, Y., and Driessen, A. J. M. (2021). Modular synthetic biology toolkit for filamentous fungi. *ACS Synth. Biol.* 10, 2850–2861. doi:10.1021/acssynbio.1c00260

Ohno, A., Maruyama, J.-I., Nemoto, T., Arioka, M., and Kitamoto, K. (2011). A carrier fusion significantly induces unfolded protein response in heterologous protein production by *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* 92, 1197–1206. doi:10.1007/s00253-011-3487-9

Oliveira, J. M., Van Der Veen, D., De Graaff, L. H., and Qin, L. (2008). Efficient cloning system for construction of gene silencing vectors in *Aspergillus niger*. *Appl. Microbiol. Biotechnol.* 80, 917–924. doi:10.1007/s00253-008-1640-x

Orejas, M., Maccabe, A. P., Pérez González, J. A., Kumar, S., and Ramón, D. (1999). Carbon catabolite repression of the *Aspergillus nidulans xlnA* gene. *Mol. Microbiol.* 31, 177–184. doi:10.1046/j.1365-2958.1999.01157.x

Ozeki, K., Kanda, A., Hamachi, M., and Nunokawa, Y. (1996). Construction of a promoter probe vector autonomously maintained in *Aspergillus* and characterization of promoter regions derived from *A. niger* and *A. oryzae* genomes. *Biosci. Biotechnol. Biochem.* 60, 383–389. doi:10.1271/bbb.60.383

Palou, L. (2014). "*Penicillium digitatum, Penicillium italicum* (green mold, blue mold)," in *Postharvest decay*. Editor S. Bautista-Baños (San Diego: Academic Press), 45–102.

Polli, F., Meijrink, B., Bovenberg, R. A. L., and Driessen, A. J. M. (2016). New promoters for strain engineering of *Penicillium chrysogenum*. *Fungal Genet. Biol.* 89, 62–71. doi:10.1016/j.fgb.2015.12.003

Robey, M. T., Caesar, L. K., Drott, M. T., Keller, N. P., and Kelleher, N. L. (2021). An interpreted atlas of biosynthetic gene clusters from 1,000 fungal genomes. *PNAS* 118, e2020230118. doi:10.1073/pnas.2020230118

Ropero-Pérez, C., Bolós, B., Giner-Llorca, M., Locascio, A., Garrigues, S., Gandía, M., et al. (2023). Transcriptomic profile of *Penicillium digitatum* reveals novel aspects of the mode of action of the antifungal protein AfpB. *Microbiol. Spectr.* 11 (3), e0484622. doi:10.1128/spectrum.04846-22

Sarrion-Perdigones, A., -Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P., et al. (2013). GoldenBraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* 162, 1618–1631. doi:10.1104/pp.113.217661

Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-Del-Carmen, A., Granell, A., et al. (2011). GoldenBraid: An iterative cloning system for standardized assembly of reusable genetic modules. *PLOS ONE* 6, e21622. doi:10.1371/journal.pone.0021622

Selma, S., Bernabé-Orts, J. M., Vazquez-Vilar, M., Diego-Martin, B., Ajenjo, M., Garcia-Carpintero, V., et al. (2019). Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-Based programmable transcriptional activator. *Plant Biotechnol. J.* 17, 1703–1705. doi:10.1111/pbi.13138

Selma, S., Sanmartín, N., Espinosa-Ruiz, A., Gianoglio, S., Lopez-Gresa, M. P., Vázquez-Vilar, M., et al. (2022). Custom-made design of metabolite composition in *N. benthamiana* leaves using CRISPR activators. *Plant Biotechnol. J.* 20, 1578–1590. doi:10.1111/pbi.13834

Sigl, C., Handler, M., Sprenger, G., Kürnsteiner, H., and Zadra, I. (2010). A novel homologous dominant selection marker for genetic transformation of *Penicillium chrysogenum:* Overexpression of squalene epoxidase-encoding *ergA*. *J. Biotechnol.* 150, 307–311. doi:10.1016/j.jbiotec.2010.09.941

Sonderegger, C., Galgóczy, L., Garrigues, S., Fizil, Á., Borics, A., Manzanares, P., et al. (2016). A *Penicillium chrysogenum*-based expression system for the production of small, cysteine-rich antifungal proteins for structural and functional analyses. *Microb. Cell Fact.* 15, 192. doi:10.1186/s12934-016-0586-4

Song, L., Ouedraogo, J. P., Kolbusz, M., Nguyen, T. T. M., and Tsang, A. (2018). Efficient genome editing using tRNA promoter-driven CRISPR/Cas9 gRNA in *Aspergillus niger*. *PLOS ONE* 13, e0202868. doi:10.1371/journal.pone.0202868

Storms, R., Zheng, Y., Li, H., Sillaots, S., Martinez-Perez, A., and Tsang, A. (2005). Plasmid vectors for protein production, gene expression and molecular manipulations in *Aspergillus niger*. *Plasmid* 53, 191–204. doi:10.1016/j.plasmid.2004.10.001

Tada, S., Gomi, K., Kitamoto, K., Takahashi, K., Tamura, G., and Hara, S. (1991). Construction of a fusion gene ComprisingThe taka-amylase A promoter and the *Escherichia coli* β-glucuronidase gene and analysis of its expression in *Aspergillus oryzae*. *Mol. Genet. Genom.* 229, 301–306. doi:10.1007/bf00272170

Takken, F. L., Van Wijk, R., Michielse, C. B., Houterman, P. M., Ram, A. F., and Cornelissen, B. J. (2004). A one-step method to convert vectors into binary vectors suited for *agrobacterium*-mediated transformation. *Curr. Genet.* 45, 242–248. doi:10.1007/s00294-003-0481-5

Tsuchiya, K., Tada, S., Gomi, K., Kitamoto, K., Kumagai, C., Jigami, Y., et al. (1992a). High level expression of the synthetic human lysozyme gene in *Aspergillus oryzae*. *Appl. Microbiol. Biotechnol.* 38, 109–114. doi:10.1007/bf00169428

Tsuchiya, K., Tada, S., Gomi, K., Kitamoto, K., Kumagai, C., and Tamura, G. (1992b). Deletion analysis of the taka-amylase A gene promoter using a homologous transformation system in *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.* 56, 1849–1853. doi:10.1271/bbb.56.1849

Vazquez-Vilar, M., Gandía, M., García-Carpintero, V., Marqués, E., Sarrion-Perdigones, A., Yenush, L., et al. (2020). Multigene engineering by GoldenBraid cloning: From plants to filamentous fungi and beyond. *Curr. Protoc. Mol. Biol.* 130, e116. doi:10.1002/cpmb.116

Vazquez-Vilar, M., Quijano-Rubio, A., Fernandez-Del-Carmen, A., Sarrion-Perdigones, A., Ochoa-Fernandez, R., Ziarsolo, P., et al. (2017). GB3.0: A platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Res.* 45, 2196–2209. doi:10.1093/nar/gkw1326

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLOS ONE* 6, e16765. doi:10.1371/journal.pone.0016765

Wiebe, M. G., Karandikar, A., Robson, G. D., Trinci, A. P. J., Candia, J.-L. F., Trappe, S., et al. (2001). Production of tissue plasminogen activator (t-PA) in *Aspergillus niger*. *Biotechnol. Bioeng.* 76, 164–174. doi:10.1002/bit.1156

Yap, A., Glarcher, I., Misslinger, M., and Haas, H. (2022). Characterization and engineering of the xylose-inducible *xylP* promoter for use in mold fungal species. *Metab. Eng. Commun.* 15, e00214. doi:10.1016/j.mec.2022.e00214

Zameitat, E., Freymark, G., Dietz, C. D., Löffler, M., and Bölker, M. (2007). Functional expression of human dihydroorotate dehydrogenase (DHODH) in *pyr4* mutants of *Ustilago maydis* allows target validation of DHODH inhibitors *in vivo*. *Appl. Environ. Microbiol.* 73, 3371–3379. doi:10.1128/aem.02569-06

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 30 October 2023
DOI 10.3389/fbioe.2023.1278601

OPEN ACCESS

EDITED BY
Dipesh Dhakal,
University of Florida, United States

REVIEWED BY
Chin-Yuan Chang,
National Chiao Tung University, Taiwan
Cuauhtemoc Licona-Cassani,
Monterrey Institute of Technology and Higher Education (ITESM), Mexico

*CORRESPONDENCE
Bin Hong,
binhong69@hotmail.com,
hongbin@imb.pumc.edu.cn
Yunying Xie,
xieyy@imb.pumc.edu.cn
Yuanyuan Shi,
luckshiyuanyuan@163.com

RECEIVED 16 August 2023
ACCEPTED 18 October 2023
PUBLISHED 30 October 2023

CITATION
Lu Y, Li Y, Fan J, Li X, Sun H, Wang L, Han X, Zhu Y, Zhang T, Shi Y, Xie Y and Hong B (2023), Expanding structural diversity of 5′-aminouridine moiety of sansanmycin via mutational biosynthesis.
Front. Bioeng. Biotechnol. 11:1278601.
doi: 10.3389/fbioe.2023.1278601

Expanding structural diversity of 5′-aminouridine moiety of sansanmycin via mutational biosynthesis

Yuan Lu[1], Yihong Li[1], Jiahui Fan[1], Xingxing Li[1], Hongmin Sun[1], Lifei Wang[1], Xingli Han[2,3,4,5], Yuting Zhu[2,3,4], Tianyu Zhang[2,3,4,5], Yuanyuan Shi[1]*, Yunying Xie[1]* and Bin Hong[1]*

[1]CAMS Key Laboratory of Synthetic Biology for Drug Innovation and NHC Key Laboratory of Biotechnology of Antibiotics, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]State Key Laboratory of Respiratory Disease, Guangzhou Institutes of Biomedicine and Health (GIBH), Chinese Academy of Sciences (CAS), Guangzhou, China, [3]China-New Zealand Joint Laboratory of Biomedicine and Health, Guangzhou Institutes of Biomedicine and Health (GIBH), Chinese Academy of Sciences (CAS), Guangzhou, China, [4]Guangdong-Hong Kong-Macau Joint Laboratory of Respiratory Infectious Diseases, Guangzhou, China, [5]University of Chinese Academy of Sciences (UCAS), Beijing, China

Sansanmycins represent a family of uridyl peptide antibiotics with antimicrobial activity specifically against *Mycobacterium tuberculosis* (including drug-resistant *M. tuberculosis*) and *Pseudomonas aeruginosa*. They target translocase I (MraY) to inhibit bacterial cell wall assembly. Given the unique mechanism of action, sansanmycin has emerged as a potential lead compound for developing new anti-tuberculosis drugs, while the 5′-aminouridine moiety plays a crucial role in the pharmacophore of sansanmycin. For expanding the structural diversity of the 5′-aminouridine moiety of sansanmycin through biosynthetic methods, we firstly demonstrated that SsaM and SsaK are responsible for the biosynthesis of the 5′-aminouridine moiety of sansanmycin *in vivo*. Using the *ssaK* deletion mutant (SS/KKO), we efficiently obtained a series of new analogues with modified 5′-aminouridine moieties through mutational biosynthesis. Based on molecular networking analysis of MS/MS, twenty-two new analogues (SS-KK-1 to -13 and SS-KK-A to -I) were identified. Among them, four new analogues (SS-KK-1 to -3 and SS-KK-C) were purified and bioassayed. SS-KK-2 showed better antibacterial activity against *E. coli ΔtolC* than the parent compound sansanmycin A. SS-KK-3 showed the same anti-TB activity as sansanmycin A against *M. tuberculosis* H37Rv as well as clinically isolated, drug-sensitive and multidrug-resistant *M. tuberculosis* strains. Furthermore, SS-KK-3 exhibited significantly improved structural stability compared to sansanmycin A. The results suggested that mutasynthesis is an effective and practical strategy for expanding the structural diversity of 5′-aminouridine moiety in sansanmycin.

KEYWORDS

sansanmycin, 5′-aminouridine, structural diversity, mutational biosynthesis, new analogues

1 Introduction

Tuberculosis (TB), an infectious disease caused by *Mycobacterium tuberculosis*, remains one of the most significant global health threats. The recently published "2022 Global Tuberculosis Report" (World Health Organization, 2022) points that TB is the second deadliest infectious disease following COVID-19. The COVID-19 pandemic has reversed the progress made against TB in recent years. Global TB deaths increased from 1.4 million in 2019 to 1.5 million in 2020 and further to 1.6 million in 2021. In the near future, TB may surpass COVID-19 again as the leading cause of death worldwide from a single infectious source as COVID-19 is gradually under control. The current approach to treating TB involves a 6-month course of quadruple therapy consisting of rifampicin (RIF), isoniazid (INH), ethambutol, and pyrazinamide (Zumla et al., 2013). While this treatment can effectively control drug-sensitive TB, it is ineffective against drug-resistant TB infections, which are becoming increasingly prevalent worldwide. Given the severity of the current TB situation and the propensity for drug resistance, the development of new drugs specifically targeting drug-resistant TB is urgently needed.

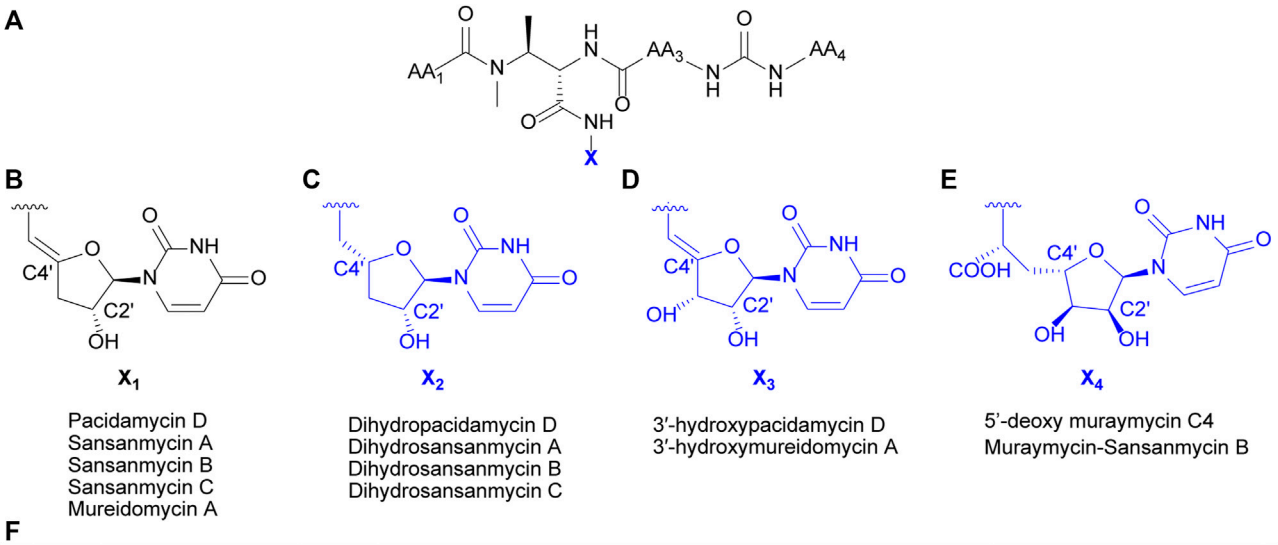
Sansanmycins (Xie et al., 2007), produced by *Streptomyces* sp. SS, belong to the family of uridyl peptide antibiotics (UPAs) including pacidamycins (Karwowski et al., 1989), napsamycins (Chatterjee et al., 1994), and mureidomycins (Inukai et al., 1989). They share a core scaffold of a 3′-deoxyuridine unit linked to a pseudo-tetra/pentapeptidyl backbone via an unusual enamide linkage. UPAs possess noteworthy antimicrobial activity against *Mycobacterium tuberculosis* (including drug-resistant *Mycobacterium tuberculosis*) (Shi et al., 2016) and *Pseudomonas aeruginosa* by targeting translocase I (MraY) to block bacterial cell wall assembly (Winn et al., 2010). As this unique mode of action has not been clinically targeted, UPA has emerged as a lead compound in the quest for new anti-TB drugs.

Extensive research has been conducted on UPAs' biosynthesis logic and structural modification. UPAs are synthesized by a series of highly dissociated non-ribosomal peptide synthetases (NRPSs) (Zhang et al., 2010). Their assembly begins with the central building block 2,3-diaminobutyric acid (DABA), then the elongation of the peptide framework includes the attachment of the *N*-terminal amino acid (AA_1) to the *β*-amino group of DABA, and the *C*-terminal ureidodipeptide (AA_3 and AA_4) linked to the *α*-amino group of DABA (Zhang et al., 2010; Zhang et al., 2011). The unusual ureidodipeptide assembly is achieved under the catalysis of a unique condensation enzyme PacN, resulting in the formation of a ureido-bond between AA_3 and AA_4 (Zhang et al., 2010; Ióca et al., 2021). The synthesis of 4′,5′-enamide-3′-deoxyuridine of UPA has been elucidated *in vitro*, which starts from uridine, and a flavin-dependent dehydrogenase Pac11 (i.e., PacK) catalyzes the oxidation of uridine to uridine-5′-aldehyde (Ragab et al., 2011). The aminotransferase Pac5 (i.e., PacE) catalyzes the transamination of the 5′-aldehyde group, and the dehydration reaction mediated by Pac13 (i.e., PacM) finally generates 4′,5′-enamide-3′-deoxyuridine (Ragab et al., 2011; Michailidou et al., 2017) (Supplementary Figure S1). Then under the catalysis of PacI, it is attached to pseudo-tetra backbone through an enamide bond, consequently producing UPAs (Zhang et al., 2011).

Previous structural modifications of UPAs mainly focused on substituting amino acid residues in peptide chains, such as obtaining *C*-terminally modified pacidamycin derivatives (Gruschow et al., 2009) and sansanmycin derivatives (Zhang et al., 2016) by precursor-directed biosynthesis. A series of new sansanmycin analogues with modification at *N*-terminal amino acid were obtained using mutational biosynthesis (Shi et al., 2016). Nevertheless, few analogue compounds showed significantly increase of the antibacterial activity. The 5′-aminouridine moiety in UPAs is believed to be essential for competing active site binding with the natural substrate UDP-*N*-acetylmuramoyl (UDP-MurNAc) pentapeptide of MraY (Walsh and Zhang, 2011). Recently, the MraY-inhibitor complex structure data show a common feature of the uridine binding pocket of naturally occurring nucleoside inhibitors including UPAs, and the residues that form the uridine binding pocket are likely involved in binding the natural substrate of MraY (Mashalidis et al., 2019). However, as an important part of pharmacophore, there have been no reported structural modifications of 5′-aminouridine through biosynthetic methods.

Currently, reported analogues with modified 5′-aminouridine are almost entirely obtained through chemical synthesis (Figure 1). Dihydropacidamycin D, the first analogue compound with modified 5′-aminouridine was obtained through the hydrogenation of C-4′ exocyclic olefin of pacidamycin D, and it showed comparable activity against *P. aeruginosa* to pacidamycin D (Boojamra et al., 2001) (Figures 1A–C,F), indicating that the hydrogenation of enamide linkage would not affect the activity of UPAs. In recent years, certain dihydrosansanmycin analogues with modified peptide chains have been reported to exhibit enhanced activity against *M. tuberculosis* (Tran et al., 2017; Tran et al., 2021) (Figures 1A–C, F). Notably, the stereochemistry at C-4′ of the ribose moiety in both dihydrosansanmycin and dihydropacidamycin has been suggested to be essential for antibacterial activity, with *S*-stereochemistry in analogues usually leading to loss or reduction of activity (Boojamra et al., 2001; Tran et al., 2017). The antibacterial activity of 3′-hydroxypacidamycin D against *P. aeruginosa* was comparable to that of pacidamycin D (Okamoto et al., 2012), while the MraY inhibition activity of 3′-hydroxymureidomycin A (Mashalidis et al., 2019) was also comparable to that of mureidomycin A (Inukai et al., 1993) (Figures 1A, D, F). These findings suggest that the introduction of a hydroxyl group at the 3′-position of the 5′-aminouridine moiety is well-tolerated for both MraY inhibition and antibacterial activity (Okamoto et al., 2012). In 2020, Niro et al. (Niro et al., 2020) reported a new hybrid structure, which consists of the peptide chain of sansanmycin B and uridine core moiety of 5′-deoxy muraymycin C4 (Figures 1A, E–F). Similar to UPA, the 5′-deoxy muraymycin C4 is also a uridine-derived nucleoside antibiotic, comprising a (6′*S*)-5′-deoxy-glycyluridine moiety, and demonstrates inhibitory activity against MraY (Koppermann et al., 2018). The muraymycin-sansanmycin B hybrid structure showed inhibition activity against MraY, but displayed no significant antibacterial activity against *P. aeruginosa* (Niro et al., 2020). This suggests that the hybrid structure resulting from the incorporation of uridine building blocks from other nucleoside antibiotics into UPAs may still exhibit inhibitory activity against MraY. Chemical synthesis of UPAs is challenging and low yielding,

10.3389/fbioe.2023.1278601



F

Isolated natural products	AA_1	AA_3	AA_4	X	**Chemical synthesized compounds**	AA_1	AA_3	AA_4	X
Pacidamycin D	Ala	Ala	Trp	X_1	**Dihydropacidamycin D**	Ala	Ala	Trp	X_2
MIC (*Pae* PAM1020) =64 μg/mL					MIC (*Pae* PAM1020) =64 μg/mL				
MIC (*Pae* PAO1) =64 μg/mL									
Sansanmycin A	*m*-Tyr	Met	Trp	X_1	**Dihydrosansanmycin A**	*m*-Tyr	Met	Trp	X_2
MIC (*Mtb* H37Rv) =16 μg/mL					MIC_{50} (*Mtb* H37Rv) =0.43 μg/mL (0.50 ± 0.05 μM)				
Sansanmycin B	*m*-Tyr	Met	Trp	X_1	**Dihydrosansanmycin B**	*m*-Tyr	Met	Trp	X_2
MIC (*Mtb* H37Rv) =8 μg/mL					MIC_{50} (*Mtb* H37Rv) =0.25 μg/mL (0.30 ± 0.07 μM)				
Sansanmycin C	*m*-Tyr	Met^{SO}	Trp	X_1	**Dihydrosansanmycin C**	*m*-Tyr	Met^{SO}	Trp	X_2
MIC (*Mtb* H37Rv) >32 μg/mL					MIC_{50} (*Mtb* H37Rv) =16 μg/mL (18.5 ± 0.70 μM)				
					3′-hydroxypacidamycin D	Ala	Ala	Trp	X_3
					MIC (*Pae* PAO1) =32 μg/mL				
Mureidomycin A	*m*-Tyr	Met	*m*-Tyr	X_1	**3′-hydroxymureidomycin A**	*m*-Tyr	Met	*m*-Tyr	X_3
IC_{50} (MraY *B. subtilis*)= 0.05 μg/mL					IC_{50} (MraY *A. aeolicus*)=0.045 μg/mL (0.052 μM)				
5’-deoxy muraymycin C4				X_4	**Muraymycin-Sansanmycin B**	*m*-Tyr	Met	Trp	X_4
IC_{50} (MraY *S. aureus*)= 0.08 μg/mL (0.095 ± 0.019 μM)					IC_{50} (MraY *S. aureus*)= 0.61 μg/mL (0.71 ± 0.17 μM)				

FIGURE 1
Chemical structures of natural UPAs (black) and synthetic analogues with modified 5′-aminouridine moiety (blue) and their corresponding inhibitory activity. **(A)** The pseudo-tetrapeptide backbone of UPAs. **(B)** The chemical structures of 5′-aminouridine of natural UPAs. **(C–E)** The chemical structures of modified 5′-aminouridine of synthetic analogues. **(F)** Detailed information on structure and inhibitory activities of natural UPAs and synthesized derivatives. *Pae*, *Pseudomonas aeruginosa*; *Mtb*, *Mycobacterium tuberculosis*; *B. subtilis*, *Bacillus subtilis*; *S. aureus*, *Staphylococcus aureus*; *A. aeolicus*, *Aquifex aeolicus*.

but it provides some insights into the structure-activity relationship of UPAs. What's more, it indicates that making appropriate modifications to 5′-aminouridine can be further explored for new UPA derivatives to improve their activity and/or druggability.

Here, we aimed to expand the structural diversity of sansanmycin and explore the structure-activity relationship of 5′-aminouridine moiety by mutational biosynthesis. Within sansanmycin biosynthetic gene cluster, SsaM, SsaK and SsaE share high homology with PacM, PacK and PacE, respectively, which have been previously confirmed to be responsible for the biosynthesis of the 5′-aminouridine of pacidamycin *in vitro*. In this study, through *in vivo* assays, we demonstrated that SsaM and SsaK are responsible for the biosynthesis of 5′-aminouridine in sansanmycin biosynthetic pathway, and confirmed SsaM acts as the dehydratase by the detection of a new hydrated derivative. On this basis, two 5′-aminouridine analogues were fed to the knockout strains of *ssaK and ssaM*, and *ssaK* knockout strain showed higher yield of derivatives. Twenty-two new derivatives were identified by the molecular networking of LC-MS/MS data. Among them, four monomers were purified and one of them were further confirmed

10.3389/fbioe.2023.1278601

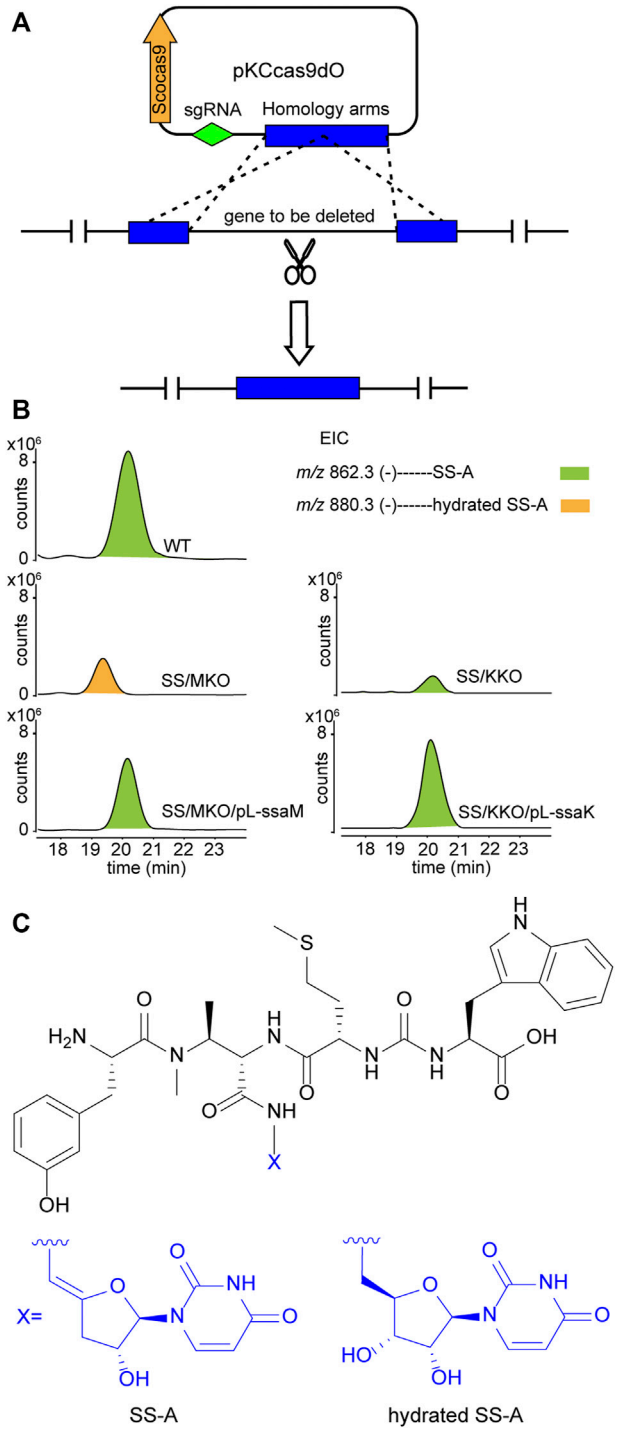


FIGURE 2
Effects of blocking 5′-aminouridine biosynthetic gene *ssaM* and *ssaK* on sansanmycin production. **(A)** Construction of mutant strains with deleted 5′-aminouridine biosynthetic genes. **(B)** LC-ESI-MS analyses of sansanmycin A (SS-A) and its derivatives in different strains. **(C)** Structures of SS-A and hydrated SS-A.

by NMR. SS-KK-2 exhibited improved antibacterial activity against *E. coli ΔtolC*. SS-KK-3 showed the same activity against *M. tuberculosis* including clinically isolated, multidrug-resistant strains with improved structural stability.

2 Results

2.1 Blocking the biosynthesis of 4′,5′-enamide-3′-deoxyuridine

Upon scrutiny of the sansanmycin biosynthetic gene cluster, SsaM, SsaK, and SsaE were found to exhibit over 80% amino acid sequence identity with PacM, PacK, and PacE (Supplementary Figures S2–4), respectively, indicating their probable role in the biosynthesis of the 5′-aminouridine moiety in sansanmycin. To investigate the function of SsaM, SsaK, and SsaE *in vivo*, the corresponding encoding genes *ssaM*, *ssaK*, and *ssaE* were in-frame deleted from *Streptomyces* sp. SS by CRISPR/Cas9 using plasmid pKCcas9dO (He et al., 2015), which contains codon-optimized *scocas9* encoding *Streptococcus* pyogenes Cas9 protein. Deletion plasmids pKC-M, pKC-K and pKC-E were constructed by replacing the sgRNA sequence and homology arm sequences in plasmid pKCcas9dO with the corresponding sequences of the genes to be knocked out, and were further introduced into *Streptomyces* sp. SS by conjugation (Figure 2A). Knock-out mutants SS/MKO, SS/KKO and SS/EKO of *ssaM*, *ssaK*, and *ssaE* were verified by PCR, and the results showed that the correct SS/MKO and SS/KKO were obtained (Supplementary Figures S5, 6). There was no correct SS/EKO obtained even after many efforts, which may be due to the off-target effect. The complete coding regions of *ssaM* and *ssaK* were cloned into the expression plasmid pL646 (Hong et al., 2007) derived from pSET152 (Bierman et al., 19922) under the control of a strong constitutive promoter, *ermE*p*, respectively. Then resulting plasmids pL-ssaM and pL-ssaK were introduced into SS/MKO and SS/KKO by conjugation, respectively, and the complementary strains SS/MKO/pL-ssaM and SS/KKO/pL-ssaK were obtained to exclude the possibility of polar effects in SS/MKO and SS/KKO.

Secondary metabolites of SS/MKO and SS/KKO, as well as the complementary strains SS/MKO/pL-ssaM and SS/KKO/pL-ssaK were analyzed by LC-MS/MS, with wild-type strain as control. In SS/MKO, sansanmycin A (SS-A), the major component compound of wild-type strain, was eliminated and instead a new hydrated sansanmycin A (hydrated SS-A) was detected (Figures 2B, C, Supplementary Figure S7). The production of hydrated SS-A confirmed that SsaM acts as the dehydratase responsible for biosynthesizing the 5′-aminouridine of sansanmycins *in vivo*. Mutant SS/KKO still produced trace amount of SS-A (Figure 2B), potentially due to its function being partially complemented by other enzymes in the production strain. Production of sansanmycins could be restored in both complementary strains (Figure 2B). All of above results demonstrated that SsaM and SsaK were responsible for the biosynthesis of 5′-aminouridine in sansanmycin biosynthetic pathway *in vivo*.

2.2 Production of new sansanmycin analogues by mutational biosynthesis

To explore the possibility of mutational biosynthesis on modification of 5′-aminouridine moiety, we intended to feed uridine analogues to the knock-out mutants. Besides the natural amide linkage between 5′-aminouridine and tetra/pentapeptides, PacI could also catalyze the formation of an oxoester in the *in vitro* assays for its relaxed substrate specificities (Zhang et al., 2011). Therefore, a series of commercially accessible uridine analogues

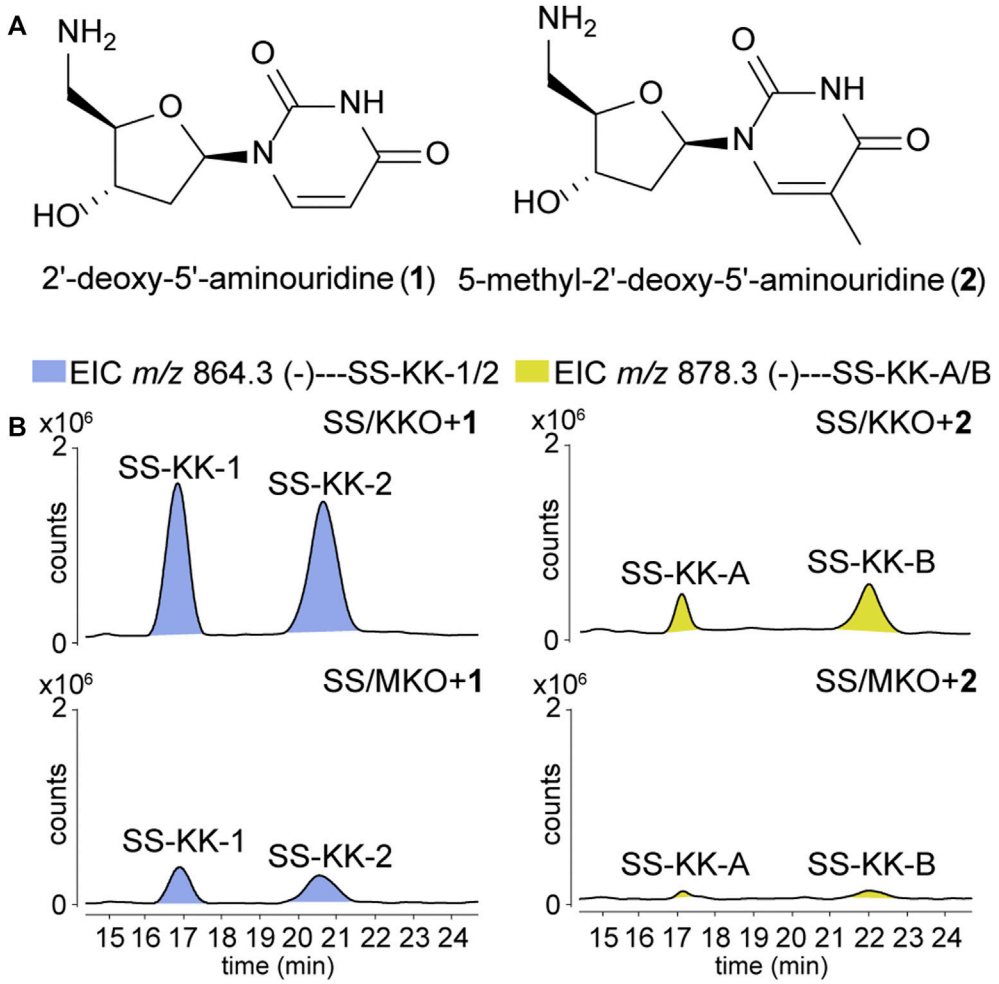


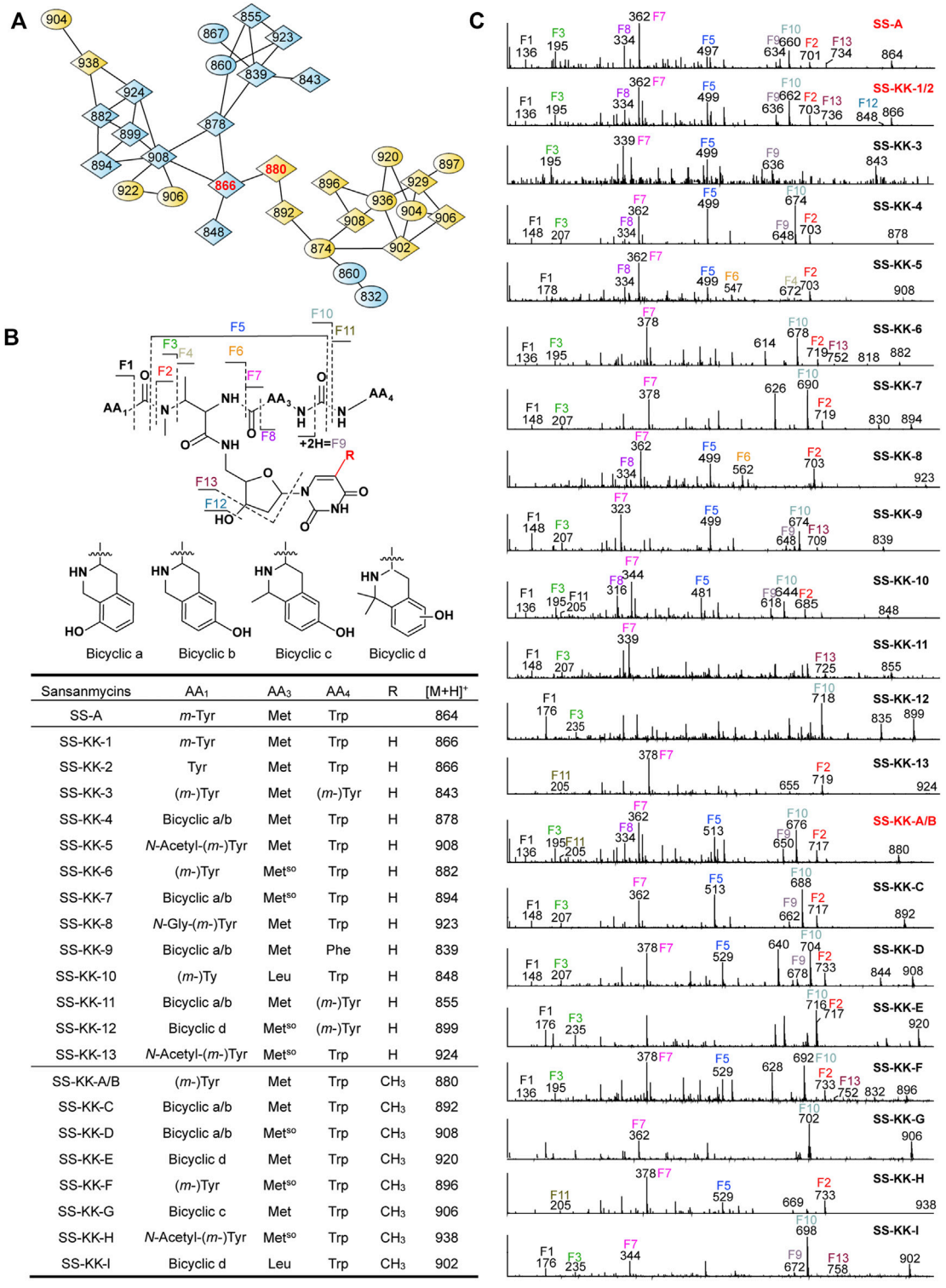
FIGURE 3
Production of new sansanmycin analogues with 5′-aminouridine modification by mutational biosynthesis. **(A)** Structures of 2′-deoxy-5′-aminouridine (**1**) and 5-methyl-2′-deoxy-5′-aminouridine (**2**). **(B)** LC-ESI-MS analyses of fermentations fed with different 5′-aminouridine analogues **1** and **2**.

(Supplementary Figure S8) were firstly fed to SS/KKO. However, no new derivatives were detected in the LC-MS/MS analysis, potentially due to the low efficiency of SsaI (the homologue of PacI) in catalyzing the formation of ester bonds *in vivo*. Then 5′-aminouridine analogues were taken into our consideration and only two commercial compounds were available, 2′-deoxy-5′-aminouridine (1) and 5-methyl-2′-deoxy-5′-aminouridine (**2**) both with the *R*-configuration of the chiral C-4' (Figure 3A). The two 5′-aminouridine analogues 1 and 2 (final concentration of 3 mM) were respectively fed to both SS/MKO and SS/KKO mutants. The fermentation broths were analyzed by extracting ion chromatograms of target compounds derived from SS-A biosynthetic pathway. It was found that two peaks (Figure 3B), with identical MS/MS fragmentations (Figure 4C), appeared in each of the two cultures fed with **1** and **2**. The new derivatives generated from feeding substrate **1** were named SS-KK-1 and SS-KK-2, while those produced from feeding substrate **2** were named SS-KK-A and SS-KK-B.

The molecular weight of SS-KK-1 and SS-KK-2 showed 2 Da larger than that of SS-A, while the molecular weights of **1** is 2 Da larger than that of original 5′-aminouridine moiety, suggesting that 5′-aminouridine of SS-KK-1 and SS-KK-2 were replaced by **1**. This hypothesis was supported by MS/MS spectra (Figure 4C), in which the characteristic fragment ions containing 5′-aminouridine portion, including F5, F6, F12, F13, were 2 Da larger than that of SS-A, and other characteristic fragment ions were consistent with those of SS-A. The F1 ion at *m/z* 136, corresponding to immonium ions of *meta*-tyrosine or tyrosine (Tyr), together with the F2 ion at *m/z* 703, corresponding to the loss of an *m*-Tyr or Tyr residue from *N*-terminus, inferred that SS-KK-1 and SS-KK-2 hold an *m*-Tyr or Tyr at AA_1 position as found in SS-A (Xie et al., 2007) and SS MX-2 (Shi et al., 2016) respectively.

The molecular weight of **2** was 14 Da larger than that of **1**. Similarly, the quasi molecular of SS-KK-A/B were also 14 Da larger than that of SS-KK-1/2, indicating that the fed precursor **2** was incorporated into the structures of SS-KK-A/B. This inference was supported by the corresponding tandem mass spectrum fragments, in which the characteristic fragment ions containing 5′-aminouridine moiety, including F2, F5, F9, F10, were found to be 14 Da larger than that of SS-KK-1/2, while other characteristic fragment ions were consistent with those of SS-KK-1/2 (Figure 4C).

10.3389/fbioe.2023.1278601



Sansanmycins	AA_1	AA_3	AA_4	R	$[M+H]^+$
SS-A	*m*-Tyr	Met	Trp		864
SS-KK-1	*m*-Tyr	Met	Trp	H	866
SS-KK-2	Tyr	Met	Trp	H	866
SS-KK-3	(*m*-)Tyr	Met	(*m*-)Tyr	H	843
SS-KK-4	Bicyclic a/b	Met	Trp	H	878
SS-KK-5	*N*-Acetyl-(*m*-)Tyr	Met	Trp	H	908
SS-KK-6	(*m*-)Tyr	Met^{so}	Trp	H	882
SS-KK-7	Bicyclic a/b	Met^{so}	Trp	H	894
SS-KK-8	*N*-Gly-(*m*-)Tyr	Met	Trp	H	923
SS-KK-9	Bicyclic a/b	Met	Phe	H	839
SS-KK-10	(*m*-)Ty	Leu	Trp	H	848
SS-KK-11	Bicyclic a/b	Met	(*m*-)Tyr	H	855
SS-KK-12	Bicyclic d	Met^{so}	(*m*-)Tyr	H	899
SS-KK-13	*N*-Acetyl-(*m*-)Tyr	Met^{so}	Trp	H	924
SS-KK-A/B	(*m*-)Tyr	Met	Trp	CH_3	880
SS-KK-C	Bicyclic a/b	Met	Trp	CH_3	892
SS-KK-D	Bicyclic a/b	Met^{so}	Trp	CH_3	908
SS-KK-E	Bicyclic d	Met	Trp	CH_3	920
SS-KK-F	(*m*-)Tyr	Met^{so}	Trp	CH_3	896
SS-KK-G	Bicyclic c	Met	Trp	CH_3	906
SS-KK-H	*N*-Acetyl-(*m*-)Tyr	Met^{so}	Trp	CH_3	938
SS-KK-I	Bicyclic d	Leu	Trp	CH_3	902

FIGURE 4
Molecular networking directed discovery of new sansanmycin analogues. **(A)** Molecular network consisting of all parent ions detected by LC–MS/MS in the extract crude of SS/KKO when fed with **1** or **2**. A constellation for potential sansanmycins was picked out using SS-KK-1/2 and SS-KK-A/B (highlight in red) as a probe from the whole molecular network (Supplementary Figure S9) and amplified for displaying. The derivatives produced by feeding SS/KKO with **1** are labeled in blue, while the derivatives produced by feeding SS/KKO with **2** are labeled in yellow. All identified sansanmycin analogues are represented as nodes with diamond symbols. **(B)** The tentative structures of SS-KK-1 to SS-KK-13 and SS-KK-A to SS-KK-I. **(C)** MS/MS analysis for corresponding compounds. The diagnostic fragments are indicated.

Therefore, it is speculated that SS-KK-A/B are similar to SS-KK-1/2 except for the replacement of 5′-aminouridine with **2**.

The production of SS-KK-1/2 and SS-KK-A/B indicates that feeding 5′-aminouridine analogues to knockout strains SS/MKO and SS/KKO can afford new derivatives. Among them, SS/KKO showed a higher yield of new derivatives (Figure 3), indicating it is a better host for expanding the structural diversity of 5′-aminouridine by mutational biosynthesis. The above results suggest the potential of using mutational biosynthesis for production of 5′-aminouridine-modified sansanmycin derivatives.

2.3 New sansanmycin analogues discovered by molecular networking

The method of tandem mass-based molecular networking was previously established for rapid identification of sansanmycin analogues (Jiang et al., 2018). Here, we applied this method for analyzing sansanmycin derivatives produced by the above two cultures fed with precursors **1** and **2**, respectively. Through this analysis, secondary metabolites showing similar fragmentation patterns in tandem mass spectra were grouped into multiple sub-networks (Supplementary Figure S9). Using SS-KK-1/2 and SS-KK-A/B as probes, a sub-network consisting of over 30 nodes was discovered in the molecular networks of extracts from the two cultures (Figure 4A; Supplementary Figure S9). The visualization of metabolites from the two cultures in different colors allowed us to easily distinguish the source of the molecules (Figure 4A; Supplementary Figure S9). Further manual analysis of the MS/MS data corresponding to these molecules reveals that the uridine moiety of the molecules from the culture fed with precursor **1** have altered from the original 5′-aminouridine moiety to **1**. Similarly, the molecules from the culture fed with precursor **2** are found to bear the unit of **2** in place of the original 5′-aminouridine moiety in sansanmycin (detailed analyses are shown in Supplementary Results).

By closer inspection of their MS/MS spectra, we found that, in addition to the alterations in the 5′-aminouridine, the amino acids comprising the pseudopeptide chain also exhibit diversity. Due to the susceptibility of these molecules to cleavage at peptide bonds, analysis of their MS/MS data can provide a clear inference of the types and positions of amino acids in the pseudopeptide chain (Figure 4 and Supplementary Results). For instance, the fragment losses for M-F10 indicate the molecular weight of the *C*-terminal amino acid residue (AA_4). A mass loss of 204 Da for M-F10 (SS-KK-1/2, SS-KK-4, SS-KK-6, SS-KK-7, SS-KK-10, SS-KK-A/B to SS-KK-G and SS-KK-I) suggests the presence of Trp at AA_4. Similarly, a shortage of 165 Da indicates that Phe is at AA_4 (SS-KK-9), while 181 Da (SS-KK-12) points to *m*-Tyr/Tyr ((m-)Tyr) in the place of AA_4. Therefore, by analyzing the MS/MS fragments and comparing them with those of known sansanmycins (Jiang et al., 2018), the structures of additional eighteen new sansanmycin analogues (SS-KK-3 to SS-KK-13 and SS-KK-C to SS-KK-I) were tentatively deduced (Figure 4 and Supplementary Results).

Due to the limited production of many sansanmycin analogues, their structures can only be deduced through MS/MS data. The molecular networking approach, based on MS/MS data, provides valuable clues towards discovering more 5′-aminouridine-modified sansanmycin derivatives.

2.4 Isolation and structural determination of new sansanmycin analogues

To further characterize the structures deduced by MS/MS analysis of sansanmycin derivatives and determine their antimicrobial activity, SS/KKO fermentations fed with **1** and **2** were scaled up. The main components were isolated as previously described (Xie et al., 2007; Xie et al., 2014), and then further purified by semi-preparative HPLC to give SS-KK-1 (2.1 mg), SS-KK-2 (0.8 mg), SS-KK-3 (0.9 mg) and SS-KK-C (0.4 mg). The purity of these compound monomers was determined by HPLC (Supplementary Figures S10–13). Subsequently, the 1D and 2D NMR spectra of SS-KK-1 were collected in DMSO-d_6 to aid in structural elucidation. However, due to the limited quantity of SS-KK-2, SS-KK-3 and SS-KK-C, their structures could not be further confirmed by NMR methods.

The NMR spectra in DMSO-d_6 for SS-KK-1 were very similar to those of SS-A (Xie et al., 2007), except the signals from sugar moiety. Further analysis of these different signals disclosed that the fed 2-deoxy-5′-aminouridine in SS-KK-1 (δH 6.05 for H-sugar-1, 2.02 for H-sugar-2, 3.99 for H-sugar-3, 3.66 for H-sugar-4, 3.21 for H-sugar-5) was in place of 5′-aminouridine in SS-A, which was further confirmed by the sequential ^{1}H–1H COSY correlations of H-sugar-1/H-sugar-2/H-sugar-3/H-sugar-4/H-sugar-5. Interpretation of the 1D and 2D NMR spectra (Supplementary Figures S14–19 and Supplementary Table S1) confirmed the proposed structure which holds an *m*-Tyr at the *N*-terminus.

Based on above MS/MS analysis, SS-KK-1 and SS-KK-2 produced the same fragmentations and was expected to hold an *m*-Tyr or Tyr at *N*-terminus, respectively. As *m*-Tyr was confirmed in the SS-KK-1, Tyr was inferred at the *N*-terminus of SS-KK-2.

2.5 Antibacterial activity and stability of new sansanmycin analogues

The antibacterial activity of sansanmycin analogues SS-KK-1, SS-KK-2, SS-KK-3, and SS-KK-C was tested against a range of bacteria (Table 1), including both Gram-negative and Gram-positive bacteria, as well as standard strain of *M. tuberculosis* H37Rv, and clinically isolated strains of *M. tuberculosis* S17, M6600, M3551 and M9483. Among them, M6600, M3551, and M9483 are multidrug-resistant strains, resistant to INH at 0.1 μg/mL and RIF at 1 μg/mL. Unexpectedly, both SS-KK-1 and SS-KK-C lost their antibacterial activity against *P. aeruginosa*, *M. tuberculosis*, and even *E. coli ΔtolC*. SS-KK-3 showed the same antibacterial activity as SS-A against *E. coli ΔtolC* and *M. tuberculosis* strains including H37Rv and clinically isolated, drug-sensitive and multidrug-resistant strains. Interestingly, although SS-KK-2 lost its inhibitory activity against *P. aeruginosa*, it showed the best inhibition activity against *E. coli ΔtolC* compared with SS-KK-3 and SS-A. The bioactivity of our newly obtained SS-KK-2 and SS-KK-3, which do not contain the 2′-OH group, suggests that it is not essential for their activity.

Poor stability is one of the factors hindering the development of sansanmycin analogues into drug candidates. Thus, we conducted stability experiments followed the method reported previously (Shi et al., 2016) on the two active sansanmycin analogues, SS-KK-2 and SS-KK-3 (Figure 5). These experiments were carried out at room temperature with the parent compound SS-A as a control. When subjected to these conditions, the residual quantities of SS-A were

TABLE 1 Activities of sansanmycin analogues.

Compounds	MIC (μg/mL)							
	E. coli ΔtolC	*P. aeruginosa* 11	*M. Phlei*	*M. tuberculosis*				
				H37Rv	S17	M6600	M3551	M9483
SS-KK-1	>32	>32	>32	>64	>64	>64	>64	>64
SS-KK-2	1	>32	>32	32	32	64	64	32
SS-KK-3	4	32	64	16	16	16	16	16
SS-KK-C	>32	>32	>32	>64	>64	>64	>64	>64
SS-A	4	16	64	16	16	16	16	16
streptomycin	2	8	16					
INH				0.03	0.03	R	R	R
RIF				0.125	0.25	R	R	R
LZD				0.25	0.25	S	S	S

M. tuberculosis H37Rv, standard strain; S17, M6600, M3551, and M9483 are clinical isolates of *M. tuberculosis*. INH, isoniazid; RIF, rifampicin; LZD, linezolid. M6600, M3551, and M9483, multidrug-resistant strains, resistant to INH, at 0.1 μg/mL and RIF, at 1.0 μg/mL, sensitive to LZD, at 2.0 μg/mL.

depleted below 70% after 9 days of incubation, whereas SS-KK-2 showed a slightly higher level of stability than SS-A. SS-KK-3 exhibited minimal degradation and substantially greater stability than both SS-A and SS-KK-2. As SS-KK-3 displayed antimicrobial activity comparable to that of SS-A, it is a more promising lead compound for anti-TB drug development.

3 Discussion

To expand the structural diversity of 5′-aminouridine moiety of sansanmycin through biosynthetic methods, the function of *ssaM* and *ssaK* as biosynthetic genes for 5′-aminouridine moiety of sansanmycins was firstly validated *in vivo*. The production of hydrated SS-A in the *ssaM*-deletion mutant confirmed the role of *ssaM* as a dehydratase. Unfortunately, due to the low yield of hydrated SS-A, the pure compound could not be obtained for activity assay. Preliminary attempt of feeding easily accessible uridine analogues containing 5′-OH *in vivo*, which can form ester bonds with tetra/pentapeptides catalyzed by PacI *in vitro* (Zhang et al., 2011), did not result in the expected derivatives by mutational biosynthesis. This could be due to either the varying efficiency of SsaI, which is the homologue of PacI, between *in vivo* and *in vitro* conditions, or to the larger substrate tolerance of PacI compared to SsaI. As expected, 5′-aminouridine analogues were shown to be readily incorporated into sansanmycin biosynthetic pathway in SS/KKO and SS/MKO mutant, with higher effectiveness in SS/KKO mutant. To the best of our knowledge, it is the first report to modify 5′-aminouridine of UPAs and obtained a range of 5′-aminouridine-modified derivatives through mutational biosynthesis, which confirms that SS/KKO obtained in this study is a good host for enhancing the structural diversity of 5′-aminouridine of UPAs. In contrast to tedious and challenging chemical synthesis, employing SS/KKO to increase the structural diversity of 5′-aminouridine via mutasynthesis is more efficient and effective.

he 5′-aminouridine moiety of UPAs, as present in other classes of naturally occurring nucleoside inhibitors such as muraymycin, caprazamycin and capuramycin, resembles the uridine moiety of UDP-MurNAc pentapeptide, the natural substrate of MraY. Recent structural biological data of MraY-inhibitor complex also showed a common feature of the uridine binding pocket of naturally occurring nucleoside inhibitors including caprazamycin, capuramycin, and mureidomycin (Mashalidis et al., 2019). The chemically synthesized dihydropacidamycin D resulting from reduction of the chemically unstable 4′-exo double bond of ribosyl moiety showed similar antibacterial activity to that of pacidamycin D (Boojamra et al., 2001). Lemoine et al. further synthesized a series of dihydropacidamycin D derivatives by chemically modifying the uracil ring of 5′-aminouridine, however, none of these derivatives were capable of inhibiting MraY, highlighting the crucial role of uracil for UPAs' activity (Lemoine et al., 2002). These early discoveries were also consistent with recent structural biological data of MraY-inhibitor complex, which revealed that the uridine pocket of MraY has a certain level of spatial tolerance. Moreover, compared to the ribosyl moiety, the modifications in the uracil moiety are more likely to interfere with the interactions between the residues forming the uridine pocket of MraY, leading to the loss of inhibitory effect (Mashalidis et al., 2019). In this study, SS-KK-C, an analogue obtained from the culture fed precursor **2** with a methyl modification on uracil, also lost its antibacterial activity against *P. aeruginosa*, *M. tuberculosis*, and even *E. coli ΔtolC*.

Previous chemically synthesized 5′-aminouridine-modified derivatives have always contained a 2′-OH on the nucleoside ribosyl group (Figure 1) (Inukai et al., 1993; Boojamra et al., 2001; Okamoto et al., 2012; Tran et al., 2017; Mashalidis et al., 2019; Niro et al., 2020; Tran et al., 2021). By generating a series of 5′-aminouridine analogues that lacks 2′-OH for the first time, this study demonstrated that the presence of 2′-OH is dispensable for the bioactivity of UPAs. In previous reports, the *R*-stereochemistry at the C-4′ position of dihydropacidamycin has been considered as playing a significant role in maintaining antibacterial activity (Boojamra et al., 2001; Tran et al., 2017). However, we noticed that SS-KK-1 with the *R*-stereochemistry at the C-4′ position of the ribose attached to the same tetrapeptide backbone as sansanmycin A completely losses

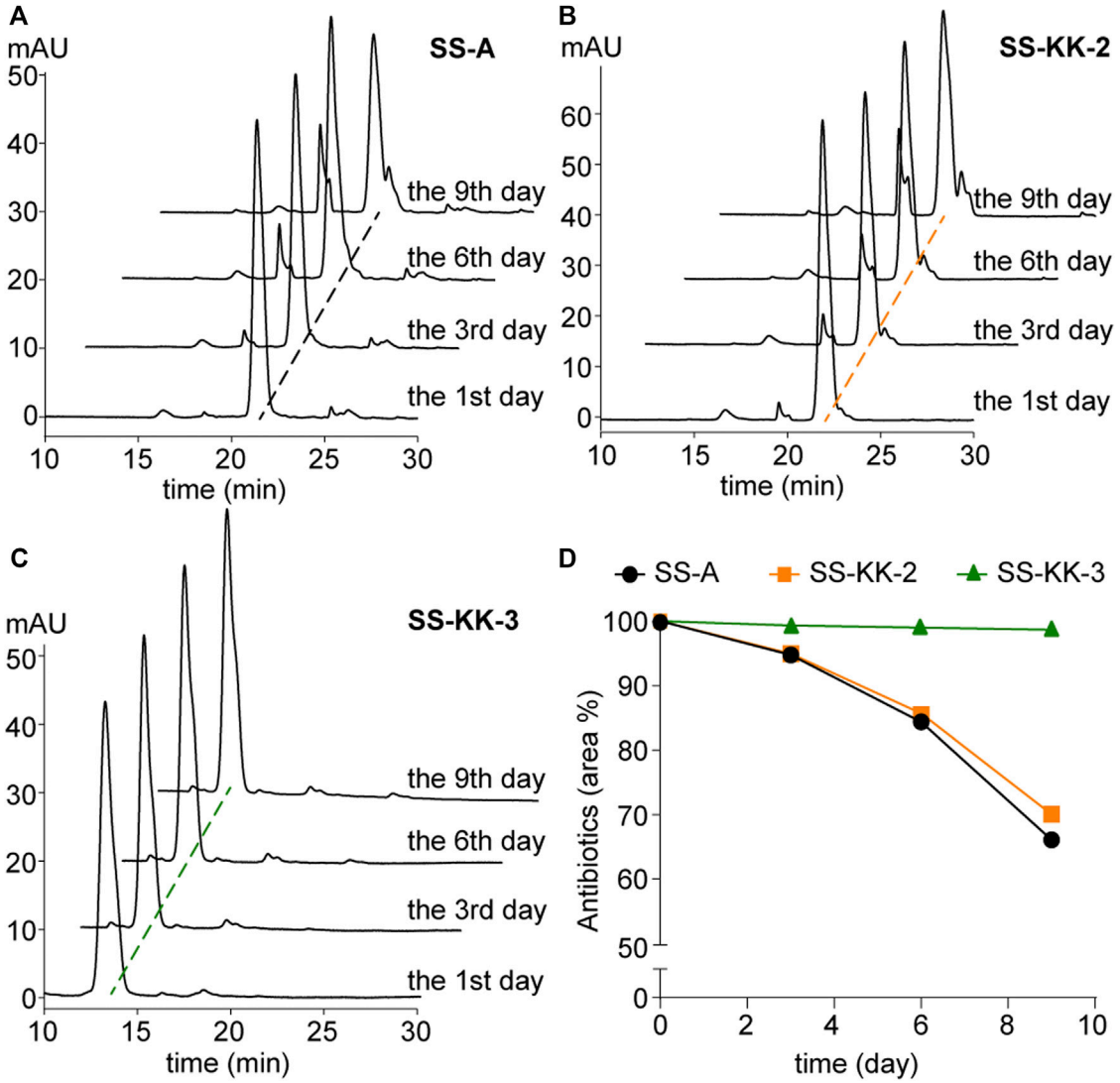


FIGURE 5
Stability of SS-A, SS-KK-2 and SS-KK-3. **(A–C)** HPLC analysis of SS-A, SS-KK-2 and SS-KK-3 at indicated days in KH_2PO_4 buffer (pH 6.0) at room temperature. **(D)** The changes of the level of SS-A, SS-KK-2 and SS-KK-3 over time. All samples were analyzed by HPLC and quantified according to the areas of peaks. Each sample underwent three parallel repeats. SS-A, black line; SS-KK-2, orange line; SS-KK-3, green line.

antibacterial activity. Further inspection of the structure of uridine unit disclosed that although the new dihydro UPA derivatives obtained in this study share the *R*-stereochemistry at the C-4′ position of the ribose with the chemically synthesized dihydro UPAs reported previously (Boojamra et al., 2001; Tran et al., 2017), their relative configurations differ. The chemically synthesized dihydro UPAs bear an *α*-L-ribose configuration, while the dihydro UPA derivatives in this study show a *β*-D-ribose configuration. It appears that the role of stereochemistry of C-4′ position required further evaluation using more UPA derivatives with structurally diverse uridine.

As uridine moiety is an important part of pharmacophore, further research should be conducted on the structure-activity relationship of the nucleoside part of UPAs. Here we show that mutational biosynthesis offers an efficient and accessible method to expand the structural diversity of 5′-aminouridine moiety of sansanmycin. In the future, by synthesizing structurally varied 5′-aminouridine analogues for mutasynthesis in SS/KKO, the diversity of UPA structures may be further increased, which would provide more valuable insights into the structure-activity relationship of the nucleoside part in UPAs. Additionally, utilizing the biosynthetic method to obtain UPAs with different 5′-aminouridine modifications may be combined with genetically manipulated biosynthesis on the pseudopeptide parts *in vivo*, which will significantly enhance the diversity of UPAs and hold promise for obtaining UPAs with improved activity and/or druggability.

4 Conclusion

This study confirmed the functions of SsaM and SsaK in the biosynthesis of 5′-aminouridine of sansanmycin through gene deletion and complementation experiments. Based on this, mutational biosynthesis enriched the diversity of nucleoside

TABLE 2 Strains and plasmids used in this study.

Strains/plasmids	Relevant characteristics	References
Strains		
Streptomyces sp. SS	Wild-type strain (sansanmycin-producing strain, CPCC200442 from China Pharmaceutical Culture Collection)	Xie et al. (2007)
SS/MKO	*Streptomyces* sp. SS with the in-frame deletion of *ssaM*	This study
SS/KKO	*Streptomyces* sp. SS with the in-frame deletion of *ssaK*	This study
SS/MKO/pL-ssaM	SS/MKO with the expression vector pL-ssaM	This study
SS/KKO/pL-ssaK	SS/KKO with the expression vector pL-ssaK	This study
Escherichia coli DH5α	General cloning host	Sambrook (2001)
Escherichia coli ET12567/pUZ8002	Strain used for *E. coli*/*Streptomyces* conjugation	Paget et al. (1999)
Escherichia coli ΔtolC	Strain for testing antimicrobial activity	Tang et al. (2013)
Pseudomonas aeruginosa 11	Strain for testing antimicrobial activity	Xie et al. (2008)
Mycobacterium phlei	Strain for testing antimicrobial activity	
Mycobacterium tuberculosis	Strain for testing antimicrobial activity	
H37Rv	Standard strain, susceptible to isoniazid and rifampicin	
S17	Clinically isolated strain, susceptible to isoniazid and rifampicin	
M6600	Clinically isolated drug-resistant strain, resistant to isoniazid and rifampicin	
M3551	Clinically isolated drug-resistant strain, resistant to isoniazid and rifampicin	
M9483	Clinically isolated drug-resistant strain, resistant to isoniazid and rifampicin	
Plasmids		
pKCcas9dO	Vector used for the construction of deletion plasmids	He et al. (2015)
pCB003	Vector used as the template for amplifying sgRNA	Jiang et al. (2015)
pKC-M	pKCcas9dO derivative plasmid with the deletion of *ssaM*	This study
pKC-K	pKCcas9dO derivative plasmid with the deletion of *ssaK*	This study
pKC-E	pKCcas9dO derivative plasmid with the deletion of *ssaE*	This study
pSET152	*Streptomyces* integrative vector	Bierman et al., 1992
pL646	pSET152 derivative containing the constitutive promoter *ermE*p*	Hong et al. (2007)
pL-ssaM	pL646 derivative plasmid containing complete coding region of *ssaM*	This study
pL-ssaK	pL646 derivative plasmid containing complete coding region of *ssaK*	This study

moiety of sansanmycins by applying two 5′-aminouridine analogues, especially in *ssaK* knockout strain. Combining manual analysis and molecular networking of MS/MS data, twenty-two new 5′-aminouridine-modified sansanmycin derivatives were identified, among which four were purified and their antibacterial activities were determined. Of note, SS-KK-3 exhibited anti-mycobacterial activity comparable to that of sansanmycin A, even to clinically isolated multidrug-resistant strains of *M. tuberculosis*. Furthermore, SS-KK-3 exhibited substantial improvements in stability compared to sansanmycin A. The enhanced druggability of this compound may spur further research into the development of new anti-TB drug candidates acting against clinically unexplored target MraY.

5 Materials and methods

5.1 Strains, plasmids and growth conditions

To propagate and transform the sansanmycin-producing strain *Streptomyces* sp. SS, obtained from China Pharmaceutical Culture Collection (CPCC 200442), the strain was grown on solid S5 medium (Wang et al., 2009) at 28°C for sporulation and in liquid fermentation medium (Xie et al., 2007) for sansanmycin production. Mannitol soya flour (MS) agar medium (Kieser et al., 2000) was used for conjugation, and liquid phage medium (Korn et al., 1978) was used for genomic DNA isolation. *Escherichia coli*

DH5α was used for general cloning experiments (Sambrook, 2001), while *E. coli* ET12567/pUZ8002 (Kieser et al., 2000) was used for conjugal transfer following established protocols. *E. coli* strains were incubated in Luria–Bertani medium (LB) (Sambrook, 2001) at 37°C. Strains were incubated with apramycin (Am, 50 μg/mL), kanamycin (Km, 25 μg/mL), and chloramphenicol (Cm, 25 μg/mL) when required. The antimicrobial activity of the compounds was tested against *E. coli ΔtolC* mutant, *P. aeruginosa* 11, *M. phlei*, and *M. tuberculosis*. All strains and plasmids used in this study see Table 2.

5.2 Construction of *Streptomyces* sp. SS ssaM and ssaK mutant

The mutant strains *ΔssaM* or *ΔssaK* were generated by CRISPR/Cas9-mediated deletion. The mutant strain *ΔssaE* was also attempted to be constructed using the same method, but the correct strain was not obtained in the end. The pKCcas9dO (He et al., 2015) which comprises a codon-optimized *cas9* for *Streptomyces* was used for gene deletion. To design the sgRNA targeting sequences, the gene sequences of interest were firstly submitted on the website http://www.rgenome.net/cas-designer/, and the output result is a series of sgRNA targeting sequences containing 20 bp located inside the gene of interest. Then, the off-target rate of sgRNA targeting sequences is assessed using CasOT (Xiao et al., 2014) to find sgRNA sequences that can be used to knock out the gene of interest. The sgRNA module plasmid pCB003 was used for amplifying M-sgRNA, K-sgRNA, and E-sgRNA. M-sgRNA, K-sgRNA, and E-sgRNA fragments were amplified with specific forward primers M-sg/K-sg/E-sg (Supplementary Table S2) which contain sgRNA targeting sequences of *ssaM*/*ssaK*/*ssaE*, and general reverse primers P2 (Supplementary Table S2), respectively. Homologous arm flanking *ssaM*, *ssaK* and *ssaE* were amplified from *Streptomyces* sp. SS genomic DNA by PCR with specific primer pairs MP3/MP4 and MP5/MP6, KP3/KP4 and KP5/KP6, and EP3/EP4 and EP5/EP6 respectively. For the deletion plasmid of *ssaM*, the fragment containing M-sgRNA and homologous arms were assembled by overlapping PCR with primer pair M-sg/MP6, then the fragment was cloned into *Spe*I-*Hin*dIII sites of pKCcas9dO to obtain the pKC-M. For the deletion plasmids of *ssaK* and *ssaE*, the fragment containing K-sgRNA or E-sgRNA and corresponding homologous arms were assembled into *Nde*I-*Bam*HI sites of pL646 (Hong et al., 2007) with NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, United States) to obtain the pL-KO and pL-EO, then the fragment obtained by pL-KO and pL-EO digested with *Spe*I/*Hin*dIII was cloned into *Spe*I-*Hin*dIII sites of pKCcas9dO to obtain the pKC-K and pKC-E. The constructed plasmids were introduced into *Streptomyces* sp. SS by conjugation from *E. coli* ET12567/pUZ8002. The resulting mutant strains were designated SS/MKO and SS/KKO, but SS/EKO was not obtained.

5.3 Construction of complementary strains

Complementation plasmids for mutant strains *ΔssaM* or *ΔssaK* were constructed using pL646 (Hong et al., 2007) as the vector. The *ssaM* and *ssaK* genes were amplified from the genomic DNA of *S.* sp. SS using ssaM-F/ssaM-R and ssaK-F/ssaK-R primers (Supplementary Table S2), respectively. These genes contained *Nde*I and *Bam*HI restriction sites and were cloned into the T-vector prior to sequencing. After correctly sequencing the *ssaM* and *ssaK* genes, they were ligated into the *Nde*I and *Bam*HI restriction sites of the pL646 (Hong et al., 2007) vector resulting in the plasmids pL-ssaM and pL-ssaK. The recombined plasmids were validated by conjugation transfer into their corresponding knockout strains resulting in complemented strains SS/MKO/pL-ssaM and SS/KKO/pL-ssaK.

5.4 Analysis of production of sansanmycins

The methods for fermentation, isolation, and high-pressure liquid chromatography (HPLC) analysis of sansanmycins were conducted in accordance with previously published procedures (Xie et al., 2007; Xie et al., 2014). Cultivation of *Streptomyces* sp. SS and its derivatives proceeded in a 100 mL volume of liquid fermentation medium at 28°C under agitation at 200 rpm. After 2 days of culture, a 5% seed culture was transferred into new fermentation medium and maintained at 28°C for an additional 5 days under agitation at 200 rpm. During the feeding experiment, exogenous substrate (uridine analogue-1 to -12 (Supplementary Figure S8) and 5′-aminouridine analogues **1**/**2** (Figure 3A)) was added to the fermentation medium at a final concentration of 3 mM. The fermented broth was clarified by centrifugation at 4 °C for 20 min at 4,000 rpm, and the supernatant was harvested and subjected to HPLC-MS analysis after extraction with a Sep-Pack C18 cartridge. HPLC-MS analysis was performed on an Agilent 6,410 (Agilent Technologies, United States) equipped with XBridge® C18 column (4.6 × 150 mm, 3.5 μm, Waters, Dublin, Ireland), using a gradient mobile phase that consisted of solvent A (water +0.1% w/v ammonium acetate) and solvent B (100% MeOH). The gradient started from 80:20 solvent A: solvent B and was changed to 40: 60 solvent A: solvent B within 30 min. The flow rate was set at 0.8 mL/min.

5.5 Tandem mass-based molecular networking

For analyzing sansanmycins through LC-MS/MS, the fermentation broth was enriched using macroporous absorbent resin D4006 (Nankai University Fine Chemical Experiment Factory, Tianjin, China) column. After elution with acetone aqueous solutions of various concentrations, sansanmycins were obtained in 20% and 30% solutions respectively. The crude extract consisting of sansanmycins was acquired after removing the acetone through rotary evaporation. LC-ESI(+) MS/MS data was collected on a Waters ACQUITY UPLC H-Class system (Milford, MA, United States) with a Waters Xevo G2-XS QTof detector (Manchester, United Kingdom). ACQUITY UPLC CSHTM C18 column (1.7 μm, 2.1 × 100 mm, Waters) was employed during UPLC analysis carried out at a flow rate of 0.3 mL/min. Solvent A consisted of 0.1% (w/v) formic acid in water, while Solvent B contained 0.1% (w/v) formic acid in acetonitrile. Over 10 min, mobile phase B increased gradually from 10% to 40%. Mass spectral data were acquired in continuum mode using the fast DDA function. The obtained raw data was

converted into.mzML format before undergoing further analysis via the molecular networking tool on the GNPS (global natural product social molecular networking) website (https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp). Various parameters defaulted and precursor ion mass was set to 1.00 Da with an MS/MS fragment ion tolerance of 0.02 Da. Finally, results were downloaded and visualized by employing Cytoscape 3.9.1.

5.6 Purification and characterization of sansanmycin analogues

Separation and purification of sansanmycin analogues followed the method reported previously (Xie et al., 2014) with some modifications.

Fermentation supernatants were subjected to macroporous absorbent resin D4006 column chromatography. Sansanmycins were eluted in 20% and 30% acetone aqueous solutions and then separately passed through DEAE-Sephadex A-25 chromatograph (GE Healthcare, United States), utilizing Tris-HCl (20 mM, pH 8.5) plus varying concentrations of NaCl, while being monitored by HPLC-UV. The effluent consisting of target compounds was collected and subjected to further purification using preparative HPLC, conducted on a SHIMADZU LC-20 HPLC machine equipped with a diode array detector (DAD). A XBridge® prep C18 column (5 μm, 250 × 10 mm, Waters) was used during preparative HPLC. NMR data from the purified samples were acquired employing Bruker spectrometers and DMSO-*d*6 as solvent.

SS-KK-1: white amorphous powder; UV (MeOH) λmax (log ε) 216 (4.77), 265 (4.26) nm $[\alpha]^{20}_{D}$ −21.05 (c0.19, MeOH).

5.7 Antibacterial assay

To determine the inhibition activity of sansanmycin analogues against *M. tuberculosis* H37Rv and clinically isolated strains, a measurement was conducted by the microplate Alamar blue assay, following the method described by (Zhang et al., 2013). The analogues were first dissolved in dimethyl sulfoxide (DMSO) and then serially diluted to final concentrations (μg/mL) ranging from 64 to 0.125 before being tested with 7H9 broth supplemented. INH, RIF and linezolid (LZD) were used as controls for *M. tuberculosis* H37Rv and clinical strains isolated by previous reported method (Hameed et al., 2022). DMSO served as a negative control in these experiments. Minimum inhibitory concentration (MIC) is defined as the lowest concentration of samples that prevents the conversion of the blue color into pink color. Triplicate wells per drug concentration were used.

The microdilution assay recommended by the Clinical and Laboratory Standards Institute (CLSI, 2009, formerly NCCLS) was employed to determine the MICs of other bacterial strains. Strains were cultured in Mueller-Hinton broth (MHB) (Weissauer-Condon et al., 1987) and the final bacterial suspension (in MHB medium) was adjusted to 10^6 cells/ml. Dilutions of test compounds were substituted with MHB medium as described above. Then, triplicate 100 μL transfers of serial dilutions were placed into 96-well plates, after which 100 μL of bacterial suspension was added to each well. MIC is defined as the lowest concentration of samples that restrains the growth of the test organism detected visually following incubation for 10 h at 37°C. Streptomycin served as a positive control.

5.8 Stability determination of sansanmycin analogues

The stability assessment of sansanmycin A and its analogues was conducted using a previously reported method (Shi et al., 2016). Sansanmycins were dissolved in 0.05 M KH_2PO_4 buffer (with pH adjusted to 6.0 using NaOH) and incubated at 28 °C for 9 days with testing conducted every 3 days. Each sample underwent three parallel repeats. All samples were analyzed by HPLC and quantified by peak area analysis.

Data availability statement

The GNPS data of this study are available from the corresponding author upon request.

Author contributions

YLu: Investigation, Data curation, Methodology, Writing–original draft. YLi: Methodology, Writing–original draft. JF: Methodology, Investigation, Writing–original draft. XL: Methodology, Writing–original draft, Funding acquisition. HS: Methodology, Writing–original draft. LW: Methodology, Writing–original draft. XH: Writing–original draft, Data curation, Investigation. YZ: Writing–original draft, Data curation. TZ: Methodology, Data curation, Writing–review and editing. YS: Data curation, Methodology, Writing–review and editing. YX: Data curation, Methodology, Writing–review and editing. BH: Writing–review and editing, Conceptualization, Funding acquisition, Investigation, Project administration, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key Research and Development Program of China (2018YFA0902000, 2021YFA1300904), National Natural Science Foundation of China (82273832, 32270106, 82104046, 81973372), and the CAMS Innovation Fund for Medical Sciences (2021-I2M-1-029).

Acknowledgments

We thank Bertolt Gust (Pharmaceutical Institute, University of Tuebingen, Germany) for kindly providing *Escherichia coli ΔtolC* mutant strain.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2023.1278601/full#supplementary-material

References

Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N., and Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* 116, 43–49. doi:10.1016/0378-1119(92)90627-2

Boojamra, C. G., Lemoine, R. C., Lee, J. C., Léger, R., Stein, K. A., Vernier, N. G., et al. (2001). Stereochemical elucidation and total synthesis of dihydropacidamycin D, a semisynthetic pacidamycin. *J. Am. Chem. Soc.* 123, 870–874. doi:10.1021/ja003292c

Chatterjee, S., Nadkarni, S. R., Vijayakumar, E. K., Patel, M. V., Ganguli, B. N., Fehlhaber, H. W., et al. (1994). Napsamycins, new Pseudomonas active antibiotics of the mureidomycin family from *Streptomyces* sp. HIL Y-82,11372. *J. Antibiot.* 47, 595–598. doi:10.7164/antibiotics.47.595

CLSI (2009). "Methods for Dilution antimicrobial susceptibility tests for bacteria that grow aerobically: approved standard—8th Edition," in *CLSI document M07-A8* (Wayne: Clinical and laboratory standards institute).

Gruschow, S., Rackham, E. J., Elkins, B., Newill, P. L., Hill, L. M., and Goss, R. J. (2009). New pacidamycin antibiotics through precursor-directed biosynthesis. *Chembiochem* 10, 355–360. doi:10.1002/cbic.200800575

Hameed, H. M. A., Fang, C., Liu, Z., Ju, Y., Han, X., Gao, Y., et al. (2022). Characterization of genetic variants associated with rifampicin resistance level in *Mycobacterium tuberculosis* clinical isolates collected in Guangzhou chest hospital, China. *Infect. Drug. Resist.* 15, 5655–5666. doi:10.2147/idr.s375869

He, H., Zheng, G., Jiang, W., Hu, H., and Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta. Biochim. Biophys. Sin.* 47, 231–243. doi:10.1093/abbs/gmv007

Hong, B., Phornphisutthimas, S., Tilley, E., Baumberg, S., and Mcdowall, K. J. (2007). Streptomycin production by *Streptomyces griseus* can be modulated by a mechanism not associated with change in the adpA component of the A-factor cascade. *Biotechnol. Lett.* 29, 57–64. doi:10.1007/s10529-006-9216-2

Inukai, M., Isono, F., Takahashi, S., Enokita, R., Sakaida, Y., and Haneishi, T. (1989). Mureidomycins A-D, novel peptidylnucleoside antibiotics with spheroplast forming activity. I. Taxonomy, fermentation, isolation and physico-chemical properties. *J. Antibiot.* 42, 662–666. doi:10.7164/antibiotics.42.662

Inukai, M., Isono, F., and Takatsuki, A. (1993). Selective inhibition of the bacterial translocase reaction in peptidoglycan synthesis by mureidomycins. *Antimicrob. Agents. Chemother.* 37, 980–983. doi:10.1128/AAC.37.5.980

Ióca, L. P., Dai, Y., Kunakom, S., Diaz-Espinosa, J., Krunic, A., Crnkovic, C. M., et al. (2021). A family of nonribosomal peptides modulate collective behavior in *Pseudovibrio* bacteria isolated from marine sponges. *Angew. Chem. Int. Ed. Engl.* 60, 15891–15898. doi:10.1002/anie.202017320

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., Yang, S., et al. (2015). Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* 81, 2506–2514. doi:10.1128/AEM.04023-14

Jiang, Z. B., Ren, W. C., Shi, Y. Y., Li, X. X., Lei, X., Fan, J. H., et al. (2018). Structure-based manual screening and automatic networking for systematically exploring sansanmycin analogues using high performance liquid chromatography tandem mass spectroscopy. *J. Pharm. Biomed. Anal.* 158, 94–105. doi:10.1016/j.jpba.2018.05.024

Karwowski, J. P., Jackson, M., Theriault, R. J., Chen, R. H., Barlow, G. J., and Maus, M. L. (1989). Pacidamycins, a novel series of antibiotics with anti-*Pseudomonas aeruginosa* activity. I. Taxonomy of the producing organism and fermentation. *J. Antibiot.* 42, 506–511. doi:10.7164/antibiotics.42.506

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., Hopwood, D. A., Charter, K., et al. (2000). *Practical* Streptomyces *genetics*. Norwich: John Innes Foundation.

Koppermann, S., Cui, Z., Fischer, P. D., Wang, X., Ludwig, J., Thorson, J. S., et al. (2018). Insights into the target interaction of naturally occurring muraymycin nucleoside antibiotics. *ChemMedChem* 13, 779–784. doi:10.1002/cmdc.201700793

Korn, F., Weingrtner, B., and Kutzner, H. J. (1978). "A study of twenty actinophages: morphology, serological relationship and host range," in *Genetics of the actinomycetales*. Editors E. Freerksen, I. Tarnok, and H. Thumin (New York: Gustav Fischer Verlag Press), 251–270.

Lemoine, R. C., Magon, A., and Hecker, S. J. (2002). Synthesis of base-modified dihydropacidamycins. *Bioorg. Med. Chem. Lett.* 12, 1121–1123. doi:10.1016/s0960-894x(02)00100-2

Mashalidis, E. H., Kaeser, B., Terasawa, Y., Katsuyama, A., Kwon, D. Y., Lee, K., et al. (2019). Chemical logic of MraY inhibition by antibacterial nucleoside natural products. *Nat. Commun.* 10, 2917. doi:10.1038/s41467-019-10957-9

Michailidou, F., Chung, C. W., Brown, M. J. B., Bent, A. F., Naismith, J. H., Leavens, W. J., et al. (2017). Pac13 is a Small, Monomeric dehydratase that mediates the formation of the 3'-deoxy nucleoside of pacidamycins. *Angew. Chem. Int. Ed. Engl.* 56, 12492–12497. doi:10.1002/anie.201705639

Niro, G., Weck, S. C., and Ducho, C. (2020). Merging natural products: muraymycin-sansanmycin hybrid structures as novel scaffolds for potential antibacterial agents. *Chemistry* 26, 16875–16887. doi:10.1002/chem.202003387

Okamoto, K., Sakagami, M., Feng, F., Togame, H., Takemoto, H., Ichikawa, S., et al. (2012). Total synthesis and biological evaluation of pacidamycin D and its 3'-hydroxy analogue. *J. Org. Chem.* 77, 1367–1377. doi:10.1021/jo202159q

Paget, M. S., Chamberlin, L., Atrih, A., Foster, S. J., and Buttner, M. J. (1999). Evidence that the extracytoplasmic function sigma factor çE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 181, 204–211. doi:10.1128/JB.181.1.204-211.1999

Ragab, A. E., Gruschow, S., Tromans, D. R., and Goss, R. J. (2011). Biogenesis of the unique 4',5'-dehydronucleoside of the uridyl peptide antibiotic pacidamycin. *J. Am. Chem. Soc.* 133, 15288–15291. doi:10.1021/ja206163j

Sambrook, J. R. D. (2001). *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Shi, Y., Jiang, Z., Lei, X., Zhang, N., Cai, Q., Li, Q., et al. (2016). Improving the *N*-terminal diversity of sansanmycin through mutasynthesis. *Microb. Cell. Fact.* 15, 77. doi:10.1186/s12934-016-0471-1

Tang, X., Gross, M., Xie, Y., Kulik, A., and Gust, B. (2013). Identification of mureidomycin analogues and functional analysis of an *N*-acetyltransferase in napsamycin biosynthesis. *Chembiochem* 14, 2248–2255. doi:10.1002/cbic.201300287

Tran, A. T., Watson, E. E., Pujari, V., Conroy, T., Dowman, L. J., Giltrap, A. M., et al. (2017). Sansanmycin natural product analogues as potent and selective anti-mycobacterials that inhibit lipid I biosynthesis. *Nat. Commun.* 8, 14414. doi:10.1038/ncomms14414

Tran, W., Kusay, A. S., Hawkins, P. M. E., Cheung, C. Y., Nagalingam, G., Pujari, V., et al. (2021). Synthetic sansanmycin analogues as potent *Mycobacterium tuberculosis* translocase I inhibitors. *J. Med. Chem.* 64, 17326–17345. doi:10.1021/acs.jmedchem.1c01407

Walsh, C. T., and Zhang, W. (2011). Chemical logic and enzymatic machinery for biological assembly of peptidyl nucleoside antibiotics. *ACS Chem. Biol.* 6, 1000–1007. doi:10.1021/cb200284p

Wang, L., Hu, Y., Zhang, Y., Wang, S., Cui, Z., Bao, Y., et al. (2009). Role of *sgcR3* in positive regulation of enediyne antibiotic C-1027 production of *Streptomyces globisporus* C-1027. *Bmc. Microbiol.* 9, 14. doi:10.1186/1471-2180-9-14

Weissauer-Condon, C., Engels, I., and Daschner, F. D. (1987). *In vitro* activity of four new quinolones in Mueller-Hinton broth and peritoneal dialysis fluid. *Eur. J. Clin. Microbiol.* 6, 324–326. doi:10.1007/BF02017630

Winn, M., Goss, R. J., Kimura, K., and Bugg, T. D. (2010). Antimicrobial nucleoside antibiotics targeting cell wall assembly: recent advances in structure-function studies and nucleoside biosynthesis. *Nat. Prod. Rep.* 27, 279–304. doi:10.1039/b816215h

World Health Organization (2022). Global tuberculosis report 2022. Available at: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022 (Accessed October 27, 2022).

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., et al. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* 30, 1180–1182. doi:10.1093/bioinformatics/btt764

Xie, Y., Cai, Q., Ren, H., Wang, L., Xu, H., Hong, B., et al. (2014). NRPS substrate promiscuity leads to more potent antitubercular sansanmycin analogues. *J. Nat. Prod.* 77, 1744–1748. doi:10.1021/np5001494

Xie, Y., Chen, R., Si, S., Sun, C., and Xu, H. (2007). A new nucleosidyl-peptide antibiotic, sansanmycin. *J. Antibiot.* 60, 158–161. doi:10.1038/ja.2007.16

Xie, Y., Xu, H., Si, S., Sun, C., and Chen, R. (2008). Sansanmycins B and C, new components of sansanmycins. *J. Antibiot.* 61, 237–240. doi:10.1038/ja.2008.34

Zhang, N., Liu, L., Shan, G., Cai, Q., Lei, X., Hong, B., et al. (2016). Precursor-directed biosynthesis of new sansanmycin analogs bearing para-substituted-phenylalanines with high yields. *J. Antibiot.* 69, 765–768. doi:10.1038/ja.2016.2

Zhang, W., Ntai, I., Bolla, M. L., Malcolmson, S. J., Kahne, D., Kelleher, N. L., et al. (2011). Nine enzymes are required for assembly of the pacidamycin group of peptidyl nucleoside antibiotics. *J. Am. Chem. Soc.* 133, 5240–5243. doi:10.1021/ja2011109

Zhang, W., Ostash, B., and Walsh, C. T. (2010). Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16828–16833. doi:10.1073/pnas.1011557107

Zhang, Y. J., Reddy, M. C., Ioerger, T. R., Rothchild, A. C., Dartois, V., Schuster, B. M., et al. (2013). Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* 155, 1296–1308. doi:10.1016/j.cell.2013.10.045

Zumla, A., Nahid, P., and Cole, S. T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. *Nat. Rev. Drug. Discov.* 12, 388–404. doi:10.1038/nrd4001

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Review
PUBLISHED 22 December 2023
DOI 10.3389/fbioe.2023.1328141

OPEN ACCESS

EDITED BY
Luan Luong Chu,
Phenikaa University, Vietnam

REVIEWED BY
Jin-Ho Lee,
Kyungsung University, Republic of Korea
Richard Kelwick,
Imperial College London,
United Kingdom
Tianwen Wang,
Anhui Polytechnic University, China

*CORRESPONDENCE
Tsan-Yu Chiu,
qiucanyu@genomics.cn

RECEIVED 26 October 2023
ACCEPTED 12 December 2023
PUBLISHED 22 December 2023

CITATION
Liu N, Dong W, Yang H, Li J-H and Chiu T-Y (2023), Application of artificial scaffold systems in microbial metabolic engineering.
Front. Bioeng. Biotechnol. 11:1328141.
doi: 10.3389/fbioe.2023.1328141

Application of artificial scaffold systems in microbial metabolic engineering

Nana Liu[1,2], Wei Dong[2], Huanming Yang[2], Jing-Hua Li[1] and Tsan-Yu Chiu[1,2]*

[1]College of Pharmaceutical Science, Zhejiang University of Technology, Hangzhou, China, [2]HIM-BGI Omics Center, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences (CAS), Hangzhou, China

In nature, metabolic pathways are often organized into complex structures such as multienzyme complexes, enzyme molecular scaffolds, or reaction microcompartments. These structures help facilitate multi-step metabolic reactions. However, engineered metabolic pathways in microbial cell factories do not possess inherent metabolic regulatory mechanisms, which can result in metabolic imbalance. Taking inspiration from nature, scientists have successfully developed synthetic scaffolds to enhance the performance of engineered metabolic pathways in microbial cell factories. By recruiting enzymes, synthetic scaffolds facilitate the formation of multi-enzyme complexes, leading to the modulation of enzyme spatial distribution, increased enzyme activity, and a reduction in the loss of intermediate products and the toxicity associated with harmful intermediates within cells. In recent years, scaffolds based on proteins, nucleic acids, and various organelles have been developed and employed to facilitate multiple metabolic pathways. Despite varying degrees of success, synthetic scaffolds still encounter numerous challenges. The objective of this review is to provide a comprehensive introduction to these synthetic scaffolds and discuss their latest research advancements and challenges.

KEYWORDS

multienzyme complexes, enzyme molecular scaffolds, reaction microcompartments, microbial cell factory, synthetic scaffold

1 Introduction

Recently, the successful synthesis of diverse natural products has been achieved through the introduction of heterologous metabolic pathways into microbial cell factories (Srinivasan and Smolke, 2020; Yuan et al., 2022; Zhang et al., 2022). These synthetic pathways are collectively built with heterologous enzymes selected from various sources and are not accompanied by their regulatory partners in the new host (Tran et al., 2023). Thus, these unregulated enzymes may not be able to channel intermediates from the input reactions to the formation of end products properly (Dueber et al., 2009; Liu et al., 2023). In contrast, metabolic enzymes of a native pathway can be formed multi-enzyme complexes (Tittes et al., 2022), enzyme molecular scaffolds (Artzi et al., 2017), reaction microchambers [e.g., arom multienzyme complexes (Lumsden and Coggins, 1977), or caveolae, etc. (Polka et al., 2016)] to mediate the catalytic cascades coordinately. The structural entities play a pivotal role in facilitating efficient substrate transfer between adjacent enzyme active sites (Castellana et al., 2014). Thus, adopting a synthetic scaffold is one of the strategies to co-ordinate the none-native enzymes in microbial cell factories. To address the challenges several synthetic

scaffolds have been devised for the precise modulation of enzyme activity. For instance, enzymes can be assembled on scaffolds made of DNA or protein, where protein-protein or DNA interactions are employed to facilitate the formation of cascading complexes among enzymes (Gad and Ayakar, 2021). These DNA or protein scaffolds are thought to form channels conducive to continuous metabolism, directing metabolic intermediates from one enzyme to another to regulate the spatial distribution of enzymes and increase their local concentrations (Tippmann et al., 2017). The utilization of synthetic scaffolds serves to significantly decrease interenzyme distances, thereby effectively restricting the diffusion of intermediate metabolitesand concurrently attenuates cellular cytotoxicity (Conrado et al., 2008). Though, the synthetic scaffolds had been successfully applied to metabolic engineering, trial and error are still the only way we may learn.

This review provides a detailed account of the applications of artificially synthesized scaffolds through specific case studies and comprehensively summarizes the latest advancements in various scaffold assembly methods. Additionally, it explores the potential challenges faced by artificially synthesized scaffolds. At the same time, the possible role of the current hot artificial intelligence (AI) technology in the application of artificial stent systems is also discussed.

2 Protein scaffold

Paired protein scaffolders fall into three main categories, including protein-peptide, peptide-peptide, and protein-protein pairs. These scaffolds can be fused directly with target enzymes to induce assembly, and they achieve enzyme assembly through non-covalent or covalent interactions between ligands and receptors in the scaffolds, with little effect on enzyme properties (Price et al., 2016; Chen et al., 2023).

2.1 Protein–peptide pair

Protein-peptide interaction recognition domains are widely present in various cells, where they participate in the assembly of intracellular complexes and play diverse cellular functions. Currently, several modular protein domains [e.g., PDZ domain, SH3 domains, GTPase binding domain (GBD), $GBD_1SH3_1PDZ_2$, $GBD_1SH3_2PDZ_4$, and $GBD_1SH3_4PDZ_4$] and their corresponding partners have been identified (Pawson, 2007). The PDZ domain (also known as GLGF repeats or DHR domains) is typically an essential component of multi-domain scaffold proteins involved in cell polarity and intercellular interactions (Fanning and Anderson, 1996). It can selectively recognize the C-terminal peptide sequences on its partner protein and then assemble them into a complex and target specific subcellular localization sites (Tonikian et al., 2008). Based on this protein-peptide interaction, Gao et al. (2014) proposed a scaffold-free self-assembly strategy. This strategy was successfully demonstrated using the NAD (H) cycle system with L-*tert*-leucine as a model, achieving scaffold-free self-assembly technology. They fused the PDZ (PSD95/Dlg1/zo-1) domain and corresponding ligands (PDZlig) from metazoan cells separately with the octameric leucine dehydrogenase (LDH, derived from *Bacillus subtilis* BEST7613) and the dimeric formate dehydrogenase (FDH, derived from *Lodderomyces elongisporus* NRRL YB4239) (Figure 1A). The fusion proteins self-assembled into extended supramolecular interaction networks, significantly enhancing the efficiency and structural stability of the coenzyme cycling system involving NAD (H). Compared to their non-assembled counterparts, they exhibited better performance (Gao et al., 2014).

The SRC Homology 3 Domain (or SH3 domain) is a small protein domain containing 60 amino acid residues that are folded into beta-barrels with five or six β-strands arranged as two tightly packed anti-parallel β sheets (Schlessinger, 1994). It typically binds to proline-rich peptides in its respective binding partner. A SH3-ligand interaction strategy was used to successfully assemble methanol dehydrogenase (Mdh), 3-hexulose-6-phosphate synthase (Hps), and 6-phospho-3-hexuloseisomerase (Phi) into highly efficient enzyme complexes, significantly improving the conversion efficiency of methanol to fructose-6-phosphate (F6P) (Price et al., 2016). Meanwhile, in *Escherichia coli*, lactate dehydrogenase was utilized as an NADH scavenger to establish an "NADH sink." By combining these two strategies, a 97-fold increase in extracellular F6P production and a 9-fold improvement in intracellular methanol consumption were successfully achieved (Price et al., 2016).

The GTPase binding domain (GBD) from the actin polymerization switch N-WASP could be recognized by the GTP-bound Cdc42. Dueber et al. (2009), used the GTPase binding domain (GBD), the SH3 domain, and the PDZ domain to build a synthetic scaffold to provide modular control over metabolic pathway flux. By varying the numbers of these three domains ($GBD_xSH3_yPDZ_z$; x, y, z = number of domain repeats) to control the co-localization ratio of the interacting catalytic enzymes (e.g., atoB, HMGS, and HMGR), the optimal scaffold quantity is $GBD_1SH3_2PDZ_2$, and this engineering strategy led to a 77-fold increase in malic acid production (Dueber et al., 2009). By constructing a self-assembly enzyme reactor in *E. coli*, the stoichiometric ratio of two enzymes in the baicalein synthesis pathway was regulated to form an enzyme complex. This strategy significantly increased the titers of baicalein and scutellarein by 6.6 and 1.4 folds, respectively (Ji et al., 2021). Wei et al. utilized *tobacco mosaic virus* (TMV) virus-like particle (VLP) as a protein scaffold and orthogonal reactive protein pairs (SpyCatcher/SpyTag and SnoopCatcher/SnoopTag) as a linking module to assemble terpene biosynthesis in *E. coli*, enabling the production of amorpha-4,11-diene (Wei et al., 2020).

2.2 Peptide–peptide pair

In nature, there are numerous examples of optimizing metabolic pathway performance by forming multienzyme complexes. A prominent example is polyketide synthases (PKSs), which are considered among the most intricate proteins in nature. PKSs are classified into types I, II, and III and are involved in the synthesis of numerous compounds (Nivina et al., 2019). Through the sequential action of multiple catalytic modules, type I modular polyketide synthases are capable of extending, modifying, and terminating polyketide peptide chains. These interrelated modules interact

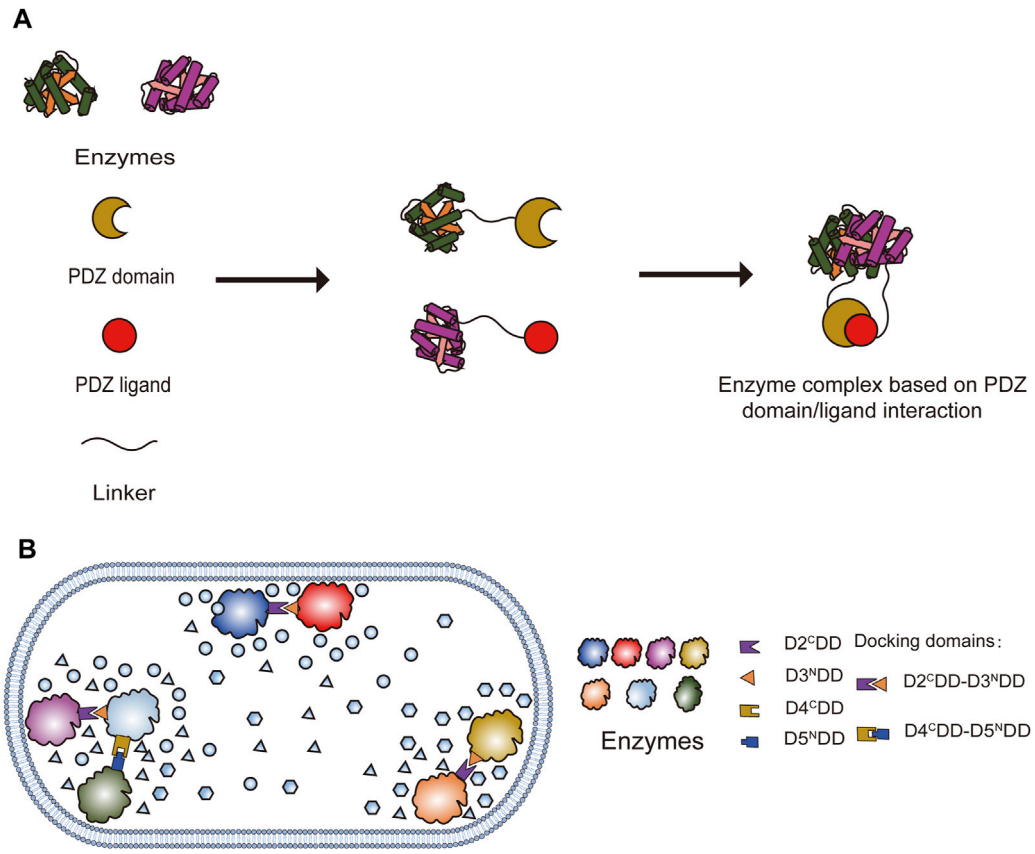


FIGURE 1
Protein scaffold; **(A)** Enzyme complex assembly based on PDZ domain and ligand interaction, adapted from (Gao et al., 2014); **(B)** Enzyme complex assembly based on mPKSeal strategy, adapted from (Sun et al., 2022).

with each other through docking domains (DDs) mediated by folding regions at the C- and N-termini (Weissman, 2016). Sun et al. (2022) utilized the DDs of type I *cis*-AT-PKS as mediators to develop a multi-enzyme assembly strategy named mimic PKS enzyme assembly line (mPKSeal), which mimics the assembly line of PKS enzymes (Figure 1B). This strategy was applied in engineered *E. coli* to enhance astaxanthin production and possesses the ability to co-locate enzymes within the cell, enabling the assembly of two or three enzyme units in different cellular environments (Sun et al., 2022). Their research also found that DDs from different PKSs but located on the same molecular evolutionary tree also possess enzyme assembly activity. The mPKSeal enzyme assembly strategy has tremendous potential for enhancing the efficiency of biocatalytic reactions by regulating the spatial positioning of enzymes without altering their abundance. These short-chain DDs have little significant impact on the catalytic activity of most enzyme assemblies, but they have a more pronounced effect on certain specific membrane proteins. Thus, this is an issue that needs to be noted when dealing with membrane proteins (Sun et al., 2022).

Kang et al. (2019) developed a scaffold-free modular enzyme assembly, which incorporated short peptide tags RIDD and RIAD derived from cAMP-dependent protein kinase (PKA) (Wong and Scott, 2004) and the A kinase-anchoring proteins (AKAPs) (Sarma et al., 2010), respectively. In *E. coli*, researchers successfully assembled enzyme complexes by combining the interaction peptides of RIAD and RIDD with the isopentenyl diphosphate isomerase (IDI) and Geranylgeranyl diphosphate synthase (CrtE), involved in the carotenoid biosynthesis pathway. This led to a significant increase in the production of carotenoids. Furthermore, in *S. cerevisiae*, the assembly of these two short peptides with the IDI and CrtE for the biosynthesis of lycopene resulted in a 58% increase in lycopene production (Kang et al., 2019). Xu et al. (2022) assembled two cytochrome P450 enzymes, ent-kaurene oxidase (KO) and kaurenoic acid 13α-hydroxylase (KAH), using RIAD and RIDD, successfully increasing the production of rubusoside and rebaudiosides in yeast. Fink et al. (2020), through orthogonally designed coiled-coil interaction domains, cluster resveratrol biosynthetic pathway enzymes, thereby increasing the yield of resveratrol in *E. coli*. The yield of resveratrol produced by this method is higher than that of direct enzyme fusion and internal protein-mediated fusion. At the same time, the biosynthesis of mevalonate in yeast was improved by this clustering method.

2.3 Protein-protein pair

Protein-protein pairs of special peptides in TatB/TatC can spontaneously interact to form aggregates. Henriques de Jesus

TABLE 1 Examples and overview of artificial scaffold systems.

Types of scaffolds	Description	Host	Application	References
Protein scaffold	PDZ and PDZ ligand	*E.coli*; *Pichia pastoris*	Enhance the biosynthesis yield of baicalein and scutellarein; facilitate the biosynthesis of ginsenoside precursors; Improve the production of itaconic acid	Zhao et al. (2016), Yang et al. (2017), Ji et al. (2021)
	SH3-ligand interaction pair	*E.coli*	Drive the conversion of methanol into H6P; promote the production of malic acid	Price et al. (2016), Somasundaram et al. (2020)
	PduA*-Multi-Enzyme Complex System	*E.coli*	Improve the synthesis efficiency of 5-aminolevulinic acid (5-ALA)	Luo et al. (2022)
	Cohesin-dockerin (Coh-Doc) pair	*S. cerevisiae*	Enhance the production rate of NADH; improve the metabolic flux of pyruvate	Liu et al. (2013), Kim et al. (2016)
	PDZ and SH3 domains	*E.coli*	Increase the effective concentration of myoinositol	Moon et al. (2010)
	RIAD and RIDD short peptide tags	*E.Coli; S. cerevisiae*	Increase carotenoid production by 5.7-fold and lycopene production by 58%; increase the yield of rubusosides and rebaudiosides	Kang et al. (2019), Xu et al. (2022)
	Artificial Protein Scaffold System (AProSS)	*S. cerevisiae*	The yield of violacein and deoxyviolacein increased by 29% and 63%, respectively, while the ratio of violacein to deoxyviolacein increased by 18%	Li et al. (2018)
	GBD, SH3, and PDZ domain	*E.coli*	Increase the production of indigoidine, gamma-aminobutyric acid, butyrate, and R-(−)-linalool; increase the production of methylhydroxybutyrate by 77 yields	Dueber et al. (2009), Baek et al. (2013), Pham et al. (2015), Pham et al. (2016), Wang et al. (2020), Wu et al. (2021)
	SpyCatcher/SpyTag and SnoopCatcher/SnoopTag pairs	*E.coli*	Increase the biosynthetic flux of carotenoids	Qu et al. (2019)
	Tobacco mosaic virus (TMV) virus-like particle (VLP), SpyCatcher/SpyTag and SnoopCatcher/SnoopTag	*E.coli*	Realization of the production of amorpha-4,11-diene	Wei et al. (2020)
	mimic PKS enzyme assembly line (mPKSeal)	*E.coli*	Improve the production of astaxanthin	Sun et al. (2022)
Nucleic acid scaffold	ADB1, ADB2, and ADB3	*Bacillus subtilis*	Increase the production of *N*-acetylglucosamine; enhance the biosynthesis of *L*-threonine	Lee et al. (2013), Liu et al. (2014)
	ADO and AAR	*E.coli*	Enhance the production potential of linear n-alkanes	Rahmana et al. (2014)
	TALEs	*E.coli*	Increase the biosynthesis of indole-3-acetic acid (IAA)	Zhu et al. (2016), Xie et al. (2019)
	ZF domains	*E.coli*	Enhance the metabolism of resveratrol, 1,2-propanediol, and mevalonate	Conrado et al. (2012)
	PCNA	*E.coli*	Increase the catalytic activity of P450 and electron transfer-associated proteins	Hirakawa and Nagamune (2010)
	dCas9, SpyCatcher, and SnoopCatcher pairs	*E.coli*	Increase the yield of reducing sugars by 2.8 folds	Berckman and Chen (2020)
	dCas9, MS2 and PP7 aptamers	*S. cerevisiae*	Regulate the expression of enzymes involved in the violacein biosynthetic pathway to control metabolic flux	Pothoulakis et al. (2022)
	2DRNA scaffolds	*E.coli*	Increase the metabolic output of the pathway for pentadecane production	Delebecque et al. (2011), Sachdeva et al. (2014)
	RNA Scaffold, MS2 and PP7 aptamers	*E.coli*	The fluorescence intensity in the GFP cleavage assay increased by 2.25-fold, while the multi-enzyme efficiency in the IAA synthesis pathway increased by 1.43-fold	Team and Chen (2015)
	CRISPR scRNA	*S. cerevisiae*	Redirecting metabolic flux in a complex branched metabolic pathway	Zalatan et al. (2015)
	MCMs	*E.coli*	Improving the production of α-farnesene	Wei et al. (2022)

(Continued on following page)

TABLE 1 (*Continued*) Examples and overview of artificial scaffold systems.

Types of scaffolds	Description	Host	Application	References
Organelle scaffold	ER-Derived Vesicles	*S. cerevisiae*	Constructing a *cis,cis*-muconic acid (CCM) biosynthetic pathway in vesicles to assess its feasibility	Reifenrath et al. (2020)
	lipid droplets (LDs)	*S. cerevisiae*	The production rate of ethyl acetate has been increased by nearly two-fold	Lin et al. (2017)
	outer membrane vesicles (OMVs)	*E.coli*	The glucose yield has increased by 23-fold compared to the free enzyme	Park et al. (2014)
	protein cages	*E.coli*	The production of lycopene has increased by 8.5-fold	Kang et al. (2022)

et al. (2017) realized the co-localization of enzymes by exchanging the membrane anchors of the dhurrin biosynthesis pathway enzymes into TatB and TatC components of the twin-arginine translocation pathway with self-assembly properties. This method achieved a 4-fold increase in dhurrin titer and reduced the amount of intermediates and side products. CipA and CipB are two small proteins that form protein crystalline inclusions (PCIs) in the cytoplasm of *Photorhabdus luminescens*. Wang Y. et al. (2017) used CipA as a protein scaffold to bring together multiple enzymes (Vio enzymes) of the violacein biosynthetic pathway to explore its application *in vivo*. They found that the violacein production in the complex was significantly increased with fewer side-products (Wang Y. et al., 2017). More recently, Park et al. (2022) applied CipB scaffold proteins to bring P450s and reductase in close proximity, facilitating electron transfer between them. The development of strains producing lutein, apigenin, (+)-nootkatone, and L-3, 4-dihydroxyphenylalanine (l-DOPA) in *E. coli* has demonstrated the universal applicability of this electronic channel strategy. By using an implicit negative design, Sahtoe et al. (2022) generated beta sheet-mediated heterodimers capable of assembling into a variety of complexes. Their implicit negative design principle makes it possible to design higher-order asymmetric polyprotein complexes by rigid fusion of components through structured helical linkers. Moreover, due to the small size of the unfused protomers, the complex can be functionalized by easily fusing with the protein of interest by subunits (Sahtoe et al., 2022).

The scaffold assembly can realize the orderly arrangement of multiple enzymes, shorten the spatial distance of enzymes, accelerate sequential catalysis, and achieve a high yield. The composite formed by the scaffold assembly strategy can significantly improve the efficiency of enzyme catalysis in the biosynthesis of natural products and has broad application prospects in the fields of metabolic engineering and synthetic biology. However, until now, only four enzymes could be assembled sequentially. The reason is that the assembly of the enzyme requires the fusion expression of the scaffold and the enzyme through the joint, and misfolding is easy to occur during the fusion process, which will affect the assembly performance. On the other hand, the assembly of multiple enzymes is affected by steric hindrance, making it difficult to achieve sequential arrangement (Chen et al., 2023).

3 Nucleic acid scaffold

3.1 DNA scaffold

In addition to using protein scaffold approaches, the DNA double helix can serve as an alternative scaffold system. Compared with protein scaffolds, nucleic acid scaffolds have higher flexibility and maneuverability. With the advancement of gene editing technology, several molecular tools are available for efficient and specific DNA targeting, such as zinc finger proteins (ZFPs), transcription activator-like effector (TALE) proteins, and CRISPR-Cas (Kim and Kim, 2014). Moreover, the plasmid DNA as a configurable, stable, and robust scaffold for arranging biosynthetic enzymes in the cytoplasm is proposed.

3.1.1 DNA scaffold based on zinc finger protein

In *E. coli*, the plasmid DNAs equipped with corresponding zinc finger protein binding sites were designed to assemble three different biosynthetic pathways to produce resveratrol, 1,2-propanediol, or mevalonate (Conrado et al., 2012). By varying the enzymatic ratios and the base pairs between each enzyme, the catalytic efficiency is improved, which leads to better production of final products. This similar approach had been applied to using zinc finger proteins (ZFPs) as adaptors to anchor the L-threonine biosynthetic genes. By using DNA scaffold assembly, the accumulation of the intermediate homoserine is reduced and significantly increases the efficiency of L-threonine biosynthesis due to the shortening of the distance between enzymes and the enhancement of the local concentration of metabolic products (Lee et al., 2013). In addition, Rahmana et al. (2014) utilized fusion proteins of chimeric acyl-ACP reductase (AAR) and aldehyde decarbonylase (ADO), or zinc finger proteins, as guides to assemble ADO/AAR with DNA scaffolds. The strain containing the fusion protein ADO-AAR showed a 4.8-fold increase in the production of branched alkanes. On the DNA scaffold, when the stoichiometric ratio of ADO to AAR was 3:1, the strain exhibited an 8.8-fold increase in production, reaching the optimal level of branched alkane synthesis (Rahmana et al., 2014). Liu et al. (2014) constructed a *B. subtilis* strain capable of producing *N*-acetylglucosamine (GlcNAc) and, for the first time, utilized DNA scaffolds to regulate the activities of glucosamine-6-phosphate synthase and GlcNAc-6-phosphate *N*-acetyltransferase, resulting in a GlcNAc titer of 4.55 g/L.

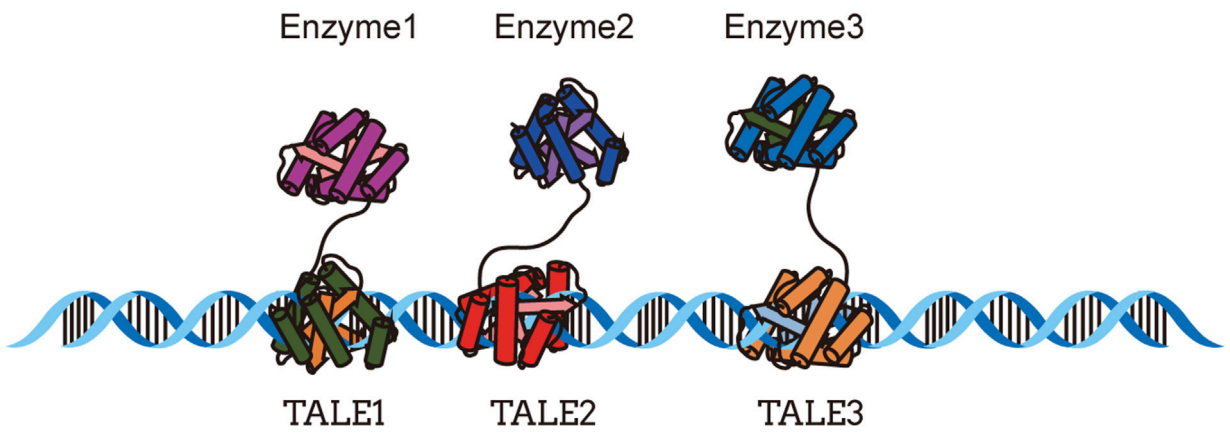


FIGURE 2
DNA scaffold system based on TALE, adapted from (Xie et al., 2019).

3.1.2 DNA scaffold based on TALEs

TALEs (transcription activator-like effectors) are effectors from the family III in *Xanthomonas* bacteria (Boch and Bonas, 2010), and different TALEs share similar structural domains. These domains are capable of binding to the host cell genome and act as transcription factors to recognize specific DNA repeat sequences (Boch et al., 2009). Based on the binding characteristics of TALEs to DNA, Zhu et al. (2016) developed a TALE-based DNA scaffold system and applied it to the biosynthesis of indole-3-acetic acid (IAA). Furthermore, in a modified TALE-DNA scaffold system, three fusion enzymes were successfully assembled in *E. coli* and significantly increased the production of a mevalonate-producing tri-enzymatic pathway (Xie et al., 2019) (Figure 2).

3.1.3 DNA scaffold based on CRISPR-Cas

CRISPR-associated (Cas) nucleases are a class of DNA-binding proteins distinct from zinc finger enzymes and transcriptional activator-like effector proteins. Through the guidance of RNA molecules complementary to DNA sequences, the CRISPR-Cas DNA scaffold achieves Cas-specific customization (Lim et al., 2020). The Cas9 protein is ideally suited for modular enzyme assembly, with a high affinity for DNA and the ability to bind to specific DNA sequences (Berckman and Chen, 2019). By combining the dCas9 nuclease (from *Streptococcus pyogenes*) with the Spycatcher-Spytag chemical binding system, successful modular assembly of five enzymatic pathways involved in violacein biosynthesis was achieved, resulting in a significant increase in violacein production (Lim et al., 2020). In another study, two orthogonal SpyCatcher and SnoopCatcher pairs were bioconjugated onto two different dCas9 proteins, enabling them to guide the enzyme assembly to the DNA scaffold, resulting in a 2.8-fold increase in reducing sugar production compared to the unassembled enzyme (Berckman and Chen, 2020).

3.1.4 DNA scaffold based on PCNA

Proliferating cell nuclear antigen (PCNA) is a trimeric ring-shaped protein (Moldovan et al., 2007) that binds to DNA as a scaffold for DNA-related enzymes. The fusion protein between the PCNA and the functional protein can act as a nanoscale part and self-assemble to form a functional nanohybrid complex. Fusion of three PCNA proteins with bacterial cytochrome P450 or one of the two electron transfer-related proteins can form a stable heterotrimer complex, resulting in increased local ferridoxin concentrations of P450 and ferredoxin reductase and high catalytic activity of electron transfer within the complex (Hirakawa and Nagamune, 2010).

3.2 RNA scaffold

By binding to the adapter, the RNA scaffold achieves highly specific binding to the target enzyme. Delebecque et al. (2011) designed and assembled multidimensional RNA structures to spatially organize proteins in cells and utilized this RNA scaffold to optimize a biosynthetic pathway for hydrogen production. A fluorescent protein library containing 8 aptamers and corresponding RNA domains was successfully assembled by fusing the active viral scaffold made of RNA with engineered proteins and specific RNA domains. The scaffold enables the co-localization of fragmented green fluorescent proteins to achieve precise measurement of cellular activity (Sachdeva et al., 2014). The application of this RNA scaffold to the synthesis pathway enzymes of pentadecane and succinic acid demonstrated the assembly of 0D, 1D, and 2D scaffolds. In the biosynthesis pathway of pentadecane, the 2D-assembled scaffold resulted in a 2.4-fold increase in pentadecane production (Sachdeva et al., 2014).

CRISPR-associated RNA scaffolds provide a powerful approach to the construction of synthetic gene programs. By inducing the expression of the dCas9 protein, we can achieve gene activation and inhibition, thereby enabling the directed expression of complex branching metabolic pathways (Zalatan et al., 2015). Pothoulakis et al. (2022) developed an RNA design approach for RNA origami scaffolds (termed sgRNAO) by recruiting activation domains from fused single-guide RNAs and RNA origami scaffolds to control gene expression in yeast. They successfully applied sgRNAOs to regulate the expression of enzymes involved in the violacein biosynthetic pathway (Pothoulakis et al., 2022).

Compared to DNA or protein-based scaffolds, RNA scaffolds, as non-coding synthetic scaffolds, offer greater flexibility. They can

control protein spatial organization, such as distances and orientations between bound proteins, chemical dosage, and complex sizes, among others (Delebecque et al., 2012). However, although RNA scaffolds have certain advantages, they also have obvious disadvantages, including high synthesis costs, easy hydrolysis by nucleases, large environmental factors, and structural instability, which limit their application. Therefore, it requires further research and development to resolve these drawbacks (Geraldi et al., 2021).

4 Cellular scaffolds

4.1 Natural cellular scaffolds

As subcellular structures within cells, the integrity and autonomy of organelles have sparked scientists interest in using them as scaffolds for enzyme assembly (Liu et al., 2023). Bacterial microcompartments (BMCs) (Kerfeld et al., 2018) are self-assembling organelles composed of enzymatic cores that participate in the metabolism of various organic compounds such as 1,2-propanediol (Petit et al., 2013), ethanolamine (Petit et al., 2013), fucose, and rhamnose (Parsons et al., 2008), playing a crucial role in carbon fixation processes (Lawrence et al., 2014). In *E. coli*, reconstitution of recombinant microcompartments can be achieved by translocating the entire propanediol utilization (Pdu) operon from *Citrobacter freundii* (Parsons et al., 2008). The *Zymomonas mobilis* enzymes pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (Adh) can be targeted to PduP of *C. freundii* to form a simple ethanol bioreactor inside the Pdu microcompartment shell (Lawrence et al., 2014). The enzymes Pdc and Adh, necessary for ethanol production, are expressed heterologously using a foreign host and targeted to the protein shell. In strains containing target enzymes, the ethanol yield significantly increases when the protein shell content is highest, including the strains producing shell proteins P18-Pdc and D18-Adh (Lawrence et al., 2014). In another study, the known Pdu (D18 and P18) targeting peptides were fused with four different 1,2-propanediol synthetic enzymes to create fusion proteins that target the empty Pdu BMC system. The fusion strategy of targeting peptides with all proteins involved in 1,2-propanediol synthesis significantly increased the product yield (Lee et al., 2016). Studies have shown that BMCs have great potential for constructing organelle scaffolds and are relatively easy to design, especially when it comes to the metabolism of toxic intermediates (Lawrence et al., 2014). Nielsen et al. (2013) transferred the biosynthetic pathway of the aromatic defense compound dhurrin [D-glucopyranosyloxy-(S)-p-hydroxymandelonitrile, a cyanogenic glucoside] to plant chloroplasts, utilizing photo-induced water splitting as the electron source to drive product synthesis in a light-dependent manner. The biosynthetic pathway of dhurrin involves three ER-localized enzymes, including two P450 enzymes, CYP79A1 and CYP71E1, as well as an NADPH cytochrome P450 oxidoreductase, POR. The chloroplast stroma provides a reducing environment for P450 enzymes, thereby enhancing their stability (Nielsen et al., 2013). Gram-negative bacteria release spherical nanoscale particles called outer membrane vesicles (OMVs) during their growth process. These vesicles have a composition similar to the bacterial outer membrane, containing lipopolysaccharides (LPSs), outer membrane proteins (OMPs), and phospholipids (Beveridge, 1999). In order to achieve the goal of hijacking the bacterial cell export pathway to simultaneously produce, package, and release an active enzyme, phosphotriesterase (PTE), Alves et al. (2015) attempted to establish synthetic linkages between enzymes and proteins known to exist in the outer membrane. They used the SpyCatcher/SpyTag (SC/ST) bioconjugated system to connect OmpA proteins present in OMVs to phosphotriesterase from *Brevundimonas diminuta*. A PTE-SpyCatcher (PTE-SC) fusion protein and a SpyTag transmembrane porin protein (OmpA-ST) were constructed. The coexpression of OmpA-ST with PTE-SC not only reduced the toxicity of PTE and improved the overall PTE production level, but also enhanced the stability of packaging enzymes against repeated freeze-thaw cycles (Alves et al., 2015). By employing a truncated ice nucleation protein anchoring motif (INP) on OMVs, a trivalent protein scaffold containing three divergent cohesin domains was utilized for site-specific expression of a three-enzyme cascade, resulting in a 23-fold increase in glucose production (Park et al., 2014) (Figure 3A). Yang et al. (2021) developed a metabolically engineered strain of *E. coli* to produce seven natural colorants, and they significantly increased the yield of seven natural colorants through cell morphological engineering, IMV and OMV formation, and fermentation optimization strategies. *S. cerevisiae* possesses various subcellular compartments, making it an ideal host for building heterologous natural product biosynthesis. Shi et al. (2021) used the PLN1 protein to target endoplasmic reticulum-localized cytochrome P450 enzymes and protopanaxadiol (PPD) synthase (PPDs) towards lipid droplets [DDs, the storage organelle for dammarenediol-II (DD)], resulting in a 394% increase in the conversion of DD to PPD and an elevated conversion rate of DD to 86.0%. Peroxisomes are organelles involved in fatty acid degradation. In yeast, significant improvements in the production of fatty acid derivatives, such as fatty alcohols, alkanes, and olefins, can be achieved by engineering peroxisomes. For example, increasing the number of peroxisomes can triple the production of fatty acid derivatives (Zhou et al., 2016).

4.2 Synthetic organelle scaffold

Cellular organelle scaffolds can be artificially designed to meet various metabolic pathway requirements. Lipids are widely present in cells and can form cell membranes, with many proteins anchored to these membrane structures. Inspired by this, researchers have attempted to use lipids as synthetic scaffolds to achieve co-assembly of lipids and target proteins. They discovered that, unlike most bacteriophages, bacteriophage $\phi6$ contains a protein nucleocapsid surrounded by a lipid envelope and several membrane proteins (Sinclair et al., 1975). During infection of the natural host (Stitt and Mindich, 1983) and in strains of *E. coli* expressing genes encoding bacteriophages $\phi6$ viral proteins (Johnson and Mindich, 1994), assembly intermediates of bacteriophage $\phi6$ with lipid-like structures were found. When only genes encoding the three viral proteins P8, P9, and P12 are expressed in *E. coli*, circular particles composed of a mixture of lipids and proteins can be observed by cryo-electron microscopy (Sarin et al., 2012). Myhrvold et al. (2016)

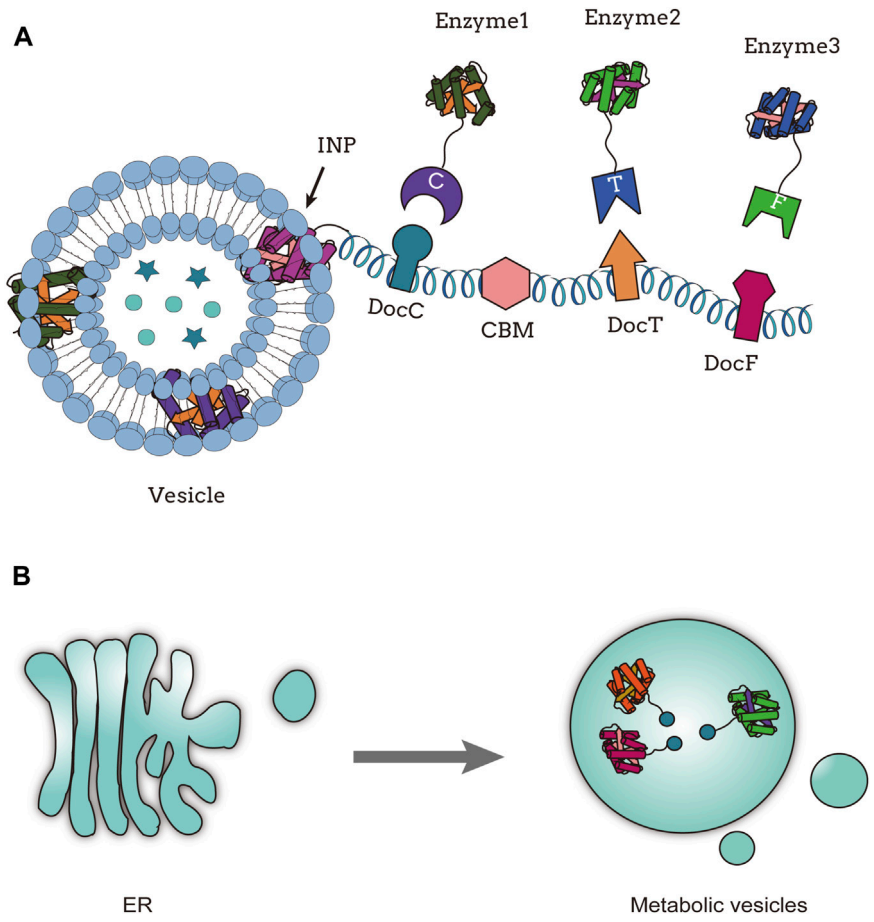


FIGURE 3
Organelle scaffold; **(A)** Functional assembly of multienzyme systems on outer membrane vesicles, adapted from (Park et al., 2014); **(B)** Compartmental assembly of endoplasmic reticulum derived metabolic vesicles, adapted from (Reifenrath et al., 2020).

engineered synthetic lipid-containing scaffolds (SLSs) in *E. coli*. The scaffold consists of the membrane protein P9 and the non-structural protein P12, which are required for the formation of the particle structure. The target proteins are fused to the C-terminus of P9 to position them on the scaffold. TnaA and FMO enzymes involved in indigo biosynthesis were fused to the C-terminus of P9 to construct lipid scaffolds for increased indigo production (Myhrvold et al., 2016). By utilizing a protein scaffold based on the plant lipid droplet protein oleosin and cohesin-dockerin interaction pairs recruited upstream enzymes, the final three reaction steps of yeast ester biosynthesis were assembled on lipid droplets (LDs) within yeast cells. This resulted in a doubling of the synthesis rate of ethyl acetate (Lin et al., 2017). There are derived vesicles containing metabolic pathway enzymes in the endoplasmic reticulum ("metabolic vesicles"). Reifenrath et al. (2020) integrate the three enzymes involved in the production of *cis,cis*-muconic acid into yeast vesicles, construct an ER-derived synthetic cell envelope containing the metabolic pathway, and map the yeast metabolism (Figure 3B). Wei et al. (2022) used β-Cav1 caveolar vesicles as an enzyme assembly scaffold to immobilize enzymes involved in the biosynthesis of α-farnesene from isopentenyl diphosphate and dimethylallyl pyrophosphate through non-covalent interactions or covalent protein reactions on the β-Cav1 caveolar vesicles. They successfully constructed a multienzyme complex called multi-enzyme caveolar membranes (MCMs), which increased the catalytic efficiency of α-farnesene by 10-fold compared to the non-assembled enzymes (Wei et al., 2022).

5 Discussion and summary

With the advancement of biotechnology, building synthetic pathways into heterologous microbial hosts became possible (Jiang et al., 2021). However, this none-native expression of catalytic enzymes still faces numerous issues and challenges, including low productivity and yield, rapid diffusion and degradation of key intermediates, and the accumulation of toxic metabolites (Jiang et al., 2021). To address these issues, scientists have been organizing these enzymes into molecular complexes in space to enhance the local concentration of enzymes and metabolites, thereby improving reaction flux (Wang S. Z. et al., 2017). Currently, various assembly strategies have been derived based on the interactions of biomacromolecules such as nucleic acids, peptides, and scaffold proteins (Table 1). In comparison, the spatial organization of biosynthetic pathway enzymes through synthetic scaffolds has proven to be an effective method for

enhancing reaction rates and biosynthetic yields while also improving host viability. Despite achieving a certain level of success, the precise prediction of artificial scaffold assembly remains challenging due to the complexity of enzyme structures. Furthermore, the catalytic efficiency of multi-enzyme complexes is influenced by numerous parameters, and the impact mechanism of linker sequences on both the multi-enzyme complex and substrate channeling effects remains undisclosed. Therefore, further research is needed to address the factors in the design-built-test-learn (DBTL) cycles of establishing artificial scaffold systems.

The computational inference of biological systems has emerged as a transformative field, driven by the confluence of advanced computational methods, machine learning techniques, and ever-expanding biological data. The best-known case is protein structure prediction, and it has been a grand challenge for decades. Since the advancement of the computational implication of a protein's spatial arrangement of atoms and computing power, several artificial intelligence (AI) systems have been successfully applied to predict the structure of protein complexes (e.g., the AlphaFold-Multimer, RoseTTAFolds, and trRosetta) (Baek et al., 2021; Du et al., 2021; Brems et al., 2022; Ivanov et al., 2022). In addition, the use of artificial intelligence computational methods can effectively find scaffold proteins from protein interactors and fully reveal their functions (Oh and Yi, 2016). Currently, we may *de novo* design protein structures for specific purposes (Watson et al., 2023).

In contrast to protein-based scaffold systems, using nucleic acid as a scaffold provides its unique advantages. The DNA scaffold systems are more stable, easier to design, and have higher stability (Chen et al., 2014; Siu et al., 2015). However, the DNA scaffolds also serve several drawbacks, such as high cost and high error rates in self-assembly, making the folding process in living cells susceptible to environmental factors such as temperature and ions and difficult to manipulate. In addition, the DNA topology is single in dimension and has the problem of supercoiling (Polka et al., 2016). DNA origami is a technology for designing complex three-dimensional DNA structures. This has been broadly applied in nanotechnology, creating nanoscale structures for various purposes, including drug delivery and molecular computing (Zhou et al., 2023). Thus, DNA origami could greatly assist in resolving the issues and unlocking the great potential of using DNA scaffolds in metabolic engineering. Alternatively, the RNA can also be used as a scaffold for the self-assembly of multi-enzyme complexes. However, its discrete structure prevents the formation of complex geometric shapes in living organisms, limiting its application compared to other types of scaffolds (Lv et al., 2020). AI models also open up new opportunities for predicting the secondary and tertiary structures of RNA molecules, providing insights into designing new functions and interactions and improving stability (Sato and Hamada, 2023).

Cellular scaffolds constructed based on different organelles or membranes have achieved varying degrees of success. However, due to the levels of complexity of biological parts (e.g., membranes, proteins, and protein interactions), the working mechanism of such scaffold systems still needs to be further explored. Synthetic biology aims to combine multidisciplinary disciplines that pursue the development new biological parts or systems. The promise of using AI technology to resolve complex biological questions is now is gradually revealed. Combining these two technologies, we can better understand the mechanisms of different types of artificial scaffolds and design streamlined artificial scaffolds that provide a broader range of engineering applications.

Author contributions

NL: Writing–review and editing, Writing–original draft. WD: Writing–review and editing, Supervision. HY: Funding acquisition, Writing–review and editing, Supervision. J-HL: Supervision, Writing–review and editing. T-YC: Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was financially supported by HIM-BGI Omics Center, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences (CAS), Hangzhou, China 310022. The funder did not participate in the designing, performing, or reporting of the current study.

Acknowledgments

Sincerely thanks to teacher T-YC for revising and polishing the whole article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alves, N. J., Turner, K. B., Daniele, M. A., Oh, E., Medintz, I. L., and Walper, S. A. (2015). Bacterial nanobioreactors--directing enzyme packaging into bacterial outer membrane vesicles. *ACS Appl. Mater Interfaces* 7 (44), 24963–24972. doi:10.1021/acsami.5b08811

Artzi, L., Bayer, E. A., and Moraïs, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* 15 (2), 83–95. doi:10.1038/nrmicro.2016.164

Baek, J. M., Mazumdar, S., Lee, S. W., Jung, M. Y., Lim, J. H., Seo, S. W., et al. (2013). Butyrate production in engineered *Escherichia coli* with synthetic scaffolds. *Biotechnol. Bioeng.* 110 (10), 2790–2794. doi:10.1002/bit.24925

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373 (6557), 871–876. doi:10.1126/science.abj8754

Berckman, E. A., and Chen, W. (2019). Exploiting dCas9 fusion proteins for dynamic assembly of synthetic metabolons. *Chem. Commun. (Camb)* 55 (57), 8219–8222. doi:10.1039/c9cc04002a

Berckman, E. A., and Chen, W. (2020). A modular approach for dCas9-mediated enzyme cascading via orthogonal bioconjugation. *Chem. Commun. (Camb)* 56 (77), 11426–11428. doi:10.1039/d0cc04196c

Beveridge, T. J. (1999). Structures of gram-negative cell walls and their derived membrane vesicles. *J. Bacteriol.* 181 (16), 4725–4733. doi:10.1128/jb.181.16.4725-4733.1999

Boch, J., and Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi:10.1146/annurev-phyto-080508-081936

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326 (5959), 1509–1512. doi:10.1126/science.1178811

Brems, M. A., Runkel, R., Yeates, T. O., and Virnau, P. (2022). AlphaFold predicts the most complex protein knot and composite protein knots. *Protein Sci.* 31 (8), e4380. doi:10.1002/pro.4380

Castellana, M., Wilson, M. Z., Xu, Y., Joshi, P., Cristea, I. M., Rabinowitz, J. D., et al. (2014). Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* 32 (10), 1011–1018. doi:10.1038/nbt.3018

Chen, R., Chen, Q., Kim, H., Siu, K. H., Sun, Q., Tsai, S. L., et al. (2014). Biomolecular scaffolds for enhanced signaling and catalytic efficiency. *Curr. Opin. Biotechnol.* 28, 59–68. doi:10.1016/j.copbio.2013.11.007

Chen, Z., Wu, T., Yu, S., Li, M., Fan, X., and Huo, Y. X. (2023). Self-assembly systems to troubleshoot metabolic engineering challenges. *Trends Biotechnol.* doi:10.1016/j.tibtech.2023.06.009

Conrado, R. J., Varner, J. D., and DeLisa, M. P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* 19 (5), 492–499. doi:10.1016/j.copbio.2008.07.006

Conrado, R. J., Wu, G. C., Boock, J. T., Xu, H., Chen, S. Y., Lebar, T., et al. (2012). DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* 40 (4), 1879–1889. doi:10.1093/nar/gkr888

Delebecque, C. J., Lindner, A. B., Silver, P. A., and Aldaye, F. A. (2011). Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333 (6041), 470–474. doi:10.1126/science.1206938

Delebecque, C. J., Silver, P. A., and Lindner, A. B. (2012). Designing and using RNA scaffolds to assemble proteins *in vivo*. *Nat. Protoc.* 7 (10), 1797–1807. doi:10.1038/nprot.2012.102

Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., et al. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* 16 (12), 5634–5651. doi:10.1038/s41596-021-00628-9

Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., et al. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27 (8), 753–759. doi:10.1038/nbt.1557

Fanning, A. S., and Anderson, J. M. (1996). Protein-protein interactions: PDZ domain networks. *Curr. Biol.* 6 (11), 1385–1388. doi:10.1016/s0960-9822(96)00737-3

Fink, T., Stevovic, B., Verwaal, R., Roubos, J. A., Gaber, R., Bencina, M., et al. (2020). Metabolic enzyme clustering by coiled coils improves the biosynthesis of resveratrol and mevalonate. *Amb. Express* 10 (1), 97. doi:10.1186/s13568-020-01031-5

Gad, S., and Ayakar, S. (2021). Protein scaffolds: a tool for multi-enzyme assembly. *Biotechnol. Rep. (Amst)* 32, e00670. doi:10.1016/j.btre.2021.e00670

Gao, X., Yang, S., Zhao, C., Ren, Y., and Wei, D. (2014). Artificial multienzyme supramolecular device: highly ordered self-assembly of oligomeric enzymes *in vitro* and *in vivo*. *Angew. Chem. Int. Ed. Engl.* 53 (51), 14027–14030. doi:10.1002/anie.201405016

Geraldi, A., Khairunnisa, F., Farah, N., Bui, L. M., and Rahman, Z. (2021). Synthetic scaffold systems for increasing the efficiency of metabolic pathways in microorganisms. *Biol. (Basel)* 10 (3), 216. doi:10.3390/biology10030216

Henriques de Jesus, M. P. R., Zygadlo Nielsen, A., Busck Mellor, S., Matthes, A., Burow, M., Robinson, C., et al. (2017). Tat proteins as novel thylakoid membrane anchors organize a biosynthetic pathway in chloroplasts and increase product yield 5-fold. *Metab. Eng.* 44, 108–116. doi:10.1016/j.ymben.2017.09.014

Hirakawa, H., and Nagamune, T. (2010). Molecular assembly of P450 with ferredoxin and ferredoxin reductase by fusion to PCNA. *Chembiochem* 11 (11), 1517–1520. doi:10.1002/cbic.201000226

Ivanov, Y. D., Taldaev, A., Lisitsa, A. V., Ponomarenko, E. A., and Archakov, A. I. (2022). Prediction of monomeric and dimeric structures of CYP102A1 using AlphaFold2 and AlphaFold multimer and assessment of point mutation effect on the efficiency of intra- and interprotein electron transfer. *Molecules* 27 (4), 1386. doi:10.3390/molecules27041386

Ji, D., Li, J., Xu, F., Ren, Y., and Wang, Y. (2021). Improve the biosynthesis of baicalein and scutellarein via manufacturing self-assembly enzyme reactor *in vivo*. *ACS Synth. Biol.* 10 (5), 1087–1094. doi:10.1021/acssynbio.0c00606

Jiang, Y., Zhang, X., Yuan, H., Huang, D., Wang, R., Liu, H., et al. (2021). Research progress and the biotechnological applications of multienzyme complex. *Appl. Microbiol. Biotechnol.* 105 (5), 1759–1777. doi:10.1007/s00253-021-11121-4

Johnson, M. D., 3rd, and Mindich, L. (1994). Plasmid-directed assembly of the lipid-containing membrane of bacteriophage phi 6. *J. Bacteriol.* 176 (13), 4124–4132. doi:10.1128/jb.176.13.4124-4132.1994

Kang, W., Ma, T., Liu, M., Qu, J., Liu, Z., Zhang, H., et al. (2019). Modular enzyme assembly for enhanced cascade biocatalysis and metabolic flux. *Nat. Commun.* 10 (1), 4248. doi:10.1038/s41467-019-12247-w

Kang, W., Ma, X., Kakarla, D., Zhang, H., Fang, Y., Chen, B., et al. (2022). Organizing enzymes on self-assembled protein cages for cascade reactions. *Angew. Chem. Int. Ed. Engl.* 61 (52), e202214001. doi:10.1002/anie.202214001

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F., and Sutter, M. (2018). Bacterial microcompartments. *Nat. Rev. Microbiol.* 16 (5), 277–290. doi:10.1038/nrmicro.2018.10

Kim, H., and Kim, J. S. (2014). A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* 15 (5), 321–334. doi:10.1038/nrg3686

Kim, S., Bae, S. J., and Hahn, J. S. (2016). Redirection of pyruvate flux toward desired metabolic pathways through substrate channeling between pyruvate kinase and pyruvate-converting enzymes in *Saccharomyces cerevisiae*. *Sci. Rep.* 6, 24145. doi:10.1038/srep24145

Lawrence, A. D., Frank, S., Newnham, S., Lee, M. J., Brown, I. R., Xue, W. F., et al. (2014). Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.* 3 (7), 454–465. doi:10.1021/sb4001118

Lee, J. H., Jung, S. C., Bui le, M., Kang, K. H., Song, J. J., and Kim, S. C. (2013). Improved production of L-threonine in *Escherichia coli* by use of a DNA scaffold system. *Appl. Environ. Microbiol.* 79 (3), 774–782. doi:10.1128/AEM.02578-12

Lee, M. J., Brown, I. R., Juodeikis, R., Frank, S., and Warren, M. J. (2016). Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*. *Metab. Eng.* 36, 48–56. doi:10.1016/j.ymben.2016.02.007

Li, T., Chen, X., Cai, Y., and Dai, J. (2018). Artificial Protein Scaffold System (AProSS): an efficient method to optimize exogenous metabolic pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* 49, 13–20. doi:10.1016/j.ymben.2018.07.006

Lim, S., Kim, J., Kim, Y., Xu, D., and Clark, D. S. (2020). CRISPR/Cas-directed programmable assembly of multi-enzyme complexes. *Chem. Commun. (Camb)* 56 (36), 4950–4953. doi:10.1039/d0cc01174f

Lin, J. L., Zhu, J., and Wheeldon, I. (2017). Synthetic protein scaffolds for biosynthetic pathway colocalization on lipid droplet membranes. *ACS Synth. Biol.* 6 (8), 1534–1544. doi:10.1021/acssynbio.7b00041

Liu, F., Banta, S., and Chen, W. (2013). Functional assembly of a multi-enzyme methanol oxidation cascade on a surface-displayed trifunctional scaffold for enhanced NADH production. *Chem. Commun. (Camb)* 49 (36), 3766–3768. doi:10.1039/c3cc40454d

Liu, M., Wang, Y., Jiang, H., Han, Y., and Xia, J. (2023). Synthetic multienzyme assemblies for natural product biosynthesis. *Chembiochem* 24 (6), e202200518. doi:10.1002/cbic.202200518

Liu, Y., Zhu, Y., Ma, W., Shin, H. D., Li, J., Liu, L., et al. (2014). Spatial modulation of key pathway enzymes by DNA-guided scaffold system and respiration chain engineering for improved N-acetylglucosamine production by Bacillus subtilis. *Metab. Eng.* 24, 61–69. doi:10.1016/j.ymben.2014.04.004

Lumsden, J., and Coggins, J. R. (1977). The subunit structure of the arom multienzyme complex of Neurospora crassa. A possible pentafunctional polypeptide chain. *Biochem. J.* 161 (3), 599–607. doi:10.1042/bj1610599

Luo, Z., Pan, F., Zhu, Y., Du, S., Yan, Y., Wang, R., et al. (2022). Synergistic improvement of 5-aminolevulinic acid production with synthetic scaffolds and system pathway engineering. *ACS Synth. Biol.* 11 (8), 2766–2778. doi:10.1021/acssynbio.2c00157

Lv, X., Cui, S., Gu, Y., Li, J., Du, G., and Liu, L. (2020). Enzyme assembly for compartmentalized metabolic flux control. *Metabolites* 10 (4), 125. doi:10.3390/metabo10040125

Moldovan, G. L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. *Cell* 129 (4), 665–679. doi:10.1016/j.cell.2007.05.003

Moon, T. S., Dueber, J. E., Shiue, E., and Prather, K. L. (2010). Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab. Eng.* 12 (3), 298–305. doi:10.1016/j.ymben.2010.01.003

Myhrvold, C., Polka, J. K., and Silver, P. A. (2016). Synthetic lipid-containing scaffolds enhance production by colocalizing enzymes. *ACS Synth. Biol.* 5 (12), 1396–1403. doi:10.1021/acssynbio.6b00141

Nielsen, A. Z., Ziersen, B., Jensen, K., Lassen, L. M., Olsen, C. E., Møller, B. L., et al. (2013). Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth. Biol.* 2 (6), 308–315. doi:10.1021/sb300128r

Nivina, A., Yuet, K. P., Hsu, J., and Khosla, C. (2019). Evolution and diversity of assembly-line polyketide synthases. *Chem. Rev.* 119 (24), 12524–12547. doi:10.1021/acs.chemrev.9b00525

Oh, K., and Yi, G. S. (2016). Prediction of scaffold proteins based on protein interaction and domain architectures. *BMC Bioinforma.*, 17 (Suppl. 6), 220. doi:10.1186/s12859-016-1079-5

Park, M., Sun, Q., Liu, F., DeLisa, M. P., and Chen, W. (2014). Positional assembly of enzymes on bacterial outer membrane vesicles for cascade reactions. *PLoS One* 9 (5), e97103. doi:10.1371/journal.pone.0097103

Park, S. Y., Eun, H., Lee, M. H., and Lee, S. Y. (2022). Metabolic engineering of *Escherichia coli* with electron channelling for the production of natural products. *Nat. Catal.* 5 (8), 726–737. doi:10.1038/s41929-022-00820-4

Parsons, J. B., Dinesh, S. D., Deery, E., Leech, H. K., Brindley, A. A., Heldt, D., et al. (2008). Biochemical and structural insights into bacterial organelle form and biogenesis. *J. Biol. Chem.* 283 (21), 14366–14375. doi:10.1074/jbc.M709214200

Pawson, T. (2007). Dynamic control of signaling by modular adaptor proteins. *Curr. Opin. Cell Biol.* 19 (2), 112–116. doi:10.1016/j.ceb.2007.02.013

Petit, E., LaTouf, W. G., Coppi, M. V., Warnick, T. A., Currie, D., Romashko, I., et al. (2013). Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PLoS One* 8 (1), e54337. doi:10.1371/journal.pone.0054337

Pham, V. D., Lee, S. H., Park, S. J., and Hong, S. H. (2015). Production of gamma-aminobutyric acid from glucose by introduction of synthetic scaffolds between isocitrate dehydrogenase, glutamate synthase and glutamate decarboxylase in recombinant *Escherichia coli*. *J. Biotechnol.* 207, 52–57. doi:10.1016/j.jbiotec.2015.04.028

Pham, V. D., Somasundaram, S., Lee, S. H., Park, S. J., and Hong, S. H. (2016). Redirection of metabolic flux into novel gamma-aminobutyric acid production pathway by introduction of synthetic scaffolds strategy in Escherichia coli. *Appl. Biochem. Biotechnol.* 178 (7), 1315–1324. doi:10.1007/s12010-015-1948-9

Polka, J. K., Hays, S. G., and Silver, P. A. (2016). Building spatial synthetic biology with compartments, scaffolds, and communities. *Cold Spring Harb. Perspect. Biol.* 8 (8), a024018. doi:10.1101/cshperspect.a024018

Pothoulakis, G., Nguyen, M. T. A., and Andersen, E. S. (2022). Utilizing RNA origami scaffolds in *Saccharomyces cerevisiae* for dCas9-mediated transcriptional control. *Nucleic Acids Res.* 50 (12), 7176–7187. doi:10.1093/nar/gkac470

Price, J. V., Chen, L., Whitaker, W. B., Papoutsakis, E., and Chen, W. (2016). Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proc. Natl. Acad. Sci. U. S. A.* 113 (45), 12691–12696. doi:10.1073/pnas.1601797113

Qu, J., Cao, S., Wei, Q., Zhang, H., Wang, R., Kang, W., et al. (2019). Synthetic multienzyme complexes, catalytic nanomachineries for cascade biosynthesis *in vivo*. *ACS Nano* 13 (9), 9895–9906. doi:10.1021/acsnano.9b03631

Rahmana, Z., Sung, B. H., Yi, J. Y., Bui le, M., Lee, J. H., and Kim, S. C. (2014). Enhanced production of n-alkanes in *Escherichia coli* by spatial organization of biosynthetic pathway enzymes. *J. Biotechnol.* 192 Pt A, 187–191. doi:10.1016/j.jbiotec.2014.10.014

Reifenrath, M., Oreb, M., Boles, E., and Tripp, J. (2020). Artificial ER-derived vesicles as synthetic organelles for *in vivo* compartmentalization of biochemical pathways. *ACS Synth. Biol.* 9 (11), 2909–2916. doi:10.1021/acssynbio.0c00241

Sachdeva, G., Garg, A., Godding, D., Way, J. C., and Silver, P. A. (2014). *In vivo* co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Res.* 42 (14), 9493–9503. doi:10.1093/nar/gku617

Sahtoe, D. D., Praetorius, F., Courbet, A., Hsia, Y., Wicky, B. I. M., Edman, N. I., et al. (2022). Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* 375 (6578), eabj7662. doi:10.1126/science.abj7662

Sarin, L. P., Hirvonen, J. J., Laurinmäki, P., Butcher, S. J., Bamford, D. H., and Poranen, M. M. (2012). Bacteriophage ϕ6 nucleocapsid surface protein 8 interacts with virus-specific membrane vesicles containing major envelope protein 9. *J. Virol.* 86 (9), 5376–5379. doi:10.1128/jvi.00172-12

Sarma, G. N., Kinderman, F. S., Kim, C., von Daake, S., Chen, L., Wang, B. C., et al. (2010). Structure of D-AKAP2:PKA RI complex: insights into AKAP specificity and selectivity. *Structure* 18 (2), 155–166. doi:10.1016/j.str.2009.12.012

Sato, K., and Hamada, M. (2023). Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery. *Brief. Bioinform* 24 (4), bbad186. doi:10.1093/bib/bbad186

Schlessinger, J. (1994). SH2/SH3 signaling proteins. *Curr. Opin. Genet. Dev.* 4 (1), 25–30. doi:10.1016/0959-437x(94)90087-6

Shi, Y., Wang, D., Li, R., Huang, L., Dai, Z., and Zhang, X. (2021). Engineering yeast subcellular compartments for increased production of the lipophilic natural products ginsenosides. *Metab. Eng.* 67, 104–111. doi:10.1016/j.ymben.2021.06.002

Sinclair, J. F., Tzagoloff, A., Levine, D., and Mindich, L. (1975). Proteins of bacteriophage phi6. *J. Virol.* 16 (3), 685–695. doi:10.1128/jvi.16.3.685-695.1975

Siu, K. H., Chen, R. P., Sun, Q., Chen, L., Tsai, S. L., and Chen, W. (2015). Synthetic scaffolds for pathway enhancement. *Curr. Opin. Biotechnol.* 36, 98–106. doi:10.1016/j.copbio.2015.08.009

Somasundaram, S., Jeong, J., Irisappan, G., Kim, T. W., and Hong, S. H. (2020). Enhanced production of malic acid by Co-localization of phosphoenolpyruvate carboxylase and malate dehydrogenase using synthetic protein scaffold in *Escherichia coli*. *Biotechnol. Bioprocess Eng.* 25 (1), 39–44. doi:10.1007/s12257-019-0269-1

Srinivasan, P., and Smolke, C. D. (2020). Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* 585 (7826), 614–619. doi:10.1038/s41586-020-2650-9

Stitt, B. L., and Mindich, L. (1983). Morphogenesis of bacteriophage phi 6: a presumptive viral membrane precursor. *Virology* 127 (2), 446–458. doi:10.1016/0042-6822(83)90157-5

Sun, X., Yuan, Y., Chen, Q., Nie, S., Guo, J., Ou, Z., et al. (2022). Metabolic pathway assembly using docking domains from type I cis-AT polyketide synthases. *Nat. Commun.* 13 (1), 5541. doi:10.1038/s41467-022-33272-2

Team, Z. J. C., and Chen, M. (2015). RNA scaffold: designed to Co-localize enzymes. *Methods Mol. Biol.* 1316, 105–112. doi:10.1007/978-1-4939-2730-2_9

Tippmann, S., Anfelt, J., David, F., Rand, J. M., Siewers, V., Uhlen, M., et al. (2017). Affibody scaffolds improve sesquiterpene production in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 6 (1), 19–28. doi:10.1021/acssynbio.6b00109

Tittes, Y. U., Herbst, D. A., Martin, S. F. X., Munoz-Hernandez, H., Jakob, R. P., and Maier, T. (2022). The structure of a polyketide synthase bimodule core. *Sci. Adv.* 8 (38), eabo6918. doi:10.1126/sciadv.abo6918

Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J. H., Reva, B., et al. (2008). A specificity map for the PDZ domain family. *PLoS Biol.* 6 (9), e239. doi:10.1371/journal.pbio.0060239

Tran, K.-N. T., Kumaravel, A., and Hong, S. H. (2023). Impact of the synthetic scaffold strategy on the metabolic pathway engineering. *Biotechnol. Bioprocess Eng.* 28 (3), 379–385. doi:10.1007/s12257-022-0350-z

Wang, L., Sun, Y., Lv, D., Liu, B., Guan, Y., and Yu, D. (2020). Protein scaffold optimizes arrangement of constituent enzymes in indigoidine synthetic pathway to improve the pigment production. *Appl. Microbiol. Biotechnol.* 104 (24), 10493–10502. doi:10.1007/s00253-020-10990-5

Wang, S. Z., Zhang, Y. H., Ren, H., Wang, Y. L., Jiang, W., and Fang, B. S. (2017a). Strategies and perspectives of assembling multi-enzyme systems. *Crit. Rev. Biotechnol.* 37 (8), 1024–1037. doi:10.1080/07388551.2017.1303803

Wang, Y., Heermann, R., and Jung, K. (2017b). CipA and CipB as scaffolds to organize proteins into crystalline inclusions. *ACS Synth. Biol.* 6 (5), 826–836. doi:10.1021/acssynbio.6b00323

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., et al. (2023). *De novo* design of protein structure and function with RFdiffusion. *Nature* 620 (7976), 1089–1100. doi:10.1038/s41586-023-06415-8

Wei, Q., He, S., Qu, J., and Xia, J. (2020). Synthetic multienzyme complexes assembled on virus-like particles for cascade biosynthesis in cellulo. *Bioconjug Chem.* 31 (10), 2413–2420. doi:10.1021/acs.bioconjchem.0c00476

Wei, Q., Wang, Y., Liu, Z., Liu, M., Cao, S., Jiang, H., et al. (2022). Multienzyme assembly on caveolar membranes in cellulo. *ACS Catal.* 12 (14), 8372–8379. doi:10.1021/acscatal.2c01906

Weissman, K. J. (2016). Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* 33 (2), 203–230. doi:10.1039/c5np00109a

Wong, W., and Scott, J. D. (2004). AKAP signalling complexes: focal points in space and time. *Nat. Rev. Mol. Cell Biol.* 5 (12), 959–970. doi:10.1038/nrm1527

Wu, J., Wang, X., Xiao, L., Wang, F., Zhang, Y., and Li, X. (2021). Synthetic protein scaffolds for improving R-(−)-Linalool production in *Escherichia coli*. *J. Agric. Food Chem.* 69 (20), 5663–5670. doi:10.1021/acs.jafc.1c01101

Xie, S. S., Qiu, X. Y., Zhu, L. Y., Zhu, C. S., Liu, C. Y., Wu, X. M., et al. (2019). Assembly of TALE-based DNA scaffold for the enhancement of exogenous multi-enzymatic pathway. *J. Biotechnol.* 296, 69–74. doi:10.1016/j.jbiotec.2019.03.008

Xu, Y., Wang, X., Zhang, C., Zhou, X., Xu, X., Han, L., et al. (2022). *De novo* biosynthesis of rubusoside and rebaudiosides in engineered yeasts. *Nat. Commun.* 13 (1), 3040. doi:10.1038/s41467-022-30826-2

Yang, D., Park, S. Y., and Lee, S. Y. (2021). Production of rainbow colorants by metabolically engineered *Escherichia coli*. *Adv. Sci. (Weinh)* 8 (13), e2100743. doi:10.1002/advs.202100743

Yang, Z., Gao, X., Xie, H., Wang, F., Ren, Y., and Wei, D. (2017). Enhanced itaconic acid production by self-assembly of two biosynthetic enzymes in *Escherichia coli*. *Biotechnol. Bioeng.* 114 (2), 457–462. doi:10.1002/bit.26081

Yuan, W., Jiang, C., Wang, Q., Fang, Y., Wang, J., Wang, M., et al. (2022). Biosynthesis of mushroom-derived type II ganoderic acids by engineered yeast. *Nat. Commun.* 13 (1), 7740. doi:10.1038/s41467-022-35500-1

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160 (1-2), 339–350. doi:10.1016/j.cell.2014.11.052

Zhang, J., Hansen, L. G., Gudich, O., Viehrig, K., Lassen, L. M. M., Schrubbers, L., et al. (2022). A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 609 (7926), 341–347. doi:10.1038/s41586-022-05157-3

Zhao, C., Gao, X., Liu, X., Wang, Y., Yang, S., Wang, F., et al. (2016). Enhancing biosynthesis of a ginsenoside precursor by self-assembly of two key enzymes in Pichia pastoris. *J. Agric. Food Chem.* 64 (17), 3380–3385. doi:10.1021/acs.jafc.6b00650

Zhou, Y., Dong, J., and Wang, Q. (2023). Fabricating higher-order functional DNA origami structures to reveal biological processes at multiple scales. *NPG Asia Mater.* 15 (1), 25. doi:10.1038/s41427-023-00470-3

Zhou, Y. J., Buijs, N. A., Zhu, Z., Gómez, D. O., Boonsombuti, A., Siewers, V., et al. (2016). Harnessing yeast peroxisomes for biosynthesis of fatty-acid-derived biofuels and chemicals with relieved side-pathway competition. *J. Am. Chem. Soc.* 138 (47), 15368–15377. doi:10.1021/jacs.6b07394

Zhu, L. Y., Qiu, X. Y., Zhu, L. Y., Wu, X. M., Zhang, Y., Zhu, Q. H., et al. (2016). Spatial organization of heterologous metabolic system *in vivo* based on TALE. *Sci. Rep.* 6, 26065. doi:10.1038/srep26065

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 05 February 2024
DOI 10.3389/fbioe.2024.1334427

OPEN ACCESS

EDITED BY
Jae Kyung Sohng,
Sun Moon University, Republic of Korea

REVIEWED BY
Tsan-Yu Chiu,
Beijing Genomics Institute (BGI), China
David Teze,
Technical University of Denmark, Denmark

*CORRESPONDENCE
Jinsong Shi,
shijs@163.com
Yan Zhang,
zhangyan@jiangnan.edu.cn

RECEIVED 07 November 2023
ACCEPTED 12 January 2024
PUBLISHED 05 February 2024

CITATION
Yang L, Yang M, Deng Z, Hou X, Zheng X, Ping Q, Rao Y, Shi J and Zhang Y (2024), Selective synthesis of rebaudioside M2 through structure-guided engineering of glycosyltransferase UGT94D1.
Front. Bioeng. Biotechnol. 12:1334427.
doi: 10.3389/fbioe.2024.1334427

Selective synthesis of rebaudioside M2 through structure-guided engineering of glycosyltransferase UGT94D1

Lifeng Yang[1,2], Mengliang Yang[3], Zhiwei Deng[3], Xiaodong Hou[3], Xiangting Zheng[1], Qian Ping[1], Yijian Rao[3], Jinsong Shi[4]* and Yan Zhang[1]*

[1]School of Life Sciences and Health Engineering, Jiangnan University, Wuxi, China, [2]School of Chemical and Material Engineering, Jiangnan University, Wuxi, China, [3]School of Biotechnology, Jiangnan University, Wuxi, China, [4]Key Laboratory of Carbohydrate Chemistry and Biotechnology, Ministry of Education, School of Life Sciences and Health Engineering, Jiangnan University, Wuxi, China

Rebaudioside M2 (Reb M2), a novel steviol glycoside derivative, has limited industrial applications due to its low synthetic yield and selectivity. Herein, we identify UGT94D1 as a selective glycosyltransferase for rebaudioside D (Reb D), leading to the production of a mono β-1,6-glycosylated derivative, Reb M2. A variant UGT94D1-F119I/D188P was developed through protein engineering. This mutant exhibited a 6.33-fold improvement in catalytic efficiency, and produced Reb M2 with 92% yield. Moreover, molecular dynamics simulations demonstrated that UGT94D1-F119I/D188P exhibited a shorter distance between the nucleophilic oxygen (OH6) of the substrate Reb D and uridine diphosphate glucose, along with an increased $O^{phosphate}$-C1-$O^{acceptor}$ angle, thus improving the catalytic activity of the enzyme. Therefore, this study provides an efficient method for the selective synthesis of Reb M2 and paves the way for its applications in various fields.

KEYWORDS

steviol glycoside, rebaudioside M2, glycosyltransferase, UGT94D1, protein engineering

1 Introduction

Steviol glycosides (SGs), which are derived from the plant *Stevia rebaudiana*, have become a popular alternative to sucrose and artificial sweeteners due to their low-calorie content and high sweetness (Ahmed et al., 2011; DuBois and Prakash, 2012; Gantait et al., 2015; Perrier et al., 2018). Importantly, toxicological studies conducted on SGs have found no evidence of teratogenic, carcinogenic or mutagenic effects, supporting their safe consumption (Ceunen and Geuns, 2013; Libik-Konieczny et al., 2021). To date, researchers have identified over 60 SGs, including steviol, rebaudioside A (Reb A), rebaudioside D (Reb D) and rebaudioside M (Reb M). Each of these SGs has a characteristic level of sweetness and taste as well as specific biological activities due to variations in the positions and quantities of sugars attached to their C-19 and/or C-13 positions (Bhardwaj et al., 2020; Ceunen and Geuns, 2013; Gerwig et al., 2016; Ohtani et al., 1992; te Poele et al., 2018; Tian et al., 2022). Among these known SGs, rebaudioside M2 (Reb M2) (Prakash, et al., 2014a), an isomer of Reb M, has a β 1→6 glycosidic bond at its C-19 position, which leads to distinct characteristics and a high economic value (Hellfritsch et al., 2012; Prakash, et al., 2014b; Olsson et al., 2016;

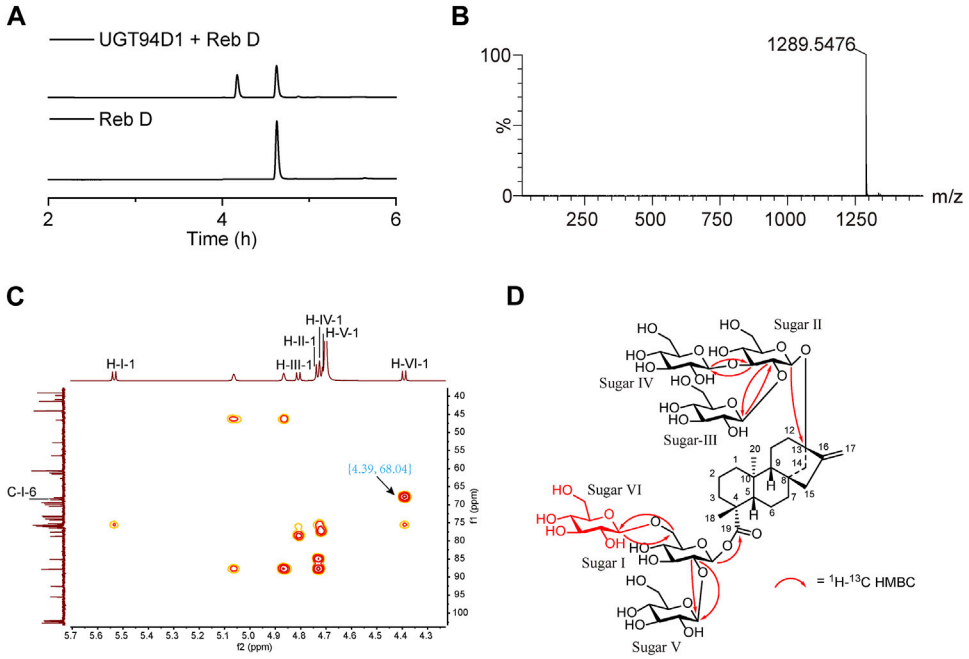


FIGURE 1
Selective synthesis of Reb M2 from Reb D catalyzed by glycosyltransferase UGT94D1. **(A)** UPLC analysis of the glycosylated product from Reb D by UGT94D1. **(B)** LC-MS analysis of the glycosylated product from Reb D by UGT94D1. **(C)** The correlation between H-1 of new sugar VI (δ_H 4.39) and C-6 of sugar I (δ_C 68.04) in ^{1}H-^{13}C heteronuclear multiple bond correlation (HMBC) spectrum of Reb M2 (D_2O). **(D)** ^{1}H-^{13}C HMBC correlations of Reb M2.

Purkayastha et al., 2016). Reb M2 was first identified as a byproduct of the bioconversion of Reb A by the glycosyltransferase UGTSL2 (Prakash, et al., 2014a). The yield of Reb M2 from this process is limited, and subsequent studies have aimed to improve the yield by utilizing UGTSL2 and a sucrose synthase (*St*SUS1). These efforts resulted in yields in excess of 10 g/L (Li et al., 2016), but the amount of Reb M2 produced was still not sufficient to meet the demands of the food industry. Therefore, there is a need for the development of more efficient and selective methods for the production of Reb M2.

Glycosyltransferases (GTs) are enzymes that catalyze the glycosylation of compounds in plant organisms (Lombard et al., 2014; Nidetzky et al., 2018). They transfer the activated sugar donor onto the sugar acceptor, resulting in the formation of glycosylated products. Uridine diphosphate glycosyltransferases (UGTs) (Yonekura-Sakakibara and Hanada, 2011; Zhang et al., 2020), which rely on uridine diphosphate sugar donors, belong to the GT1 multigene family (Paquette et al., 2003), comprising approximately half of the entire GTs family. UGTs have demonstrated exceptional stereo- and regio-selectivity in glycosylation of various natural products (Caputi et al., 2012; Wen et al., 2018; Li et al., 2020; Teze et al., 2021), showcasing their significant potential in the preparation of steviol glycoside derivatives. In our previous research (Ping et al., 2022), the glycosylation of Reb A to synthesize Reb D2 through the construction of a β 1→6 glycosidic bond at the C-19 position of Reb A was facilitated by a UGT known as UGT94D1, which was originally identified from *Sesamum indicum* (Noguchi et al., 2008; Ono et al., 2020; Brandt et al., 2021). Hence, it is plausible to achieve a highly selective and efficient synthesis of Reb M2 through the forming of a similar β 1→6 glycosidic bond using Reb D as the substrate.

In this study, we demonstrate that Reb M2 can be selectively synthesized by glycosyltransferase UGT94D1 from Reb D via the construction of a β 1→6 glycosidic bond at the C-19 position. Its conversion efficiency was greatly improved by the creation of the variant UGT94D1-F119I/D188P through protein engineering, which resulted in a 6.33-fold increase. Efficient production of Reb M2 was realized by coupling this mutant with sucrose synthase (Schmölzer et al., 2016) *At*SuSy in a cascade reaction system to realize the regeneration the glycosyl donor uridine diphosphate glucose (UDPG), resulting in a yield of 92% and generating 23.08 mM (29.79 mg/mL) of Reb M2. Therefore, in this study, we not only developed an efficient method for the selective synthesis of Reb M2 but also paved the way for its future applications.

2 Materials and methods

2.1 Construction of plasmids and bacterial strains

A list of the bacterial strains and plasmids utilized in this study can be found in Supplementary Table S1. The open reading frames encoding glycosyltransferase UGT94D1 (accession number: XP_011076907.1) from *S. indicum* and sucrose synthase *At*SuSy (accession number: NP_001031915) from *Arabidopsis thaliana* were synthesized and codon-optimized for expression in *E. coli* by Exsyn-bio (Wuxi, China). *Escherichia coli* Top10 was utilized for plasmid construction, while *E. coli* BL21 (DE3) strain was used for protein expression. The expression vectors used were the plasmids pET-21b (+) and pACYCDuet-1. Mutations in UGT94D1 were

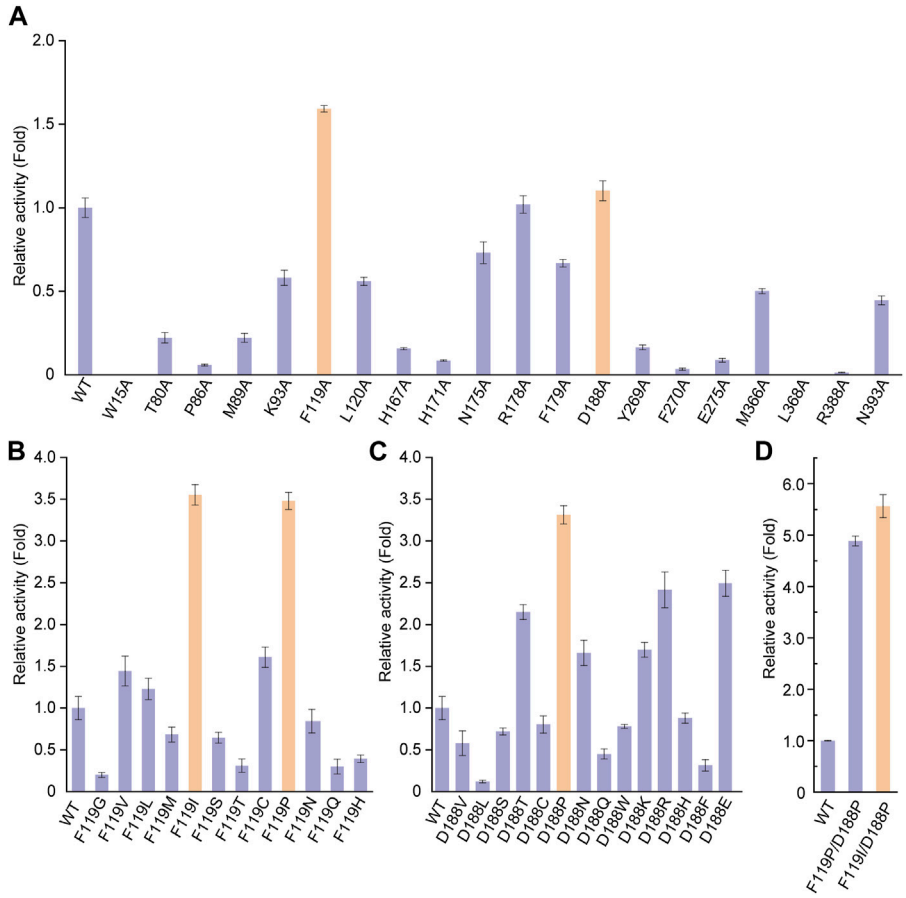


FIGURE 2
Protein engineering of UGT94D1. **(A)** Relative activities of alanine-scanning mutants of selected residues. Orange columns mean the mutants with improvement in the catalytic activity of UGT94D1. **(B)** Relative activities of different F119 mutants. **(C)** Relative activities of different D188 mutants. **(D)** Relative activities of different combinatorial mutants. UGT94D1 is used as the parent enzyme for mutagenesis, and its activity is set to 1. Orange column means the mutant with the highest catalytic activity. Error bars represent the standard deviation of three duplications.

accomplished through polymerase chain reaction, with the relevant primers listed in Supplementary Table S2. Sequencing was conducted to confirm all insertions and mutations in this process. Macklin Biotechnology (China) supplied Reb D, and other chemicals were obtained from Energy Chemical (Shanghai, China) or China National Pharmaceutical Group Corporation (Shanghai, China).

2.2 Protein expression and purification

Recombinant protein expression was performed essentially as described previously (Ping et al., 2022). *Escherichia coli* BL21 (DE3) with the relevant plasmids were cultured in 2 × YT medium with 100 mg/L ampicillin, incubated at 37°C with shaking. Protein expression was induced by adding 0.1 mM isopropyl-*β*-D-thiogalactopyranoside when the optical density at 600 nm reached 0.6 to 0.8, followed by incubation at 18°C for 8 h. After centrifugation, the cells were collected and suspended in a buffer solution containing 250 mM NaCl, 5% glycerol, 10 mM imidazole, and 50 mM Tris-HCl at pH 8.0, with a final concentration of 0.1 g/mL. Upon high-pressure homogenization-mediated disruption of the suspended cells, the proteins were purified using a Ni-nitrilotriacetate column. The proteins were eluted using a buffer comprising 250 mM NaCl, 250 mM imidazole, 5% glycerol, and 50 mM Tris-HCl pH 8. Imidazole was removed with a desalting column. Protein analyses were performed using sodium dodecyl sulfate polyacrylamide gel electrophoresis. For subsequent enzymatic assays, protein samples were concentrated to 10 mg/mL.

2.3 Enzymatic activity determination

The glycosylation specificity and catalytic activity of UGT94D1 or its mutants were assessed by incubating 5 μM of the enzyme with 0.5 mM Reb D in 200 μL reaction solutions containing 50 mM Tris-HCl at pH 8.0, 10 mM $MnCl_2$, and 5 mM UDPG. Enzymatic reactions were carried out at 35°C for 4 h, followed by heating at 95°C for 5 min. 400 μL methanol was then added to the system and the mixture was centrifuged at 20,000 *g* for 5 min. The resulting supernatants were filtered and prepared for analysis using ultra-performance liquid chromatography (UPLC) or liquid chromatography-mass spectrometry (LC-MS).

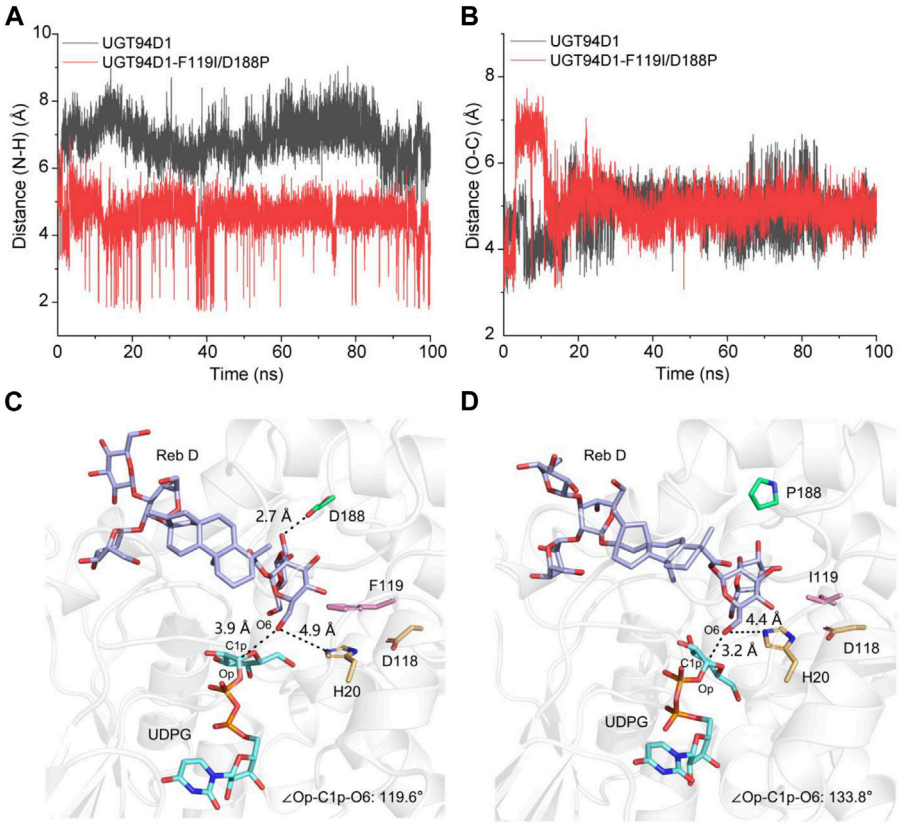


FIGURE 3
Variations of distance and the representative catalytic conformation of UGT94D1-UDPG-Reb D and UGT94D1-F119I/D188P-UDPG-Reb D in MD stimulations. **(A)** Distance between the nitrogen atom at ε position of catalytic residue H20 and the hydrogen atom of 6-OH of sugar I along 100 ns MD simulations. **(B)** Distance between C1p of UDPG and the hydroxyl oxygen atom in C-6 of sugar I along 100 ns MD simulations. **(C)** Representative catalytic conformation of UGT94D1 in MD simulations. Residues D188 and F119 are represented as green and pink sticks, respectively. The active sites (H20 and D118) are indicated in yellow sticks. **(D)** Representative catalytic conformation of UGT94D1-F119I/D188P in MD simulations. The black dash lines indicate the distances. Residues D188 and F119 are represented as green and pink sticks, respectively. Catalytic sites (H20 and D118), Reb D, and UDPG are shown as yellow, purple, and cyan sticks, respectively. The atoms oxygen and nitrogen are shown in the colors red and blue, respectively.

For the determination of kinetic parameters of UGT94D1 toward Reb D, reaction mixtures (200 μL) consisted of 50 mM Tris-HCl pH 8.0, 10 mM $MnCl_2$, 5 mM UDPG, varying concentrations of Reb D (0–2 mM), and purified enzyme (1–10 μg). After completing the above operations, analysis was conducted using a Waters Acquity UPLC system at 40°C. The detection wavelength was 210 nm. Elution solvents included acetonitrile (A) and 11.5 mM NaH_2PO_4 buffer (pH 2.6, B). A linear gradient elution method was used (0–1 min: 15% A, 6 min: 40% A, and 7–8 min: 15% A) with a flow rate of 0.3 mL/min.

2.4 Molecular docking and molecular dynamics simulations

The structural models of UGT94D1 and UDT94D1-F119I-D188P were built using AlphaFold (Jumper et al., 2021), and then UDPG and Reb D were docked into UGT94D1 with the Glide module of Schrödinger 2021. The reasonable conformation of UGT94D1-UDPG-Reb D and UDT94D1-F119I-D188P-UDPG-Reb D with the best score was chosen for subsequent molecular dynamics (MD) simulations. The protein was simulated using the FF14SB force field (Maier et al., 2015), while UDPG and Reb D were modeled using the General Amber force field (Wang et al., 2004; Wang et al., 2006). The protein complex was then solvated with a TIP3P water box (Jorgensen et al., 1983), which has at least 12 Å from the protein to the boundary of the water box. All MD simulations were implemented using the pmemd module of Amber20 (Case, et al., 2020). The procedure of MD simulations was performed as our previous study (Guo et al., 2022; Yang et al., 2023).

2.5 Optimization of cascade reaction conditions

Cascade reaction systems (200 μL reaction volumes) were prepared in which the following conditions were varied: pH (5.5–10), sucrose concentration (100–1,000 mM), concentration of UGT94D1-F119I/D188P (1–15 μM), concentration of *At*SuSy (1–15 μM) and reaction temperature (25°C–45°C). The reactions were performed using 50 mM buffer, 2.5 mM Reb D and 0.6 mM UDP and were incubated for 20 min. The reactions were terminated at 95°C for 5 min, followed by the addition of 10 times the volume of methanol, and analysis by UPLC-MS.

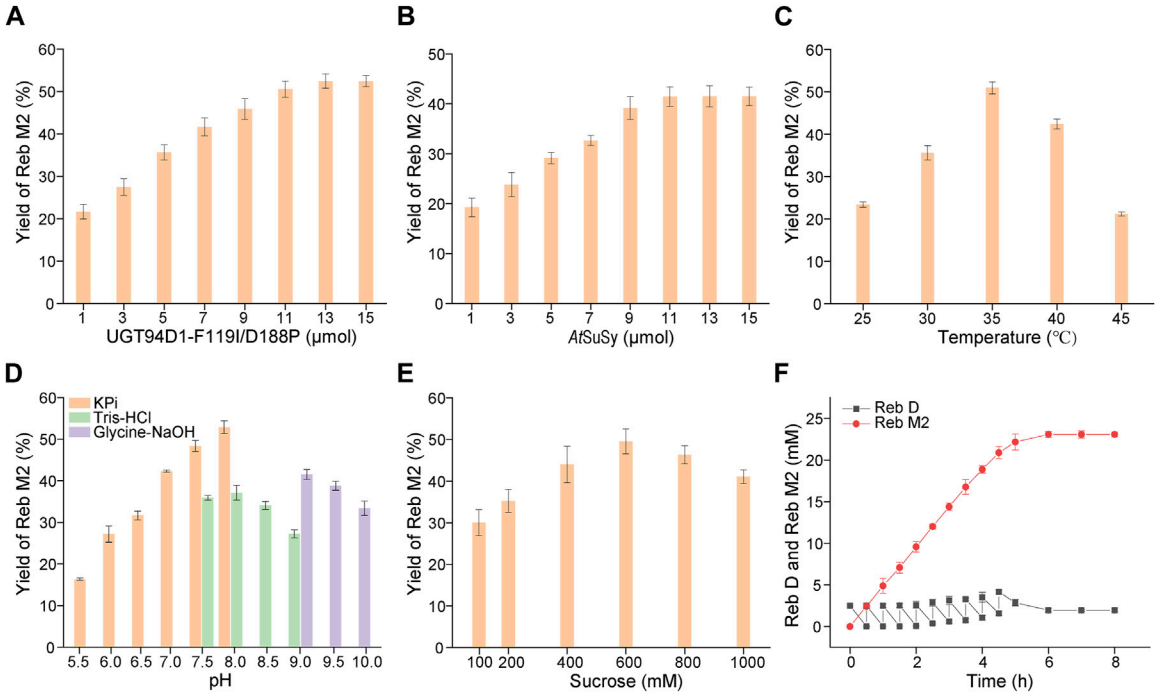


FIGURE 4
Optimization of the cascade reaction conditions for UGT94D1-F119I/D188P-*At*SuSy. **(A)** Optimization of the concentration of UGT94D1-F119I/D188P. **(B)** Optimization of the concentration of *At*SuSy. **(C)** Temperature optimization. **(D)** pH optimization. **(E)** Optimization of the concentration of sucrose. **(F)** Synthesis of Reb M2 in a cascade reaction with gradual addition of Reb D. The red line corresponds to the concentration of Reb M2, while the black line represents the concentration of Reb D. Error bars represent the standard deviation of three duplications.

2.6 Cell-free cascade reaction for the preparation of Reb M2

Under optimal conditions, cell-free cascade reactions (2 mL) were conducted. The reaction system included 13 μM UGT94D1-F119I/D188P, 11 μM *At*SuSy, 50 mM KPI (pH 8.0), 600 mM sucrose, 2.5 mM Reb D and 0.6 mM UDPG. The reactions were performed at 35°C for 8 h. Reb D (50 μL, 100 mM) was added incrementally at different time points (0.5–4.5 h, with 30 min intervals). Samples were taken at different time points (0–5 h with 30 min intervals and 6–8 h with 1 h intervals). The reactions were terminated at 95°C for 5 min, followed by the addition of 20 times the volume of methanol, and analysis by UPLC-MS.

For structure confirmation, the sample was purified with semi prepared HPLC, and the instrument and operative parameters were consistent with our previous work (Ping et al., 2022). The purified sample was further analyzed with nuclear magnetic resonance (NMR) spectroscopy.

3 Results and discussion

3.1 Selective glycosylation of Reb D to synthesize Reb M2 by UGT94D1

The activity of purified UGT94D1 was evaluated with UDPG as the glycosyl donor (Supplementary Figure S1). UPLC analysis shows the emergence of a single product (Figure 1A). This product was further analyzed by LC-MS (Figure 1B), and a peak at *m/z* 1,289.5476 ([M-H]$^-$ ion) was observed. This peak corresponds to a compound with the molecular formula $C_{56}H_{90}O_{33}$, which is consistent with a mono-glycosylated derivative of Reb D. The structure of this product was further investigated by 1D and 2D NMR spectroscopy (Figures 1C, D; Supplementary Figures S2–S8; Supplementary Table S3), and it was identified as 13-[(2-O-*β*-D-glucopyranosyl-3-O-*β*-D-glucopyranosyl-*β*-D-glucopyranosyl)oxy] ent-kaur-16-en-19-oic acid-[(2-O-*β*-D-glucopyranosyl-6-O-*β*-D-glucopyranosyl-*β*-D-glucopyranosyl) ester] (Figures 1C, D), demonstrating the formation of a 1→6 glycosidic bond attached to the 6-OH of sugar I of Reb D. Our structural characterization data were thus consistent with the structure of Reb M2 as reported in the literature (Prakash, et al., 2014a).

Next, the enzymatic and kinetic properties of UGT94D1 toward Reb D were investigated under the optimal reaction conditions (pH 8.0, 35°C) (Supplementary Figure S9). The K_m of UGT94D1 toward Reb D was determined to be 0.89 ± 0.05 mM, and the k_{cat} was determined to be 0.33 ± 0.08 min^{-1}. The resulting relatively low value of k_{cat}/K_m suggested the necessity to enhance the catalytic activity of UGT94D1 through protein engineering.

3.2 Improving the catalytic activity of UGT94D1 toward Reb D through protein engineering

To improve the catalytic activity of UGT94D1 toward Reb D, structure-guided engineering of UGT94D1 was performed. The

structural model of UGT94D1 was first built using AlphaFold (Jumper et al., 2021) and then UDPG and Reb D were docked into the predicted model (Supplementary Figures S10, S11). Comprehensive analysis revealed that the substrate binding pocket of UGT94D1 is located on the surface of the protein, and it features a relatively large cavity for accommodating the sugar moieties. The surrounding amino acids (P86, M89, H171, N175, R178, Y269, E275, M366, L368, H367, R388 and N393) may be involved in hydrophobic interactions with these sugar moieties, thereby facilitating the binding of Reb D (Supplementary Figure S12A). Meanwhile, residue D188 is positioned to form a hydrogen bond interaction with Reb D (Supplementary Figure S12B). In addition to these potentially critical binding residues, other residues (W15, T80, K93, F119, L120, F179, F270) located within 4 Å of the substrate binding pocket (Supplementary Figure S12C) were also analyzed via alanine scanning mutagenesis. As shown in Figure 2A, mutants F119A and D188A exhibited catalytic activities that were enhanced by 1.59- and 1.10-fold, respectively. These two residues were selected for saturation mutagenesis to further enhance the catalytic activity of UGT94D1. We found that mutation of F119 to isoleucine (F119I) and proline (F119P) led to increases of catalytic activity of 3.55- and 3.48-fold, respectively (Figure 2B). In addition, several mutations of D188 were found to increase the catalytic activity. Among them, mutant D188P exhibited the highest catalytic activity, which was increased by 3.31-fold compared to the wild-type (Figure 2C). Upon combinatorial mutagenesis at positions 119 and 188 (Figure 2D), we found that the catalytic activities of mutants F119I/D188P and F119P/D188P were further improved, as these enzymes exhibited catalytic activities that were 6.33 and 4.91 times that of UGT94D1, respectively. Therefore, UGT94D1-F119I/D188P was chosen for the scale-preparation of Reb M2.

3.3 Understanding the improved catalytic activity of UGT94D1-F119I/D188P by MD simulations

To study the molecular mechanism behind the catalytic enhancement of UGT94D1-F119I/D188P toward Reb D, we conducted molecular docking and 100 ns unconstrained molecular dynamics (MD) simulations of UGT94D1-F119I/D188P-UDPG-Reb D and UGT94D1-UDPG-Reb D. Based on preliminary mechanism research (Breton et al., 2006; Rahimi et al., 2019; Lin et al., 2020), the distance between the nitrogen atom at ε position of catalytic residue H20 and the hydroxyl hydrogen atom at C-6 of sugar I in Reb D and the distance between C1p of UDPG and the hydroxyl oxygen atom at C-6 of sugar I in Reb D were analyzed (Figures 3A, B). The results revealed that the distance between the Nε of His20 and the H of OH6 was shorter in UGT94D1-F119I/D188P compared to that in UGT94D1 (Figure 3A), while the distance between the electrophile C1p of UDPG and the nucleophilic oxygen of OH6 did not exhibit significant changes (Figure 3B).

To investigate the difference in conformation between UGT94D1-F119I/D188P-UDPG-Reb D and UGT94D1-UDPG-Reb D, representative conformations were obtained through MD simulations (Figures 3C, D). The obtained conformations revealed that mutations to the amino acids at positions 119 and 188 reshaped the hydrophobic binding pocket of the enzyme, making substrate Reb D binding more favorable. As a result, the distance between the nitrogen atom at ε position of catalytic residue H20 and the hydroxyl hydrogen atom at C-6 of sugar I in Reb D decreased from 4.9 to 4.4 Å. In addition, the Op-C1p-O6 angle in the UGT94D1-F119P/D188P (133.8°) is closer to the ideal position than that in UGT94D1 (119.6°), indicating a more favorable catalytic conformation (Teze et al., 2021). These conformational changes observed are highly favorable for the deprotonation of the hydroxy group at C-6 of sugar I in Reb D, thus improving the catalytic activity of the enzyme.

3.4 Cascade reaction for scale-preparation of Reb M2

To develop a scale-preparation of Reb M2, the variant UGT94D1-F119I/D188P was combined with *At*SuSy from *A. thaliana*, which was utilized to recycle UDPG from UDP using sucrose as a cost-effective sugar donor, to establish a cascade reaction. In order to optimize the production, the concentrations of the two enzymes, UGT94D1-F119I/D188P and *At*SuSy, were first investigated, and the maximum yield of Reb M2 was achieved using 13 μM UGT94D1-F119I/D188P and 11 μM *At*SuSy (Figures 4A, B). Then, the effects of temperature, pH and sucrose concentration were also examined. The optimal conditions for the cascade reaction were determined to be a reaction temperature of 35°C, pH 8.0 and sucrose concentration 600 mM (Figures 4C–E). Notably, higher concentrations of sucrose led to decreased yields of Reb M2. These optimized conditions were employed for a scale-preparation of Reb M2. Considering the low solubility of Reb D, a strategy of gradually adding the substrate Reb D was employed (Figure 4F). It was found that Reb D underwent glycosylation at a rapid rate, with almost all of it being converted to Reb M2 within 5 h. Subsequently, as a result of product inhibition and enzyme activity loss, the production rate of Reb M2 gradually decreased. Ultimately, after 8 h of stepwise addition of Reb D, a total of 23.08 mM (29.79 mg/mL) of Reb M2 was produced, achieving a yield of 92% from the initial concentration of 25 mM (28.23 mg/mL) Reb D in the optimized cascade reaction system.

4 Conclusion

In summary, the selective synthesis of a new SG, Reb M2, was achieved using the glycosyltransferase UGT94D1. Through structure-guided evolution, a mutant form of UGT94D1 with significantly enhanced catalytic activity was created; this mutant, UGT94D1-F119I/D188P exhibited a 6.33-fold improvement in catalytic activity as compared to UGT94D1. Molecular dynamics simulations demonstrated that the enhancement of its activity is attributed to the reduction in the distance between the substrate Reb D and the catalytic residues, as well as the increase in the Op-C1p-O6 angle, which favors the deprotonation of the hydroxyl group on Reb D and the formation of the glycosidic bond. By coupling this mutant with sucrose synthase *At*SuSy in a cascade reaction system, efficient production of Reb M2 was realized, and 23.08 mM (29.79 mg/mL) of Reb M2 was generated, representing a yield of 92%. Therefore, this study provides an efficient method for the selective synthesis of Reb M2 to support its applications in various fields.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. The accession numbers mentioned in this article include https://www.ncbi.nlm.nih.gov/protein/XP_011076907.1 and https://www.ncbi.nlm.nih.gov/protein/NP_001031915.

Ethics statement

Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

LY: Data curation, Formal Analysis, Investigation, Methodology, Writing–original draft. MY: Data curation, Formal Analysis, Investigation, Methodology, Writing–original draft. ZD: Software, Writing–original draft. XH: Software, Writing–original draft. XZ: Methodology, Writing–original draft. QP: Investigation, Methodology, Writing–original draft. YR: Conceptualization, Funding acquisition, Writing–review and editing. J-SS: Conceptualization, Funding acquisition, Supervision, Writing–review and editing. YZ: Conceptualization, Funding acquisition, Supervision, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key R&D Program of China (2018YFA0901700) and the Natural Science Foundation of Jiangsu Province (BK20202002).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2024.1334427/full#supplementary-material

References

Ahmed, J., Preissner, S., Dunkel, M., Worth, C. L., Eckert, A., and Preissner, R. (2011). SuperSweet-a resource on natural and artificial sweetening agents. *Nucleic Acids Res.* 39, D377–D382. doi:10.1093/nar/gkq917

Bhardwaj, V., Singh, R., Singh, P., Purohit, R., and Kumar, S. (2020). Elimination of bitter-off taste of stevioside through structure modification and computational interventions. *J. Theor. Biol.* 486, 110094. doi:10.1016/j.jtbi.2019.110094

Brandt, W., Schulze, E., Liberman-Aloni, R., Bartelt, R., Pienkny, S., Carmeli-Weissberg, M., et al. (2021). Structural modeling of two plant UDP-dependent sugar-sugar glycosyltransferases reveals a conserved glutamic acid residue that is a hallmark for sugar acceptor recognition. *J. Struct. Biol.* 213, 107777. doi:10.1016/j.jsb.2021.107777

Breton, C., Snajdrova, L., Jeanneau, C., Koca, J., and Imberty, A. (2006). Structures and mechanisms of glycosyltransferases. *Glycobiology* 16, 29r–37r. doi:10.1093/glycob/cwj016

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* 69, 1030–1042. doi:10.1111/j.1365-313X.2011.04853.x

Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, et al. (2020). *Amber 2020*. San Francisco, CA: University of California.

Ceunen, S., and Geuns, J. M. C. (2013). Steviol glycosides: chemical diversity, metabolism, and function. *J. Nat. Prod.* 76, 1201–1228. doi:10.1021/np400203b

DuBois, G. E., and Prakash, I. (2012). Non-caloric sweeteners, sweetness modulators, and sweetener enhancers. *Annu. Rev. Food Sci. Technol.* 3, 353–380. doi:10.1146/annurev-food-022811-101236

Gantait, S., Das, A., and Mandal, N. (2015). Stevia: a comprehensive review on ethnopharmacological properties and *in vitro* regeneration. *Sugar Tech.* 17, 95–106. doi:10.1007/s12355-014-0316-3

Gerwig, G. J., te Poele, E. M., Dijkhuizen, L., and Kamerling, J. P. (2016). Stevia glycosides: chemical and enzymatic modifications of their carbohydrate moieties to improve the sweet-tasting quality. *Adv. Carbohydr. Chem. Biochem.* 73, 1–72. doi:10.1016/bs.accb.2016.05.001

Guo, B. D., Deng, Z. W., Meng, F., Wang, Q. F., Zhang, Y., Yuan, Z. B., et al. (2022). Enhancement of rebaudioside M production by structure-guided engineering of glycosyltransferase UGT76G1. *J. Agric. Food Chem.* 70, 5088–5094. doi:10.1021/acs.jafc.2c01209

Hellfritsch, C., Brockhoff, A., Stahler, F., Meyerhof, W., and Hofmnann, T. (2012). Human psychometric and taste receptor responses to steviol glycosides. *J. Agric. Food Chem.* 60, 6782–6793. doi:10.1021/jf301297n

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Li, J., Yang, J. G., Mu, S. C., Shang, N., Liu, C., Zhu, Y. M., et al. (2020). Efficient O-glycosylation of triterpenes enabled by protein engineering of plant glycosyltransferase UGT74AC1. *ACS Catal.* 10, 3629–3639. doi:10.1021/acscatal.9b05232

Li, Y., Chen, K. Q., Zhou, F. F., Hao, N., and Ouyang, P. K. (2016). *A method for the production of Rebaudioside M2 through the enzymatic catalysis of a recombinant bacterium*. CN201611142851.5.

Libik-Konieczny, M., Capecka, E., Tuleja, M., and Konieczny, R. (2021). Synthesis and production of steviol glycosides: recent research trends and perspectives. *Appl. Microbiol. Biotechnol.* 105, 3883–3900. doi:10.1007/s00253-021-11306-x

Lin, M., Wang, F., and Zhu, Y. S. (2020). Modeled structure-based computational redesign of a glycosyltransferase for the synthesis of rebaudioside D from rebaudioside A. *Biochem. Eng. J.* 159, 107626. doi:10.1016/j.bej.2020.107626

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi:10.1093/nar/gkt1178

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255

Nidetzky, B., Gutmann, A., and Zhong, C. (2018). Leloir glycosyltransferases as biocatalysts for chemical production. *ACS Catal.* 8, 6283–6300. doi:10.1021/acscatal.8b00710

Noguchi, A., Fukui, Y., Iuchi-Okada, A., Kakutani, S., Satake, H., Iwashita, T., et al. (2008). Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. *Plant J.* 54, 415–427. doi:10.1111/j.1365-313X.2008.03428.x

Ohtani, K., Aikawa, Y., Kasai, R., Chou, W. H., Yamasaki, K., and Tanaka, O. (1992). Minor diterpene glycosides from sweet leaves of rubus-suavissimus. *Phytochemistry*, 31, 1553–1559. doi:10.1016/0031-9422(92)83105-8

Olsson, K., Carlsen, S., Semmler, A., Simon, E., Mikkelsen, M. D., and Moller, B. L. (2016). Microbial production of next-generation stevia sweeteners. *Microb. Cell Factories* 15, 207. doi:10.1186/s12934-016-0609-1

Ono, E., Waki, T., Oikawa, D., Murata, J., Shiraishi, A., Toyonaga, H., et al. (2020). Glycoside-specific glycosyltransferases catalyze regio-selective sequential glucosylations for a sesame lignan, sesaminol triglucoside. *Plant J.* 101, 1221–1233. doi:10.1111/tpj.14586

Paquette, S., Møller, B. L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62, 399–413. doi:10.1016/S0031-9422(02)00558-7

Perrier, J. D., Mihalov, J. J., and Carlson, S. J. (2018). FDA regulatory approach to steviol glycosides. *Food Chem. Toxicol.* 122, 132–142. doi:10.1016/j.fct.2018.09.062

Ping, Q., Yang, L. F., Jiang, J. J., Yuan, J. C., Ai, S., Sun, S. Q., et al. (2022). Efficient synthesis of rebaudioside D2 through UGT94D1-catalyzed regio-selective glycosylation. *Carbohydr. Res.* 522, 108687. doi:10.1016/j.carres.2022.108687

Prakash, I., Bunders, C., Devkota, K. P., Charan, R. D., Ramirez, C., Priedemann, C., et al. (2014a). Isolation and characterization of a novel rebaudioside M isomer from a bioconversion reaction of rebaudioside A and NMR comparison studies of rebaudioside M isolated from Stevia rebaudiana Bertoni and Stevia rebaudiana Morita. *Biomolecules* 4, 374–389. doi:10.3390/biom4020374

Prakash, I., Chaturvedula, V. S., and Markosyan, A. (2014b). Structural characterization of the degradation products of a minor natural sweet diterpene glycoside Rebaudioside M under acidic conditions. *Int. J. Mol. Sci.* 15, 1014–1025. doi:10.3390/ijms15011014

Purkayastha, S., Markosyan, A., Prakash, I., Bhusari, S., Pugh, G., Lynch, B., et al. (2016). Steviol glycosides in purified stevia leaf extract sharing the same metabolic fate. *Regul. Toxicol. Pharmacol.* 77, 125–133. doi:10.1016/j.yrtph.2016.02.015

Rahimi, S., Kim, J., Mijakovic, I., Jung, K. H., Choi, G., Kim, S. C., et al. (2019). Triterpenoid-biosynthetic UDP-glycosyltransferases from plants. *Biotechnol. Adv.* 37, 107394. doi:10.1016/j.biotechadv.2019.04.016

Schmölzer, K., Gutmann, A., Diricks, M., Desmet, T., and Nidetzky, B. (2016). Sucrose synthase: a unique glycosyltransferase for biocatalytic glycosylation process development. *Biotechnol. Adv.* 34, 88–111. doi:10.1016/j.biotechadv.2015.11.003

te Poele, E. M., Devlamynck, T., Jager, M., Gerwig, G. J., Van de Walle, D., Dewettinck, K., et al. (2018). Glucansucrase (mutant) enzymes from Lactobacillus reuteri 180 efficiently transglucosylate Stevia component rebaudioside A, resulting in a superior taste. *Sci. Rep.* 8, 1516. doi:10.1038/s41598-018-19622-5

Teze, D., Coines, J., Fredslund, F., Dubey, K. D., Bidart, G. N., Adams, P. D., et al. (2021). O-/N-/S-Specificity in glycosyltransferase catalysis: from mechanistic understanding to engineering. *ACS Catal.* 11, 1810–1815. doi:10.1021/acscatal.0c04171

Tian, X. Y., Zhong, F., and Xia, Y. X. (2022). Dynamic characteristics of sweetness and bitterness and their correlation with chemical structures for six steviol glycosides. *Food Res. Int.* 151, 110848. doi:10.1016/j.foodres.2021.110848

Wang, J. M., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260. doi:10.1016/j.jmgm.2005.12.005

Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.*, 25, 1157–1174. doi:10.1002/jcc.20035

Wen, C., Huang, W., Zhu, X. L., Li, X. S., Zhang, F., and Jiang, R. W. (2018). UGT74AN1, a permissive glycosyltransferase from *Asclepias curassavica* for the regiospecific steroid 3-O-glycosylation. *Org. Lett.* 20, 534–537. doi:10.1021/acs.orglett.7b03619

Yang, S., Hou, X. D., Deng, Z. W., Yang, L. F., Ping, Q., Yuan, Z. B., et al. (2023). Improving the thermostability of glycosyltransferase YojK by targeting mutagenesis for highly efficient biosynthesis of rebaudioside D. *Mol. Catal.* 535, 112898. doi:10.1016/j.mcat.2022.112898

Yonekura-Sakakibara, K., and Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* 66, 182–193. doi:10.1111/j.1365-313X.2011.04493.x

Zhang, P., Zhang, Z., Zhang, L. J., Wang, J. J., and Wu, C. S. (2020). Glycosyltransferase GT1 family: phylogenetic distribution, substrates coverage, and representative structural features. *Comput. Struct. Biotechnol. J.* 18, 1383–1390. doi:10.1016/j.csbj.2020.06.003

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Review
PUBLISHED 11 October 2024
DOI 10.3389/fbioe.2024.1324757

OPEN ACCESS

EDITED BY
Dipesh Dhakal,
University of Florida, United States

REVIEWED BY
Tianwen Wang,
Anhui Polytechnic University, China
Richard Kelwick,
Imperial College London, United Kingdom

*CORRESPONDENCE
Hirohito Yamazaki,
hirohitoyamazaki@vos.nagaokaut.ac.jp

RECEIVED 20 October 2023
ACCEPTED 25 September 2024
PUBLISHED 11 October 2024

CITATION
Yamazaki H, Sugawara R and Takayama Y (2024) Development of label-free light-controlled gene expression technologies using mid-IR and terahertz light.
Front. Bioeng. Biotechnol. 12:1324757.
doi: 10.3389/fbioe.2024.1324757

Development of label-free light-controlled gene expression technologies using mid-IR and terahertz light

Hirohito Yamazaki[1,2]*, Ryusei Sugawara[2] and Yurito Takayama[2]

[1]Top Runner Incubation Center for Academia-Industry Fusion, Nagaoka University of Technology, Nagaoka, Japan, [2]Department of Mechanical Engineering, Nagaoka University of Technology, Nagaoka, Japan

Gene expression is a fundamental process that regulates diverse biological activities across all life stages. Given its vital role, there is an urgent need to develop innovative methodologies to effectively control gene expression. Light-controlled gene expression is considered a favorable approach because of its ability to provide precise spatiotemporal control. However, current light-controlled technologies rely on photosensitive molecular tags, making their practical use challenging. In this study, we review current technologies for light-controlled gene expression and propose the development of label-free light-controlled technologies using mid-infrared (mid-IR) and terahertz light.

KEYWORDS

light-regulated technology, mid-infrared light, gene expression, label-free, photocage, photoswitch, terahertz light

1 Introduction

Light-controlled technologies have enabled researchers to control gene expression with spatiotemporal precision, leading to remarkable advancements in fundamental biological research and medical and pharmaceutical applications (Ryu et al., 2021; Jia and Sletten, 2022). This approach is ideal for studying biology because traditional methods that rely on chemical inducers or genetic modifications have inherent limitations in terms of specificity, reversibility, and control of the timing and location of gene expressions (Khalil, 2020; Ziegler et al., 2022). Photosensitive molecules have been modified into nucleotides, peptides, and small molecules to control *in vivo* and *in vitro* gene expression (Hoorens and Szymanski, 2018). Classically, ultraviolet (UV) light has been the main light source for activating photosensitive molecules because of its high energy and wide range. However, UV light exhibits shallow penetration in tissues and damages biological molecules, making it less ideal for biomedical applications (Sharma and Friedman, 2021). The recent development of visible light-reactive molecules has led to improvements in biocompatibility and expanded applications for cells (Dong et al., 2015; Weinstain et al., 2020). Moreover, much weaker light, such as near-infrared light (NIR), is promising for controlling biological activities due to less damage of biological molecules and better light penetration in tissues (Feng et al., 2021; Chen et al., 2020).

Although light-controlled gene expression-modulating technologies have become favorable for the regulation and study of biological systems, there are challenges. One of the challenges in conventional approaches is the requirement for the chemical conjugation of photosensitive molecules to target biological molecules. Specifically,

researchers need to examine the target biological molecule structure and select suitable photosensitive molecules for the target process, as well as design a protocol for photosensitive molecule attachment that can yield a sufficient amount of the high-purity product. Another challenge is that the addition of photosensitive molecules for regulating gene expression using cell-free systems or living cells may interfere with enzyme activity, necessitating system optimization. For example, an appropriate laser power has to be chosen because a weak power may not trigger a photoreaction and a strong power can damage biological molecules, potentially leading to undesired photocatalysis. Therefore, optimization of parameters such as the chemical components and incubation conditions is necessary.

The main requirements of light-controlled gene expression within a completely natural and untreated system are a chemical component that can uniquely absorb the excitation light and that the light itself does not damage biological molecules. Considering these, a wavelength in the range of mid-infrared (mid-IR) to terahertz would be desirable because this range of light is absorbed in molecular vibration modes, which are specific to the molecules themselves. In this study, to aid the development of label-free light-controlled gene expression regulation systems, we review the potential of using mid-IR–terahertz wavelengths for regulating gene expression for fundamental research and practical applications. First, we discuss how light can interact with materials as the underlying concept in light-regulated technologies and introduce the common photosensitive molecule modification approaches, including photocages and photoswitches. Second, we outline the use of visible light and the NIR region for activating photosensitive DNA/RNA, proteins, and small molecules for regulating gene expression and discuss the challenges of the current light-controlled technologies, not at the life level, but at the basic conceptual level. Finally, we introduce the recently developed light-controlled technology utilizing mid-IR and terahertz wavelengths and address the challenges of this approach.

2 Photosensitive molecules: photocages and photoswitches

2.1 Role of light and photosensitive molecules in gene expression regulation

Light is an electromagnetic wave that can interact with the molecules in the medium that it traverses (Gutzler et al., 2021). Light can be specifically made to interact with photoactive molecules to modulate gene expression (Hoorens and Szymanski, 2018). Upon the absorption of light, these molecules undergo specific photochemical reactions including cleavage and crosslinking, leading to chemical modifications, bond formation, and changes in their molecular structure (Tavakoli and Min, 2022). Light-controlled technologies provide a non-invasive approach to study and manipulate gene expression with high precision and selectivity. In this approach, a wide range of light can be applied, from UV to NIR and beyond. The photochemical reaction mechanism depends on the wavelength of light: UV and visible light (200–900 nm) leads to electronic transitions, while mid-IR/terahertz light (2.5 μm–3 mm) results in vibrational transitions, affecting intramolecular/intermolecular vibrations (Figure 1A). Two common strategies have been developed using photosensitive molecules: photocages and photoswitches. The following sections describe these strategies. For the ideal design of photosensitive molecules, the following characteristics are required (Singh et al., 2021): first, photosensitive molecules should show high light absorption at wavelengths that are not absorbed by or are not damaging to other biological molecules. In addition, the photoreaction should show high efficiency. Second, photosensitive molecules should have low intrinsic activity and should be stable in the reaction medium before light irradiation. Third, the by-products of the photoreaction should be transparent to the photoreactive trigger light to suppress other competitive reactions that may interfere with the designed photoreaction, and they should not react with molecules in the reaction medium. Furthermore, some photochemical reactions undergo several steps that involve excited/ground-state intermediates; therefore, the detailed mechanism of the photochemical reactions should be studied to understand the correct light irradiation.

Gene expression involves various biological processes, including transcription, where the DNA sequence of a gene is copied into complementary RNA, and translation, where messenger RNA (mRNA) serves as a template to assemble amino acids into polypeptide chains to form proteins. Proteins act as enzymes that modulate cellular metabolism and chemical reactions. Although light itself does not play an active role in this process, it can, with the assistance of photosensitive molecules, regulate a wide range of biological processes, including enzymatic activity. Figure 1B shows an example of this concept using photosensitive molecules of DNA, RNA, and proteins for light-controlled gene expression, which is further reviewed in Section 3.

2.2 Photocage

A photocage is a photoremovable protective molecule that blocks the bioactivity of the biological molecule to which it is conjugated. Photocage molecules absorb specific wavelengths of light, typically in the UV or visible range, and undergo a photoreaction that results in their removal or cleavage. This process releases the bioactive molecules and initiates their bioactivity. Popular photocages include boron–dipyrromethene (BODIPY), heptamethine cyanine (Cy7), and coumarin.

BODIPY is a fluorescence dye commonly used in bioimaging for biological studies and medical applications because it has low biotoxicity, is highly stable in various medium conditions, and has outstanding optical properties, such as high quantum yield, high absorption coefficient, narrow fluorescence spectrum, and long fluorescence lifetime (Singh et al., 2021). Additionally, structural modifications of BODIPYs can tune the absorption band to include the NIR region (Liu et al., 2019). However, the photolysis of some BODIPYs occasionally results in reduced photoreactivity (Singh et al., 2021), and real-time monitoring of the released active biomolecules is not feasible (Paul et al., 2019). Figure 2A shows an example of the photochemical reaction of BODIPY, where light irradiation initiates the cleavage of the cargo, resulting in a carbocation intermediate with a solvent-assisted nucleophilic attack (Peterson et al., 2018). A variety of BODIPY derivatives have been discussed in previous reviews (Singh et al., 2021; Cheng et al., 2023; Bobrov et al., 2021). Various types of caging groups can

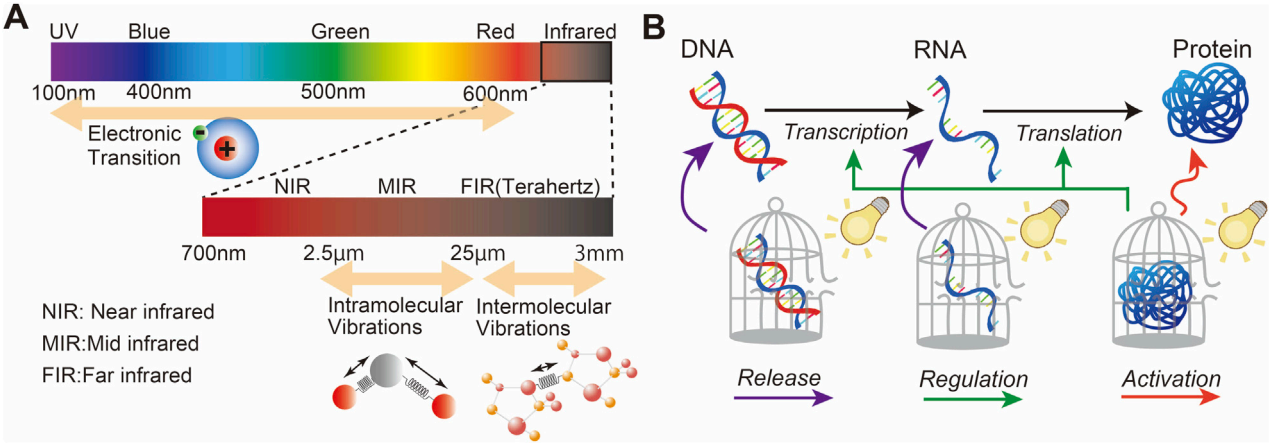


FIGURE 1
(A) Light wavelength for the excitation of electron transition, intramolecular vibrations, and intermolecular vibrations. **(B)** Overview of the basic biological process underlying gene expression and possible pathways for light-controlled gene expression.

be used to create BODIPY photocages depending on the specific application and desired photorelease mechanisms. Some caging groups that are commonly attached to the BODIPY core include carboxylic acid, amine, alcohol, thiol, halide, xanthane, and phenol (Vohradská et al., 2018; Dockalova et al., 2020; Sitkowska et al., 2018).

Cy7 is another dye that is frequently used in bioimaging because it binds to biological molecules, including DNA, RNA, and proteins (Luciano et al., 2019; Rozovsky et al., 2019; Šmidlehner et al., 2018). The advantages of Cy7 are its low toxicity, low background absorbance interference, and high NIR light absorption (Yamamoto et al., 2019). Some Cy7 derivatives have been engineered to function *in vivo* by developing NIR-absorbing photocage molecules that are less harmful to tissues (Alachouzos et al., 2022; Janeková et al., 2022). The general photoreaction of the Cy7 photocage is shown in Figure 2B, where photooxidative cleavage at the C-C bond and hydrolysis result in the release of the caged molecule (Weinstain et al., 2020; Gorka et al., 2014).

Coumarins are widely used because they can be easily synthesized and rapidly released from substrates. As an example, the reaction of a photoactivatable phosphate-releasing group using 7-methoxycoumarin derivatives is shown in Figure 2C (Tavakoli and Min, 2022; Givens and Matuszewski, 1984). Using structural modifications, the photophysical properties of coumarins such as quantum yield and aqueous solubility can be improved (Bardhan and Deiters, 2019). The advantages of coumarin are its high absorption coefficient, high photoresponse efficiency, fast photolysis kinetics, and suitability for engineering two-photon experiments. Typically, coumarins can be connected to caging groups such as carbonates, alkoxides, carbamates, thiols, sulfonates, azides, halides, phosphates, and carboxylates (Weinstain et al., 2020).

2.3 Photoswitch

Photoswitches are photochromic molecules that undergo reversible conversion between two or more stable states under light irradiation. This photochemical isomerization process involves structural changes that result in distinct differences in the UV–visible absorption spectra of the isomeric states, making them photochromic. Mostly, photoswitch molecules show positive photochromism, with the generated species showing a higher maximum absorption (λ_{max}) than the initial state (Bouas-Laurent and Dürr, 2001). However, when the initial molecules undergo bleaching upon photoisomerization, the photoswitch molecules show negative photochromism. Examples of such photoswitch molecules include azobenzenes, stilbenes, and spiropyrans.

Azobenzene is a well-studied photoswitch that exists in two distinct isomeric states: *trans* and *cis* (Axelrod et al., 2022; Volarić et al., 2022). They are widely used in biological applications because of their ease of synthesis, high quantum yield, low photobleaching, high photostationarity, and fast isomerization (Fuentes et al., 2020). Notably, *trans*-azobenzene is approximately 10 kcal/mol more stable than *cis*-azobenzene. The UV–visible absorption of *trans*-azobenzene exhibits two maximum peaks: a strong absorption peak near 320 nm due to a π–π* transition and a weaker peak absorption near 440 nm related to an n–π* transition (Gelabert et al., 2023; Kuntze et al., 2021). *cis*-azobenzene has a stronger absorption band of the π–π* transition near 400 nm and two shorter absorption bands at 280 and 250 nm (Dias et al., 1992). Absorption near 320 nm induces the rotation of the nitrogen double bond, leading to the formation of the non-polar *cis* isomer, whereas absorption at 440 nm is associated with *trans*-to-*cis* isomerization through various pathways (Giles et al., 2021). Figure 3A shows the structures of *trans*- and *cis*-azobenzene (Cattaneo and Persico, 1999). Thermal stimulation or visible light irradiation can induce a switch from the *cis* to *trans* isomer, which requires ~95 kJ mol^{-1} of activation energy (Garcia-Amorós et al., 2018).

Stilbene is another photoswitch molecule that undergoes *trans–cis* isomerization (Villarón et al., 2021). It is a hydrocarbon comprising two phenyl rings connected by an ethylene (–CH=CH–) bridge that forms a central double bond. The photoisomerization mechanism of stilbene has been reported to be slightly different from that of azobenzene. The main reason is that the excited state of the *trans*-stilbene, initiated through the π–π* transition, is metastable

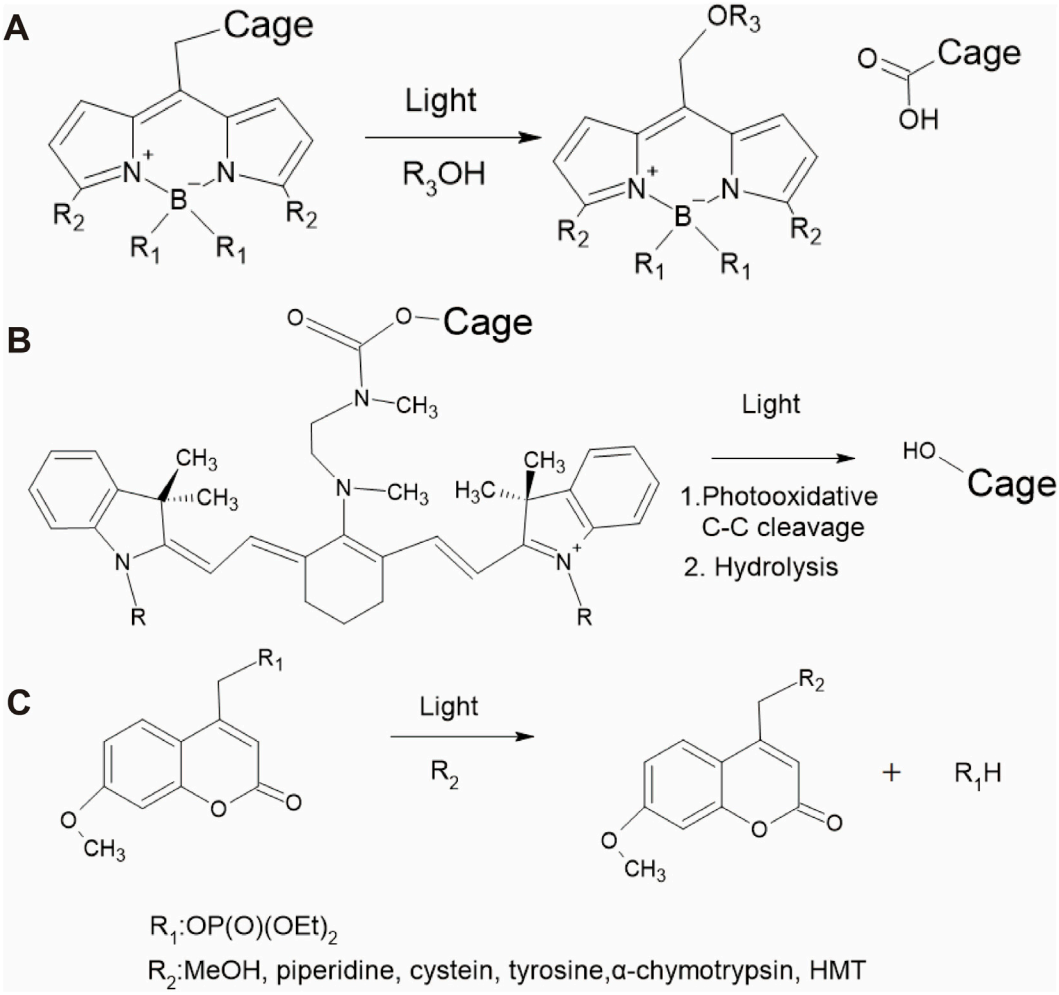


FIGURE 2
Reaction pathways of different photocage molecules. **(A)** BODIPY, **(B)** heptamethine cyanine (Cy7), and **(C)** 7-methoxycoumarin derivatives.

(Han et al., 2002). Isomerization of unsubstituted stilbene occurs under irradiation near 300 nm, resulting in the formation of the *cis* isomer. An example of this transition is shown in Figure 3B. The thermal *cis*-to-*trans* reisomerization requires 41–46 kcal/mol, which cannot be achieved at room temperature. In terms of biological applications, the disadvantage of stilbene is its irreversible cyclization and oxidization to the *cis* isomer.

Spiropyran undergoes a photochromic reaction under UV irradiation, which induces heterocyclic cleavage at the C-O bond, leading to the formation of a zwitterionic structure, as shown in Figure 3C, an example of the transition pathway (Kortekaas and Browne, 2019). This isomerization process produces a significant polarity shift (8–15D) (Mukherjee et al., 2022; Shiraishi et al., 2023). More importantly, this process is reversible upon thermal stimulation and photochemical activation under visible light irradiation (>460 nm). The importance of spiropyran is that photoisomerization causes a significant change in polarity, which influences its hydrophilicity/hydrophobicity (Wang et al., 2021). Spiropyran has been reported to strongly interact with certain biological molecules (Keyvan Rad et al., 2022; Ali et al., 2019).

2.4 Challenges in using photosensitive molecules

There are many candidates for photosensitive molecules that can be attached to target molecules. However, the challenge lies in the need to use specific attachment reactions, which limits the number of available pairs of photosensitive molecules and target molecules, or to implement tag modification. For example, prior to attachment, many photoswitch molecules are pre-labeled with protein tags such as SNAP tags, which can fuse with any target protein (Leng et al., 2017). Additionally, the modification process can be quite time-consuming. As an example, we analyzed the duration of photosensitive molecular modification based on protocol journal papers by picking up all related processes, as shown in Figure 4 (Becker et al., 2018; Cheong et al., 2024; Coelho et al., 2020; Conic et al., 2018; Erlendsson et al., 2019; Fields et al., 2019; Forero-Quintero et al., 2022; Gáspár et al., 2018; Horii et al., 2017; Kajimoto and Nakamura, 2018; Klena et al., 2023; Koch et al., 2021; Luo et al., 2023; Mai-Morente et al., 2021; Mazzucco et al., 2020; Mu et al., 2019; Munson and Ganley, 2016; Paweletz et al., 2022; Rawat and

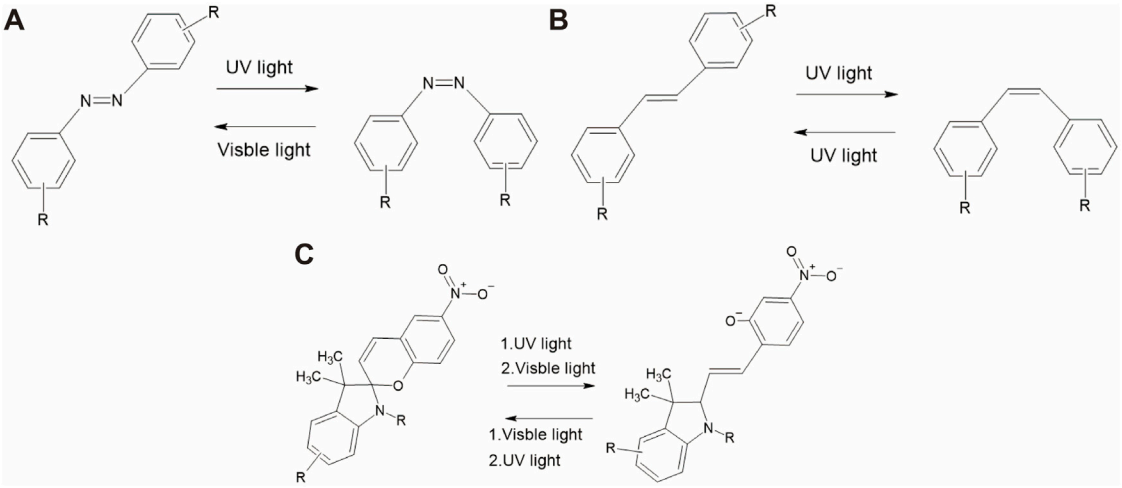


FIGURE 3
Reaction pathways of different photoswitch molecules. **(A)** Azobenzene, **(B)** stilbene, and **(C)** spiropyran.

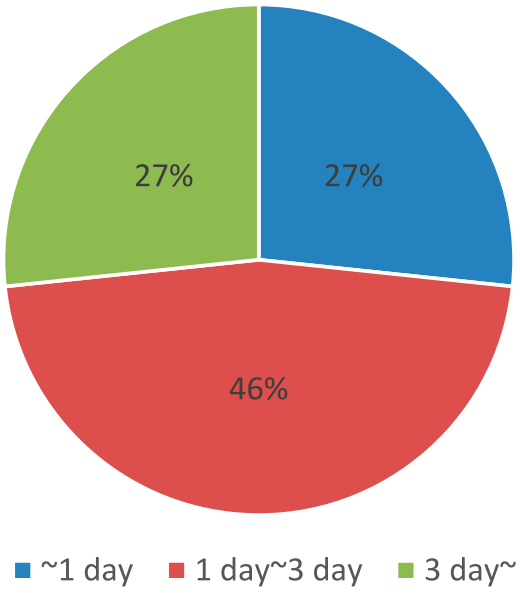


FIGURE 4
Summary of time required for photosensitive molecular modification, obtained from references (Becker et al., 2018; Cheong et al., 2024; Coelho et al., 2020; Conic et al., 2018; Erlendsson et al., 2019; Fields et al., 2019; Forero-Quintero et al., 2022; Gáspár et al., 2018; Horii et al., 2017; Kajimoto and Nakamura, 2018; Klena et al., 2023; Koch et al., 2021; Luo et al., 2023; Mai-Morente et al., 2021; Mazzucco et al., 2020; Mu et al., 2019; Munson and Ganley, 2016; Paweletz et al., 2022; Rawat and Sharma, 2024; Remenyi et al., 2019; Roberts et al., 2018; Sharma et al., 2019; Tanimizu, 2015; Teng et al., 2019; Tronnet and Oswald, 2018; Udupa et al., 2022; Wirth et al., 2020; Xu and Liu, 2019; York and Milush, 2015).

and Milush, 2015). The results indicate that the entire process, including preparation and related procedures, typically takes at least a day, with some cases requiring even more time.

Additionally, light absorption by photosensitive molecules is limited (Welleman et al., 2020). For practical use, sufficient light penetration through the cells and tissues without causing damage is ideal. Nevertheless, most photosensitive molecules have not yet been engineered to operate at longer wavelengths, such as NIR light (Jia and Sletten, 2022).

Furthermore, some photosensitive molecules show inherent poor water solubility, which might cause biomolecular aggregation and result in interference with gene expression (Deng et al., 2021; Poryvai et al., 2022; Berdnikova, 2019). For instance, traditional BODIPY dyes dissolve only in organic solvents, which is why, to date, tremendous efforts have been made to improve their water solubility by introducing hydrophilic groups such as phosphonates, quaternary ammonium salts, and sulfonates (Koh et al., 2019; Mao et al., 2020; Zhou et al., 2020).

Finally, the development of a new photosensitive molecule design is likely serendipitous (Deng et al., 2021; Dcona et al., 2020). Therefore, laborious efforts are often required to develop photosensitive molecules that are suitable for any untried gene expression system of interest (Buglioni et al., 2022; Roßmann et al., 2023).

3 Current light-controlled gene expression technology

In this section, we review how light irradiation can be used to control gene expression. As shown in Figure 1B, light irradiation can activate photosensitive DNA/RNA, proteins, and small molecules that influence gene expression; these categories are discussed in the following sections.

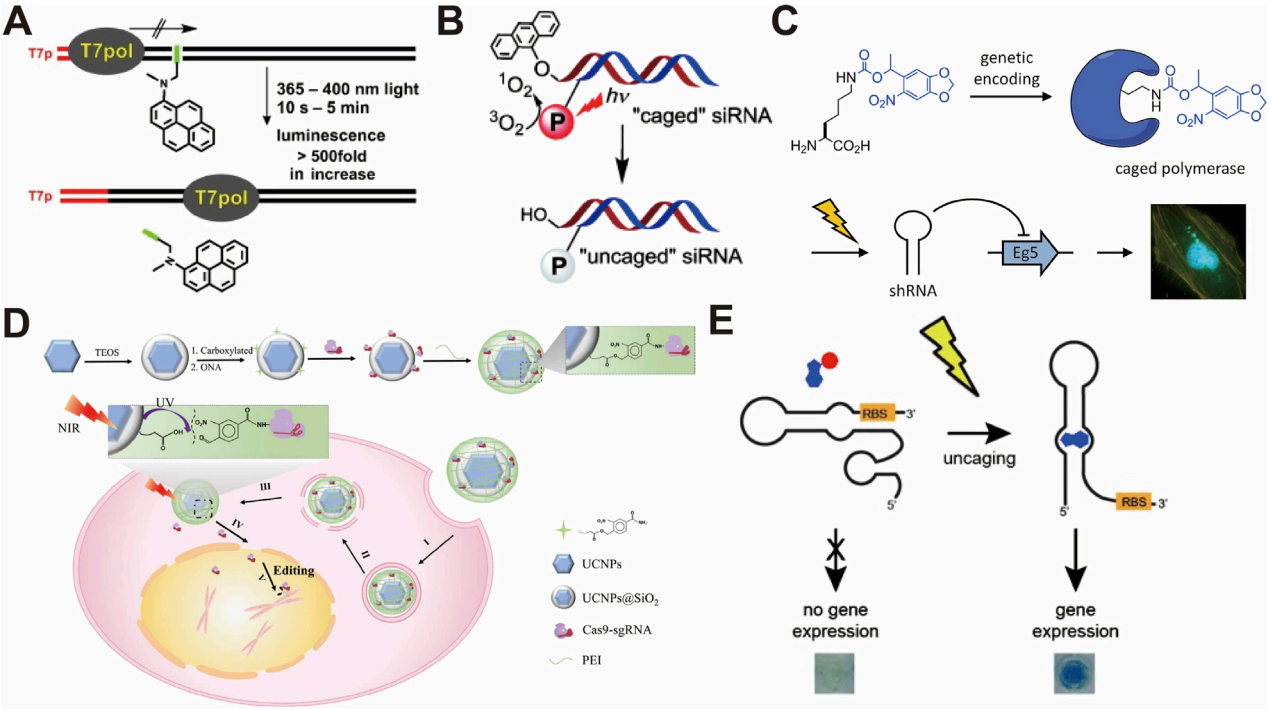


FIGURE 5
(A) Selective activation of psoralen-cross-linked DNA with UV and blue light (Stafforst and Stadler, 2013). **(B)** Photoactivation of siRNAs with red light for a non-toxic cellular approach (Meyer and Mokhir, 2014). **(C)** Spatiotemporal control of gene expression using photocaged T7 RNA polymerase (Hemphill et al., 2013). **(D)** Light-responsive nanocarrier for precise CRISPR-Cas9-mediated gene editing (Pan et al., 2019). **(E)** Spatiotemporal control of gene expression using riboswitches and photocaged ligands (Walsh et al., 2014). All figures have been adapted with permission from John Wiley and Sons Copyright (2013) and (2014), American Chemical Society Copyright (2013), and Springer Nature Copyright (2019).

3.1 DNA/RNA-based light-controlled gene expression

The photocontrol of gene expression using nucleic acids is typically facilitated by photosensitive molecule-modified DNA or RNA, as shown in Figure 1B. There are two main strategies for this approach: light-induced gene activation and light-induced knockdown using photocage molecule-modified oligonucleotides.

Active light control of transcription and translation has been demonstrated by modifying oligonucleotides with photocage molecules. The modifications in the oligonucleotides can be introduced into the phosphate backbone, Watson–Crick face, or nucleotide base. For example, gene expression in cells and synthetic cells has been controlled using photolabile groups on the phosphate backbone of the DNA and RNA modified with photosensitive molecules, including coumarin (Kaufmann et al., 2023; Ando et al., 2001), 2-nitroveratryl bromide (Hartmann and Booth, 2023), and thioether-enol phosphate (Feng et al., 2017; Wang et al., 2016). Other than at the DNA backbone, psoralen cross-linking at the Watson–Crick face of the DNA promoter impedes the unwinding of the double helix, resulting in the blocking of the transcription process (Figure 5A) (Stafforst and Stadler, 2013; Stadler and Stafforst, 2014). Photosensitive molecules can also be introduced at nucleotide bases. Nucleotides modified with photosensitive molecules, such as benzophenones (Anhäuser et al., 2020), diethylaminocoumarin (Menge and Heckel, 2011), 2-nitrobenzyl bromide (Chakrapani et al., 2020), and 6-nitropiperonyloxymethyl group (Lee et al., 2021), have been integrated into DNA or RNA templates for light-controlled transcription and translation. When photoswitch molecules are modified into DNA/RNA, they can reversibly regulate the transcription and translation processes based on light irradiation. For instance, photoswitch molecules, including azobenzene (Tian et al., 2016; Xing et al., 2015; Dudek et al., 2018), stilbenes (O'Hagan et al., 2019; O'Hagan et al., 2020), arylstilbazolium (Czerwinska and Juskowiak, 2012), and 8-pyrenylvinyl deoxyguanosine (Ogasawara, 2018), which contain G-quadruplexes in a hyperstable state, can effectively inhibit transcription and translation processes. Light irradiation induces an unstable state, allowing activity, and additional light irradiation in the unstable state recovers hyperstable G-quadruplexes.

One approach to knock down gene expression is to use photocage-modified antisense oligonucleotides (ASOs), which are short single-stranded nucleotides that specifically bind to target DNA or RNA to stop transcription and translation. Modification of ASOs with 2-nitrobenzyl-caged thymidines has been shown to cause the photoactive knockdown of cancer-related genes (Govan et al., 2013; Tang et al., 2007; Sakamoto et al., 2014) and developmental genes (Tallafuss et al., 2012; Deiters et al., 2010). The inclusion of 2-nitrobenzyl can be achieved in the base-pairing region, nucleobase, and backbone of the complementary strand. Circularized ASOs can also be formed with photocleavable linkers,

such as coumarin- (Yamazoe et al., 2014), 2-nitrobenzyl- (Yang et al., 2018), quinoline- (O'Connor et al., 2019), and Ru-BEP (Griepenburg et al., 2015), which allows the control of gene expression using light irradiation. Another approach is to use small interfering RNAs (siRNAs), which are approximately 20 base pairs of dsRNAs, that specifically suppress gene expression by destroying mRNA. By attaching 2-nitrobenzyl on the phosphate backbone and 5′ or 3′phosphate termini of siRNAs, light-controlled knockdown could be achieved in zebrafish (Blidner et al., 2008), mammalian cells (Zhang L. et al., 2018), and HeLa cells (Yu et al., 2018; Kala et al., 2014). When a 9-alkoxyanthracenyl fragment as an O_2-sensitive moiety is incorporated in photocage molecules, the selectivity of the activation wavelength can be tuned by attaching photosensitizers to the 3′-terminus (Figure 5B) (Meyer and Mokhir, 2014). The use of light-responsive nanoparticles (NPs) is also a common approach for spatiotemporally manipulating siRNA functions. Typically, irradiation with NIR light triggers the release of thiol-modified siRNAs attached to gold NPs, resulting in the NIR-controlled knockdown of gene expression in cells (Huang et al., 2015; Riley et al., 2018; Braun et al., 2009). In addition, siRNA/mPEG-b-P(APNBMA) NPs containing o-nitrobenzyl moieties have been engineered to initiate siRNA release post-UV light irradiation (Foster et al., 2015).

3.2 Protein-based light-controlled gene expression

Photocontrol of proteins can affect the regulation of DNA/RNA synthesis or activity and the direct activation of protein function, as shown in Figure 1B, using photosensitive molecule-modified proteins.

The specific incorporation of amino acids with photosensitive molecules is required to generate photosensitive molecule-modified proteins in cells (Arbely et al., 2012; Chou and Deiters, 2011). One approach to demonstrate this incorporation is to select an amber stop codon as the insertion site for amino acids with photosensitive molecules (Wang et al., 2012; Neumann, 2012). Examples of this approach include the modification of a nitrobenzyl group on RNA polymerases to allow light-controlled gene expression in cells, as shown in Figure 5C (Hemphill et al., 2013; Chou et al., 2010). Additionally, other light-activated gene-editing tools, such as Cre recombinases (Edwards et al., 2009) and zinc finger (ZF) nucleases (Chou and Deiters, 2011), have been engineered by modifying o-nitrobenzyl on essential residues in catalytic sites, enabling activation upon light irradiation (Chou and Deiters, 2011; Luo et al., 2016). By fusing the GIGANTEA (GI) protein with the ZF protein, blue light irradiation induces heterodimerization between the GI and the light–oxygen–voltage (LOV) domain, which assists the transcriptional activation domain VP16 in initiating the transcription of the gene of interest (Polstein and Gersbach, 2012). Similarly, when the light-sensitive cryptochrome 2 (CRY2) protein is fused with transcription activator-like effector (TALE) proteins, blue light-induced conformational changes in CRY2 recruit a CIB1–effector domain, which exerts the active control of transcription in the same endogenous genome (Konermann et al., 2013). Furthermore, 2-nitrobenzyl-modified tamoxifen is used in light-controlled Cre systems to effectively regulate gene expression in cells (Inlay et al., 2013). Optical dimerization of the CRY2–CIB system has been widely used to reconstitute the split recombinase of Cre and Flp, which enables gene expression control in cells (Cautereels et al., 2024) and mouse brains (Jung et al., 2019).

The CRISPR-Cas9 system is also a well-known gene editing tool that can delete, replace, or insert any part of the genome sequence and has been extensively explored for the light-activated control of gene expression. Light-activated control of the CRISPR-Cas9 system has been achieved by engineering light-responsive Cas9 (Hemphill et al., 2015), gRNA (Moroz-Omori et al., 2020; Jain et al., 2016), and transcription factors (Shao et al., 2018). An example of an engineered Cas9 nuclease is the incorporation of a 2-nitrobenzyl-modified lysine amino acid, which reversibly affects gene function (Hemphill et al., 2015). As shown in Figure 5D, when Cas9 is bound to lanthanide-doped upconversion NPs by a 2-nitrobenzyl photocage, NIR irradiation produces local UV light, resulting in a photocleavage reaction and the release of the CRISPR-Cas9 system (Pan et al., 2019). Other candidate proteins, such as light-activated phosphorylation (Nguyen et al., 2018), dimeric green fluorescent protein (Zhou et al., 2018), and cyclic diguanylate monophosphate signaling cascades (Shao et al., 2018), have also been reported for use in light-activated CRISPR-Cas9 systems.

3.3 Small-molecule-based light-controlled gene expression

In addition to DNA, RNA, and proteins, small-molecule compounds are also general candidates for the photoreactive control of gene expression. Popular small molecules include nucleotides, peptides, and ligands. Nucleotides are the basic building blocks that constitute DNA and RNA; thus, the photocontrol of nucleotides is a straightforward approach for controlling gene expression. In this regard, [7-(diethylamino) coumarin-4-yl]methyl (DEACM)-modified ATP and 2-nitrobenzyl-modified UTP/GTP have been demonstrated as molecules to photocontrol the transcription reaction *in vitro* (Shao et al., 2018; Pinheiro et al., 2008). The modification of ligands with 2-nitrobenzyl and azobenzene, which effectively interact with DNA/RNA, is a useful approach for controlling gene expression (Walsh et al., 2014; Young et al., 2009; Paul et al., 2021). An example molecule is theophylline, which can specifically bind to the mRNA riboswitch; the photocontrol of gene expression could be achieved by modifying theophylline with 2-nitrobenzyl (Figure 5E) (Walsh et al., 2014). The binding affinity of tamoxifen and cyclophen to the estrogen receptor (ER)-binding domain can be utilized for photoactive regulation by modifying tamoxifen and cyclophen with 2-nitrobenzyl, coumarin, and cyanine derivatives (Zhang W. et al., 2018; Wong et al., 2017; Cruz et al., 2000). The tetracycline (Tet) system in mammalian cells is controlled by nitrobenzyl-modified doxycyclines (Cambridge et al., 2006). Many other small molecules, such as coumarin-modified cAMP-response element-binding protein (CREB) inhibitor (Imoto et al., 2020) and 2-nitrobenzyl-modified ecdysone and nuclear hormones (Link et al., 2004; Lin et al., 2002), have also been engineered to regulate gene expression in cells.

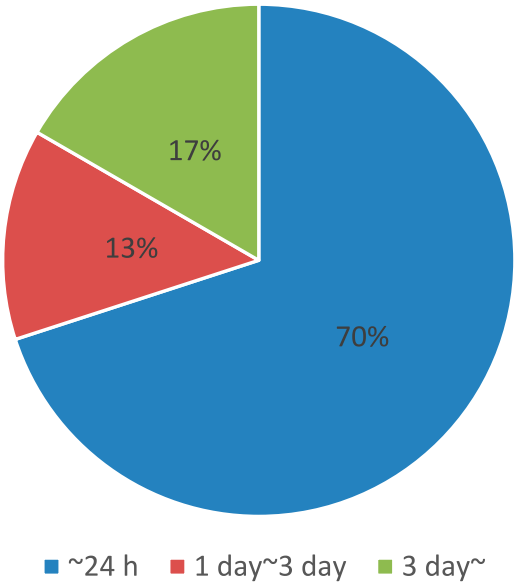


FIGURE 6
Summary of time required for electroporation, obtained from references (Alfastsen et al., 2021; Biase and Schettini, 2024; Chan et al., 2023; Clements et al., 2024; Crestani et al., 2022; Denoth-Lippuner et al., 2022; Egashira et al., 2023; Flumens et al., 2023; Gao et al., 2022; Gesuita et al., 2022; Heinz et al., 2020; Lankford and Hulleman, 2024; Laprie-Sentenac et al., 2023; Lax et al., 2022; Layden et al., 2021; Li et al., 2023; Ling et al., 2022; Meka et al., 2024; Nagahama et al., 2021; Nomura et al., 2020; Sakamoto et al., 2022; Sanketi et al., 2023; Skruber et al., 2021; Turchetto et al., 2020; Wang et al., 2020; Williams and Sauka-Spengler, 2021; Wilson et al., 2020; Wu et al., 2023; Xu et al., 2021; Xu et al., 2020; Zhou et al., 2021).

3.4 Challenges with current light-controlled gene expression approaches

To understand the nature of cells, photosensitive molecules can be used to probe fundamental aspects, such as molecular reactions, molecular kinetics, and live cell dynamics. However, the introduction of photosensitive molecules into cells or cell-free systems might influence molecular activity, thereby biasing the outcomes of the study. For example, exogenous photosensitive molecules in live cells can potentially harm normal cellular activities, including unintended effects on cell behavior, signaling pathways, and cellular homeostasis (Chen et al., 2022). Some photosensitive molecules may exhibit cytotoxic effects that affect cell viability and overall cellular health (Dcona et al., 2020). Therefore, researchers must carefully assess the concentration and exposure time required to minimize cytotoxicity.

As mentioned in Section 2.4, the introduction of photosensitive molecules is also a lengthy process. As shown in Figure 6, which summarizes the total time required for molecular introduction into cells using electroporation based on protocol journal papers by picking up all related processes, the process can be done within a day. However, in certain cases, it can take a significantly longer time (Alfastsen et al., 2021; Biase and Schettini, 2024; Chan et al., 2023; Clements et al., 2024; Crestani et al., 2022; Denoth-Lippuner et al., 2022; Egashira et al., 2023; Flumens et al., 2023; Gao et al., 2022; Gesuita et al., 2022; Heinz et al., 2020; Lankford and Hulleman, 2024; Laprie-Sentenac et al., 2023; Lax et al., 2022; Layden et al., 2021; Li et al., 2023; Ling et al., 2022; Meka et al., 2024; Nagahama et al., 2021; Nomura et al., 2020; Sakamoto et al., 2022; Sanketi et al., 2023; Skruber et al., 2021; Turchetto et al., 2020; Wang et al., 2020; Williams and Sauka-Spengler, 2021; Wilson et al., 2020; Wu et al., 2023; Xu et al., 2021; Xu et al., 2020; Zhou et al., 2021).

Furthermore, achieving high specificity and selectivity for target molecules can be challenging. Ideally, photosensitive molecules should interact only with the intended targets; however, off-target effects may occur, leading to the misinterpretation of the experimental results (Laczi et al., 2022). Moreover, efficient delivery of light-sensitive molecule-tagged biomolecules to specific locations in live cells is required to achieve the desired control of gene expression (Deng et al., 2020).

In addition, light irradiation can induce cellular stress and damage, which is known as phototoxicity. This is particularly relevant in live-cell imaging studies, in which prolonged exposure to light may lead to alterations in cellular behavior (Talone et al., 2021). Different cell types respond differently to photosensitive molecules (Powell et al., 2018). It is important to consider the specific characteristics of the cell type under investigation and validate the applicability of the chosen photosensitive molecule to the cell system of interest. The above discussion highlights that the development of light-regulated translation technologies requires addressing many challenges through enormous efforts, testing, and optimization to improve the effectiveness of the methodologies (Welleman et al., 2020).

4 Use of mid-IR–terahertz light for label-free light-controlled gene expression

Based on the challenges discussed above, we propose that mid-IR and terahertz light can be used as solutions for light-controlled gene expression regulation; this section discusses this proposal in detail.

4.1 Potential for using mid-IR–terahertz light for gene expression regulation

In principle, as described in Section 2.1, mid-IR and terahertz light (2.5 μm–3 mm) has a strong interaction with intramolecular/intermolecular vibrations (López-Lorente et al., 2016; Tonouchi, 2007; Ma et al., 2019; Hoshina et al., 2020). Therefore, mid-IR and terahertz light has been used as light sources for imaging with high special selectivity and sensitivity and molecular detection (Jain et al., 2020). For example, mid-IR photothermal imaging is an emerging technology that can effectively excite the target vibration modes in biological molecules. This results in local property changes, such as the refractive index, molecular volume, and vibration modes, and monitoring these changes allows a high spatial resolution with low photodamage to biological molecules (Samolis et al., 2023; Ishigane et al., 2023).

Since mid-IR and terahertz light can interact with specific molecular vibrations, the irradiation of these light sources can influence biological molecular activity; thus, there is potential to control specific gene expression. The strong light energy at this

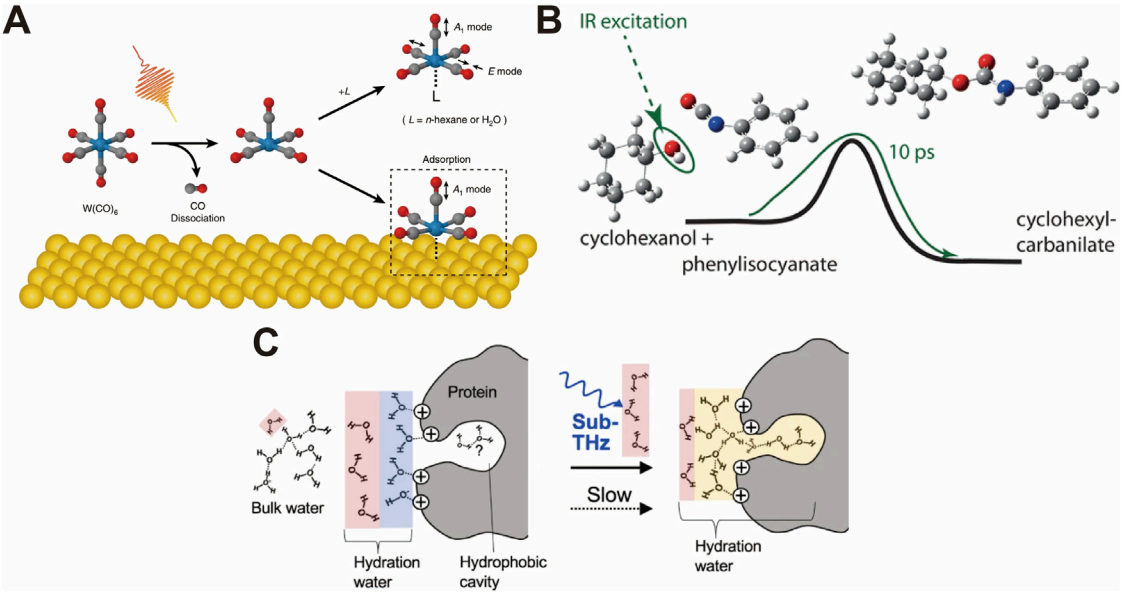


FIGURE 7
(A) Plasmonic enhanced mid-IR light field dissociation of CO of n-hexane (Morichika et al., 2019). **(B)** Selective vibrational excitation-induced ground-state reactions by mid-IR light (Heyne and Kühn, 2019). **(C)** Terahertz light promoting the hydrophobic hydration around the protein (Sugiyama et al., 2023). All figures have been adapted with permission from American Chemical Society Copyright (2019) and Springer Nature Copyright (2019), (2023).

wavelength can influence molecular activity in a label-free manner (Morichika and Ashihara, 2022; Park et al., 2018; Windhorn et al., 2002). However, to the best of our knowledge, mid-IR and terahertz light has not yet been commonly used for the light control of gene expression. One of the reasons for this is the high light absorption of water in this wavelength range, which leads to an increase in ambient temperature and hampers biomolecular activity (Holmstrom and Nesbitt, 2010; Ashwood et al., 2020). A short-pulse laser can break this fundamental barrier because cyclical ultrashort-term heating and cooling suppress the increase in ambient temperature (Toda et al., 2019; Zhang et al., 2019). Interestingly, when analyzing the energy of mid-IR and terahertz light (2.5 μm–3 mm), which corresponds to 7.95×10^{-20} to 6.62×10^{-23} joule/photon, the energy of biological processes, such as hydrolysis of a peptide bond (-1.39×10^{-20} to -2.78×10^{-20} joule/molecule), ATP hydrolysis (-5.21×10^{-20} joule/molecule), and folding energies of RNA secondary structures (-2.78×10^{-20} to -1.11×10^{-19} joule/molecule), is not very different (Jia et al., 2021). In Figure 7A, Morichika et al. (2019) showed that the irradiation of a plasmonic structure using a mid-IR pulse laser enhanced the localized light field, resulting in CO dissociation of n-hexane mediated by vibrational ladder climbing (Morichika et al., 2019). Other plasmonic studies have showed that strong mid-IR light in a plasmon nanocavity perturbs a few-nm-thick shell of water, predicting the mid-IR and terahertz light interaction with the molecular reaction (Chikkaraddy et al., 2022). Even without plasmonic enhancement, pulsed mid-IR light can selectively induce vibrational excitation, allowing the activation energy barriers to be overcome and facilitating ground-state reactions with minimal heat generation in Figure 7B (Stensitzki et al., 2018). This can result in bimolecular alcoholysis reactions, bidirectional tautomerization of thiotropolone, and the formation of urethane and polyurethane (Stensitzki et al., 2018; Heyne and Kühn, 2019; Nunes et al., 2020).

Although understanding the detailed mechanism of mid-IR and terahertz light irradiation on live cells is complicated, Toyama et al. (2022) showed that strong pulsed-light irradiation of *E. coli* cells in the mid-IR absorption region can affect their growth rate and survival. Furthermore, the much weaker energy of terahertz light influences the biological molecular activities. Recently, Sugiyama et al. (2023) discovered that the irradiation of pulsed terahertz light on proteins enhances hydrophobic hydration, leading to an increase in the number of hydrogen bonds at the hydration layer in Figure 7C, suggesting that even terahertz light can influence gene expression (Sugiyama et al., 2023). Although previous reports did not show clear findings, some previous studies have already predicted the influence of protein activity and expression (Sun et al., 2021). For example, Tan et al. found the downregulation of SYN expression in primary hippocampal neurons and PSD95 expression in cortical neurons under 0.16–0.17-THz irradiation, with a positive correlation with the exposure time and laser power (Zhi et al., 2019).

4.2 Challenges associated with mid-IR–terahertz light technology

Although some previous studies have shown the effect of mid-IR and terahertz light at the molecular level and presented some molecular reactions, many challenges remain for the active

utilization of these wavelengths of light. First, the vibrational mode of molecules is influenced by a variety of factors, such as temperature, humidity, and substrates. Thus, an absorption spectrum shift of target biological molecules can occur in the environment, and researchers may need to calibrate the excitation wavelength. Second, as mentioned above, water absorption of mid-IR and terahertz light is quite high ($100–10^4$ cm^{-1}); therefore, limited light penetration might be a critical issue when target samples are a large volume of biological molecules or tissues (Prahl, 2017). Third, instruments and optical elements for the mid-IR and terahertz regions are often expensive and not widely available for commercial purchase, necessitating researchers to develop a setup if cost considerations come into play. Fourth, targeting specific molecules is required to obtain absorption spectrum profiles using Fourier-transform infrared spectroscopy, which requires the preparation of a large volume of purified biological samples.

5 Conclusion

Here, we provide an overview of light-controlled gene expression and propose a label-free light control approach using mid-IR and terahertz light. First, we explain the interaction between light and the materials, offering insights into the selection of appropriate wavelengths. Second, we introduce common techniques for modifying photosensitive molecules, such as photocages and photoswitches, and review previous studies on light-controlled gene expression via the manipulation of photosensitive DNA/RNA, proteins, and small molecules. In addition, we discuss the technological challenges associated with the current technologies utilizing photosensitive molecules for the regulation of gene expression. Looking ahead, we explore the exciting prospect of extending label-free light-controlled technologies using the mid-IR and terahertz wavelength ranges, which are traditionally limited by water absorption, in biological applications. Because it allows the photocontrol of gene expression without any photosensitive molecular tag, this uncharted territory shows potential for groundbreaking innovations in the field of gene expression, contributing to medical and pharmaceutical development.

Author contributions

HY: conceptualization, funding acquisition, writing–original draft, and writing–review and editing. RS: writing–original draft, writing–review and editing. YT: writing–original draft, and writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. HY acknowledges the support of the Exploratory Research for Advanced Technology (ERATO, JPMJER2002) from the Japan Science and Technology Agency (JST).

Acknowledgments

We appreciate the support of the Exploratory Research for Advanced Technology and editing support of academic papers in English at Nagaoka University of Technology. We acknowledges Editage (https://campaigns.editage.jp/) and ChatGTP(https://chatgpt.com) to improve the language and readability of this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alachouzos, G., Schulte, A. M., Mondal, A., Szymanski, W., and Feringa, B. L. (2022). Computational design, synthesis, and photochemistry of cy7-PPG, an efficient NIR-activated photolabile protecting group for therapeutic applications. *Angew. Chem. Int. Ed.* 61, e202201308. doi:10.1002/anie.202201308

Alfastsen, L., Peng, X., and Bhoobalan-Chitty, Y. (2021). Genome editing in archaeal viruses and endogenous viral protein purification. *Star. Protoc.* 2, 100791. doi:10.1016/j.xpro.2021.100791

Ali, A. A., Kang, M., Kharbash, R., and Kim, Y. (2019). Spiropyran as a potential molecular diagnostic tool for double-stranded RNA detection. *BMC Biomed. Eng.* 1, 6. doi:10.1186/s42490-019-0008-x

Ando, H., Furuta, T., Tsien, R. Y., and Okamoto, H. (2001). Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat. Genet.* 28, 317–325. doi:10.1038/ng583

Anhäuser, L., Klöcker, N., Muttach, F., Mäsing, F., Špaček, P., Studer, A., et al. (2020). A benzophenone-based photocaging strategy for the N7 position of guanosine. *Angew. Chem. Int. Ed.* 59, 3161–3165. doi:10.1002/anie.201914573

Arbely, E., Torres-Kolbus, J., Deiters, A., and Chin, J. W. (2012). Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* 134, 11912–11915. doi:10.1021/ja3046958

Ashwood, B., Lewis, N. H. C., Sanstead, P. J., and Tokmakoff, A. (2020). Temperature-jump 2D IR spectroscopy with intensity-modulated CW optical heating. *J. Phys. Chem. B* 124, 8665–8677. doi:10.1021/acs.jpcb.0c07177

Axelrod, S., Shakhnovich, E., and Gómez-Bombarelli, R. (2022). Excited state non-adiabatic dynamics of large photoswitchable molecules using a chemically transferable machine learning potential. *Nat. Commun.* 13, 3440. doi:10.1038/s41467-022-30999-w

Bardhan, A., and Deiters, A. (2019). Development of photolabile protecting groups and their application to the optochemical control of cell signaling. *Curr. Opin. Struct. Biol.* 57, 164–175. doi:10.1016/j.sbi.2019.03.028

Becker, Y., Green, K. A., Scott, B., and Becker, M. (2018). Artificial inoculation of epichloë festucae into lolium perenne, and visualisation of endophytic and epiphyllous fungal growth. *Bio-protocol* 8, e2990. doi:10.21769/bioprotoc.2990

Berdnikova, D. V. (2019). Design, synthesis and investigation of water-soluble hemi-indigo photoswitches for bioapplications. *Beilstein J. Org. Chem.* 15, 2822–2829. doi:10.3762/bjoc.15.275

Biase, F. H., and Schettini, G. (2024). Protocol for the electroporation of CRISPR-Cas for DNA and RNA targeting in *Bos taurus* zygotes. *Star. Protoc.* 5, 102940. doi:10.1016/j.xpro.2024.102940

Blidner, R. A., Svoboda, K. R., Hammer, R. P., and Monroe, W. T. (2008). Photoinduced RNA interference using DMNPE-caged 2′-deoxy-2′-fluoro substituted nucleic acids*in vitro* and *in vivo*. *Mol. Biosyst.* 4, 431–440. doi:10.1039/b801532e

Bobrov, A. V., Kishalova, M. V., Merkushev, D. A., and Marfin, Y. S. (2021). BODIPY in matrices: brief review. *J. Phys. Conf. Ser.* 1822, 012020. doi:10.1088/1742-6596/1822/1/012020

Bouas-Laurent, H., and Dürr, H. (2001). Organic photochromism (IUPAC technical report). *Pure Appl. Chem.* 73, 639–665. doi:10.1351/pac200173040639

Braun, G. B., Pallaoro, A., Wu, G., Missirlis, D., Zasadzinski, J. A., Tirrell, M., et al. (2009). Laser-activated gene silencing via gold Nanoshell–siRNA conjugates. *ACS Nano* 3, 2007–2015. doi:10.1021/nn900469q

Buglioni, L., Raymenants, F., Slattery, A., Zondag, S. D. A., and Noël, T. (2022). Technological innovations in photochemistry for organic synthesis: flow chemistry, high-throughput experimentation, scale-up, and photoelectrochemistry. *Chem. Rev.* 122, 2752–2906. doi:10.1021/acs.chemrev.1c00332

Cambridge, S. B., Geissler, D., Keller, S., and Cürten, B. (2006). A caged doxycycline analogue for photoactivated gene expression. *Angew. Chem. Int. Ed.* 45, 2229–2231. doi:10.1002/anie.200503339

Cattaneo, P., and Persico, M. (1999). An abinitio study of the photochemistry of azobenzene. *Phys. Chem. Chem. Phys.* 1, 4739–4743. doi:10.1039/a905055h

Cautereels, C., Smets, J., De Saeger, J., Cool, L., Zhu, Y., Zimmermann, A., et al. (2024). Orthogonal LoxPsym sites allow multiplexed site-specific recombination in prokaryotic and eukaryotic hosts. *Nat. Commun.* 15, 1113. doi:10.1038/s41467-024-44996-8

Chakrapani, A., Vaňková Hausnerová, V., Ruiz-Larrabeiti, O., Pohl, R., Krásný, L., and Hocek, M. (2020). Photocaged 5-(hydroxymethyl)pyrimidine nucleoside phosphoramidites for specific photoactivatable epigenetic labeling of DNA. *Org. Lett.* 22, 9081–9085. doi:10.1021/acs.orglett.0c03462

Chan, D. K. H., Collins, S. D., and Buczacki, S. J. A. (2023). Generation and immunofluorescent validation of gene knockouts in adult human colonic organoids using multi-guide RNA CRISPR-Cas9. *Star. Protoc.* 4, 101978. doi:10.1016/j.xpro.2022.101978

Chen, G., Cao, Y., Tang, Y., Yang, X., Liu, Y., Huang, D., et al. (2020). Advanced near-infrared light for monitoring and modulating the spatiotemporal dynamics of cell functions in living systems. *Adv. Sci.* 7, 1903783. doi:10.1002/advs.201903783

Chen, W., Li, C., Liang, W., Li, Y., Zou, Z., Xie, Y., et al. (2022). The roles of optogenetics and technology in neurobiology: a review. *Front. Aging Neurosci.* 14, 867863. doi:10.3389/fnagi.2022.867863

Cheng, H.-B., Cao, X., Zhang, S., Zhang, K., Cheng, Y., Wang, J., et al. (2023). BODIPY as a multifunctional theranostic reagent in biomedicine: self-assembly, properties, and applications. *Adv. Mater.* 35, 2207546. doi:10.1002/adma.202207546

Cheong, S.-S., Luis, T. C., Hind, M., and Dean, C. H. (2024). A novel method for floxed gene manipulation using TAT-cre recombinase in *ex vivo* precision-cut lung slices (PCLS). *Bio-protocol* 14, e4980. doi:10.21769/bioprotoc.4980

Chikkaraddy, R., Xomalis, A., Jakob, L. A., and Baumberg, J. J. (2022). Mid-infrared-perturbed molecular vibrational signatures in plasmonic nanocavities. *Light Sci. and Appl.* 11, 19. doi:10.1038/s41377-022-00709-8

Chou, C., and Deiters, A. (2011). Light-activated gene editing with a photocaged zinc-finger nuclease. *Angew. Chem. Int. Ed.* 50, 6839–6842. doi:10.1002/anie.201101157

Chou, C., Young, D. D., and Deiters, A. (2010). Photocaged T7 RNA polymerase for the light activation of transcription and gene function in pro- and eukaryotic cells. *ChemBioChem* 11, 972–977. doi:10.1002/cbic.201000041

Clements, M., Simpson Ragdale, H., Garcia-Diaz, C., and Parrinello, S. (2024). Generation of immunocompetent somatic glioblastoma mouse models through *in situ* transformation of subventricular zone neural stem cells. *Star. Protoc.* 5, 102928. doi:10.1016/j.xpro.2024.102928

Coelho, C., Vij, R., Smith, D. Q., Brady, N. R., Hamacher-Brady, A., and Casadevall, A. (2020). Study of microbial extracellular vesicles: separation by density gradients, protection assays and labelling for live tracking. *Bio-protocol* 10, e3502. doi:10.21769/bioprotoc.3502

Conic, S., Desplancq, D., Tora, L., and Weiss, E. (2018). Electroporation of labeled antibodies to visualize endogenous proteins and posttranslational modifications in living metazoan cell types. *Bio-protocol* 8, e3069. doi:10.21769/bioprotoc.3069

Crestani, M., Dini, T., Gauthier, N. C., and Monzo, P. (2022). Protocol to assess human glioma propagating cell migration on linear micropatterns mimicking brain invasion tracks. *Star. Protoc.* 3, 101331. doi:10.1016/j.xpro.2022.101331

Cruz, F. G., Koh, J. T., and Link, K. H. (2000). Light-activated gene expression. *J. Am. Chem. Soc.* 122, 8777–8778. doi:10.1021/ja001804h

Czerwinska, I., and Juskowiak, B. (2012). Photoisomerizable arylstilbazolium ligands recognize parallel and antiparallel structures of G-quadruplexes. *Int. J. Biol. Macromol.* 51, 576–582. doi:10.1016/j.ijbiomac.2012.06.027

Dcona, M. M., Mitra, K., and Hartman, M. C. T. (2020). Photocontrolled activation of small molecule cancer therapeutics. *RSC Med. Chem.* 11, 982–1002. doi:10.1039/d0md00107d

Deiters, A., Garner, R. A., Lusic, H., Govan, J. M., Dush, M., Nascone-Yoder, N. M., et al. (2010). Photocaged morpholino oligomers for the light-regulation of gene function in zebrafish and Xenopus embryos. *J. Am. Chem. Soc.* 132, 15644–15650. doi:10.1021/ja1053863

Deng, P., Xiao, F., Wang, Z., and Jin, G. (2021). A novel BODIPY quaternary ammonium salt-based fluorescent probe: synthesis, physical properties, and live-cell imaging. *Front. Chem.* 9, 650006. doi:10.3389/fchem.2021.650006

Deng, Z., Wang, N., Liu, Y., Xu, Z., Wang, Z., Lau, T.-C., et al. (2020). A photocaged, water-oxidizing, and nucleolus-targeted Pt(IV) complex with a distinct anticancer mechanism. *J. Am. Chem. Soc.* 142, 7803–7812. doi:10.1021/jacs.0c00221

Denoth-Lippuner, A., Royall, L. N., Gonzalez-Bohorquez, D., Machado, D., and Jessberger, S. (2022). Injection and electroporation of plasmid DNA into human cortical organoids. *Star. Protoc.* 3, 101129. doi:10.1016/j.xpro.2022.101129

Dias, A. R., Minas Da Piedade, M. E., Martinho Simões, J. A., Simoni, J. A., Teixeira, C., Diogo, H. P., et al. (1992). Enthalpies of formation of cis-azobenzene and trans-azobenzene. *J. Chem. Thermodyn.* 24, 439–447. doi:10.1016/s0021-9614(05)80161-2

Dockalova, V., Sanchez-Carnerero, E. M., Dunajova, Z., Palao, E., Slanska, M., Buryska, T., et al. (2020). Fluorescent substrates for haloalkane dehalogenases: novel probes for mechanistic studies and protein labeling. *Comput. Struct. Biotechnol. J.* 18, 922–932. doi:10.1016/j.csbj.2020.03.029

Dong, M., Babalhavaeji, A., Samanta, S., Beharry, A. A., and Woolley, G. A. (2015). Red-shifting azobenzene photoswitches for *in vivo* use. *Accounts Chem. Res.* 48, 2662–2670. doi:10.1021/acs.accounts.5b00270

Dudek, M., Deiana, M., Pokladek, Z., Mlynarz, P., Samoc, M., and Matczyszyn, K. (2018). Light-driven chiroptical photoswitchable DNA assemblies mediated by bioinspired photoresponsive molecules. *Nanoscale* 10, 11302–11306. doi:10.1039/c8nr01784k

Edwards, W. F., Young, D. D., and Deiters, A. (2009). Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chem. Biol.* 4, 441–445. doi:10.1021/cb900041s

Egashira, T., Nakagawa-Tamagawa, N., Abzhanova, E., Kawae, Y., Kohara, A., Koitabashi, R., et al. (2023). *In vivo* two-photon calcium imaging of cortical neurons in neonatal mice. *Star. Protoc.* 4, 102245. doi:10.1016/j.xpro.2023.102245

Erlendsson, S., Thorsen, T. S., and Madsen, K. L. (2019). Supported cell membrane sheets to monitor protein assembly. *Bio-protocol* 9, e3368. doi:10.21769/bioprotoc.3368

Feng, M., Ruan, Z., Shang, J., Xiao, L., Tong, A., and Xiang, Y. (2017). Photocaged G-quadruplex DNAzyme and aptamer by post-synthetic modification on phosphodiester backbone. *Bioconjugate Chem.* 28, 549–555. doi:10.1021/acs.bioconjchem.6b00646

Feng, Z., Tang, T., Wu, T., Yu, X., Zhang, Y., Wang, M., et al. (2021). Perfecting and extending the near-infrared imaging window. *Light Sci. and Appl.* 10, 197. doi:10.1038/s41377-021-00628-0

Fields, C., Sheng, P., Miller, B., Wei, T., and Xie, M. (2019). Northern blot with IR fluorescent probes: strategies for probe preparation. *Bio-protocol* 9, e3219. doi:10.21769/bioprotoc.3219

Flumens, D., Campillo-Davo, D., Janssens, I., Roex, G., De Waele, J., Anguille, S., et al. (2023). One-step CRISPR-Cas9-mediated knockout of native TCRαβ genes in human T cells using RNA electroporation. *Star. Protoc.* 4, 102112. doi:10.1016/j.xpro.2023.102112

Forero-Quintero, L. S., Raymond, W., Munsky, B., and Stasevich, T. J. (2022). Visualization, quantification, and modeling of endogenous RNA polymerase II phosphorylation at a single-copy gene in living cells. *Bio-protocol* 12, e4482. doi:10.21769/bioprotoc.4482

Foster, A. A., Greco, C. T., Green, M. D., Epps, T. H., III, and Sullivan, M. O. (2015). Light-mediated activation of siRNA release in diblock copolymer assemblies for controlled gene silencing. *Adv. Healthc. Mater.* 4, 760–770. doi:10.1002/adhm.201400671

Fuentes, E., Gerth, M., Berrocal, J. A., Matera, C., Gorostiza, P., Voets, I. K., et al. (2020). An azobenzene-based single-component supramolecular polymer responsive to multiple stimuli in water. *J. Am. Chem. Soc.* 142, 10069–10078. doi:10.1021/jacs.0c02067

Gao, L., Wang, F., Wang, Y., Hu, L., and Mao, J. (2022). A protocol for the generation of patient-specific iPSC lines from peripheral blood mononuclear cells. *Star. Protoc.* 3, 101530. doi:10.1016/j.xpro.2022.101530

Garcia-Amorós, J., Stopa, G., Stochel, G., van Eldik, R., Martínez, M., and Velasco, D. (2018). Activation volumes for cis-to-trans isomerisation reactions of azophenols: a clear mechanistic indicator? *Phys. Chem. Chem. Phys.* 20, 1286–1292. doi:10.1039/c7cp07349f

Gáspár, I., Wippich, F., and Ephrussi, A. (2018). Terminal deoxynucleotidyl transferase mediated production of labeled probes for single-molecule FISH or RNA capture. *Bio-protocol* 8, e2750. doi:10.21769/bioprotoc.2750

Gelabert, R., Moreno, M., and Lluch, J. M. (2023). Predicting the electronic absorption band shape of azobenzene photoswitches. *Int. J. Mol. Sci.* 24, 25. doi:10.3390/ijms24010025

Gesuita, L., Argunsah, A. Ö., and Karayannis, T. (2022). Sparse postnatal labeling and quantification of superficial cortical cell synapses in the mouse neocortex. *Star. Protoc.* 3, 101837. doi:10.1016/j.xpro.2022.101837

Giles, L. W., Faul, C. F. J., and Tabor, R. F. (2021). Azobenzene isomerization in condensed matter: lessons for the design of efficient light-responsive soft-matter systems. *Mater. Adv.* 2, 4152–4164. doi:10.1039/d1ma00340b

Givens, R. S., and Matuszewski, B. (1984). Photochemistry of phosphate esters: an efficient method for the generation of electrophiles. *J. Am. Chem. Soc.* 106, 6860–6861. doi:10.1021/ja00334a075

Gorka, A. P., Nani, R. R., Zhu, J., Mackem, S., and Schnermann, M. J. (2014). A near-IR uncaging strategy based on cyanine photochemistry. *J. Am. Chem. Soc.* 136, 14153–14159. doi:10.1021/ja5065203

Govan, J. M., Uprety, R., Thomas, M., Lusic, H., Lively, M. O., and Deiters, A. (2013). Cellular delivery and photochemical activation of antisense agents through a nucleobase caging strategy. *ACS Chem. Biol.* 8, 2272–2282. doi:10.1021/cb400293e

Griepenburg, J. C., Rapp, T. L., Carroll, P. J., Eberwine, J., and Dmochowski, I. J. (2015). Ruthenium-caged antisense morpholinos for regulating gene expression in zebrafish embryos. *Chem. Sci.* 6, 2342–2346. doi:10.1039/c4sc03990d

Gutzler, R., Garg, M., Ast, C. R., Kuhnke, K., and Kern, K. (2021). Light–matter interaction at atomic scales. *Nat. Rev. Phys.* 3, 441–453. doi:10.1038/s42254-021-00306-5

Han, W.-G., Lovell, T., Liu, T., and Noodleman, L. (2002). Density functional studies of the ground- and excited-state potential-energy curves of stilbene cis–trans isomerization. *ChemPhysChem* 3, 167–178. doi:10.1002/1439-7641(20020215)3:2<167::aid-cphc167>3.0.co;2-g

Hartmann, D., and Booth, M. J. (2023). Accessible light-controlled knockdown of cell-free protein synthesis using phosphorothioate-caged antisense oligonucleotides. *Commun. Chem.* 6, 59. doi:10.1038/s42004-023-00860-2

Heinz, M. C., Oost, K. C., and Snippert, H. J. G. (2020). Introducing the stem cell ASCL2 reporter STAR into intestinal organoids. *Star. Protoc.* 1, 100126. doi:10.1016/j.xpro.2020.100126

Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A., and Deiters, A. (2015). Optical control of CRISPR/Cas9 gene editing. *J. Am. Chem. Soc.* 137, 5642–5645. doi:10.1021/ja512664v

Hemphill, J., Chou, C., Chin, J. W., and Deiters, A. (2013). Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. *J. Am. Chem. Soc.* 135, 13433–13439. doi:10.1021/ja4051026

Heyne, K., and Kühn, O. (2019). Infrared laser excitation controlled reaction acceleration in the electronic ground state. *J. Am. Chem. Soc.* 141, 11730–11738. doi:10.1021/jacs.9b02600

Holmstrom, E. D., and Nesbitt, D. J. (2010). Real-time infrared overtone laser control of temperature in picoliter H2O samples: "nanobathtubs" for single molecule microscopy. *J. Phys. Chem. Lett.* 1, 2264–2268. doi:10.1021/jz100663e

Hoorens, M. W. H., and Szymanski, W. (2018). Reversible, spatial and temporal control over protein activity using light. *Trends Biochem. Sci.* 43, 567–575. doi:10.1016/j.tibs.2018.05.004

Horii, Y., Matsuda, S., Watari, K., Nagasaka, A., Kurose, H., and Nakaya, M. (2017). An assay to determine phagocytosis of apoptotic cells by cardiac macrophages and cardiac myofibroblasts. *Bio-protocol* 7, e2553. doi:10.21769/bioprotoc.2553

Hoshina, H., Saito, Y., Furuhashi, T., Shimazaki, T., Sawada, M., Hioki, Y., et al. (2020). Terahertz spectroscopy for characterization of hydrogen bonding and cross-linked structure dynamics in polyurethane. *J. Infrared, Millim. Terahertz Waves* 41, 265–275. doi:10.1007/s10762-019-00667-0

Huang, X., Hu, Q., Braun, G. B., Pallaoro, A., Morales, D. P., Zasadzinski, J., et al. (2015). Light-activated RNA interference in human embryonic stem cells. *Biomaterials* 63, 70–79. doi:10.1016/j.biomaterials.2015.06.006

Imoto, T., Kawase, A., Minoshima, M., Yokoyama, T., Bito, H., and Kikuchi, K. (2020). Photolytic release of a caged inhibitor of an endogenous transcription factor enables optochemical control of CREB-mediated gene expression. *Org. Lett.* 22, 22–25. doi:10.1021/acs.orglett.9b03568

Inlay, M. A., Choe, V., Bharathi, S., Fernhoff, N. B., Baker, J. R., Weissman, I. L., et al. (2013). Synthesis of a photocaged tamoxifen for light-dependent activation of Cre-ER recombinase-driven gene modification. *Chem. Commun.* 49, 4971–4973. doi:10.1039/c3cc42179a

Ishigane, G., Toda, K., Tamamitsu, M., Shimada, H., Badarla, V. R., and Ideguchi, T. (2023). Label-free mid-infrared photothermal live-cell imaging beyond video rate. *Light Sci. and Appl.* 12, 174. doi:10.1038/s41377-023-01214-2

Jain, P. K., Ramanan, V., Schepers, A. G., Dalvie, N. S., Panda, A., Fleming, H. E., et al. (2016). Development of light-activated CRISPR using guide RNAs with photocleavable protectors. *Angew. Chem. Int. Ed.* 55, 12440–12444. doi:10.1002/anie.201606123

Jain, R. K., Hoffman, A. J., Jepsen, P. U., Liu, P. Q., Turchinovich, D., and Vitiello, M. S. (2020). Mid-infrared, long-wave infrared, and terahertz photonics: introduction. *Opt. Express* 28, 14169–14175. doi:10.1364/oe.395165

Janeková, H., Russo, M., Ziegler, U., and Štacko, P. (2022). Photouncaging of carboxylic acids from cyanine dyes with near-infrared light. *Angew. Chem. Int. Ed.* 61, e202204391. doi:10.1002/anie.202204391

Jia, B., Wang, T., and Lehmann, J. (2021). Peptidyl transferase center decompaction and structural constraints during early protein elongation on the ribosome. *Sci. Rep.* 11, 24061. doi:10.1038/s41598-021-02985-7

Jia, S., and Sletten, E. M. (2022). Spatiotemporal control of biology: synthetic photochemistry toolbox with far-red and near-infrared light. *ACS Chem. Biol.* 17, 3255–3269. doi:10.1021/acschembio.1c00518

Jung, H., Kim, S.-W., Kim, M., Hong, J., Yu, D., Kim, J. H., et al. (2019). Noninvasive optical activation of Flp recombinase for genetic manipulation in deep mouse brain regions. *Nat. Commun.* 10, 314. doi:10.1038/s41467-018-08282-8

Kajimoto, T., and Nakamura, S.-i. (2018). Quantitative analysis of cargo density in single-extracellular vesicles by imaging. *Bio-protocol* 8, e3111. doi:10.21769/bioprotoc.3111

Kala, A., Jain, P. K., Karunakaran, D., Shah, S., and Friedman, S. H. (2014). The synthesis of tetra-modified RNA for the multidimensional control of gene expression via light-activated RNA interference. *Nat. Protoc.* 9, 11–20. doi:10.1038/nprot.2013.165

Kaufmann, J., Wolf, J., and Heckel, A. (2023). Extending the palette of green coumarin photocages – oligonucleotide fragmentation and superior 5'-caps. *Chem. – A Eur. J.* 29, e202300390. doi:10.1002/chem.202300390

Keyvan Rad, J., Balzade, Z., and Mahdavian, A. R. (2022). Spiropyran-based advanced photoswitchable materials: a fascinating pathway to the future stimuli-responsive devices. *J. Photochem. Photobiol. C Photochem. Rev.* 51, 100487. doi:10.1016/j.jphotochemrev.2022.100487

Khalil, A. M. (2020). The genome editing revolution: review. *J. Genet. Eng. Biotechnol.* 18, 68. doi:10.1186/s43141-020-00078-y

Klena, N. T., Maltinti, G., Batman, U., Pigino, G., Guichard, P., and Hamel, v. (2023). An in-depth guide to the ultrastructural expansion microscopy (U-ExM) of chlamydomonas reinhardtii. *Bio-protocol* 13, e4792. doi:10.21769/bioprotoc.4792

Koch, A. L., Morisaki, T., and Stasevich, T. J. (2021). A multi-color bicistronic biosensor to compare the translation dynamics of different open reading frames at single-molecule resolution in live cells. *Bio-protocol* 11, e4096. doi:10.21769/bioprotoc.4096

Koh, X. Y., Koh, X. H., Hwang, L.-A., Ferrer, F. J., Rahmat, S. A. B., Lama, D., et al. (2019). Therapeutic anti-cancer activity of antibodies targeting sulfhydryl bond constrained epitopes on unglycosylated RON receptor tyrosine kinase. *Oncogene* 38, 7342–7356. doi:10.1038/s41388-019-0946-8

Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Le, C., et al. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472–476. doi:10.1038/nature12466

Kortekaas, L., and Browne, W. R. (2019). The evolution of spiropyran: fundamentals and progress of an extraordinarily versatile photochrome. *Chem. Soc. Rev.* 48, 3406–3424. doi:10.1039/c9cs00203k

Kuntze, K., Isokuortti, J., Siiskonen, A., Durandin, N., Laaksonen, T., and Priimagi, A. (2021). Azobenzene photoswitching with near-infrared light mediated by molecular oxygen. *J. Phys. Chem. B* 125, 12568–12573. doi:10.1021/acs.jpcb.1c08012

Laczi, D., Johnstone, M. D., and Fleming, C. L. (2022). Photoresponsive small molecule inhibitors for the remote control of enzyme activity. *Chem. – Asian J.* 17, e202200200. doi:10.1002/asia.202200200

Lankford, K. P., and Hulleman, J. D. (2024). Protocol for HiBiT tagging endogenous proteins using CRISPR-Cas9 gene editing. *Star. Protoc.* 5, 103000. doi:10.1016/j.xpro.2024.103000

Laprie-Sentenac, M., Cretet-Rodeschini, C., and Menger, L. (2023). Optimized protocol to generate genome-wide inactivated Cas9-expressing murine T cells. *Star. Protoc.* 4, 101922. doi:10.1016/j.xpro.2022.101922

Lax, C., Navarro-Mendoza, M. I., Pérez-Arques, C., Navarro, E., Nicolás, F. E., and Garre, V. (2022). Transformation and CRISPR-Cas9-mediated homologous recombination in the fungus Rhizopus microsporus. *Star. Protoc.* 3, 101237. doi:10.1016/j.xpro.2022.101237

Layden, H. M., Eleuteri, N. A., Hiebert, S. W., and Stengel, K. R. (2021). A protocol for rapid degradation of endogenous transcription factors in mammalian cells and identification of direct regulatory targets. *Star. Protoc.* 2, 100530. doi:10.1016/j.xpro.2021.100530

Lee, Y.-H., Yu, E., and Park, C.-M. (2021). Programmable site-selective labeling of oligonucleotides based on carbene catalysis. *Nat. Commun.* 12, 1681. doi:10.1038/s41467-021-21839-4

Leng, S., Qiao, Q.-L., Gao, Y., Miao, L., Deng, W.-G., and Xu, Z.-C. (2017). SNAP-tag fluorogenic probes for wash free protein labeling. *Chin. Chem. Lett.* 28, 1911–1915. doi:10.1016/j.cclet.2017.03.034

Li, Z., Reint, G., and Haapaniemi, E. M. (2023). Protocol for editing fibroblasts with *in vitro* transcribed Cas9 mRNA and profile off-target editing by optimized GUIDE-seq. *Star. Protoc.* 4, 102662. doi:10.1016/j.xpro.2023.102662

Lin, W., Albanese, C., Pestell, R. G., and Lawrence, D. S. (2002). Spatially discrete, light-driven protein expression. *Chem. and Biol.* 9, 1347–1353. doi:10.1016/s1074-5521(02)00288-0

Ling, X., Chang, L., Chen, H., and Liu, T. (2022). Efficient generation of locus-specific human CAR-T cells with CRISPR/cCas12a. *Star. Protoc.* 3, 101321. doi:10.1016/j.xpro.2022.101321

Link, K. H., Cruz, F. G., Ye, H.-F., O'Reilly, K. E., Dowdell, S., and Koh, J. T. (2004). Photo-caged agonists of the nuclear receptors RARγ and TRβ provide unique time-dependent gene expression profiles for light-activated gene patterning. *Bioorg. and Med. Chem.* 12, 5949–5959. doi:10.1016/j.bmc.2004.08.022

Liu, M., Ma, S., She, M., Chen, J., Wang, Z., Liu, P., et al. (2019). Structural modification of BODIPY: improve its applicability. *Chin. Chem. Lett.* 30, 1815–1824. doi:10.1016/j.cclet.2019.08.028

López-Lorente, Á. I., Wang, P., Sieger, M., Vargas Catalan, E., Karlsson, M., Nikolajeff, F., et al. (2016). Mid-infrared thin-film diamond waveguides combined with tunable quantum cascade lasers for analyzing the secondary structure of proteins. *Phys. status solidi (a)* 213, 2117–2123. doi:10.1002/pssa.201600134

Luciano, M. P., Crooke, S. N., Nourian, S., Dingle, I., Nani, R. R., Kline, G., et al. (2019). A nonaggregating heptamethine cyanine for building brighter labeled biomolecules. *ACS Chem. Biol.* 14, 934–940. doi:10.1021/acschembio.9b00122

Luo, H., Zhang, J., Yang, A., Ouyang, W., Long, S., Lin, X., et al. (2023). Large-scale isolation of exosomes derived from NK cells for anti-tumor therapy. *Bio-protocol* 13, e4693. doi:10.21769/bioprotoc.4693

Luo, J., Arbely, E., Zhang, J., Chou, C., Uprety, R., Chin, J. W., et al. (2016). Genetically encoded optical activation of DNA recombination in human cells. *Chem. Commun.* 52, 8529–8532. doi:10.1039/c6cc03934k

Ma, J., Qin, Z., Xie, G., Qian, L., and Tang, D. (2019). Review of mid-infrared mode-locked laser sources in the 2.0 μm–3.5 μm spectral region. *Appl. Phys. Rev.* 6. doi:10.1063/1.5037274

Mai-Morente, S. P., Irigoyen, J. P., Carriquiry, V. M., Marset, V. M., Di Doménico, M., Isasi, E., et al. (2021). Pericyte mapping in cerebral slices with the far-red fluorophore TO-PRO-3. *Bio-protocol* 11, e4222. doi:10.21769/bioprotoc.4222

Mao, Z., Yang, X., Mizutani, S., Huang, Y., Zhang, Z., Shinmori, H., et al. (2020). Hydrogen sulfide mediates tumor cell resistance to thioredoxin inhibitor. *Front. Oncol.* 10, 252. doi:10.3389/fonc.2020.00252

Mazzucco, M. R., Vartanian, T., and Linden, J. R. (2020). *In vivo* blood-brain barrier permeability assays using *Clostridium perfringens* epsilon toxin. *Bio-protocol* 10, e3709. doi:10.21769/bioprotoc.3709

Meka, D. P., Richter, M., Rücker, T., Voss, H., Rissiek, A., Krisp, C., et al. (2024). Protocol for differential multi-omic analyses of distinct cell types in the mouse cerebral cortex. *Star. Protoc.* 5, 102793. doi:10.1016/j.xpro.2023.102793

Menge, C., and Heckel, A. (2011). Coumarin-caged dG for improved wavelength-selective uncaging of DNA. *Org. Lett.* 13, 4620–4623. doi:10.1021/ol201842x

Meyer, A., and Mokhir, A. (2014). RNA interference controlled by light of variable wavelength. *Angew. Chem. Int. Ed.* 53, 12840–12843. doi:10.1002/anie.201405885

Morichika, I., and Ashihara, S. (2022). Ultrafast infrared plasmonics advances vibrational spectroscopy. *Photonics Rev.* 2022. doi:10.11470/photo.220204

Morichika, I., Murata, K., Sakurai, A., Ishii, K., and Ashihara, S. (2019). Molecular ground-state dissociation in the condensed phase employing plasmonic field enhancement of chirped mid-infrared pulses. *Nat. Commun.* 10, 3893. doi:10.1038/s41467-019-11902-6

Moroz-Omori, E. V., Satyapertiwi, D., Ramel, M.-C., Høgset, H., Sunyovszki, I. K., Liu, Z., et al. (2020). Photoswitchable gRNAs for spatiotemporally controlled CRISPR-cas-based genomic regulation. *ACS Central Sci.* 6, 695–703. doi:10.1021/acscentsci.9b01093

Mu, W., Homann, S., Hofmann, C., Gorin, A., Huynh, D., Yang, O. O., et al. (2019). A flow cytometric method to determine transfection efficiency. *Bio-protocol* 9, e3244. doi:10.21769/bioprotoc.3244

Mukherjee, D., Chakraborty, G., Hasan, M. N., Pal, U., Singh, P., Rakshit, T., et al. (2022). Reversible photoswitching of spiropyran in biomolecular interfaces: a combined spectroscopy and computational study. *J. Photochem. Photobiol. A Chem.* 430, 113958. doi:10.1016/j.jphotochem.2022.113958

Munson, M. J., and Ganley, I. G. (2016). Determination of cellular phosphatidylinositol-3-phosphate (PI3P) levels using a fluorescently labelled selective PI3P binding domain (PX). *Bio-protocol* 6, e1903. doi:10.21769/bioprotoc.1903

Nagahama, K., Fujino, S., Watanabe, T., Uesaka, N., and Kano, M. (2021). Combining electrophysiology and optogenetics for functional screening of pyramidal neurons in the mouse prefrontal cortex. *Star. Protoc.* 2, 100469. doi:10.1016/j.xpro.2021.100469

Neumann, H. (2012). Rewiring translation – genetic code expansion and its applications. *FEBS Lett.* 586, 2057–2064. doi:10.1016/j.febslet.2012.02.002

Nguyen, N. T., He, L., Martinez-Moczygemba, M., Huang, Y., and Zhou, Y. (2018). Rewiring calcium signaling for precise transcriptional reprogramming. *ACS Synth. Biol.* 7, 814–821. doi:10.1021/acssynbio.7b00467

Nomura, T., Yoshikawa, M., Suzuki, K., and Mochida, K. (2020). Highly efficient CRISPR-associated protein 9 ribonucleoprotein-based genome editing in euglena gracilis. *Star. Protoc.* 1, 100023. doi:10.1016/j.xpro.2020.100023

Nunes, C. M., Pereira, N. A. M., Reva, I., Amado, P. S. M., Cristiano, M. L. S., and Fausto, R. (2020). Bond-breaking/bond-forming reactions by vibrational excitation: infrared-induced bidirectional tautomerization of matrix-isolated thiotropolone. *J. Phys. Chem. Lett.* 11, 8034–8039. doi:10.1021/acs.jpclett.0c02272

O'Connor, M. J., Beebe, L. L., Deodato, D., Ball, R. E., Page, A. T., VanLeuven, A. J., et al. (2019). Bypassing glutamic acid decarboxylase 1 (Gad1) induced craniofacial defects with a photoactivatable translation blocker morpholino. *ACS Chem. Neurosci.* 10, 266–278. doi:10.1021/acschemneuro.8b00231

Ogasawara, S. (2018). Transcription driven by reversible photocontrol of hyperstable G-quadruplexes. *ACS Synth. Biol.* 7, 2507–2513. doi:10.1021/acssynbio.8b00216

O'Hagan, M. P., Haldar, S., Duchi, M., Oliver, T. A. A., Mulholland, A. J., Morales, J. C., et al. (2019). A photoresponsive stiff-stilbene ligand fuels the reversible unfolding of G-quadruplex DNA. *Angew. Chem. Int. Ed.* 58, 4334–4338. doi:10.1002/anie.201900740

O'Hagan, M. P., Peñalver, P., Gibson, R. S. L., Morales, J. C., and Galan, M. C. (2020). Stiff-stilbene ligands target G-quadruplex DNA and exhibit selective anticancer and antiparasitic activity. *Chem. – A Eur. J.* 26, 6224–6233. doi:10.1002/chem.201905753

Pan, Y., Yang, J., Luan, X., Liu, X., Li, X., Yang, J., et al. (2019). Near-infrared upconversion–activated CRISPR-Cas9 system: a remote-controlled gene editing platform. *Sci. Adv.* 5, eaav7199. doi:10.1126/sciadv.aav7199

Park, W., Lee, Y., Kang, T., Jeong, J., and Kim, D.-S. (2018). Terahertz-driven polymerization of resists in nanoantennas. *Sci. Rep.* 8, 7762. doi:10.1038/s41598-018-26214-w

Paul, A., Biswas, A., Sinha, S., Shah, S. S., Bera, M., Mandal, M., et al. (2019). Push–pull stilbene: visible light activated photoremovable protecting group for alcohols and carboxylic acids with fluorescence reporting employed for drug delivery. *Org. Lett.* 21, 2968–2972. doi:10.1021/acs.orglett.9b00124

Paul, A., Huang, J., Han, Y., Yang, X., Vuković, L., Král, P., et al. (2021). Photochemical control of bacterial gene expression based on trans encoded genetic switches. *Chem. Sci.* 12, 2646–2654. doi:10.1039/d0sc05479h

Paweletz, L. C., Veit, S., and Pomorski, T. G. (2022). A fluorescence-based approach utilizing self-labeling enzyme tags to determine protein orientation in large unilamellar vesicles. *Bio-protocol* 12, e4542. doi:10.21769/bioprotoc.4542

Peterson, J. A., Wijesooriya, C., Gehrmann, E. J., Mahoney, K. M., Goswami, P. P., Albright, T. R., et al. (2018). Family of BODIPY photocages cleaved by single photons of visible/near-infrared light. *J. Am. Chem. Soc.* 140, 7343–7346. doi:10.1021/jacs.8b04040

Pinheiro, A. V., Baptista, P., and Lima, J. C. (2008). Light activation of transcription: photocaging of nucleotides for control over RNA polymerization. *Nucleic Acids Res.* 36, e90. doi:10.1093/nar/gkn415

Polstein, L. R., and Gersbach, C. A. (2012). Light-inducible spatiotemporal control of gene activation by customizable Zinc finger transcription factors. *J. Am. Chem. Soc.* 134, 16480–16483. doi:10.1021/ja3065667

Poryvai, A., Galkin, M., Shvadchak, V., and Slanina, T. (2022). Red-shifted water-soluble BODIPY photocages for visualisation and controllable cellular delivery of signaling lipids. *Angew. Chem. Int. Ed.* 61, e202205855. doi:10.1002/anie.202205855

Powell, C. E., Gao, Y., Tan, L., Donovan, K. A., Nowak, R. P., Loehr, A., et al. (2018). Chemically induced degradation of anaplastic lymphoma kinase (ALK). *J. Med. Chem.* 61, 4249–4255. doi:10.1021/acs.jmedchem.7b01655

Prahl, S. (2017). *Light absorption by water*.

Rawat, S., and Sharma, M. (2024). CD8α-CI-M6PR particle motility assay to study the retrograde motion of CI-M6PR receptors in cultured living cells. *Bio-protocol* 14, e4979. doi:10.21769/bioprotoc.4979

Remenyi, R., Li, R., and Harris, M. (2019). On-demand labeling of SNAP-tagged viral protein for pulse-chase imaging, quench-pulse-chase imaging, and nanoscopy-based inspection of cell lysates. *Bio-protocol* 9, e3177. doi:10.21769/bioprotoc.3177

Riley, R. S., Dang, M. N., Billingsley, M. M., Abraham, B., Gundlach, L., and Day, E. S. (2018). Evaluating the mechanisms of light-triggered siRNA release from nanoshells for temporal control over gene regulation. *Nano Lett.* 18, 3565–3570. doi:10.1021/acs.nanolett.8b00681

Roberts, S. K., Hirsch, M., McStea, A., Zanetti-Domingues, L. C., Clarke, D. T., Claus, J., et al. (2018). Cluster analysis of endogenous HER2 and HER3 receptors in SKBR3 cells. *Bio-protocol* 8, e3096. doi:10.21769/bioprotoc.3096

Rozovsky, A., Patsenker, L., and Gellerman, G. (2019). Bifunctional reactive pentamethine cyanine dyes for biomedical applications. *Dyes Pigments* 162, 18–25. doi:10.1016/j.dyepig.2018.09.064

Roßmann, K., Gonzalez-Hernandez, A. J., Bhuyan, R., Börjesson, K., Levitz, J., and Broichhagen, J. (2023). Deuteration provides a general strategy to enhance azobenzene-based photopharmacology. *Biorxiv*. 2023.11.09.566420. doi:10.1101/2023.11.09.566420

Ryu, K. A., Kaszuba, C. M., Bissonnette, N. B., Oslund, R. C., and Fadeyi, O. O. (2021). Interrogating biological systems using visible-light-powered catalysis. *Nat. Rev. Chem.* 5, 322–337. doi:10.1038/s41570-021-00265-6

Sakamoto, M., Ota, K., Kondo, Y., Okamura, M., Fujii, H., and Bito, H. (2022). *In utero* electroporation and cranial window implantation for *in vivo* wide-field two-photon calcium imaging using G-CaMP9a transgenic mice. *Star. Protoc.* 3, 101421. doi:10.1016/j.xpro.2022.101421

Sakamoto, T., Shigeno, A., Ohtaki, Y., and Fujimoto, K. (2014). Photo-regulation of constitutive gene expression in living cells by using ultrafast photo-cross-linking oligonucleotides. *Biomaterials Sci.* 2, 1154–1157. doi:10.1039/c4bm00117f

Samolis, P. D., Zhu, X., and Sander, M. Y. (2023). Time-resolved mid-infrared photothermal microscopy for imaging water-embedded axon bundles. *Anal. Chem.* 95, 16514–16521. doi:10.1021/acs.analchem.3c02352

Sanketi, B. D., Sivakumar, A., and Kurpios, N. A. (2023). Visualizing and manipulating the production and accumulation of hyaluronan for functional assessment in chicken embryos. *Star. Protoc.* 4, 102200. doi:10.1016/j.xpro.2023.102200

Shao, J., Wang, M., Yu, G., Zhu, S., Yu, Y., Heng, B. C., et al. (2018). Synthetic far-red light-mediated CRISPR-dCas9 device for inducing functional neuronal differentiation. *Proc. Natl. Acad. Sci.* 115, E6722–E6730. doi:10.1073/pnas.1802448115

Sharma, M., and Friedman, S. H. (2021). The issue of tissue: approaches and challenges to the light control of drug activity. *ChemPhotoChem* 5, 611–618. doi:10.1002/cptc.202100001

Sharma, R., Sharma, A., Kumar, A., and Jaganathan, B. G. (2019). Phospho-protein analysis in adherent cells using flow cytometry. *Bio-protocol* 9, e3395. doi:10.21769/bioprotoc.3395

Shiraishi, Y., Yomo, K., and Hirai, T. (2023). Polarity-Driven isomerization of a hydroxynaphthalimide-containing spiropyran at room temperature. *ACS Phys. Chem. Au* 3, 290–298. doi:10.1021/acsphyschemau.2c00067

Singh, P. K., Majumdar, P., and Singh, S. P. (2021). Advances in BODIPY photocleavable protecting groups. *Coord. Chem. Rev.* 449, 214193. doi:10.1016/j.ccr.2021.214193

Sitkowska, K., Feringa, B. L., and Szymański, W. (2018). Green-light-sensitive BODIPY photoprotecting groups for amines. *J. Org. Chem.* 83, 1819–1827. doi:10.1021/acs.joc.7b02729

Skruber, K., Read, T.-A., and Vitriol, E. A. (2021). Delivering defined amounts of purified protein with high precision into living cells. *Star. Protoc.* 2, 100272. doi:10.1016/j.xpro.2020.100272

Šmidlehner, T., Kurutos, A., Slade, J., Belužić, R., Ang, D. L., Rodger, A., et al. (2018). Versatile click cyanine amino acid conjugates showing one-atom-influenced recognition of DNA/RNA secondary structure and mitochondrial localisation in living cells. *Eur. J. Org. Chem.* 2018, 1682–1692. doi:10.1002/ejoc.201701765

Stadler, J. M., and Stafforst, T. (2014). Pyrene chromophores for the photoreversal of psoralen interstrand crosslinks. *Org. and Biomol. Chem.* 12, 5260–5266. doi:10.1039/c4ob00603h

Stafforst, T., and Stadler, J. M. (2013). Photoactivation of a psoralen-blocked luciferase gene by blue light. *Angew. Chem. Int. Ed.* 52, 12448–12451. doi:10.1002/anie.201306150

Stensitzki, T., Yang, Y., Kozich, V., Ahmed, A. A., Kössl, F., Kühn, O., et al. (2018). Acceleration of a ground-state reaction by selective femtosecond-infrared-laser-pulse excitation. *Nat. Chem.* 10, 126–131. doi:10.1038/nchem.2909

Sugiyama, J.-i., Tokunaga, Y., Hishida, M., Tanaka, M., Takeuchi, K., Satoh, D., et al. (2023). Nonthermal acceleration of protein hydration by sub-terahertz irradiation. *Nat. Commun.* 14, 2825. doi:10.1038/s41467-023-38462-0

Sun, L., Zhao, L., and Peng, R.-Y. (2021). Research progress in the effects of terahertz waves on biomacromolecules. *Mil. Med. Res.* 8, 28. doi:10.1186/s40779-021-00321-8

Tallafuss, A., Gibson, D., Morcos, P., Li, Y., Seredick, S., Eisen, J., et al. (2012). Turning gene function ON and OFF using sense and antisense photo-morpholinos in zebrafish. *Development* 139, 1691–1699. doi:10.1242/dev.072702

Talone, B., Bazzarelli, M., Schirato, A., Dello Vicario, F., Viola, D., Jacchetti, E., et al. (2021). Phototoxicity induced in living HeLa cells by focused femtosecond laser pulses: a data-driven approach. *Biomed. Opt. Express* 12, 7886–7905. doi:10.1364/boe.441225

Tang, X., Swaminathan, J., Gewirtz, A. M., and Dmochowski, I. J. (2007). Regulating gene expression in human leukemia cells using light-activated oligodeoxynucleotides. *Nucleic Acids Res.* 36, 559–569. doi:10.1093/nar/gkm1029

Tanimizu, N. (2015). Clonal culture of mouse liver progenitor cells. *Bio-protocol* 5, e1624. doi:10.21769/bioprotoc.1624

Tavakoli, A., and Min, J.-H. (2022). Photochemical modifications for DNA/RNA oligonucleotides. *RSC Adv.* 12, 6484–6507. doi:10.1039/d1ra05951c

Teng, I. T., Bu, X., and Chung, I. (2019). Conjugation of fab' fragments with fluorescent dyes for single-molecule tracking on live cells. *Bio-protocol* 9, e3375. doi:10.21769/bioprotoc.3375

Tian, T., Song, Y., Wang, J., Fu, B., He, Z., Xu, X., et al. (2016). Small-molecule-triggered and light-controlled reversible regulation of enzymatic activity. *J. Am. Chem. Soc.* 138, 955–961. doi:10.1021/jacs.5b11532

Toda, K., Tamamitsu, M., Nagashima, Y., Horisaki, R., and Ideguchi, T. (2019). Molecular contrast on phase-contrast microscope. *Sci. Rep.* 9, 9957. doi:10.1038/s41598-019-46383-6

Tonouchi, M. (2007). Cutting-edge terahertz technology. *Nat. Photonics* 1, 97–105. doi:10.1038/nphoton.2007.3

Toyama, T., Fujioka, J., Watanabe, K., Yoshida, A., Sakuma, T., Inaba, K., et al. (2022). Investigation of bactericidal effect of a mid-infrared free electron laser on *Escherichia coli*. *Sci. Rep.* 12, 18111. doi:10.1038/s41598-022-22949-9

Tronnet, S., and Oswald, E. (2018). Quantification of colibactin-associated genotoxicity in HeLa cells by in cell western (ICW) using γ-H2AX as a marker. *Bio-protocol* 8, e2771. doi:10.21769/bioprotoc.2771

Turchetto, S., Broix, L., and Nguyen, L. (2020). *Ex vivo* recording of axonal transport dynamics on postnatal organotypic cortical slices. *Star. Protoc.* 1, 100131. doi:10.1016/j.xpro.2020.100131

Udupa, S., Nagaraja, V., and Karambelkar, S. (2022). Binding affinity quantifications of the bacteriophage Mu DNA modification protein mom using microscale thermophoresis (MST). *Bio-protocol* 12, e4472. doi:10.21769/bioprotoc.4472

Villarón, D., Duindam, N., and Wezenberg, S. J. (2021). Push-pull stiff-stilbene: proton-gated visible-light photoswitching and acid-catalyzed isomerization. *Chem. – A Eur. J.* 27, 17346–17350. doi:10.1002/chem.202103052

Vohradská, N., Sánchez-Carnerero, E. M., Pastierik, T., Mazal, C., and Klán, P. (2018). Controlled photorelease of alkynoic acids and their decarboxylative deprotection for copper-catalyzed azide/alkyne cycloaddition. *Chem. Commun.* 54, 5558–5561. doi:10.1039/c8cc03341b

Volarić, J., Buter, J., Schulte, A. M., van den Berg, K.-O., Santamaría-Aranda, E., Szymanski, W., et al. (2022). Design and synthesis of visible-light-responsive azobenzene building blocks for chemical biology. *J. Org. Chem.* 87, 14319–14333. doi:10.1021/acs.joc.2c01777

Walsh, S., Gardner, L., Deiters, A., and Williams, G. J. (2014). Intracellular light-activation of riboswitch activity. *ChemBioChem* 15, 1346–1351. doi:10.1002/cbic.201400024

Wang, K., Schmied, W. H., and Chin, J. W. (2012). Reprogramming the genetic code: from triplet to quadruplet codes. *Angew. Chem. Int. Ed.* 51, 2288–2297. doi:10.1002/anie.201105016

Wang, Q., Wu, Z., Qin, P., Ji, J., Lai, L., and Yin, M. (2021). Photoregulated morphological transformation of spiropyran derivatives achieving the tunability of interfacial hydrophilicity. *Langmuir* 37, 11170–11175. doi:10.1021/acs.langmuir.1c02053

Wang, X., Feng, M., Xiao, L., Tong, A., and Xiang, Y. (2016). Postsynthetic modification of DNA phosphodiester backbone for photocaged DNAzyme. *ACS Chem. Biol.* 11, 444–451. doi:10.1021/acschembio.5b00867

Wang, Y., Wang, Z., and Ji, Q. (2020). CRISPR-Cas9-Based genome editing and cytidine base editing in acinetobacter baumannii. *Star. Protoc.* 1, 100025. doi:10.1016/j.xpro.2020.100025

Weinstain, R., Slanina, T., Kand, D., and Klán, P. (2020). Visible-to-NIR-Light activated release: from small molecules to nanomaterials. *Chem. Rev.* 120, 13135–13272. doi:10.1021/acs.chemrev.0c00663

Welleman, I. M., Hoorens, M. W. H., Feringa, B. L., Boersma, H. H., and Szymański, W. (2020). Photoresponsive molecular tools for emerging applications of light in medicine. *Chem. Sci.* 11, 11672–11691. doi:10.1039/d0sc04187d

Williams, R. M., and Sauka-Spengler, T. (2021). Ex ovo electroporation of early chicken embryos. *Star. Protoc.* 2, 100424. doi:10.1016/j.xpro.2021.100424

Wilson, C., Rozés-Salvador, V., and Cáceres, A. (2020). Protocol for evaluating neuronal polarity in murine models. *Star. Protoc.* 1, 100114. doi:10.1016/j.xpro.2020.100114

Windhorn, L., Witte, T., Yeston, J. S., Proch, D., Motzkus, M., Kompa, K. L., et al. (2002). Molecular dissociation by mid-IR femtosecond pulses. *Chem. Phys. Lett.* 357, 85–90. doi:10.1016/s0009-2614(02)00444-x

Wirth, R., Gao, P., Nienhaus, G. U., Sunbul, M., and Jäschke, A. (2020). Confocal and super-resolution imaging of RNA in live bacteria using a fluorogenic silicon rhodamine-binding aptamer. *Bio-protocol* 10, e3603. doi:10.21769/bioprotoc.3603

Wong, P. T., Roberts, E. W., Tang, S., Mukherjee, J., Cannon, J., Nip, A. J., et al. (2017). Control of an unusual photo-claisen rearrangement in coumarin caged tamoxifen through an extended spacer. *ACS Chem. Biol.* 12, 1001–1010. doi:10.1021/acschembio.6b00999

Wu, Y., Sidharta, M., Zhong, A., Persily, B., Li, M., and Zhou, T. (2023). Protocol for the design, conduct, and evaluation of prime editing in human pluripotent stem cells. *Star. Protoc.* 4, 102583. doi:10.1016/j.xpro.2023.102583

Xing, X., Wang, X., Xu, L., Tai, Y., Dai, L., Zheng, X., et al. (2015). Correction: light-driven conformational regulation of human telomeric G-quadruplex DNA in physiological conditions. *Org. and Biomol. Chem.* 13, 4613. doi:10.1039/c5ob90054a

Xu, H., Kita, Y., Bang, U., Gee, P., and Hotta, A. (2021). Optimized electroporation of CRISPR-Cas9/gRNA ribonucleoprotein complex for selection-free homologous recombination in human pluripotent stem cells. *Star. Protoc.* 2, 100965. doi:10.1016/j.xpro.2021.100965

Xu, J., and Liu, Y. (2019). Imaging higher-order chromatin structures in single cells using stochastic optical reconstruction microscopy. *Bio-protocol* 9, e3160. doi:10.21769/bioprotoc.3160

Xu, Z., Li, Y., and Yan, A. (2020). Repurposing the native type I-F CRISPR-cas system in *Pseudomonas aeruginosa* for genome editing. *Star. Protoc.* 1, 100039. doi:10.1016/j.xpro.2020.100039

Yamamoto, T., Caldwell, D. R., Gandioso, A., and Schnermann, M. J. (2019). A cyanine photooxidation/β-elimination sequence enables near-infrared uncaging of aryl amine payloads. *Photochem. Photobiol.* 95, 951–958. doi:10.1111/php.13090

Yamazoe, S., Liu, Q., McQuade, L. E., Deiters, A., and Chen, J. K. (2014). Sequential gene silencing using wavelength-selective caged morpholino oligonucleotides. *Angew. Chem. Int. Ed.* 53, 10114–10118. doi:10.1002/anie.201405355

Yang, L., Kim, H. B., Sul, J.-Y., Yeldell, S. B., Eberwine, J. H., and Dmochowski, I. J. (2018). Efficient synthesis of light-triggered circular antisense oligonucleotides targeting cellular protein expression. *ChemBioChem* 19, 1250–1254. doi:10.1002/cbic.201800012

York, V. A., and Milush, J. M. (2015). *Ex vivo* human natural killer (NK) cell stimulation and intracellular IFNγ and CD107a cytokine staining. *Bio-protocol* 5, e1501. doi:10.21769/bioprotoc.1501

Young, D. D., Garner, R. A., Yoder, J. A., and Deiters, A. (2009). Light-activation of gene function in mammalian cells viaribozymes. *Chem. Commun.*, 568–570. doi:10.1039/b819375d

Yu, L., Jing, N., Yang, Z., Zhang, L., and Tang, X. (2018). Caged siRNAs with single folic acid modification of antisense RNA for photomodulation of exogenous and endogenous gene expression in cells. *Org. and Biomol. Chem.* 16, 7029–7035. doi:10.1039/c8ob01952e

Zhang, D., Lan, L., Bai, Y., Majeed, H., Kandel, M. E., Popescu, G., et al. (2019). Bond-selective transient phase imaging via sensing of the infrared photothermal effect. *Light Sci. and Appl.* 8, 116. doi:10.1038/s41377-019-0224-0

Zhang, L., Chen, C., Fan, X., and Tang, X. (2018). Photomodulating gene expression by using caged siRNAs with single-aptamer modification. *ChemBioChem* 19, 1259–1263. doi:10.1002/cbic.201700623

Zhang, W., Hamouri, F., Feng, Z., Aujard, I., Ducos, B., Ye, S., et al. (2018). Control of protein activity and gene expression by cyclofen-OH uncaging. *ChemBioChem* 19, 1232–1238. doi:10.1002/cbic.201700630

Zhi, T. S., Cheng, T. P., Qing, L. L., Liang, C. Y., Long, -.Y. Z., Long, Z. X., et al. (2019). - Exposure effects of terahertz waves on primary neurons and neuron-like cells under nonthermal conditions. *Biomed. Environ. Sci.* - 32–739. doi:10.3967/bes2019.094

Zhou, L., Cheng, Z.-Q., Li, N., Ge, Y.-X., Xie, H.-X., Zhu, K., et al. (2020). A highly sensitive endoplasmic reticulum-targeting fluorescent probe for the imaging of endogenous H2S in live cells. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* 240, 118578. doi:10.1016/j.saa.2020.118578

Zhou, X. X., Zou, X., Chung, H. K., Gao, Y., Liu, Y., Qi, L. S., et al. (2018). A single-chain photoswitchable CRISPR-cas9 architecture for light-inducible gene editing and transcription. *ACS Chem. Biol.* 13, 443–448. doi:10.1021/acschembio.7b00603

Zhou, Z., Ma, L., and Zhang, X. (2021). Protocol for genome-scale CRISPR screening in engineered lineage reporter hPSCs to study cell fate determination. *Star. Protoc.* 2, 100548. doi:10.1016/j.xpro.2021.100548

Ziegler, M. J., Yserentant, K., Dunsing, V., Middel, V., Gralak, A. J., Pakari, K., et al. (2022). Mandipropamid as a chemical inducer of proximity for *in vivo* applications. *Nat. Chem. Biol.* 18, 64–69. doi:10.1038/s41589-021-00922-3

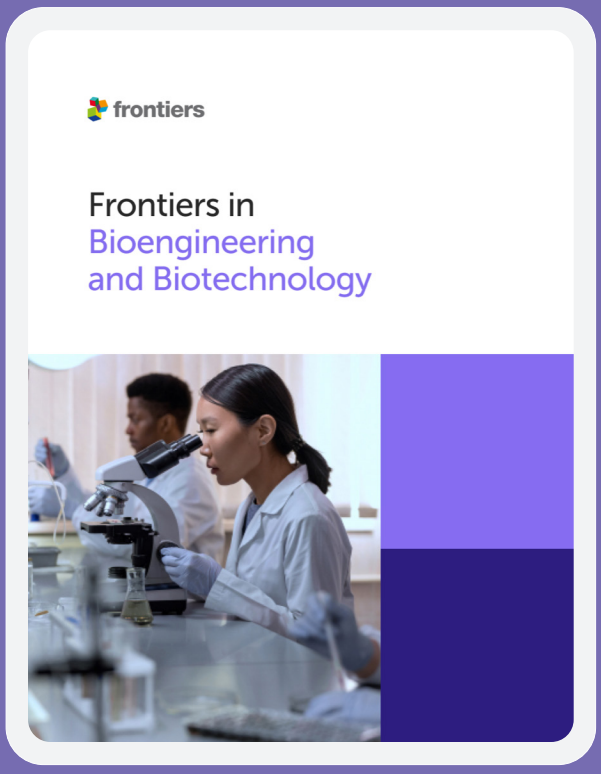
Frontiers in
Bioengineering and Biotechnology

Accelerates the development of therapies, devices, and technologies to improve our lives

A multidisciplinary journal that accelerates the development of biological therapies, devices, processes and technologies to improve our lives by bridging the gap between discoveries and their application.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact