



Cite this: *Chem. Soc. Rev.*, 2021, 50, 8003

Received 19th December 2020

DOI: 10.1039/d0cs01575j

rsc.li/chem-soc-rev

Recent trends in biocatalysis

Dong Yi, † Thomas Bayer, † Christoffel P. S. Badenhorst, † Shuke Wu, † Mark Doerr, † Matthias Höhne † and Uwe T. Bornscheuer *

Biocatalysis has undergone revolutionary progress in the past century. Benefited by the integration of multidisciplinary technologies, natural enzymatic reactions are constantly being explored. Protein engineering gives birth to robust biocatalysts that are widely used in industrial production. These research achievements have gradually constructed a network containing natural enzymatic synthesis pathways and artificially designed enzymatic cascades. Nowadays, the development of artificial intelligence, automation, and ultra-high-throughput technology provides infinite possibilities for the discovery of novel enzymes, enzymatic mechanisms and enzymatic cascades, and gradually complements the lack of remaining key steps in the pathway design of enzymatic total synthesis. Therefore, the research of biocatalysis is gradually moving towards the era of novel technology integration, intelligent manufacturing and enzymatic total synthesis.

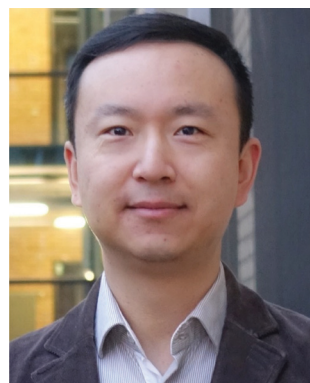
Introduction

More than a century ago, Rosenthaler synthesised (*R*)-mandelonitrile from benzaldehyde and hydrogen cyanide using a plant extract containing a hydroxynitrile lyase and opened the door to

a completely new research field: biocatalysis.¹ Over the next hundred years, this research field had developed rapidly and experienced three distinct waves,² advancing from the use of crude extract to purified enzymes and then to recombinant enzyme systems. The screening of biocatalysts had evolved from natural source extraction to gene mining with bioinformatic methods. Finally, sequence–structure–function analysis benefited from the developments in structural biology. The engineering of biocatalysts had shifted from modifying reaction conditions to suit enzymatic properties to directed

Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, University Greifswald, Felix-Hausdorff-Str. 4, D-17487 Greifswald, Germany.
E-mail: uwe.bornscheuer@uni-greifswald.de

† Equal contribution.



Dong Yi

Dong Yi studied chemistry supported by the Deutscher Akademischer Austauschdienst (DAAD) with a doctoral scholarship at the TU Darmstadt (Germany) and obtained his PhD (summa cum laude) under supervision of Prof. Wolf-Dieter Fessner in 2012. He worked as Postdoctoral Teaching Fellow at the East China University of Science and Technology (China) from 2013 and was involved in establishing a pharmaceutical company. In 2018,

he joined the Bornscheuer group at the University of Greifswald (Germany) as a postdoc and worked within the EU funded project “SynBio4Flav”. He currently works as the Head of Systems Biosynthesis in the Shanghai Institute of Pharmaceutical Industry (China).



Thomas Bayer

Thomas Bayer graduated in the Molecular Biology programme from the University of Vienna and received his PhD in (Bio)-organic Chemistry at the TU Wien (Austria) before joining the Bornscheuer group as a post-doctoral fellow. Thomas Bayer received the Erwin Schrödinger fellowship by The Austrian Science Fund (FWF), which included scientific stays at the University of Greifswald (Germany) and the Albert Einstein College of

Medicine in New York City (USA) in 2019–2020. Since 2021, he works at the Institute of Molecular Biotechnology at the TU Graz (Austria).



evolution of enzymes to adapt to reaction conditions and molecular structures of substrates.² These technological revolutions all profited from the development and integration of multidisciplinary technologies and theories which cover most scientific fields including chemistry, biology, pharmaceuticals, food, physics, mathematics, computer science, automation and engineering, and prospered the in-depth study of single-step enzymatic reactions, including enzymatic properties, reaction mechanisms, and protein engineering. As a result, a large number of enzymes used for biocatalysis have been identified from nature including almost all enzyme classes: oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases. The reactions catalysed by these enzymes cover common chemical reaction types, such as redox, substitution, addition, elimination, rearrangement and pericyclic reactions. Based on understanding enzymatic catalytic mechanisms well, these common enzymes and basic enzymatic reactions can be combined

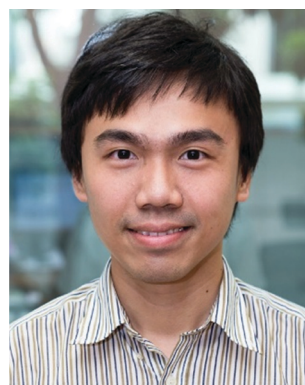
into a biocatalytic network, in which the upper edges and nodes are connected in series to form multi-step enzymatic cascades. These cascades made the biosynthesis of chemicals with complex structures feasible and thus have greatly enriched the achievements and application scope of biocatalysis. Therefore, more and more studies have focused on novel enzymatic reactions and advanced cascades for the biosynthesis of complex natural and non-natural products, which led to a remarkable research trend turning from enzymatic synthesis of compounds with simple structures to enzymatic total synthesis of challenging targets.

However, since the natural synthetic pathways of many natural products have not been fully elucidated, coupled with the challenges of heterologous expression of enzymes, it is not easy to achieve the enzymatic total synthesis of natural products *in vitro* or in model host cells. Complex drug compounds are mostly artificially designed molecular structures, which require well designed enzymatic synthesis pathways and heavy



Christoffel P. S. Badenhorst

Chris Badenhorst received a BSc in biochemistry and microbiology from the University of South Africa. He then joined Prof. Albie van Dijk's laboratory at the Potchefstroom Campus of the North West University, where he obtained his BSc Honours, MSc, and PhD degrees in biochemistry. He was subsequently awarded an Alexander von Humboldt Research fellowship to join the Bornscheuer group, where he has been working on the development of new ultrahigh-throughput screening methods.



Shuke Wu

Shuke Wu obtained his PhD from the National University of Singapore (2015) with Prof. Zhi Li (NUS) and Prof. Daniel I. C. Wang (MIT). In 2017, he moved to Switzerland to work with Prof. Thomas R. Ward at the University of Basel sponsored by a Swiss government excellence scholarship. Until 2020, he worked in the Bornscheuer group at the University of Greifswald as an Alexander von Humboldt fellow. Now he is a professor at the Huazhong Agricultural University in China.



Mark Doerr

Mark Doerr studied chemistry and biology at the Universities of Mainz and Jena. He received his PhD at the Institute of Inorganic Chemistry at the University in Jena, Germany (2004). He then joined research projects in the groups of Peter Nielsen in Copenhagen, Denmark (2006–2009) and Steen Rasmussen in Odense, Denmark (2009–2012) working on fundamental bio-/inorganic-chemical questions as a postdoctoral research associate.

Since 2012 he is staff scientist in the Bornscheuer group working on high-throughput robotic screening, lab automation, molecular modelling and the combination of machine learning with protein engineering.



Matthias Höhne

Matthias Höhne studied biochemistry and obtained his PhD in biotechnology in 2010 at Greifswald University under the supervision of Prof. Uwe Bornscheuer. He then moved to Heidelberg for his postdoc to work with Prof. Andres Jäschke and then started his Junior Professorship focussing on enzyme discovery and application with a focus on chiral amine synthesis. Since 2018, he continues with an ERC starting grant as group leader

at the Institute of Biochemistry in Greifswald and orients his research towards photobiocatalysis and the characterisation of binding and transport proteins involved in marine glycan degradation.



protein engineering to reduce the synthetic steps and solve the problem of low enzyme activity towards non-natural substrates. Comfortingly, innovative results of novel enzymes, enzymatic reaction mechanisms, and enzymatic cascades make the *de novo* biocatalytic synthesis of natural products and their derivatives easier. Especially, at present, the fourth scientific and technological revolution characterised by informatization, networking, and intellectualization is in full swing. Following Moore's Law,³ the integration and computing power of electronic chips are being continuously doubled, which enable big data mining, network relational processes, and *in silico* protein design that require heavy computation to be realised. Machine learning (ML), which has recently entered the era of deep learning (DL), has opened up the intelligence of data analysis and network construction on big data processing and extended their application in genomics, proteomics and metabolomics.^{4–6} This brings us effective tools for the discovery of novel enzymes, biocatalytic reactions, and enzymatic synthetic pathways from massive gene, protein and chemical data analysis. In addition, ML also guides the rational design and protein engineering to create artificially evolved enzymes.^{7,8} Moreover, the rapid development of quantum computers, automation and ultrahigh-throughput (UHTP) screening technology has brought unlimited possibilities for biocatalysis research. Impressively, quantum computing has been oriented towards practical use for calculating quantum chemistry models at an incredibly high speed.⁹ Automated screening robots have gained popularity in many laboratories. The next generation experiment robot has begun to serve chemical experiments with an efficiency hundreds of times higher than that of humans.¹⁰ Combined with fluorescence-activated cell sorting (FACS) and microfluidics, these advancements increase the screening throughput of enzyme mutagenesis libraries by several orders of magnitude.^{11–13} It is foreseeable that biocatalysis research will enter a new era characterised by the integration of novel technologies and intelligent manufacturing in the next decade. Therefore, enzymes and enzymatic pathway

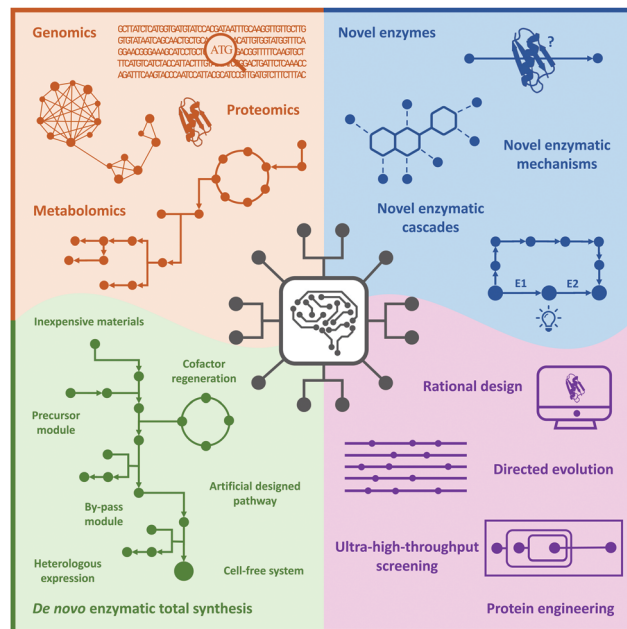


Fig. 1 New trends in biocatalysis: integration of novel technologies, intelligent manufacturing, and enzymatic total synthesis.

designs by artificial intelligence, as well as adaptive evolution assisted by UHTP screening, have simplified synthetic pathways of complex compounds and already have been adapted to model host cells. Together with the latest achievements in synthetic biology, several important natural products, their derivatives, and non-natural compounds with pharmaceutical value have been fully enzymatically synthesised, which has elevated the field of biocatalysis substantially. Eventually, historical borders between classical biocatalysis – using isolated and often immobilised enzymes – and biotransformation – using (engineered) whole-cell microorganisms – will disappear since nowadays all available tools can be integrated to create product-oriented designer pathways and only efficiency in terms of productivity in grams per litre and time will decide which type of reaction system is superior.

In this review, we briefly summarize the application of novel advanced technologies applied in biocatalysis, highlight the identification of novel enzymes and enzymatic reactions/cascades, and collect the innovative artificial enzymatic total synthesis of natural products and valuable drug compounds to elaborate the features of the new trends in biocatalysis: integration of novel technologies, intelligent manufacturing, and enzymatic total synthesis of complex molecules (Fig. 1).

Advanced technologies for the discovery and design of novel enzymes and enzymatic cascades

Machine learning

The rapid development of technologies related to high-throughput (HTP) gene sequencing, structural biology, screening, and



Uwe T. Bornscheuer

Uwe Bornscheuer studied chemistry and received his PhD in 1993 at Hannover University followed by a postdoc at Nagoya University (Japan). In 1998, he completed his Habilitation at Stuttgart University about the use of lipases and esterases in organic synthesis. He has been Professor at the Institute of Biochemistry at Greifswald University since 1999. In 2008, he received – beside other awards – the BioCat2008 Award. He was

just recognised as 'Chemistry Europe Fellow'. His current research interest focuses on the discovery and engineering of enzymes from various classes for applications in organic synthesis, lipid modification, degradation of plastics or complex marine polysaccharides.



bioinformatics, has accumulated a series of huge databases for genomics, proteomics, and metabolomics. This treasure trove of data provides us a wealth of research materials, but effectively analysing and revealing the interrelationships between the data is a current challenge. Our human brain is very good at recognising simple two- or three-dimensional (2D or 3D) relationships. However, it is powerless to deal with big data processing on a higher-dimensional level. Computing machines, in contrast, can be very well instructed to evaluate and find much higher dimensional relations, incomprehensible to most human minds. Therefore, ML has become an effective tool for biological data processing. Here, we would like to highlight some complementary aspects that are underestimated by non-ML experts.

Most ML algorithms assist in finding rules or patterns in very high dimensional data, *e.g.*, millions of 3D molecular coordinates of atoms of amino acid residues, and reduce these rules to likely (low dimensional) predictions of the behaviour of an investigated system, *e.g.*, a (single) enzyme activity. The downside of current computer systems is that they have no experience and understanding of our experimental work, the objects of investigation, the relation of parts of an experiment and the conditions under which a certain experiment was performed and data acquired. All this information about the details and the process of experiments that lead to a certain outcome is called “meta-data”. Meta-data is therefore all information required for an uninstructed entity (like a machine) to reproduce a scientific experiment: exactly what components (*e.g.*, down to the lot of a chemical), what devices (*e.g.*, exact type and firmware version), what experimental conditions (*e.g.*, temperature, pH, pressure, solvents, metabolites, metal ions), the sequence of operations, and additional information about the experiment itself, such as who did the experiments where and when, is meta-information that helps a machine to relate the outcome of different experiments and judge possible sources of error. The predictive power of “learning” machines is highly related to the quality and reliability of the data – and especially the meta-data – it “learned” from.^{14,15} Therefore, the future success of the application of ML strategies, also in protein engineering and biocatalysis, will heavily rely on proper data generation in conjunction with association of as much meta-data as possible. Meta-data and data linkage are the keys to good predictions since they describe many aspects under which a certain data set was generated and puts data into a context that can be “understood” by a machine.

If not instructed, machines have no notion of meaning, the “semantics”, of these objects and relations. With a formal language derived from the theories of logics and reasoning, we can provide a tiny subset of descriptive relations to objects and operations of experiments that enable the machine to perform a limited amount of logical reasoning and “understanding” of an underlying experiment or data set. This subset is called “ontology”, which is derived from the ancient Greek terms “being” and “logical discourse” – so a logical discourse about the being, *e.g.*, of an experiment or measurement. Creating more and more fine grained logical descriptive term networks

(“ontologies” – Google.com also calls them “knowledge graphs”) and associating them with the performed, biochemical experiments will help machines to more and more “understand” the underlying experiments and perform logical reasoning and relating information from very different sources (*e.g.*, international databases such as UniProt, PDB, BRENDA, KEGG, PubMed, and PubChem) to particular experiments. Therefore, the development of precise and consistent ontologies in the realm of protein engineering or biocatalysis and connecting these terms with ontological terms from related disciplines, such as chemistry, molecular biology, and biology, is paramount for a very powerful aggregation of data to further improve the predictive power of machine-based reasoning and ML approaches. This also requires not only a standardised terminology and (fitting) ontologies, but also standardised means of (meta-)data transfer and a standardised query language. Good sources for actively developed ontologies are the EMBL-EBI Ontology Lookup Service (<https://www.ebi.ac.uk/ols/index>), ontologies of the Pistoia Alliance (<https://www.pistoiaalliance.org/>), and Allotrope foundation (<https://www.allotrope.org/ontologies>). Furthermore, a very recent initiative of the German Research Foundation (DFG), NFDI4Cat, is currently assembling a workgroup to develop and refine ontologies for (bio-)catalytic processes, which stresses their importance. In the case of database query language development, the international W3C consortium gathered a standardization workgroup, which is developing the Graph Query Language (GQL, <https://www.gqlstandards.org/>), a very promising candidate to aggregate data from large, public chemistry and biochemistry databases.

A foreboding of the power of combining knowledge of different databases is given by the recently published Enzyme-Miner webtool (<https://loschmidt.chemi.muni.cz/enzymeminer/>) of the Damborsky group.¹⁶ Based on ML predictions, Enzyme-Miner proposes a selection of promising enzymes originating from a broad range of organisms that have a potentially similar activity as a provided template and are very likely solubly expressible.¹⁷ This can dramatically enhance the exploration of new enzyme types that were not in the focus of current research and knowledge.

Another webtool of this kind in the realm of synthetic biology pathway design is Galaxy SynbioCAD (<https://galaxy-synbiocad.org/>).¹⁸ SynbioCAD starts with a retrosynthetic pathway of a target compound. It then proposes the best biosynthetic pathways for the transformation, cloning strategies of the corresponding enzymes into target vectors and finally liquid handling protocols to execute the assembly of the complex vector constructs.

In current ML-based protein engineering approaches, ML experts need to work tightly with biochemists to first extract the relevant features from a protein engineering question and build a numeric feature set from that to train the ML algorithms. This requires a lot of experience and deep knowledge in modelling, the right questions from the ML side as well as some general understanding from the biochemist's side. With good ontological systems in place, many aspects of this model building



processes might be possibly automated in the future, so that ML can become a standard tool for enzyme engineers.

Semantics, derived from ontologies and logical reasoning, is currently a very fast-growing part to improve the information retrieval for (bio-)catalytic information, but it is only one aspect. Another very related aspect that is also rapidly changing and which will influence how we engineer novel enzymes in the future is the development of lab automation standards and data file formats and data exchange protocols. It is very important that these standards are open, freely accessible and royalty free to achieve a global adoption overcoming the monopolism of proprietary island solutions of large instrument vendors. Such free and open lab automation and data standards are developed by the SiLA (<https://sila-standard.com/>) and AnIML (<https://animl.org/>) workgroups to support reproducible and reliable data transfer and automated documentation. The quality of the collected data also highly relies on a reasonable number of replicates, consistency checks, personal responsibility of researchers, a systematic way of reporting data (e.g., Standards for reporting Enzyme Data – STREND¹⁹). It should be also stressed that reports about experiments with negative results are of a very high value for the community – not only that they reduce the number of redundant trials, but also enable machines to learn what did not work. It should also be noted that results might get a completely different context, if our knowledge proceeds: unexplainable results might become very understandable in the future, when new knowledge is created. Results of our biocatalytic exploration should therefore be handled neutral, avoiding human bias.

If ML shall be successful in the future, the machines need standardised, consistent data with the semantics to interpret them. Therefore, a shift in paradigm in the way how we design experiments is required: a high level of automation – with validated, reproducible devices, to get good statistics about the intrinsic variation of a measurement (e.g., three technical replicates of an activity measurement are not enough). Variations related to the expression host, growth medium, composition, expression vector, *etc.*, need to be much more explored.

A shift in thinking will also be necessary in modelling ML questions for protein engineering. Most aspects of current ML in protein engineering have been covered by very recent and exhaustive reviews from the groups of Frances Arnold,⁸ Jiri Damborsky,¹⁵ and Manfred Reetz.¹⁴ Briefly, ML has been successfully applied to improve enzyme activities,²⁰ stereo-selectivity by the AR algorithm²¹ which is related to the promising Adaptive Substituent Reordering Algorithm (ASRA),²² enzyme thermostabilities,¹⁷ and soluble expression using the SoluProt tool.¹⁶ Currently, most ML approaches use too crude assumptions about an enzymatic system despite experimental experience that in many cases, enzyme reactions rely on a very fine-tuned arrangement and dynamics of residues in the active site and very small levels of energy differences play crucial roles in their activities. Enzymes are complex multi-dimensional objects (at least three spatial and one temporal). The exchange of one single amino acid can sometimes completely alter the reactivity or stereo-/enantio-selectivity of an enzyme.²³

Predicting these small effects on a quantum mechanical level seemed for a long time inaccessible. Very recent developments of ML derived/generated force fields with quantum chemical accuracy might very soon lead to much better predictions of enzymatic activities and lead to a deeper understanding of enzyme mechanisms.²⁴

A very contrasting, holistic approach is followed by Alley *et al.*,²⁵ who applied deep learning to approximately 24 million unlabelled natural amino-acid sequences to extract important features, like protein stability and – to some extent – activity.

These unified representations contain short and long term “memories” of a protein sequence, allowing forecasting of structural features with high accuracy even of completely *de novo* proteins unknown in nature. The model contains a hidden knowledge of protein architecture and could be by that a very powerful tool to annotate or even improve protein activity. Very powerful algorithms for predicting protein structures – one foundation of enzyme activity prediction – have been developed by DeepMind, a Google.com company: AlphaFold 2,²⁶ a community version of the AlphaFold algorithms with open source code is maintained by Wendy Billings *et al.*¹⁸ AlphaFold and related algorithms learn and predict distances between pairs of amino acids and bond angles within amino acid residues from large protein structure databases, like the PDB, by feeding deep neural networks. AlphaFold 2 won the 2020 Critical Assessment of Protein Structure Prediction contest (CASP14) with a highest Global Distance Test (GDT) score of 92.4. One of the currently fastest implementations of folding algorithms, based on Recurrent Geometric Networks (RGN), was published by Mohammed AlQuraishi.²⁷ His algorithm can predict protein structures in milliseconds and accuracies in similar orders as the AlphaFold algorithm.

One important aspect of recent ML-based protein engineering is the question of epistatic mutations²⁸ and their pivotal role as enablers for the emergence of new enzymatic activities and means of leaving local activity optima towards higher enzymatic activities. More experimental data exploring folding and enzyme activity transitions *via* epistatic mutations should be extended to explore the full power of this concept.

Enzymes in a biological system, either a cell or an *in vitro* system, also have a very dynamically changing environment, such as different metal ions, protonation states, electron transfers, interactions with small molecules, oligomers and other proteins or (poly-)nucleic acids, which modulate their activities. Therefore, reducing enzymes to the bare one-dimensional sequence information is just a tribute to our present limitations in computational power – and the lack of knowledge of the structural dynamics, molecular interactions and electro-dynamics of the associated components. Advances in high-performance computing and quantum computing might overcome these current restrictions in the future.

Microfluidics

Screening novel catalysts from rapidly growing gene and protein databases and large enzyme mutagenesis libraries requires HTP and UHTP screening platforms, which have thus become



an important topic in protein engineering and biocatalyst discovery. The microtiter plate format (96- and 384-well) currently dominates screening assays and will likely continue to do so for some time. The reasons for this are that microtiter plate readers are commonly available, microtiter plates offer unmatched flexibility in sample manipulation, a vast collection of protocols have been published, and it is easy to increase throughput by robotic automation. In practice, microtiter plate assays are limited to screening several tens of thousands of clones. This can be sufficient if single site-saturation or small combinatorial libraries have to be screened.^{29,30} While these methods are still applicable to larger random and combinatorial libraries, they are not capable of screening a significant fraction of the library diversity.³¹ One solution to this problem is to make the wells smaller, going through 1536-well plates to microcapillary

arrays. Chen *et al.* reported the use of microcapillary arrays for HTP screening. While the flexibility of microtiter plates is lost, this method potentially has much higher throughput than using plates.^{32,33} Continuing this miniaturisation down a third dimension turns the capillaries into droplets. Droplets of an aqueous phase in an oil phase can be created by simple vortexing (polydisperse) or using sophisticated pumps and microfluidic chips.^{34,35} Microfluidic droplets are highly monodisperse, commonly with less than 2% variation between droplets. This allows enzyme activity to be quantitatively measured in droplets that are only a few picolitres in volume.^{35,36} Based on activity measurements, droplets can then be sorted, at rates up to 30 kHz, into active and inactive pools by application of dielectrophoretic force using embedded electrodes (Fig. 2).^{36,37} Fluorescence is the most commonly used readout and it is even possible to simultaneously

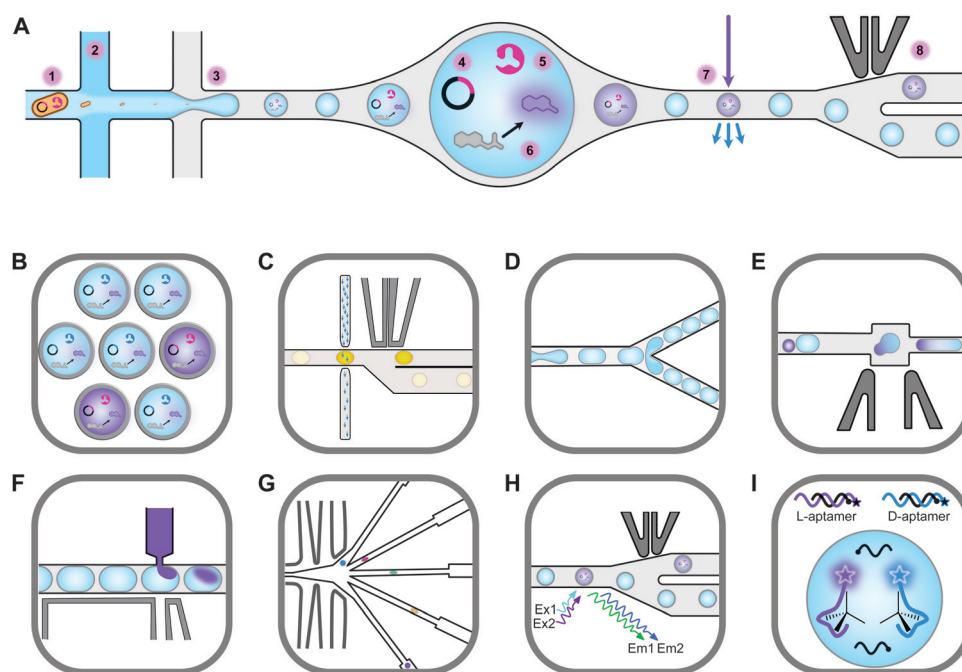


Fig. 2 An overview of droplet microfluidic technologies relevant to ultrahigh-throughput screening. (A) A schematic representation of droplet formation using two aqueous components and subsequent fluorescence activated droplet sorting (FADS). (1) Single cells are injected simultaneously with (2) a mixture of assay components, usually containing a fluorogenic substrate and a lysis agent. (3) A fluorinated oil containing a surfactant is injected into the third inlet. This breaks up the aqueous mixture into monodisperse water-in-oil droplets. (4) The single-cell lysate contains both the genotype (plasmid) and (5) the phenotype (enzyme). (6) Active variants convert the substrate to a fluorescent product, which (7) can be detected using the appropriate lasers and photomultiplier tubes (not shown). (8) A fluorescence signal triggers an electric pulse, delivered through imbedded electrodes. The dielectrophoretic force pulls the droplet from the path of least resistance into the "sorted" channel. DNA is subsequently recovered from sorted droplets by plasmid isolation or PCR (not shown). (B) Droplets can be reinjected into a second chip to make double emulsions. In this setup, inlet 1 would contain the droplets, inlet 2 some oil to space the droplets, and inlet 3 the outer aqueous phase containing a surfactant like Tween.⁴³ The resulting double emulsion droplets can be sorted using flow cytometers, similar to FACS. (C) As an alternative to fluorescence measurements, imbedded fibre optic cables have been used to measure the absorbance of droplets. An increased absorbance signal triggers an electric pulse that moves the droplet into the sorting channel.^{44,45} (D) Droplets can be split, evenly or unevenly, into smaller droplets. This is useful for example if a destructive detection method like mass spectrometry is used, so the first droplet is sacrificed, and the second droplet sorted.^{46,47} (E) Droplets can also be fused, which is useful, for example, for delivering larger volumes of substrate or lysis agents. The chip first aligns droplets and then uses an electric pulse to merge them.⁴⁸ (F) Smaller volumes of liquid can also be added to droplets using a method called picoinjection. Droplets flow past an aqueous inlet and an electric field is used to combine the two aqueous phases.⁴⁹ (G) Droplets can also be sorted into multiple channels³⁸ based on (H) at least two fluorescence measurements.⁴⁰ (I) An interesting and promising approach allows the label free detection of molecules in droplets using RNA or DNA aptamers.^{50,51} A unique feature of aptamers is that they allow different enantiomers to be detected using the L- and D-forms of the nucleic acid aptamers. In the example shown, fluorescently labelled DNA aptamers are hybridised with antisense oligonucleotides labelled with quenchers. Binding to the target molecule separates the fluorophores and quenchers, resulting in a fluorescence signal. This allows both the concentration and enantiopurity of the substance to be determined. Importantly, if an aptamer for one enantiomer is available, the other enantiomer is easily detectable by synthesising the opposite enantiomer of the aptamer.⁵⁰ This figure was inspired by Kintses *et al.* and Neun *et al.*^{52,53}



sort droplets into multiple populations.³⁸ Equally important is the ability to sort droplets based on more than one fluorescence measurement, which facilitates for example the evolution of enantioselective enzymes (Fig. 2G and H).^{39,40} The combination of multiple fluorescence measurements with multiple sorting channels should enable very sophisticated screening, but this has not been applied to directed evolution.^{32,38,40–42} The major drawback of these advanced methods is that they require complex setups that are intimidating if not unavailable. Few laboratories have the necessary expertise in microfluidics, optics, electronics, and programming to set up fully functional sorting rigs. Fortunately, there is an alternative that is accessible to a much larger number of laboratories. The actual droplet generation step is relatively straight forward, and the necessary droplet-generation chips are commercially available. Instead of using electronic sorting chips, water-in-oil droplets can be subjected to a second droplet formation step, generating water-in-oil-in-water or “double emulsion” droplets (Fig. 2B). The advantage of these droplets is that they are stable enough to be sorted using standard flow cytometry equipment, just like cells.⁴³

Despite the technical simplicity of double emulsion sorting by flow cytometry, droplet microfluidics has made a relatively minor impact on biocatalyst discovery and engineering. While droplets have been used to screen hydrolase, oxidoreductase, aldolase, transferase, and isomerase activities, the hydrolases dominate by far (lipase, esterase, phosphatase, phosphonate hydrolase, sulfatase, β -glucosidase, β -galactosidase, and more).^{42,53–56} The reason for this is simply that most droplet sorting systems require a fluorescent signal and that it is relatively simple to design and synthesise fluorogenic hydrolase substrates. Four years ago, the Hollfelder group broadened the applicability of droplet sorting by introducing absorbance-activated droplet sorting (AADS) (Fig. 2B). They used a chip with embedded optic fibres to measure the absorbance of and sort individual $\sim 80\ \mu\text{m}$ droplets at a rate of about 300 Hz. This enabled them to evolve an NAD^+ -dependent amino acid dehydrogenase.⁴⁴ AADS has attracted significant attention and the paper has been highly cited. Just recently, the Hollfelder group published the first follow up AADS papers, albeit for the same type of reaction and detection system.^{45,57} As far as we know, no other group has published the use of AADS for directed evolution. There might be several reasons for this. The droplets used are rather large ($\sim 80\ \mu\text{m}$ compared to $\sim 20\ \mu\text{m}$ for fluorescence sorting, representing a >50 -fold larger volume), resulting in a much lower final enzyme concentration and making the method unsuitable for the detection of very low activities. Furthermore, the detection limit was about $10\ \mu\text{M}$ for a strongly absorbing formazan dye (extinction coefficient $>37\,000\ \text{M}^{-1}\ \text{cm}^{-1}$), meaning that much higher concentrations of dyes with lower extinction coefficients, like 4-nitrophenolate ($18\,500\ \text{M}^{-1}\ \text{cm}^{-1}$), would be needed. Unfortunately, this brings us to a much bigger problem, which affects not only absorbance assays but also assays based on fluorescence or any other type of detection system. Despite being predominantly charged at alkaline pH, 4-nitrophenolate is known to “leak” between droplets, meaning that it can transiently enter the oil phase and then move to

neighbouring droplets.⁵⁸ Some dyes leak within seconds (aminocoumarin), some in minutes (rhodamine 6G), some over hours (resorufin) and some over days (fluorescein).^{59,60} While adding charged groups to a dye is known to dramatically slow leaking (from seconds to days), the case of 4-nitrophenolate demonstrates that this is a complex and often counterintuitive phenomenon.^{58,60} Not only hydrophilicity but also size matters, demonstrated by the ability of water molecules to diffuse over the fluorinated oil barrier rather easily.⁴³ For simple hydrophobic molecules like haloalkanes there is a correlation between hydrophobicity ($\log P$) and a tendency to partition into the oil phase.⁶¹ However, for more complex structures like fluorescent dyes, leakage seems to be related to surfactant concentration, with lower surfactant concentrations slowing down the leaking process, probably due to lower rates of micellar transport.⁵⁹ Unfortunately, surfactants are needed to facilitate droplet formation and to stabilise droplets during incubation. Sindy Tang's group introduced the use of amphiphilic silica nanoparticles to address this problem. The partially fluorophilic nanoparticles adsorb to the aqueous phase, forming very stable pickering emulsions.^{62,63} Unlike surfactants, the nanoparticles cannot escape from the droplet surface and can therefore not facilitate the transport of molecules between droplets. This exciting technology has been commercialised by Dolomite microfluidics as a ready-to-use mix in HFE-7500 called Fluoro-Phase. While this approach, combined with alternative oils like perfluoro(methyldecalin), has been demonstrated to significantly (but not completely) reduce leakage, we are not yet aware of a publication describing the use of Fluoro-Phase in high-throughput screening (HTS).^{62,63}

Beyond leakage, limited detection options are certainly one of the most serious challenges of microfluidic droplet sorting. As mentioned before, fluorescence and absorbance assays are not always applicable, and when they are, the use of chromogenic surrogate substrates tends to bias the screening outcome (“you get what you screen for”).⁶⁴ Therefore, there is intense interest in developing alternative detection strategies. Surface-enhanced Raman spectroscopy, light scattering, image analysis, mass spectrometry, impedance measurements, electrochemical detection, and even NMR have been used to detect droplet contents. Wang *et al.* recently reported Raman-activated droplet sorting (RADS) at a frequency of about 1–2 Hz. They achieved sensitive detection of intracellular triacylglycerols by using an electric field to temporarily halt a moving cell, allowing enough time for an accurate Raman measurement. They claim to have detected intracellular TAG levels previously undetectable using fluorescent stains like Nile Red. However, because they do measurements on single cells before droplet encapsulation, it is not clear whether this method will find use for analysis of products that do not accumulate intracellularly.⁶⁵ Mass spectrometry (MS) has been used to detect droplet contents at rates up to 30 Hz,⁶⁶ and sorting was reported at rates of about 6 Hz, using electrospray ionization (ESI)-MS to analyse $\sim 15\,000$ droplets in 6 h.⁴⁶ While droplets containing an *in vitro* expressed transaminase could be sorted, the large droplet volume (25 nanolitre compared to 2 picolitre for



fluorescence activated droplet sorting (FADS)) means that this detection strategy would not be applicable to biocatalysts derived from single DNA molecules or single cells.⁵³ However, recent advances in *in vitro* DNA amplification, transcription, translation, and assay in microfluidic droplets could address this issue.⁶⁷ Another limitation of MS detection is that it is destructive, so droplets have to be split before analysis. Droplets stored in a delay line are later sorted based on the outcome of the MS analysis.⁴⁶ This means that the sequence of droplets is critically important and that the fusion of any two droplets would desynchronise the droplet trains, resulting in the loss of all hits. While this problem can be dealt with by co-injection of reference droplets, the system is far from straight-forward. An excellent summary by Neun *et al.* shows that despite all the progress, only fluorescence, absorbance, and electrochemical detection has been used for the sorting of actual directed evolution-derived or metagenome libraries.⁵³ Furthermore, fluorescence detection is still unique in being the only format that uses droplets of only a few picolitres (high enzyme concentration) and is capable of detecting low nanomolar product concentrations and sorting at frequencies of several kHz. High-speed absorbance measurements on picolitre and femtolitre droplets is possible using differential detection photothermal interferometry (DDPI). This recent technology allows 100 picolitre droplets to be analysed at 1 kHz, with a detection limit of 1.4 μM for Erythrosin B ($82\,500\text{ M}^{-1}\text{ cm}^{-1}$).⁶⁸ DDPI therefore has the potential to dramatically expand the scope of AADS. However, due to its complexity, the technique will likely remain limited to a few specialist laboratories.

Smart libraries, powered by advances in rational design and DNA synthesis, reduce but have yet to abolish the need for UHTP screening in directed evolution. Furthermore, HTP screening will remain important for the identification of novel biocatalysts from metagenome libraries.^{42,69} Until DNA synthesis becomes significantly cheaper than it currently is, functional metagenomics will remain essential for exploring the rich functional diversity of nature, which is more relevant to this review than UHTP screening of mutant libraries. Therefore, continued research and development is critical, and it is important that the basic techniques become more accessible to larger numbers of researchers. Commercialisation of key technologies would certainly facilitate this process. Affordable commercially available devices capable of generating and sorting droplets based on fluorescence measurements, combined with simple and user-friendly software, would significantly encourage more researchers to start working with droplet microfluidics.

Biosensors

Traditional HTP screening methods for enzyme activity mainly use absorption, ultraviolet, or fluorescence spectroscopy to determine the concentrations of substrates, products, or cofactors. However, many of these methods cannot meet the requirements of UHTP screening and it is also difficult to accurately evaluate the performance of each enzymatic step in a multi-enzyme cascade in real-time. Likewise, efficient manufacturing by engineered microbial cell factories regularly suffers from metabolic imbalances

arising from the introduction and expression of (heterologous) pathways, impaired microbial growth due to the drain of cellular resources, the cytotoxicity of substrates and metabolites, as well as unintended contextual effects, ultimately lowering yields and product titres.^{70–72} Hence, fine-tuning of (heterologous) pathway expression, maintaining cell viability, and maximising metabolic fluxes towards the desired product are equally important.^{70,72} The performance of microbial cell factories is usually assessed by the qualitative and quantitative analysis of the metabolites produced through chromatographical methods. However, these can be time-consuming at only low to moderate sample throughput, therefore, impeding the development and optimisation of microbial cell factories.^{73–75}

This bottleneck has been addressed by the implementation of genetically encoded biosensors such as (allosteric) transcription factors (TFs) or riboswitches (Fig. 3). Although TFs have already long been used to construct inducible gene expression systems for different prokaryotic and eukaryotic hosts,^{76–79} their added value as biosensors for the detection of small molecules has only been recognised in the last decade.^{80–82} Since then, genetically encoded biosensors facilitated the directed evolution of enzymes^{83–86} and the engineering of (natural) metabolic pathways by the high-throughput detection of metabolites,^{87–89} as well as the dynamic regulation of genetic circuits to improve overall pathway performance, among other applications.^{71,80,90–94}

To function as biosensors, TFs contain a ligand-binding domain (LBD) and a DNA-binding domain (DBD). The LBD detects the presence of a chemical compound in the environment or a metabolite inside the cell, whereas the DBD facilitates the association with the cognate nucleotide sequence or the dissociation of the TF upon binding of a ligand (Fig. 3A). TFs can act as transcriptional activators or repressors and have been successfully identified by gene expression and protein profiling in the presence of a desired small molecule through combined transcriptome and proteome analyses^{95–97} and the computer-assisted mining of databases.^{74,98,99} Furthermore, (microbial) TFs can be responsive to different but structurally related compounds, which has inspired both random mutagenesis and rational design of various LBDs and DBDs.^{100–105} For example, the TtgR regulatory protein from *Pseudomonas putida* was engineered by directed evolution and subjected to repeating rounds of FACS, yielding variants with enhanced response to resveratrol.^{102,103} The Keasling group employed a chemoinformatic approach inspired by small molecule drug discovery. By scouting catabolisable chemicals with molecular shapes similar to the metabolic engineering target and subsequent gene cluster analysis, the ChnR/Pb TF-promoter pair was identified as a suitable biosensor for lactams.¹⁰⁶ Besides the identification of TF-based biosensors, improving their performance in terms of selectivity, sensitivity, and operational range can be challenging since additional regulatory elements in the 5' and 3' untranslated region (UTR) including (natural and synthetic) promoters, the context of RBS, and transcriptional terminators will influence the functionality of the biosensor.^{71,72,80} The operational range is defined as the concentration



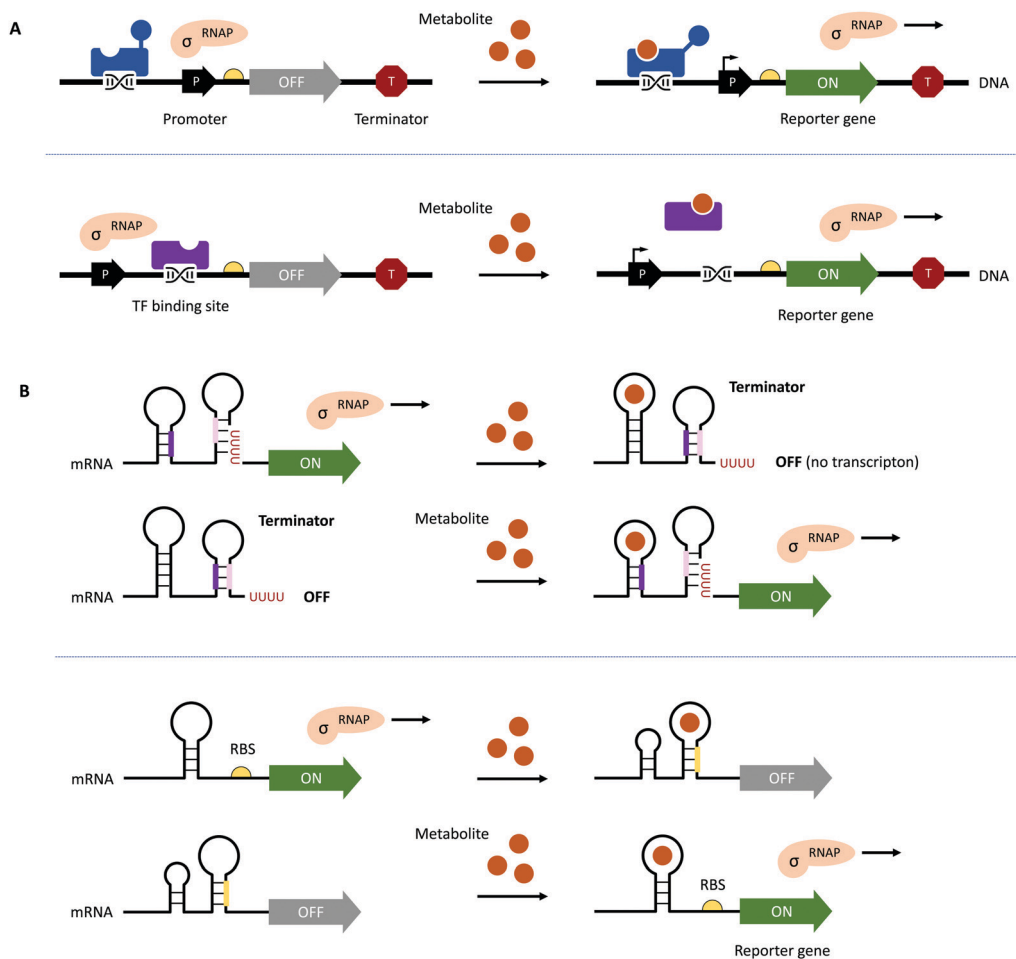


Fig. 3 Genetically encoded biosensors. (A) (Allosteric) TFs which bind metabolites can act as transcriptional activators (shown in blue at the top) or repressors (shown in purple at the bottom). TFs can also recruit other activators or repressors to regulate the activity of RNA polymerase (RNAP; not shown).⁷¹ (B) Upon binding a ligand or metabolite, riboswitches can act on the levels of transcription and translation by the formation/resolution of a terminator hairpin (top) and the sequestration/release of the ribosome binding site (RBS, bottom, shown in yellow).

of the metabolite of interest (*i.e.*, the input signal) required for the biosensor to provide a significant change in the output signal (*e.g.*, fluorescence).⁸⁹ It is not only affected by the attributes of TF including affinity for the ligand and the cognate DNA sequence and its concentration to saturate operator binding sites; expression levels of the reporter have to be carefully adjusted as well. Alternatively, the output signal can be amplified by the implementation of an enzymatic reporter.^{71,89} The dynamic range of the output signal can be expanded by TF, riboswitch, promoter, and RBS engineering.^{94,107–114} Curated databases for prokaryotic TFs, their associated regulatory elements and target genes may be helpful for initial designs but might not avoid the necessity of iterative rounds permutating different combinations of genetic parts.^{71,99,115,116} All these strategies are often time-consuming and results non-intuitive. Berepiki *et al.* addressed this issue and used a design of experiments (DoE) methodology to efficiently map gene expression levels and provide biosensors for protocatechuic acid and ferulic acid with maximised signal output, improved dynamic range, expanded sensing range and sensitivity.¹¹⁷

In contrast to the multicomponent design of TF-based biosensors, riboswitches comprise of an RNA aptamer – an

oligonucleotide sequence with a length of 30–80 nucleobases located in the 5'- or 3'-UTR of mRNA – specifically binding a target molecule (Fig. 3B). Due to the physical proximity, binding of the ligand leads to a conformational change, which can directly affect the binding of ribosomes to RBS on the mRNA upstream of a reporter gene or facilitate the formation of a terminator.^{71,107} Although riboswitch-based biosensors seem to have a simple architecture and exclusively act at the post-transcriptional level, their rational design is still in its infancy due to the limited understanding of ligand-induced structural changes and the frequently encountered small operational window of riboswitches.^{118–120} Accessible sources for riboswitches are the RiboD,¹²¹ Rfam,¹²² and RiboGap¹²³ databases, compiling information about prokaryotic riboswitches and their ligands, sequence alignments and conserved secondary structures, and intergenic regions harboring noncoding RNAs and Rho-independent terminators, respectively.

Recently, Calero *et al.* connected a synthetic fluoride-responsive riboswitch (FRS) to the induction of artificial metabolic pathways for the biosynthesis of fluoronucleotides and fluorosugars in engineered *P. putida* using inorganic fluoride as



both the only fluorine source (*i.e.*, substrate) and as the inducer of the genetic circuit.¹²⁴ The FRS post-transcriptionally (Fig. 3B, bottom) binds fluoride ions, which triggers the translation of the orthogonal T7 RNA polymerase, subsequently enabling the T7 promoter-controlled production of fluorinases and a purine nucleotide phosphorylase.

Regarding the construction of artificial riboswitches, the systematic evolution of ligands by exponential enrichment (SELEX) has been successfully employed. During SELEX, a library of oligonucleotides specifically binding a target ligand or ligands are produced, selected, and enriched *in vitro*.^{73,125,126} Furthermore, natural riboswitches can be engineered like TFs and enzymes.¹²⁷ Examples include the engineering of a set of riboswitch-based genetic devices to enable the control of gene expression according to changes in the environmental pH¹²⁸ and the switching of a thiamine pyrophosphate-sensing riboswitch from a device for the repression of downstream genes to an activator.¹²⁹ Lastly, the physicochemical stability of DNA can be used to detect natural products such as biotin, vitamin D, and folate at nanomolar levels by strand displacement reaction-based biosensors, which have been shown to exhibit increased sensitivity, low interference, and high controllability.^{130,131}

The application of (small-molecule) biosensors and the development and engineering of new sensory devices is certainly of interest for different industries to meet performance criteria through the directed evolution of enzymes,^{132,133} for the optimisation of microbial cell factories,^{134–136} and the real-time monitoring of the production of target molecules¹³⁷ including (aromatic) alcohols, aldehydes, and acids,^{112,113,138–141} precursors for the synthesis fatty acids and their derivatives,^{84,92,93,142–148} isoprene and terpenoids,^{149,150} steroids, as well as flavonoids. Biosensor systems for the last two will be highlighted in the following.

Steroids are polycyclic and highly functionalised compounds and their production is of high interest due to their broad significance as active pharmaceutical ingredients (APIs). However, the synthesis of steroids is demanding and often low-yielding and advanced bio-based procedures are desirable as addressed later in this review. The analysis of steroids usually involves time-consuming sample preparation and analysis by chromatographic methods that limits sample throughput and the efficient development of production strains. Consequently, the development of biosensors for steroids offers advantages to established methods. Mazumder and McMillen constructed a dual-mode promoter in yeast that comprises five steroid hormone responsive elements and one *lac* operator upstream and downstream of the TATA box, respectively, in a minimal cytochrome *C* promoter. This dual controller is activated by testosterone (see also Scheme 8) and repressed by IPTG.¹⁵¹ More recently, Chamas *et al.* created biosensors for the detection of estrogens, progestogens (see also Scheme 8), and androgens in *Arxula adenivorans* yeast strains by coupling human hormone receptors and different fluorescent reporter proteins.¹⁵² A complementary approach was followed by Maser and Xiong and put into perspective of alternative steroid-sensing methods.¹⁵³ Their *Comamonas testosteroni* steroid-sensor (COSS) system is based on the insertion of a green fluorescent protein (*gfp*) gene upstream

of the regulatory region of the *hsdA* gene encoding a 3α -hydroxysteroid dehydrogenase/carbonyl reductase. Upon steroid exposure, GFP is produced. Disadvantages of the COSS assay were the high background of fluorescence observed in both cellular and cell-free assays. Lastly, the Galagan group identified a progesterone-sensing TF from *Pimelobacter simplex* by exposure of cultures to different steroids, subsequent RNA sequencing, and bioinformatic analysis. The allosteric TF was ultimately implemented into an optical biosensor consisting of quantum dots coated with the TF and oligonucleotides. The latter resemble the TF binding site and are conjugated to a fluorescence resonance energy transfer (FRET) acceptor. Upon ligand binding to the TF, the DNA probe is released and the FRET signal quenched, corresponding to the concentration of progesterone.¹⁵⁴ Optical and other emerging strategies for the design of biosensors for the detection of natural products were recently reviewed by Piroozmand and co-workers.¹⁵⁵

For similar reasons, the synthesis and sensing of flavonoids has gained attention in the last decades. Hence, studies aimed at the design of flavonoid-biosensors and the improvement of microbial production. Siedler and colleagues developed a FdeR-based biosensor for naringenin and a QdoR-based sensing device to detect quercetin and kaempferol in real-time.¹⁵⁵ Recently, De Paepe *et al.* followed two strategies for the development of chimeric LysR-type biosensors with customised ligand specificities towards the flavonoids naringenin, apigenin, and luteolin. The first strategy involved the construction of chimeric promoter regions to tune TF binding; the second approach created chimeric TFs by engineering and customization of the LBDs.¹⁵⁶ Although DBDs and linker sequences connecting them to the LBDs as well as chimeric TFs were constructed previously,^{101,105,157,158} the combination of both strategies certainly points towards the expansion of the repertoire of (chimeric) biosensors for the detection of flavonoids and other natural products.¹⁵⁹

Thus, biosensors are a powerful tool not only for the engineering of enzymes and the set-up of HTS by monitoring the presence of metabolites in real-time; biosensor systems can time and precisely control the expression levels of pathway enzymes.⁷¹ Most of the selected examples of biosensor applications in living cells sensed target molecules in the cytosol. However, recent efforts have been made to sense natural products secreted into the extracellular environment as well.^{137,146,160–162} An elegant biosensor set-up was realised by Mukherjee *et al.* who coupled a medium-chain fatty acid (MCFA)-responsive G protein to a receptor on the cellular membrane, enabling the transduction of subsequent signals in the presence of extracellular MCFAs.¹⁴⁷ Similarly, the group of Peralta-Yahya engineered a human serotonin G protein-coupled receptor to detect serotonin secreted by a serotonin producing yeast strain.¹⁶³

Remaining challenges involve the contextualization of novel biosensor designs in terms of their operational range and the functional implementation in heterologous hosts, especially transferring prokaryotic TFs into eukaryotic hosts.^{164,165} The continuous advancements in bioinformatics and synthetic



biology provide a solid foundation to discover TFs and riboswitches and their cognate natural genetic parts or rationally combine them with artificial regulatory elements. DoE methodologies have already shortened this process.¹¹⁷ Furthermore, the combination of TFs and riboswitches that complement each other's shortcomings have emerged as exemplified by Wang *et al.* who reported a hybrid controller consisting of a riboswitch-based detector and a protein regulator for compensating the low dynamic range of the riboswitch.¹¹¹ Current and future strategies address feedback control and aim at synchronizing a cell population, reducing the metabolic burden, and balancing the expression of multiple pathway genes depending on the input signals.^{166–169} Ceroni *et al.* designed a dCas9-based feedback-regulation system in which the promoter automatically adjusts the downstream gene expression in response to burden¹⁶⁷ and Liu *et al.* constructed quorum sensing-controlled CRISPRi systems, which can dynamically program bacterial consortia.¹⁶⁶ To control multigene expression by one chemical signal, Cunningham-Bryant and co-workers reported a genetic controller that consists of catalytically inactive Cas9 and an RNA-binding protein fused to an inducible TF.¹⁶⁸ These last examples not only highlight the versatility and variations of the CRISPR/Cas9 technology (which won the Nobel Prize in Chemistry in 2020) but showcase how far our understanding of biosensors as integral parts in regulatory networks and associated metabolic pathways has already advanced.

Novel enzymes and enzymatic reactions leading to total synthesis of complex natural products

Natural products have a wide range of pharmacological activities to humans. However, the extraction of natural products from original organisms often requires huge planting and breeding resources with limitations to season and region. Therefore, heterologous biosynthesis of natural products *in vitro* or in model microorganisms is a more promising way and has been a research hotspot for decades. At present, the use of heterologous pathways for natural products still mainly replicates the biosynthetic pathways present in the original hosts. However, due to the dependence of enzyme expression and coenzyme/energy cycles in the original host organism and the intracellular environment, the same synthetic pathway is hard to replicate in model microorganisms or *in vitro*. Therefore, some optimisation strategies from the viewpoint of biochemistry and organic chemistry are widely used, including but not limited to (i) discovery of novel enzymes and protein interactions in the original metabolic pathways with well-studied catalytic and interaction mechanisms to provide novel enzyme tools for synthetic pathway design, (ii) substitution of enzymes in natural synthetic pathways with isoenzymes derived from microorganisms, (iii) replacement of natural synthetic pathways with microbial degradation pathways and artificial enzymatic cascades, (iv) design of artificial coenzyme regeneration pathways, and (v) protein engineering of key enzymes by rational design and/or directed

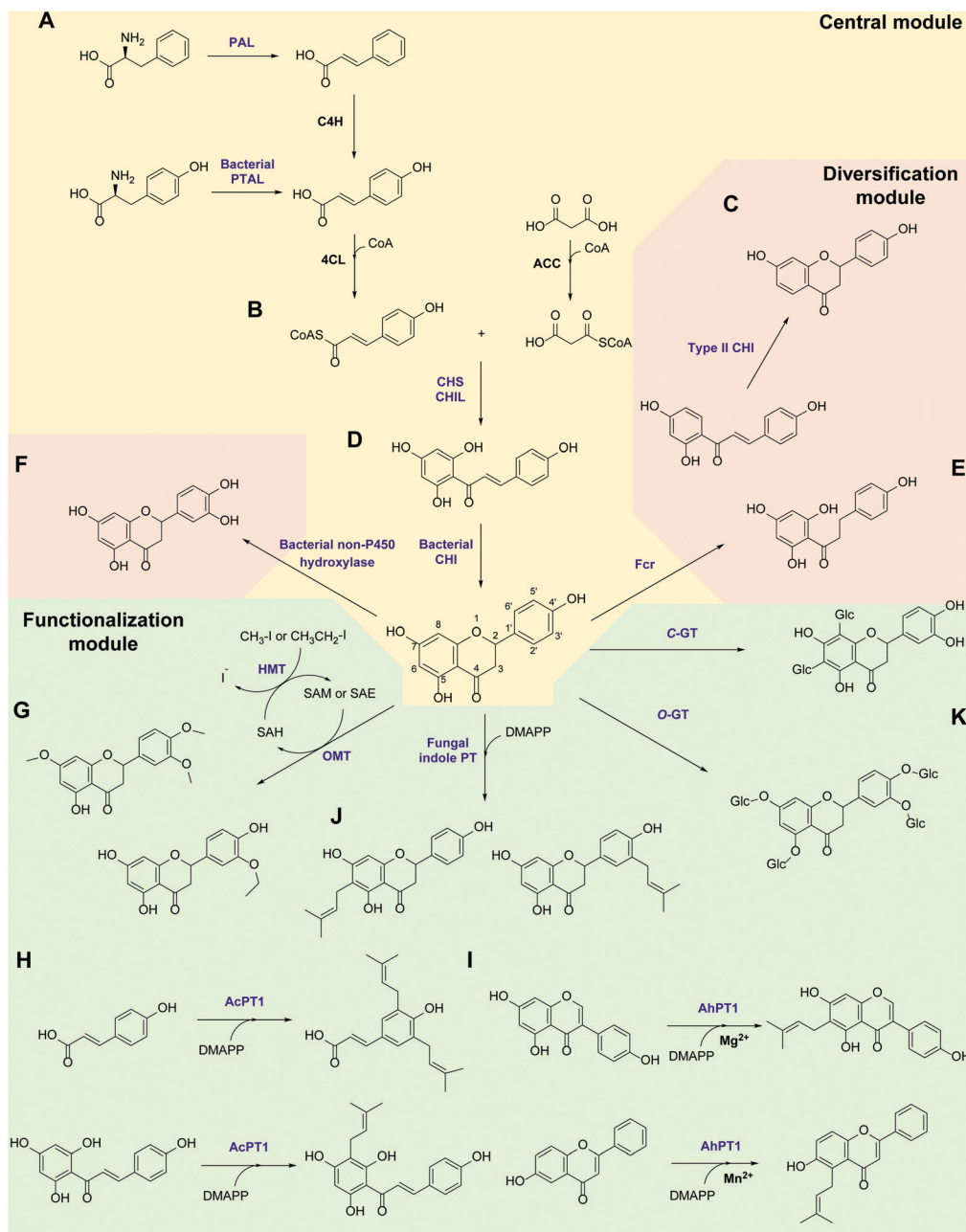
evolution to improve their activity, specificity, substrate scope as well as expression level and stability. Here, we only concentrate on some excellent examples to elaborate the key effects of novel enzymes, novel enzymatic mechanisms, and novel enzymatic cascades on the optimisation of synthetic pathways for the heterologous and *in vitro* biosynthesis of typical natural products.

Flavonoids

Flavonoids are some of the main secondary metabolites of plants, which are a class of polyphenol compounds mainly containing a C₆–C₃–C₆ backbone with more than 9000 structures found in nature.¹⁷⁰ Besides their important roles in the physiology of plants, flavonoids also have a variety of pharmacological effects on humans and are one of the main active substance classes in herbal medicine.¹⁷¹ At present, the heterologous biosynthetic pathway of flavonoids mainly contains the precursor module for the synthesis of *p*-coumaric acid (*p*-CA) and malonate, a central and diversification module for the synthesis of dihydrochalcones, flavanones, flavanonols, flavones and flavanols, and a functionalization module for hydroxylation, methylation, prenylation, glycosylation, acetylation, sulfonation or glucuronidation (Scheme 1).^{172–176} However, the metabolic pathways of flavonoids are not limited to plants. Biosynthesis and biodegradation pathways of flavonoids have also been found in fungi and bacteria, bringing novel enzyme tools and *de novo* synthetic pathways to the heterologous biosynthesis of flavonoids.

In the precursor synthesis pathway, the biosynthesis of *p*-CA mainly relies on phenylalanine or tyrosine as starting materials (Scheme 1A). Aromatic ammonia lyases catalyse the deamination of phenylalanine or tyrosine to form cinnamic acid or *p*-CA, respectively. Cinnamic acid can be further hydroxylated by cinnamate 4-hydroxylase (C4H) to obtain *p*-CA. Phenylalanine ammonia lyase (PAL) and bifunctional phenylalanine/tyrosine ammonia lyase (PTAL) have been identified in dicotyledonous plants and some monocots, respectively.^{177,178} Both have quite high activity towards phenylalanine, and the latter can also accept tyrosine but with lower affinity.¹⁷⁷ Plant-derived PALs and PTALs are good biocatalyst candidates for the synthesis of *p*-CA,¹⁷⁸ but may require chaperones for better expression in prokaryotic hosts, which is a limiting factor for their synthetic application.¹⁷⁹ On the other hand, PAL, PTAL and monofunctional tyrosine ammonia lyase (TAL) have been reported in microorganisms, especially fungi and bacteria that produce antibiotic phenylpropanoids or utilize phenylalanine and tyrosine as carbon and nitrogen sources.^{177,180} Among them, the PTAL from *Rhodotorula glutinis* (RgPTAL) shows impressive activity towards tyrosine¹⁸¹ and thus has been used for the biosynthesis of several phenylpropanoids.^{182–184} Furthermore, a mutant of RgPTAL (S9N/A11T/E518V), obtained through random mutagenesis, probably anchors the flexible loop region (Glu325–Arg336) to maintain the active-site pocket opening which ensures easy access by tyrosine and thus significantly improved its activity and the yield of *p*-CA.¹⁸⁵ The regioselectivity of ammonia lyases to phenylalanine and tyrosine is also





Scheme 1 Novel enzymes, enzymatic mechanisms and cascades in flavonoid synthesis. PAL: phenylalanine ammonia lyase; C4H: cinnamate 4-hydroxylase; PTAL: bifunctional phenylalanine/tyrosine ammonia lyase; CHS: chalcone synthase; CHIL: non-catalytic chalcone isomerase-like protein; CHI: chalcone isomerase; Fcr: flavanone- and flavanone-cleaving reductase; OMT: O-methyltransferase; HMT: halide methyltransferase; PT: aromatic prenyltransferase; GT: glycosyltransferase.

affected by the amino acid residue that binds to the *para*-hydroxyl group on the benzene ring. Substitution of this position by polar residues can obviously increase the specificity of PAL towards tyrosine.¹⁸⁶ Recently, two novel TALs were identified from actinomycetes and achieved the productivity of *p*-CA up to 2.88 g L⁻¹ h⁻¹ using recombinant *Escherichia coli* as a whole-cell biocatalyst, which currently represents the highest efficiency for microbial production of *p*-CA.¹⁸⁷ Therefore, these microbial-derived ammonia lyases and their mutants have become the preferred enzymes for the synthesis of *p*-CA in

the precursor pathway with significantly higher expression level, catalytic activity and *p*-CA yield.

In the central synthesis pathway, chalcone synthase (CHS) catalyses the synthesis of a chalcone, 2',4,4',6'-tetrahydroxy-chalcone (THC), from *p*-coumaroyl-CoA and malonyl-CoA (Scheme 1B).¹⁸⁸ CHS is a plant-specific promiscuous type III polyketide synthase (PKS) which also produces other polyketides such as the *p*-coumaroyltriacytic acid lactone (CTAL). Therefore, how to improve its product specificity is of great significance for optimising metabolic flux and increasing



flavonoid production. Very recently, a conserved strategy was uncovered by which non-catalytic chalcone isomerase-like proteins (CHILs), which are ubiquitous in plants, are able to bind to CHS a rectifier and increased the k_{cat} value (2–15 times higher) for the THC production potentially through binding to the tetraketide-CoA intermediate in an energetically favourable manner, and thus enhance THC production and decrease CTAL formation.¹⁸⁹ Since CHILs perform macromolecular interaction with other enzymes in plant specialised metabolism, this result brings us a revelation that protein–protein interactions could be widespread in the biosynthesis of natural products by a broader effect on promoting the activity and specificity of enzymes and the regulation of metabolic flux, which can provide an important tool for optimising heterologous biosynthesis.

Chalcone isomerase (CHI) catalyses the intramolecular cyclisation of THC and generates the flavanone naringenin which is a key intermediate for the structural differentiation to other flavonoids. Plant CHIs are considered to have evolved from fatty acid binding proteins,¹⁹⁰ which shows the key role of protein evolution in modifying the catalytic mechanism of enzymes and broadening the source of novel enzymes by mutagenesis.¹⁹¹ According to substrate selectivity and catalytic mechanism, plant CHIs can be divided into type I and II. Both can accept 6'-hydroxychalcones as substrates, while the latter also has high activity towards 6'-deoxychalcones (Scheme 1C).¹⁹² Recently, the reaction mechanisms of enantioselective oxa-Michael cyclisation performed by type I and II CHIs have been revealed by X-ray crystal structure and molecular dynamics simulations, wherein the guanidinium ion of a conserved arginine positions the nucleophilic phenoxide and activates the electrophilic enone for cyclisation through Brønsted and Lewis acid interactions.¹⁹³ This mechanism presents a new enantioselective Michael-type reaction in natural product biosynthesis that efficiently constructs C–O bonds. The crystal structure of type II CHI also revealed two unique water molecules in the active pocket which form an ordered hydrogen bond network with the polar amino acids in the pocket. This extended hydrogen bond network supports the role of ordered water in the destruction of the intramolecular interaction between ketone oxygen and 2'-OH and further provides a ring flip of 6'-deoxychalcone. Therefore, the catalytic efficiency towards 6'-deoxychalcone has been greatly improved.¹⁹³ These results provide a theoretical basis for screening novel CHIs and broadening the substrate tolerance of CHIs through mutagenesis.

Besides in plants, CHIs also occur in some anaerobic intestinal bacteria as key enzymes for the degradation of flavonoids. The first bacterial CHI was isolated and cloned from *Eubacterium ramulus* (*ErCHI*), which has activity towards THC, isoliquiritigenin, butein, eriodictyol chalcone, and hesperetin chalcone (Scheme 1D).^{194,195} However, this bacterial CHI has no homology to plant CHIs and is even rare in protein databases, which shows its unique evolutionary origin.¹⁹¹ The protein structure of *ErCHI* consists of two ferredoxin domains as catalytic domains and a solvent-exposed domain.¹⁹⁶ Unlike plant CHIs, the intramolecular cyclisation of chalcones is catalysed by bacterial CHI *via* a reversible Michael addition catalysed by histidine.¹⁹⁶

Therefore, *ErCHI* can also catalyse the isomerisation of the flavanonol taxifolin to the auronol alaphitonin.¹⁹⁷ The study of these novel enzymatic mechanisms shows the impressive diversity of isoenzymes from various sources related to flavonoid metabolic pathways. In-depth studies on bacterial enzymes in the degradation pathway of flavonoids may enable them to replace plant-derived enzymes and enable the design of novel biosynthetic pathways for flavanones. For example, a flavanone- and flavanonol-cleaving reductase (*Fcr*) was recently identified from *E. ramulus*, which is an iron-sulfur flavoprotein containing an intramolecular electron transfer chain. It performs a cofactor-mediated hydride transfer from nicotinamide adenine dinucleotide (NADH) onto C2 of the respective substrate *via* flavin adenine dinucleotide (FAD), a 4Fe–4S cluster, and flavin mononucleotide (FMN), and further directly attaches the C2 of flavanones and flavanonols and cleaves the heterocyclic C-ring, which provides a novel pathway to synthesise dihydrochalcones and dihydroflavonols from flavanones and flavanonols, respectively (Scheme 1E).¹⁹⁸

Hydroxylation and methylation greatly extend the structural differentiation of flavonoids. The hydroxylation mainly occurs on C3 of the A-ring and *para*- and *ortho*-positions of the B-ring catalysed by plant-derived flavanone 3-hydroxylase, 3'-hydroxylase, and 3',5'-hydroxylase, respectively. Because of the low expression of P450 enzymes and the lack of effective electron transport systems in prokaryotic host cells, bacterial hydroxylases, such as an endogenous non-P450 hydroxylase complex from *E. coli* (HpaBC), have shown their advantages in cell factory construction and have been reported to additionally hydroxylate the *ortho*-position of the B-ring to achieve conversion of naringenin and afzelechin to eriodictyol and catechin, respectively, with high yields (Scheme 1F).¹⁹⁹ Moreover, *O*-methylation of hydroxyl groups is a common modification of flavonoids catalysed by *O*-methyltransferases (OMTs) using *S*-adenosyl-L-methionine (SAM) as cofactor for providing the methyl group, which mostly takes place on the 7,3',4',5'-hydroxyl groups of flavonoids and the 7,4'-hydroxyl groups of isoflavonoids. OMTs have been widely found in plants, showing diverse substrate specificity and regioselectivity.²⁰⁰ Meanwhile, some flavonoid OMTs were also discovered in microorganisms, such as *Bacillus* and *Streptomyces*.^{201–203} Many OMTs have been recombinantly produced and used for the biosynthesis of flavonoids due to their superior chemo-, regio- and stereo-selectivity.^{204,205} However, the bulk demand of the methyl group donating cofactor SAM has hindered the industrial applications of OMTs.²⁰⁴ Therefore, *in situ* regeneration of SAM is one of the key factors affecting methylation biosynthesis. A early attempt of SAM regeneration was a complex SAM recycling cascade involving five additional enzymes on the basis of the physiological cycle of the metabolites in cells.^{206,207} *S*-adenosyl-L-homocysteine (SAH) produced after transferring the methyl group is hydrolysed to adenosine and homocysteine by a SAH hydrolase, after which adenosine is sequentially phosphorylated by adenosine kinase, polyphosphate kinases 2 I and II, producing adenosine triphosphate (ATP). After that, SAM is reproduced from ATP and L-methionine by a methionine adenosyltransferase. Although this



is a feasible way to regenerate SAM, such a long and energy-consuming coenzyme regeneration pathway is not suitable for biocatalytic methylation, at least *in vitro*. A newly established and more efficient SAM recycling system consists of only one enzyme, halide methyltransferase (HMT), which produces SAM directly from SAH with methyl iodide as donor (Scheme 1G).²⁰⁸ This novel cascade shows that a simpler cofactor regeneration system can be designed and realised by introducing non-natural donors and off-path tool enzymes, which helps to partly depart from the original metabolic pathway and simplify the biosynthetic pathways of natural products. In addition to methylation, the structural and functional diversity of flavonoids can be dramatically expanded *via* hydroxyl group bioalkylation with SAM analogues and promiscuous MTs. SAM analogues containing different alkyl substituents can be produced *via* chemical method or by a chemoenzymatic methods using L-methionine analogues catalysed by methionine adenosyltransferases or halogenases.^{209–211} A more advanced way is to explore promiscuous or engineered HMTs for the production of SAM analogues and to achieve flavonoid bioalkylation on the basis of the MT-HMT cofactor regeneration system.^{30,212} This artificial cofactor regeneration pathway provides a novel inspiration and solution for solving the problem of low efficiency of SAM regeneration in biosynthesis.

Prenylation is another structural modification for the functionalization of flavonoids catalysed by aromatic prenyltransferases (PTs). Plant-derived PTs generally have high regio-specificity, transferring the prenyl moiety on the C6 and C8 of the final flavonoid skeleton, as well as the C3' of chalcones and the C3 and C5 of *p*-CA in the intermediate biosynthetic step (Scheme 1H).¹⁷⁴ Recently, a novel di-PT was isolated from *Artemisia capillaris* which can accept *p*-CA as its specific substrate and transfers two prenyl residues stepwise to yield artepillin C.²¹³ This is the first plant PT involved in the biosynthesis of phenylpropanes and capable to introduce multiple prenyl residues to native substrates with different regiospecificity. The plant-derived PTs are transmembrane enzymes. Due to the lack of high-resolution protein crystal structures, the substrate binding pocket and catalytic mechanism of PTs are currently unclear, which limits the protein engineering studies of PTs, such as widening the donor-binding pocket to accept longer chain prenyl donors and thereby broadening the diversity of product structures.¹⁷⁴ Plant PTs prefer magnesium ions (Mg^{2+}) to stabilize the pyrophosphate group of the donor. However, a recent study reveals that metal ions can change the substrate specificity of a flavonoid PT from *Artocarpus heterophyllus* (AhPT1). AhPT1 could catalyse 6-*C*-prenylation of genistein when Mg^{2+} served as cofactor but without any activity towards 6-hydroxyflavone. However, 5-*C*-prenylation of 6-hydroxyflavone was identified by AhPT1 when Mn^{2+} was used (Scheme 1I).²¹⁴ This new discovery shows that metal ions play a key role in the substrate specificity, prenylation sites and catalytic mechanism of PTs, rather than just stabilizing the donor. Besides, the prenylation products on the *O*-site have also been found in plants. However, *O*-specific PTs have not been discovered yet, which revealed that *O*-specific PTs might have no homology with the *C*-specific ones.¹⁷⁴ Therefore, the intelligent analysis of genomic, proteomic and metabolomic data could most

likely bring new opportunities for the discovery of PTs with *O*-specificity. In addition, soluble PTs from bacteria show their catalytic capability towards flavonoids and prenylation specificity. For example, indole PT 7-DMATS from the fungus *Aspergillus fumigatus* accepted chalcones, isoflavonoids, and flavanones, and mainly catalysed prenylation at C6, while another indole PT, AnaPT, prefers prenylations at C6 or C3' of flavanones and isoflavones (Scheme 1J). These fungal PTs have replaced plant-derived PTs for the heterologous biosynthesis of prenylated flavonoids.²¹⁵ In addition, dimethylallyl diphosphate (DMAPP) is the preferred donor for PTs and is synthesised through the mevalonate (MEV) pathway and the methylerythritol phosphate (MEP) pathway *in vivo*. Ensuring an adequate donor supply is one of the limiting conditions of prenylation. Besides the optimisation of the natural donor synthesis pathway, one step phosphorylation of dimethylallyl alcohol by acid phosphatase and isopentenylphosphate kinase with ATP as high-energy phosphate donor offers a simplified pathway to improve the efficiency of prenylation.²¹⁶

Glycosylation is a major structural modification for flavonoids to increase their solubility, reduce toxicity, and improve bioavailability. Glycosylation takes place mainly on the multi-hydroxyl groups of the flavonoid structure (3-OH, 5-OH, 7-OH, 3'-OH, 4'-OH and 5'-OH) with glucose, mannose or galactose and their 6-deoxy derivatives, arabinose, apiose, and xylose (Scheme 1K). Some dideoxyhexosides, such as pyranoside and bovine pyranoside, have also been reported as sugar moieties.²¹⁷ In addition, the carbon atoms of the benzene ring can also be glycosylated to form *C*-glycosides. Glycosylation is mainly catalysed by glycosyltransferases (GTs), which generally have high regio- and stereo-selectivity towards donors and acceptors. Therefore, mining novel GTs in whole genomes and the CAZy database (<http://www.cazy.org>) *via* bioinformatic methods is the main concept to find novel enzymes with specificity towards flavonoids.²¹⁸ Meanwhile, protein engineering has been carried out on GTs to excise the transmembrane domain of GTs to improve soluble expression, optimize the substrate binding pocket to extend substrate scope and improve the efficiency of glycosyl transfer, and to reduce the flexibility of enzyme structures to improve the stability of GTs.²¹⁸ The use of transglycosylation activity catalysed by glycosidases is another way to achieve the *O*-glycosylation of flavonoids. Recently, an amylosucrase obtained from *Deinococcus geothermalis* (DgASase) exhibited its unique transglycosylation activity towards various hydroxyflavones and hydroxyflavanones with high site specificity at the 6-OH and 4'-OH positions, leaving the 3-OH and 7-OH positions unchanged.²¹⁹ This provides a reference for catalytic mechanism and glycosylation site specificity for predicting and screening more glycosidases. On the other hand, as more *C*-glycoside flavonoids have been discovered from plants and show great medicinal potential, *C*-glycosylation has become a hot topic in the study of flavonoid glycosylation. Some *C*-GTs derived from plants and fungi were cloned and confirmed to catalyse the *C*-glycosylation of flavonoids at positions C6 and C8 (Scheme 1K).^{220–225} Very recently, a promiscuous *C*-GT from *Aloe barbadensis* was identified to be capable of *C*-glycosylating



scaffolds lacking an acyl group. With dihydrochalcones as substrates, di-*C*-glycosylation can occur at the C6 and C8 positions.²²⁶ Remarkably, a promiscuous *C*-glycosyltransferase from *Trollius chinensis* can accept multiple structures of flavones, flavonols, flavanones, flavanonols and dihydrochalcones, and introduce a glycosyl moiety at the C8 position. Meanwhile, it showed *O*-glycosylation activity on the C7 position when the C8 is already substituted by methoxyl or prenyl moieties.²²⁰ The study of the catalytic mechanism and site mutagenesis at two positions (I94E and G284K) switched its *C*- to *O*-glycosylation, which provides an important reference for the rational design and directed evolution of *C*- to *O*-GTs for synthetic purposes.²²⁰

As natural products with the most extensive pharmacological activity, the potential medical use of flavonoids has recently been expanded to treat infections by the coronavirus.²²⁷ The discovery of each enzymatic step in the natural synthesis of flavonoids and the replacement of designed *de novo* enzymatic reactions/cascades are completing the map of heterologous flavonoid synthesis. Under the guidance of the Design–Build–Test–Learn (DBTL) concept and the application of ML,²²⁷ the construction of *in vitro* biosynthesis and establishing new cell factories for flavonoid syntheses provides an efficient biosynthesis program for natural flavonoids and their novel structural derivatives.

Alkaloids

Alkaloids are a large class of nitrogen-containing natural products with the heterocycle nitrogen atom derived from an amino acid (e.g., tyrosine, tryptophan, lysine or ornithine).^{228–230} The majority of alkaloids originate higher plants with natural functions to interact with other organisms. Thus, many alkaloids are privileged compounds exhibiting biological and pharmacological activities (e.g., analgesic, anticancer, antibacterial, stimulant, etc.). Although the majority of important alkaloids are currently produced and extracted from the native or engineered plants or plant tissue culture, it is more promising to produce them in heterologous microbes (*in vivo*) or *in vitro* with biosynthetic enzymes because of (i) advanced technologies for engineering and optimising microbial production; (ii) much cleaner targeted products with less byproducts, and (iii) the potential to access analogues and derivatives.^{231–236} The recent advances of enzyme discovery/engineering and synthetic biology for several important groups of alkaloids are discussed below.

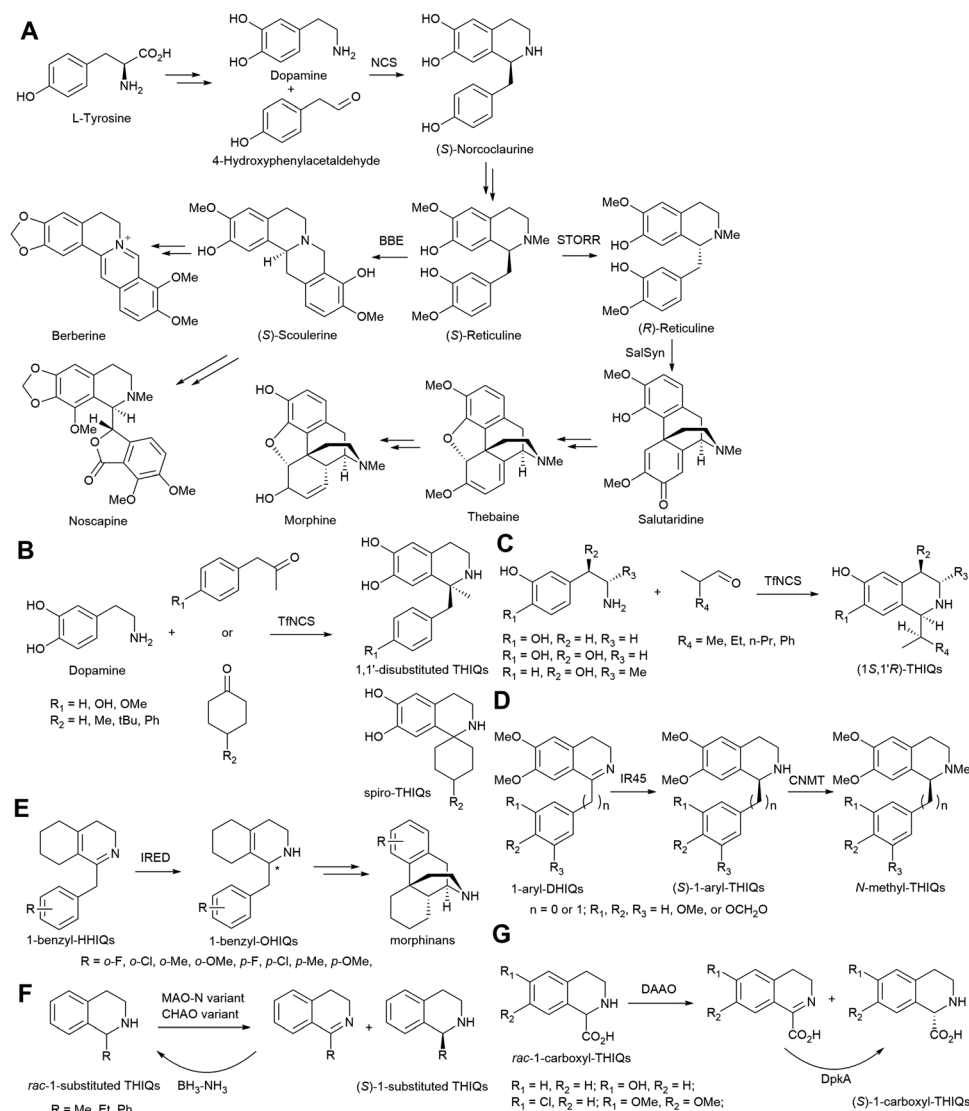
Benzylisoquinoline alkaloids (BIAs) are one of the most important plant alkaloids derived from tyrosine or phenylalanine. Because of several important drugs, including morphine, codeine, berberine and noscapine, the biosynthesis pathways of BIAs have been intensively investigated and almost all the key steps were elucidated in opium poppy recently (Scheme 2A).²³⁷ In brief, dopamine and 4-hydroxyphenylacetaldehyde (both derived from tyrosine) undergo a Pictet–Spengler reaction with a Norcoclaurine synthase (NCS) to yield (*S*)-Norcoclaurine, the first committed intermediate in the BIAs biosynthesis. (*S*)-Norcoclaurine is subjected to a hydroxylation and three methylation steps to form (*S*)-reticuline, the pivotal intermediate of many BIAs. The further synthesis of key BIAs

branches here. For the synthesis of berberine and noscapine, a Berberine bridge enzyme (BBE) catalyses the oxidative C–C bond formation to give (*S*)-scoulerine, which is further transformed into berberine or noscapine *via* enzymes from a 10-gene cluster in opium poppy.²³⁸ For the synthesis of morphine, (*S*)-reticuline is epimerised to (*R*)-reticuline by a unique (*S*)- to (*R*)-reticuline epimerase (STORR) with a fused cytochrome P450 monooxygenase (CYP) domain and a reductase domain, which was revealed by three research groups in 2015.^{239–241} (*R*)-Reticuline is transformed to salutaridine *via* a C–C phenol-coupling mediated by salutaridine synthase (SalSyn), and then further converted to thebaine and morphine *via* several tailoring enzymes.²⁴² The almost full elucidation of enzymes in the biosynthetic pathway of several BIAs has significantly facilitated the engineering of fast-growing microbes for efficient heterogenous production.²⁴³

In 2015, the Smolke group reported the complete biosynthesis of opioids from glucose in yeast, a major milestone in heterogenous BIA production.²⁴⁰ In total, more than 20 enzymes from plants, mammals, bacteria, and yeast were over-expressed to access thebaine and hydrocodone. Although the final titres of thebaine and hydrocodone were only on the 0.3–6.4 $\mu\text{g L}^{-1}$ scale, it represents a ground-breaking advance for the total biosynthesis of complex natural products. One year later, the total biosynthesis of opiates was also achieved *via* stepwise conversion using four engineered *E. coli* strains giving thebaine in 2.1 mg L^{-1} .²⁴⁴ For the other branch of BIAs, the Smolke group reported the first total biosynthesis of the anticancer drug noscapine (2.2 mg L^{-1}) in yeast *via* expression of >30 enzymes from various sources.²⁴⁵ In addition, by feeding 3-halogenated tyrosines, the yeast produced several 8-halogenated (*S*)-*N*-methylcoclaurines and (*S*)-reticulines. A much more practical synthesis of the BIA intermediate was reported very recently: 4.6 g L^{-1} of (*S*)-reticuline was successfully produced from sugar *via* extensive engineering of yeast and using more efficient key enzymes (e.g., NCS).²⁴⁶ Furthermore, by feeding dopamine and different L-amino acids (precursors for aldehydes) to a simplified version of yeast, an array of non-natural tetrahydroisoquinolines (THIQs) were produced, illustrating the broad substrate scope of NCS and methylation enzymes.

Besides engineering BIA pathways in heterogenous hosts, many enzymes (especially the C–C bond forming NCS and BBE) in the biosynthetic pathway could be engineered and evolved for the *in vitro* synthesis of novel THIQs.^{247–249} NCS is well-known for its broad scope for accepting different aldehydes to give THIQs. In 2017, Lichman *et al.* discovered that the TfNCS from *Thalictrum flavum* catalysed the Pictet–Spengler reaction between dopamine and ketones, leading to novel chiral 1,1'-disubstituted and spiro-THIQs (Scheme 2B).²⁵⁰ The 1,1'-disubstituted THIQ was featured with a chiral quaternary carbon centre, which is challenging to form in organic chemistry and unattainable through imine reductases (IREDS) or monoamine oxidases (MAOs).²⁵¹ Several variants of TfNCS were explored to ensure high conversion and preparation of these unique THIQs. The Ward and Hailes groups continued to





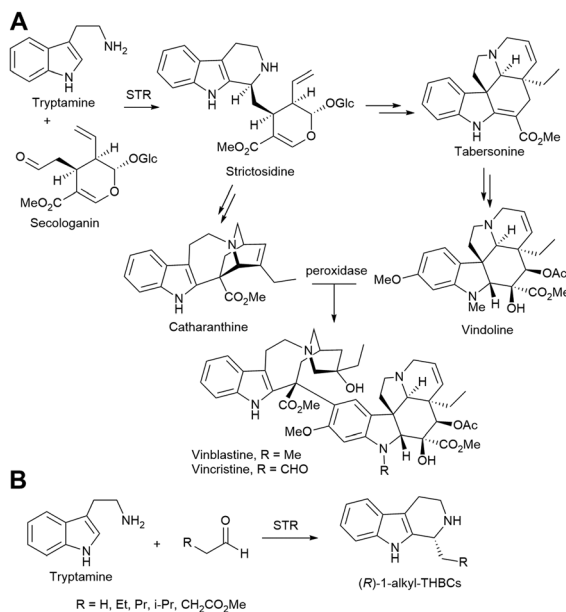
Scheme 2 Novel enzymes, enzymatic mechanisms and cascades for the synthesis of benzylisoquinoline alkaloids and related tetrahydroisoquinolines. NCS: norcoclaurine synthase; BBE: berberine bridge enzyme; STORR: (S)- to (R)-reticuline epimerase; SalSyn: salutaridine synthase; TfNCS: norcoclaurine synthase from *Thalictrum flavum*; IR45: imine reductase from *Streptomyces aurantiacus*; CNMT: coclaurine N-methyltransferase; IRED: imine reductase; MAO-N: monoamine oxidase from *Aspergillus niger*; CHAO: cyclohexylamine oxidase; DAAO: D-amino acid oxidase; DpkA: piperidine-2-carboxylate reductase from *Pseudomonas putida*.

explore TfNCS and its variants for kinetic resolution of α -methyl aldehydes leading to (1S,1'R)-THIQs with two chiral centres in a single step (Scheme 2C).²⁵² The broad scope (aldehyde and ketone), enantioselectivity and mechanism of NCS were investigated and explained in a recent quantum chemical study.²⁵³ Furthermore, NCS was combined with other enzymatic or chemical transformations as *in vitro* (chemo)-enzymatic cascades for the synthesis of THIQ analogues: (i) a carbonylase-transaminase-NCS cascade to access chiral 1,3,4-trisubstituted THIQs;²⁵⁴ (ii) an NCS-catalysed Pictet-Spengler reaction and Na₂CO₃-mediated cyclisation to afford (S)-trolline;²⁵⁵ (iii) a network of tyrosinase, decarboxylase, transaminase and NCS for efficient synthesis of several natural and non-natural BIAs.²⁵⁶ Besides NCS, another important C–C bond forming enzyme, BBE, has been explored for synthetic purposes, such as preparation of

(S)-scoulerine and its analogues *via* kinetic resolution²⁵⁷ or deracemization²⁵⁸ of the corresponding THIQs, and enantioselective dealkylation of N-ethyl THIQs.²⁵⁹

Given the importance of the chiral THIQ scaffolds, many other biocatalytic approaches (besides NCS and BBE in the BIA pathway) have also been developed. One facile approach is direct asymmetric reduction of chemically synthesised 3,4-dihydroisoquinolines (DHIQs) with natural IREDs^{260,261} or artificial transferhydrogenases.^{262,263} Many natural IREDs were able to enantioselectively reduce DHIQs with a 1-methyl- or simple 1-alkyl substituent.²⁶⁴ To produce chiral bulky 1-aryl-THIQs, the Qu group assayed a large number of diverse IREDs and found several (R)-selective IREDs and one unique (S)-selective IRED (IR45) converting chloro-, methyl-, and methoxyl-benzyl DHIQ into the corresponding (R)- or (S)-THIQ in high-to-excellent





Scheme 4 Novel enzymes, enzymatic mechanisms and cascades for the synthesis of monoterpene indole alkaloids. STR: strictosidine synthase.

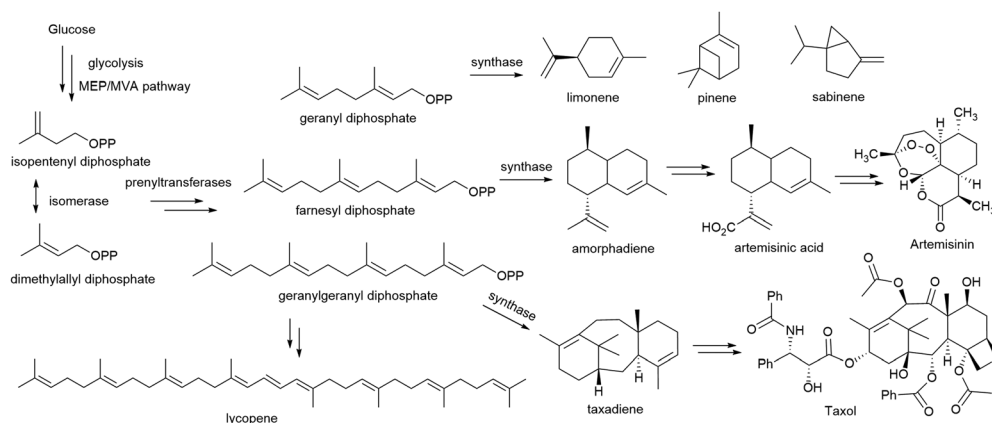
benzaldehydes and produce (*R*)-1-aryl-THBCs.²⁸⁶ Besides STR, IREDs have been used to produce simple 1-methyl-THBCs from the corresponding imines²⁸⁷ and MAO-N mediated dera-cemization has been explored to access a variety of chiral 1-substituted-THBCs.²⁸⁸

Terpenoids

Terpenoids (isoprenoids) consisting of C₅ isoprene units are the most diverse group of natural products, with more than 80 000 known structures to date.^{289,290} They are prevalent in all kingdoms of living organisms, and many of them possess indispensable biological functions and activities (*e.g.*, light-harvesting, electron transfer, membrane constituents) with broad potential applications as fragrances, nutraceuticals, pharmaceuticals, *etc.* Despite their astonishing structural diversity and widespread occurrence, the canonical biosynthetic

route of terpenoids is highly modular (Scheme 5):^{291,292} two key C₅ precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), are synthesised from common primary metabolites *via* either the MEP (methylerythritol phosphate) pathway in bacteria and plant plastids or the MVA pathway in animals, fungi, and the cytosol of plants;²⁹³ IPP and DMAPP are assembled by prenyltransferases to give poly-isoprenoid diphosphates, such as geranyl diphosphate (GPP, C₁₀), farnesyl diphosphate (FPP, C₁₅), or geranylgeranyl diphosphate (GGPP, C₂₀); polyisoprenoid diphosphates are usually cyclised with terpene synthases to generate terpenoid skeletons;²⁹⁰ structures are further tailored with various enzymes, such as P450s.²⁹⁴ Due to the importance of many terpenoids and existence of native MEP/MVA pathways in microbes, terpenoids have been very popular targets in metabolic engineering and synthetic biology for more than 20 years.^{295–298} Several early studies focused on engineering *E. coli* to produce lycopene and carotenoids due to easy detection and lack of complex tailoring.²⁹⁹ The engineering of *S. cerevisiae* to produce artemisinic acid (precursor for the anti-malarial drug artemisinin) by the Keasling group and Amyris is a landmark in metabolic engineering/synthetic biology.³⁰⁰ More recent progress includes engineering *S. cerevisiae* for the production of farnesene (bulk chemical)³⁰¹ and tocotrienols (vitamin E),³⁰² as well as engineering *E. coli* for the production of taxadiene (precursor for Taxol)³⁰³ and viridiflorol (fine chemical).³⁰⁴ However, achieving a productive synthesis of terpenoids (*e.g.*, > 10 g L^{−1}) in microbes is often very challenging and requires extensive engineering and optimisation efforts due to the toxicity of intermediates/products and the complex regulation of native pathways.

Complementary to the biosynthesis of terpenoids in microbes, cell-free *in vitro* multi-enzymatic synthesis avoids many problems in cells (*e.g.*, toxicity, complex regulation), minimizes side reactions and genetic engineering efforts, and often offers higher yields of final products in a cleaner reaction system.^{305–308} These advantages were clearly demonstrated in the pioneering work on the cell-free one-pot production of monoterpenes from glucose by the Bowie group.³⁰⁹ Simply combining standard Embden–Meyerhof–Parnas glycolysis and



Scheme 5 Novel enzymes, enzymatic mechanisms and cascades for the synthesis of terpenoids. MEP: methylerythritol phosphate; MVA: mevalonate.

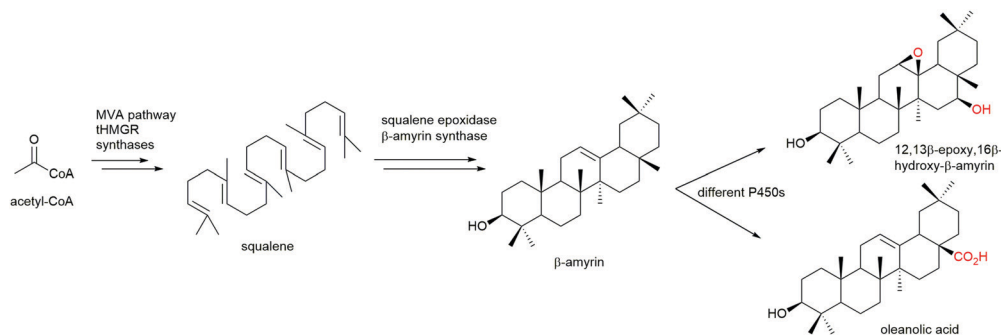


the MVA pathway leads to imbalance of cofactors: the glycolysis generates ATP and NADH (excess), while the MVA pathway consumes ATP and NADPH. The authors cleverly tackled this issue by creating enzymatic purge valve nodes:³¹⁰ an NAD⁺-utilizing glyceraldehyde-3-phosphate dehydrogenase (GAPDH), an NADP⁺-utilizing mutated GAPDH and an NADH oxidase. Mathematic modelling was used to identify the potential bottlenecks (hexokinase, pyruvate dehydrogenase, and phosphate), and these key parameters were experimentally optimised. With careful consideration and optimisation, the final *in vitro* system comprising 27 enzymes converted glucose (500 mM) to limonene (12.5 g L⁻¹) with a theoretical yield of 88% over 7 days. By replacing the terpene synthase, pinene and sabinene were also produced at 14.9 and 15.9 g L⁻¹ and almost quantitative yields, respectively. These product titres are at least 10 times higher than those from microbial production, far exceeding the toxicity limits of these compounds. *In vitro* multi-enzymatic synthesis has been reported for other terpenoids, such as the production of amorphadiene from mevalonic acid (6 steps with ATP recycling)³¹¹ and the production of geosmin and patchouliol from acetic acid (10 steps with cofactor regeneration).³¹² Cell-free *in vitro* multi-enzymatic synthesis has made great progress recently, but it is still difficult to work on some complex and difficult enzymes (*e.g.*, membrane enzymes, P450s), and large-scale application is hampered by relatively high costs (*e.g.*, enzyme purification, cofactors).

Currently, the biosynthetic pathways for the majority of natural terpenoids have not been fully elucidated. Thanks to the advances in genomic sequencing and bioinformatics, many putative terpenoid synthetic enzymes could be identified *in silico*. To verify and characterize these putative enzymes, heterologous expression in suitable (engineered) hosts could often enable efficient production of terpenoid products for characterisation. For rather simple and small bacterial terpenes, heterologous expression in *E. coli* is often sufficient for rapid and facile characterisation.^{313,314} While for more complex plant-origin terpenoids, heterologous expression of the enzymes in yeast or plants is necessary.^{315–317} For example, plant diterpene labdanes and clerodanes are often synthesised by a pair of distinct monofunctional class I and class II diterpene synthases (diTPSS). By mimicking the modular diterpene biosynthesis, the Hamberger

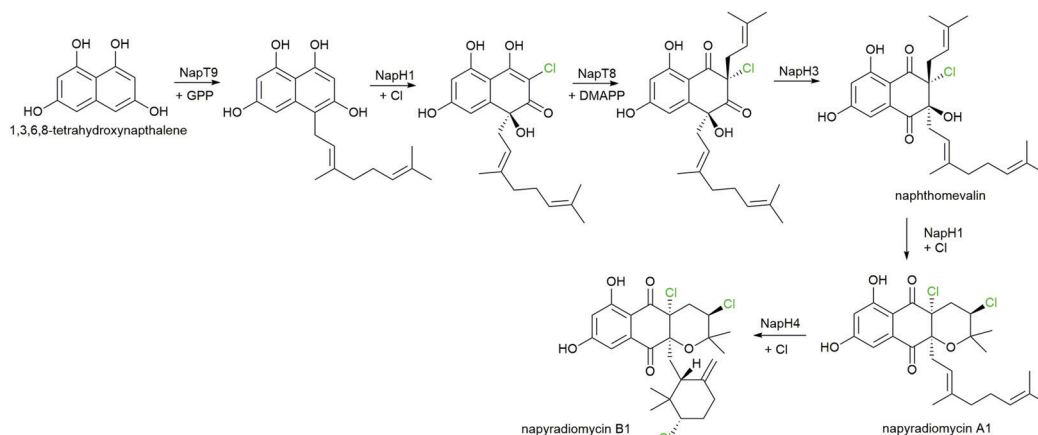
group tested every combination of 9 class I and 11 class II diTPS from 10 plant species in *N. benthamiana* by *A. tumefaciens*-mediated transient expression.³¹⁸ 51 Diterpene skeletons were stereo-selectively biosynthesised, including 41 new-to-nature ones. By engineering *S. cerevisiae*, four useful diterpenes were produced at a scale relevant for industrial applications. To quickly access highly diverse oxygenated plant triterpenes (>20 000 reported so far),³¹⁹ the Osbourn group developed a translational synthetic biology platform based on transient expression in the whole plant of *N. benthamiana* (Scheme 6).³²⁰ Initially, a feedback-insensitive version of an HMG-CoA reductase (tHMGR) in the MVA pathway was found to significantly boost the production of β -amyrin when co-expressing with the oat β -amyrin synthase (SAD1) in *N. benthamiana*. To provide a gram-scale synthesis of triterpenes, the authors developed a vacuum agro-infiltration system for transient expression in the whole plant rather than individual leaves. Co-expression of tHMGR and SAD1 in about 460 *N. benthamiana* plants and cultivation for 5 days allowed successful isolation of 800 mg of β -amyrin with >98% purity. This platform (tHMGR and SAD1 in *N. benthamiana*) was combined with one or a pair of five different β -amyrin-oxidising P450 enzymes to offer 41 different oxygenated triterpenes (some new-to-nature). A handful of them were isolated on 10 mg scale and further evaluated for antiproliferative and anti-inflammatory activities. Besides the proper functional expression of plant enzymes, another advantage of transient expression in *N. benthamiana* is that multiple genes can be co-expressed by simply co-infiltrating multiple *A. tumefaciens* strains. This feature allows a quick test of combinations of enzymes either to generate new products (as shown in the two examples above) or to elucidate the biosynthetic pathway, which was demonstrated in a recent study of root triterpenes in *A. thaliana*.³²¹

Merochlorin A and B are common C₄-prenylated meroterpenes, while certain vanadium-dependent haloperoxidases mediate an α -hydroxyketone rearrangement, leading to naphthomevalin with a unique C₃-prenylation pattern (see also Section Meroterpenoids, Scheme 14).³²² These enzymes were recently employed for total enzymatic syntheses of antimicrobial and cytotoxic meroterpenoids, napyradiomycin A1 and napyradiomycin B1 from 1,3,6,8-tetrahydroxynaphthalene, GPP and DMAPP (Scheme 7).³²³



Scheme 6 Novel enzymes, enzymatic mechanisms and cascades for the synthesis of oxygenated plant triterpenes. tHMGR: feedback-insensitive version of HMG-CoA reductase; P450: cytochrome P450 monooxygenase.





Scheme 7 Novel enzymes, enzymatic mechanisms and cascades for the synthesis of Napyradiomycins. GPP: geranyl diphosphate; DMAPP: dimethylallyl diphosphate; NapH1, H3, and H9: vanadium-dependent haloperoxidases; NapT8 and T9: aromatic prenyltransferases.

By applying two aromatic prenyltransferases (NapT8 and T9) and two vanadium-dependent haloperoxidases (NapH1 and H3) in one pot, napyradiomycin A1 was synthesised in 22% yield. With the addition of vanadium-dependent haloperoxidase NapH4, napyradiomycin B1 was synthesised in 18% yield.

Besides the vanadium-dependent haloperoxidases, other halogenating enzymes^{324–327} could also provide plenty of opportunities in the enzymatic syntheses of terpenoids (and other natural products) as well as their derivatives.

Steroids

Steroids are a large group of biologically active compounds that share a common structure typically containing 17 carbon atoms fused in a tetracyclic system known as gonane (Scheme 8). This framework derives from the MEV pathway, where isoprene units are linked and the cyclisation of the triterpenoid squalene leads to basic precursors like lanosterol (in animals and fungi) and cycloartenol (in plants),^{328–330} but alternative biosynthetic pathways exist in bacteria and certain plants that yield the steroid core (Scheme 8).^{330–332} Besides being integral components of cellular membranes, steroids function as signaling molecules in metabolic pathways, cellular survival, reproduction, and disease. To precisely execute these diverse biological roles, (subtle) modifications of the four-ring scaffold are introduced by an equally diverse subset of enzymes with high stereo- and regio-selectivity that is hard to achieved by chemical synthetic methods. Hence, the over 250 known natural steroid compounds differ in functional group decoration and the oxidation state of the steroid nucleus and many of them are considered essential precursors for the manufacturing of drugs exhibiting antifungal, antimicrobial, antiviral, immune-modulating, antitumor, and anticonvulsant activities.³³³ Indeed, steroid-based drugs represent the second-highest marketed category of APIs – after antibiotics – with the annual global market exceeding 10 billion USD and more than 300 clinically approved natural and (semi-)synthetic steroidal compounds.^{334–336}

Traditionally, steroidal APIs have been synthesised through chemical processes that are characterised by the requirement

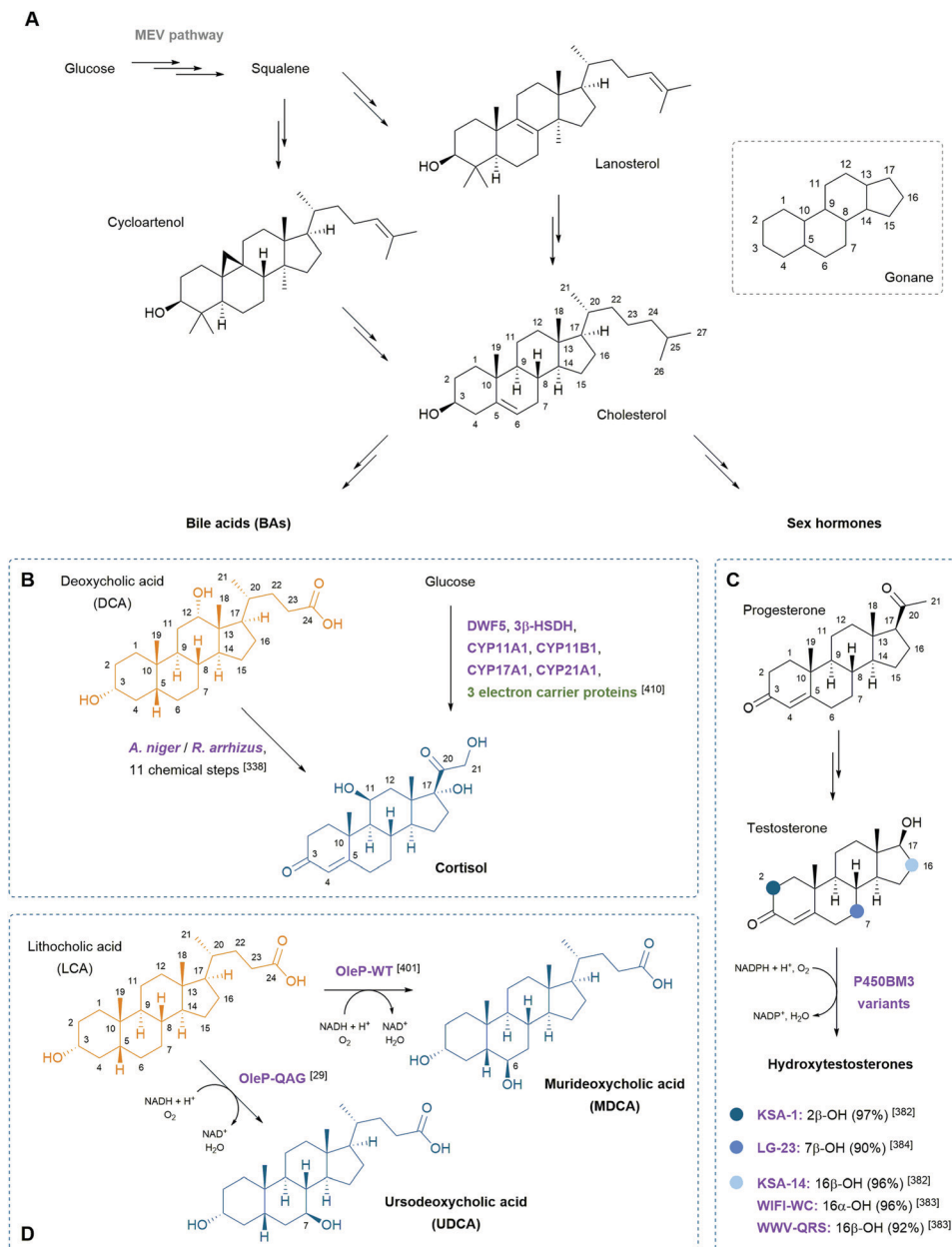
for multiple sequential steps that offer only poor control over the stereo- and regioselectivity and very low yields.³³⁷ In the 1950s, the corticosteroid hormone cortisone was synthesised from the bile acid (BA) deoxycholic acid (DCA) over 31 steps with a yield of 0.16%. By including a fermentation step with the fungi *Rhizopus arrhizus* and *A. niger*, the number of chemical steps could be reduced to 11, markedly reducing production costs (Scheme 8).³³⁸ To date, many typical steroidal APIs are manufactured chemically but involve microbial biotransformations for the preparation of key intermediates^{335,337,339} or the late-stage functionalisation of steroids.^{340–343} The latter regularly involves stereo- and regioselective hydroxylations by P450s and will be described with focus on recombinant applications below.^{333,340,344,345}

Although microorganisms have long been used for precursor synthesis and steroid modifications, fully microbial processes featuring efficient platform strains or recombinant microbial cell factories are scarce. Limitations are directly caused by the intrinsic properties of steroidal compounds such as low solubility in aqueous media and cellular toxicity.²⁹ Some of these issues have been addressed by the emulsification of substrates with surfactants or the utilization of two-phase systems with organic solvents.^{333,339,346} Furthermore, steroid transformations in well-characterised recombinant host cells (*e.g.*, *E. coli*, *S. cerevisiae*) regularly yield low product titres simply because steroid-modifying enzymes only poorly express or function outside their native hosts, combined with insufficient substrate uptake and unintended metabolism.^{75,335,347}

In the following, the bio-based synthesis of steroid precursors and (recombinant) functionalization will be highlighted as well as trends that project towards the *de novo* synthesis of steroidal APIs and their customization for future applications.

Important substrates, in the context of microbial production of API precursors, are sterols.^{330,335} This sub-group of steroidal compounds, bearing a 3-hydroxyl group, include cholesterol, lanosterol, phytosterol, ergosterol, and BAs like DCA and lithocholic acid (LCA; Scheme 8). Sterols are intermediates of both anabolic and catabolic steroid pathways and are available from





Scheme 8 Steroidal precursors and APIs. (A) The biosynthesis of steroids through the MEV pathway leads to precursors like lanosterol, cycloartenol, and cholesterol, an important intermediate towards BAs and sex hormones. (B) Cortisol production from DCA involving fermentations with *A. niger*, *R. arrhizus*, and 11 chemical steps in total or from glucose through a heterologous pathway mimicking human steroid biosynthesis. DWF5: 7-dehydrocholesterol reductase. (C) Progesterone and testosterone are important APIs; the latter can be hydroxylated by P450BM3 variants in the positions 2, 7, and 16 (shades of blue), for example. Regiospecificity given in parentheses for KSA-1, KSA-14, and LG-23; stereospecificity for WIFI-WC and WWV-QRS. (D) OleP wild-type (OleP-WT) exclusively yields MDCA from LCA by 6 β -hydroxylation. The triple mutant OleP-QAG yields the 7 β -product UDCA, a reaction new to the biocatalytic toolbox of single enzymes. Important steroid substrates are numbered and modifications in target products indicated. Structures of BA precursors are highlighted in orange, products in blue, key enzymes in purple and accessory proteins in green.

plant-based and animal feedstocks for biotechnological applications.^{335,348–352} Various strains from the genera *Mycobacterium*, *Nocardia*, and *Rhodococcus* were identified to transform sterols and, importantly, have been metabolically engineered to reroute fluxes towards the accumulation of value-added steroids. Examples feature the production of C-19 steroids such as 4-androstene-3,17-dione (AD), 1,4-androstadiene-3,17-dione (ADD), and testosterone in mutant strains of *Mycobacterium smegmatis*

(*M. smegmatis*) starting from cholesterol. Single and multiple gene deletions including *kstD* and *ksh*, encoding 3-ketosteroid- Δ^1 -dehydrogenase and 3-ketosteroid-9 α -dehydrogenase, accumulated AD and ADD, respectively; the heterologous overexpression of 17 β -hydroxysteroid dehydrogenases (HSDHs) from different (bacterial) sources successfully converted AD into testosterone.^{339,353}

Whereas the biotransformation of sterols into AD requires multiple reaction steps carried out by endogenous host enzymes,



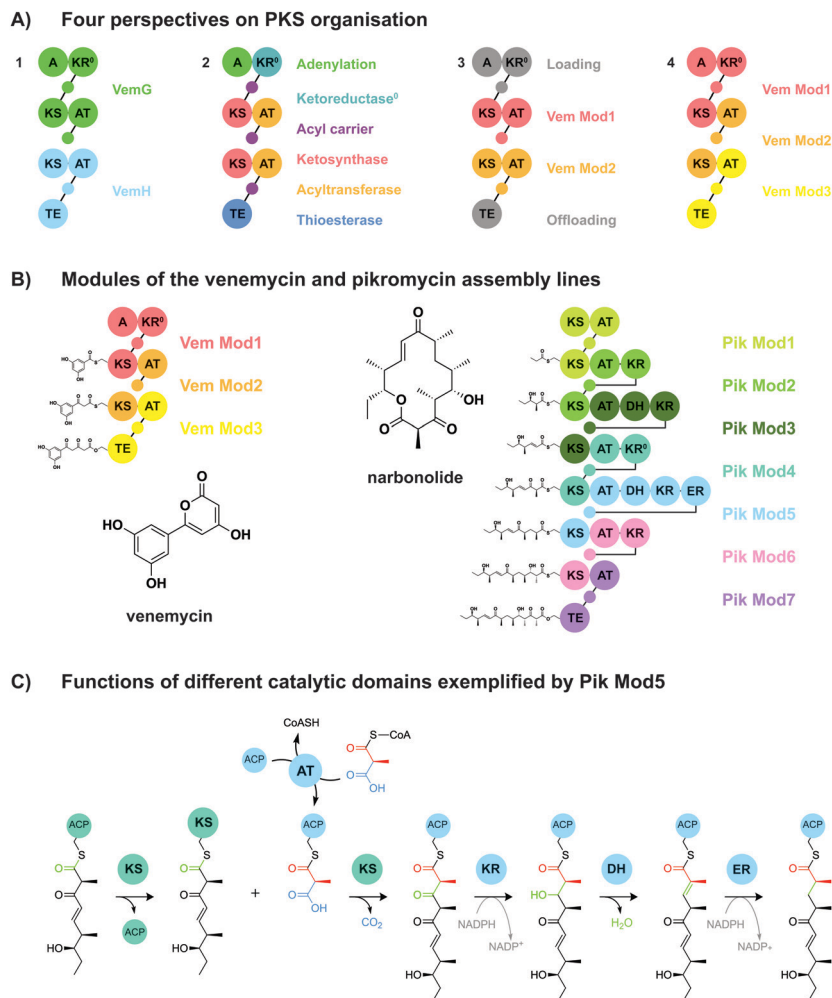


Fig. 4 Traditional and updated assembly-line polyketide synthase module definitions. (A) Four different perspectives on the venemycin assembly-line PKS. (1) The synthase can be divided into two polypeptide chains called VemG (232 kDa) and VemH (140 kDa). (2) The PKS can also be viewed in terms of its domains. The adenylation domain (A) accepts the 3,5-dihydroxybenzoyl starter unit. The inactive ketoreductase domain (KR⁰) probably plays a structural role. The ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (small circle) domains are the core components of PKSs. The thioesterase domain finally releases the polyketide from the synthase, often by cyclisation to form a macrolactone. (3) In the traditional view, the module boundaries are the N-terminus of the ketosynthase domain and the C-terminus of the acyl carrier protein domain. (4) In the updated definition, modules end at the C-termini of ketosynthase domains, reflecting the evolutionary co-migration of domains. (B) The venemycin and pikromycin assembly lines depicted using the new module definitions. The assembly-line steps and products are also shown. The pikromycin assembly line additionally includes dehydratase (DH) and enoylreductase (ER) domains. Note that modules can be split over different polypeptide chains. For example, Vem Mod2 is split between the VemG and VemH proteins. (C) The functions of the different catalytic domains exemplified by Pik Mod5, which has a full set of reductive domains. The domains are represented by spheres coloured by module as in (B). The process starts with a tetraketide intermediate attached to an acyl carrier protein, which is transferred to a cysteine residue on the ketosynthase domain. An acyltransferase domain loads the downstream ACP with an acyl-CoA-derived extender unit, in this case methylmalonyl-CoA. The ketosynthase catalyses decarboxylative condensation of the tetraketide intermediate and the extender unit to form an ACP-linked pentaketide intermediate. This intermediate is then subjected to reductive reactions by the ketoreductase, dehydratase, and enoylreductase domains (the carbonyl subjected to reduction is coloured green). All three reactions are optional so that other modules may stop at either the β -keto, β -hydroxy, or α,β -alkene intermediates. The resulting ACP-linked intermediate is substrate to either the next KS domain or the terminal thioesterase, which usually results in cyclisation. This figure was simplified and redrawn from the Miyazawa *et al.*^{440,441}

assembly-line PKS consists of at least a ketosynthase, an acyltransferase, and an acyl-carrier protein domain (Fig. 4C).⁴⁴³ The acyltransferase domain of a loading module first transfers an acyl group, usually from acetyl-CoA, to the phosphopantetheinyl arm of its acyl-carrier protein. The ketosynthase domain of the downstream module catalyses both translocation of the acyl group and carbon-carbon bond formation by a decarboxylative Claisen

condensation of the translocated acyl group with a malonyl derivative (bound to the acyl-carrier protein as a thioester). This reaction forms a 3-ketoacyl intermediate and releases carbon dioxide, which thermodynamically drives the process.⁴³⁷ This intermediate may then be reduced, and further functionalised by an optional “reducing loop” composed of a keto-reductase, a dehydratase, and an enoyl-reductase.⁴⁴⁴ All these reactions are



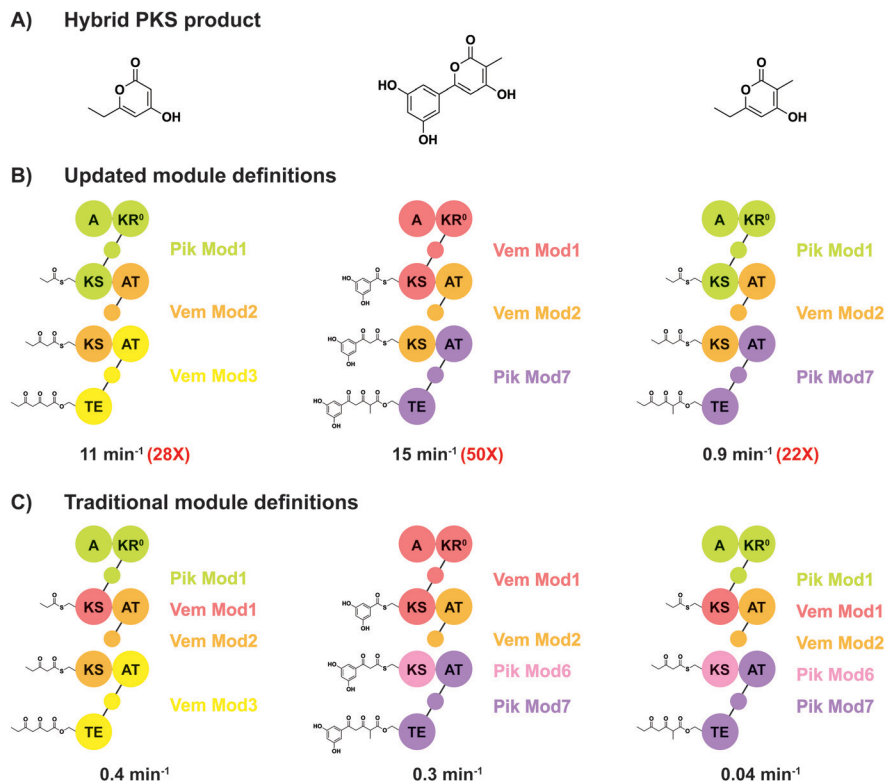


Fig. 5 Hybrid polyketide synthases constructed using both the updated and the traditional module definitions. (A) The products of three hybrid assembly lines. (B) Three assembly lines constructed using the updated module definition are over an order of magnitude more productive (red numbers) than (C) assembly lines constructed using the traditional modules. Despite the success of assembly lines based on the new module definitions, the native VemG-VemH assembly line is still more than double as active (36 min⁻¹). The domain abbreviations are as for Fig. 4 and again the acyl carrier proteins are represented by small circles. This figure was simplified and redrawn from Miyazawa *et al.*⁴⁴⁰

and neoareothin, two compact and highly functionalised polyketides produced by homologous biosynthetic gene clusters. They successfully morphed the neoareothin assembly line to produce the aureothin backbone by deletion of two modules. They found that the KS-AT linker is well suited for both insertion and deletion of modules, further supporting the alternative domain definition.⁴⁴⁶

Despite advances in modelling natural evolutionary processes and computational prediction of optimal splice sites, a major barrier to successful engineering has been a lack of experimental data for guiding optimal selection of splice sites for generating functional chimeric PKSs.^{447,450,451,456,457} Experimentally accelerated molecular evolution based on homologous recombination is a recently introduced strategy for gaining valuable information on optimal splice sites for functional chimeras. Homologous recombination based on naturally occurring stretches of sequence similarity can be used for the assembly of novel chimeric PKSs, similar to natural PKS evolution.⁴⁴² Chemler *et al.* used homologous recombination between the erythromycin (DEBS) and pikromycin coding sequences in *Saccharomyces cerevisiae* to generate hybrid libraries containing many functional chimeras.⁴⁵⁶ This method has the potential for generating large libraries rich in functional variants. Wlodek *et al.* recently described a method for adding, removing, and replacing modules, based on recombination

between regions of high sequence homology within a PKS gene cluster. Rather than using yeast, they harnessed the homologous recombination machinery of a *Streptomyces* strain, rapidly generating diverse and highly productive assembly lines by 'accelerating' the natural evolution of modular polyketide synthases.⁴⁴⁷ They generated 17 rapamycin synthase and 9 tyllactone synthase chimeras, many of which were highly active, producing titres comparable to the wild-type strain.^{442,447} Sherman's group described the use of PKS modules *in vitro* to convert a chemically synthesised thiophenyl-activated analogue of the hexaketide intermediate of tyllactone biosynthesis. The intermediate was accepted by the ketosynthase of the JuvEIV PKS module and further processed by the JuvEIV and JuvEV modules to form tyllactone. Macrolactonization was followed by *in vivo* glycosylation, *in vitro* P450-mediated oxidation, and chemical oxidation, resulting in the total synthesis of a range of macrolide antibiotics from the juvenimicin, M-4365, and rosamicin family.⁴⁵⁸ Analogues of tyllactone intermediates accessed by homologous recombination-based genetic engineering could be valuable alternative starting points for chemoenzymatic late-stage modification, enabling structural diversification to an even larger number of macrolide antibiotics.^{447,458}

Khosla's group recently reported the use of *in vitro* reconstitution to decode the orphan polyketide assembly line responsible for producing the nocardiosis-associated polyketide (NOCAP).



The reconstituted PKS, in the presence of octanoyl-CoA, malonyl-CoA, NADPH, and SAM, produced octaketide and heptaketide products that could partially be structurally elucidated by MS and NMR.⁴⁵⁹ More recently, they reconstituted the entire NOCAP assembly line both *in vitro* and in *E. coli* to fully “deorphanise” the NOCAP synthase, independent of its genetically challenging and hazardous natural host.⁴⁶⁰ These approaches for studying multi-megadalton assembly lines are by far not standardised. For example, to overcome heterologous expression problems due to the exceptional size of the NOCAP PKS (3 MDa homodimer), multi-modular proteins had to be dissected into mono- or bimodular units that could be more easily expressed in *E. coli*.^{438,460} Optimisation or modifications of polyketide synthesis in the native host strain is often desirable because the gene cluster is already functionally expressed. However, as the NOCAP case demonstrates, the hosts may be hazardous or otherwise hard to culture and manipulate. One of the key advantages of *E. coli* is that its metabolic background is not cluttered by complex natural product biosynthesis pathways, avoiding crosstalk with heterologously introduced PKSs and facilitating the identification of novel polyketides. The metabolism and molecular biology of *E. coli* are also well understood, and the genetic toolkit is unrivalled.⁴³⁶ The *E. coli* genome has not only been extensively sequenced but also completely recoded.^{461,462} Therefore, future advances in systems level understanding could be rapidly translated into genomically-reprogrammed hosts. Despite the advantages,^{443,460} only a few assembly-line PKSs have been functionally reconstituted in *E. coli*, because *E. coli* is not an ideal “universal host”.⁴³⁶ Nontrivial engineering is necessary for biosynthesis of precursors and performing critical post-translational modifications. No assembly-line PKS has yet been completely deorphanised solely by reconstitution in *E. coli*. This achievement would revolutionize natural product discovery if robustly implemented.⁴³⁶

As in other protein engineering endeavours, the number of PKS variants interrogated is limited by the throughput of the available screening methodologies. Chromatographic analyses are time consuming but due to the complexity of the molecules, simple colorimetric or fluorometric assays are not typically applicable to PKS screening.⁴³⁴ While biosensors for detecting enhanced precursor (*e.g.*, malonyl-CoA) formation have been reported,^{143,463} relatively few high-throughput methods for polyketide products are available. Kasey *et al.* recently showed that engineered variants of the promiscuous erythromycin-sensing transcription factor MphR could be used to detect related ligands like clarithromycin. This work demonstrated the potential of engineered biosensors to facilitate the directed evolution of macrolide synthases, but little work on this topic has been published recently.^{71,434,464} While biosensors are promising tools for HTS, they are often limited to naturally occurring biosensors. Unfortunately, biosensor engineering is challenging in itself, effectively doubling the engineering effort if an off-the-shelf biosensor is not available.⁴⁶⁵ However, for a problem as complex as PKS engineering, it might well be worth to first engineer a biosensor for screening PKS libraries. Rapid advances in both protein structure prediction⁴⁶⁶ and *de novo* design of bioactive protein switches^{165,467} might make biosensor

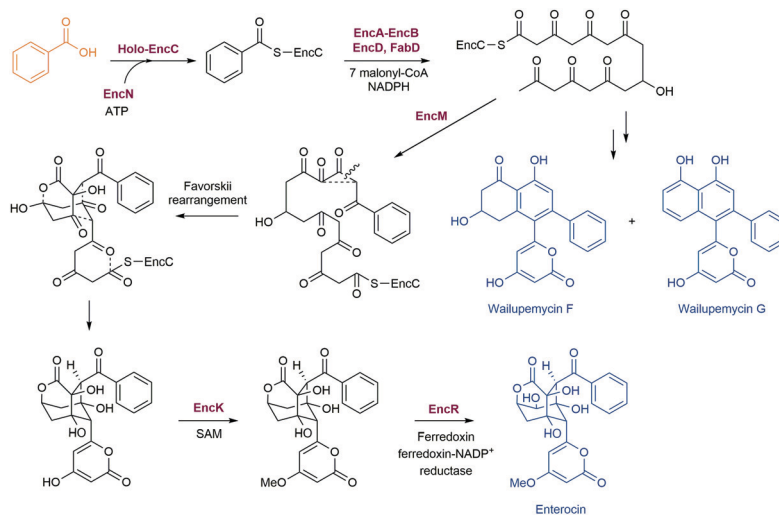
design much simpler in the near future. Unfortunately, these advances cannot entirely solve the problem, since in many cases the structure of the target polyketide will not be known in advance, making activity-based screening indispensable. While some (*e.g.*, antibiotic) activities are relatively easy to screen for using agar plate or microfluidic droplet-based UHTP screening technologies, others are not. In an interesting approach to screen for proapoptotic compounds, Theodorou *et al.* mixed bacteria in microfluidic droplets with mammalian cells, which could be assayed for apoptosis markers.⁴⁶⁸ For the foreseeable future, these complicated screening problems will have to be solved on a case-by-case basis.

Yuzawa *et al.* recently reported engineered strains of *Streptomyces albus* harbouring engineered hybrid polyketide synthases capable of converting plant biomass to methyl and ethyl ketones at titres of over 1 g L⁻¹.⁴⁶⁹ While these product titres are too low to compete with fossil fuels, they demonstrate that engineered hybrid polyketide synthases can produce g L⁻¹ titres, which is significant for more valuable pharmaceutical compounds. Probably the greatest challenge for the field is increasing production rates to industrially competitive values since both time and titre are relevant to the space-time yield. The ability to rationally exchange assembly line modules has long been one of the holy grails in biochemistry. While linking a loading domain to more than one extension module without losing functionality is still a major challenge, recent trends in evolution-guided engineering strategies suggest that there is hope for the future. Updated module definitions (Fig. 4 and 5), supported by high-throughput experimental recombination and characterisation of chimeras and machine learning, may well make this goal attainable.⁴⁷⁰ Advances in DNA synthesis and assembly coupled with automated and unbiased PKS characterisation is necessary to realise the full potential of machine learning in PKS research. The hurdles to designer polyketide synthases are expected to be overcome taking advantage of the newly arising tools available for biocatalysis.

De novo enzymatic total synthesis of complex drug compounds

Biocatalytic synthesis of drug compounds is a main research topic in the biopharmaceutical industry. The previous studies have mostly focused on single-step enzymatic reactions or simple multi-step enzymatic cascades to synthesise intermediates with high value. At present, more attempts are involved in the total biosynthesis of drug compounds with complex structures. Since drug molecules mostly contain non-natural structures, the design of their biosynthetic pathways has more challenges in comparison to natural products. An effective strategy is to introduce artificially designed enzymatic cascades on the basis of related natural synthetic pathways, which can reduce the complexity of pathway design. Moreover, *de novo* biosynthetic pathways require screening of novel enzymes or modification of existing enzymes through protein engineering to obtain biocatalysts with sufficient activity, specificity, and stability that

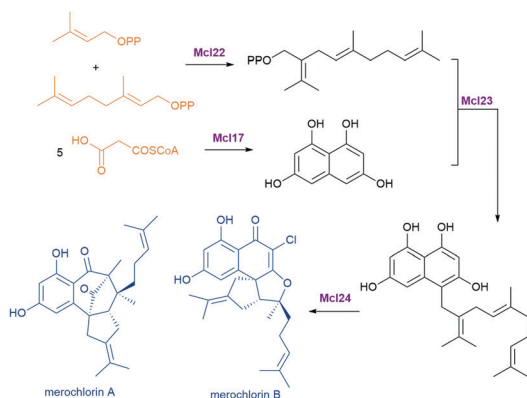




Scheme 13 Enzymatic total synthesis of enterocin and wailupemycin. Starting materials are highlighted in orange. The products are highlighted in blue. Key enzymes in this synthetic route are highlighted in purple.

Meroterpenoids

Meroterpenoids are hybrid terpenoid–polyketide natural products with many desired bioactivities such as antimicrobial, antioxidant, and anticancer. However, they are particularly challenging complex molecules for enzymatic total synthesis. The Moore group elucidated the biosynthesis of the halogenated meroterpenoids merochlorin A and B, and demonstrated the first total enzymatic synthesis of a meroterpenoid (Scheme 14).⁴⁸¹ A type III PKS, Mcl17, catalysed the formation of 1,3,6,8-tetrahydroxynaphthalene from 5 malonyl-CoA substrates. The terpenoid part of the molecule was synthesised by a unique prenyl diphosphate synthase, Mcl22, *via* a head-to-torso coupling of GPP with DMAPP. These two parts were connected by an aromatic prenyltransferase, Mcl23. The most surprising part came from the vanadium-dependent chloroperoxidase, Mcl24, catalyzing a cascade of naphthol chlorination



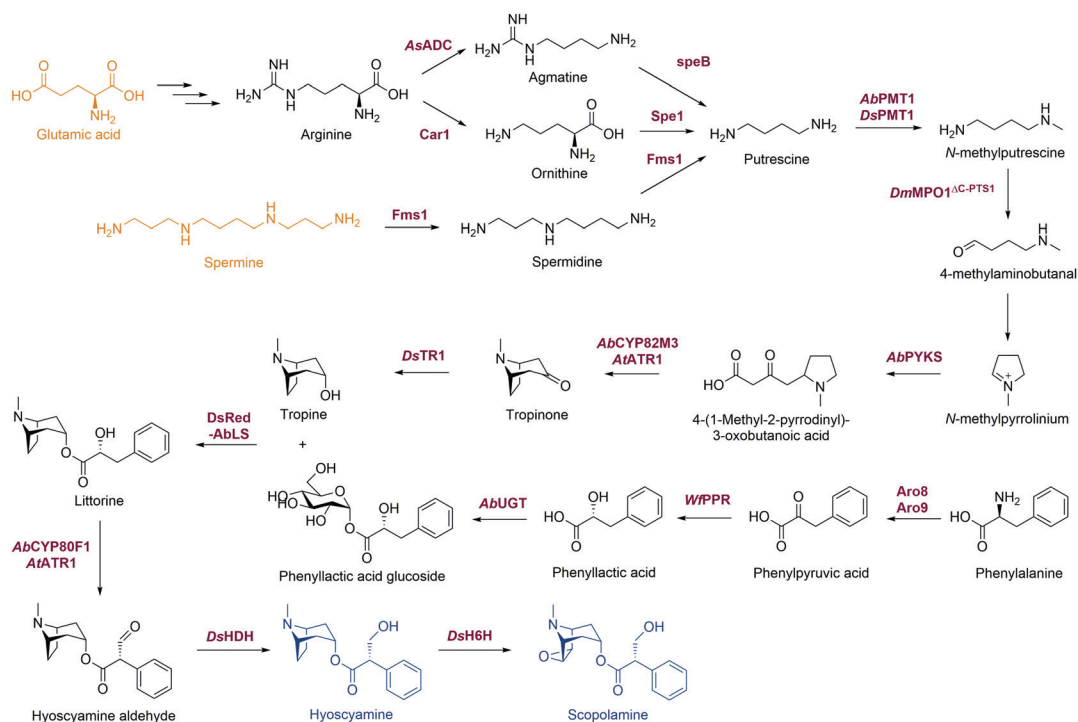
Scheme 14 Enzymatic total synthesis of the meroterpenoids merochlorin A and B. Mcl17: type III polyketide synthase; Mcl22: prenyl diphosphate synthase; Mcl23: aromatic prenyltransferase; Mcl24: vanadium-dependent chloroperoxidase. Starting materials are highlighted in orange. The products are highlighted in blue. Key enzymes in this synthetic route are highlighted in purple.

and oxidative dearomatization/cyclisation steps.⁴⁸² The highly complex halogenated meroterpenoids merochlorin A and B were synthesised by applying merely these four enzymes (Mcl17, Mcl22, Mcl23, and Mcl24) and simple substrates (malonyl CoA, DMAPP, GPP), showing the power of enzymes to generate complex structures.

Hyoscyamine and scopolamine

Hyoscyamine and scopolamine are widely used drugs to control neuropathic pain and treat neuromuscular disorders, such as nerve agent poisoning, digestive and urinary tract spasm, and Parkinson's disease. At present, the production of these two compounds can only rely on extraction from nightshade plants, which poses a challenge to guarantee the supply. Recently, the total synthesis of hyoscyamine and scopolamine in yeast has been achieved by Smolke and Srinivasan with simple amino acids as starting materials (Scheme 15).⁴⁸³ The most critical step in the enzymatic synthesis of hyoscyamine and scopolamine is to use tropine and phenyllactic acid (PLA) glucoside as acceptor and donor, respectively, to generate littorine through transacylation, followed by reduction and epoxidation to produce the final products. Therefore, the whole synthesis pathway is composed of five modules: (i) putrescine module and (ii) tropine module for the synthesis of the acyl acceptor tropine, (iii) PLA glucoside module for the synthesis of the acyl donor PLA glucoside, (iv) tropane alkaloid (TA) scaffold module for transacylation, and (v) medicinal TA module for the post-modification of the littorine structure. In the putrescine module, arginine is generated through the arginine synthesis pathway with glutamic acid as starting material, which is then converted to putrescine catalysed by arginase (Car1) and ornithine decarboxylase (Spe1). In addition, spermine present in cells can be converted to putrescine by one enzyme, polyamine oxidase (Fms1). By enhancing the expression of key enzymes such as glutamate *N*-acetyltransferase (Arg2), Car1, Spe1 and Fms1, and introducing arginine decarboxylase from



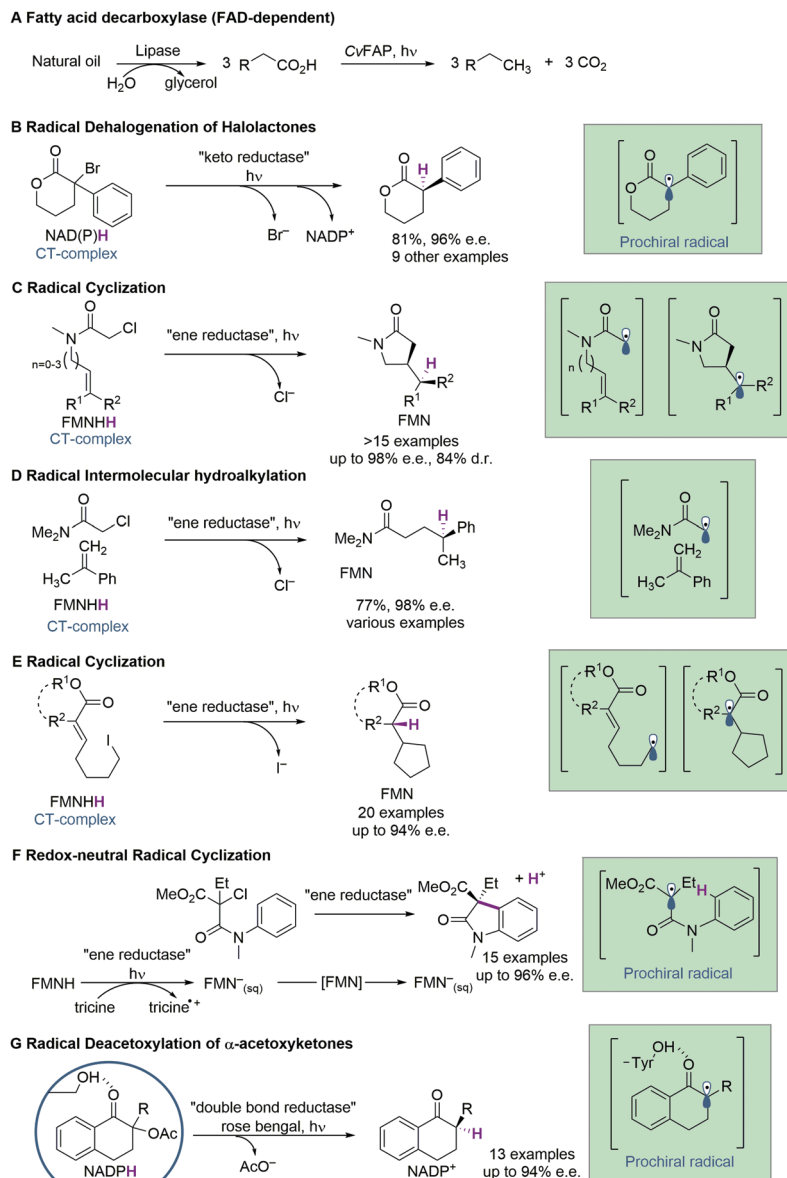


Scheme 15 Enzymatic total synthesis of hyoscyamine and scopolamine in yeast. AsADC: arginine decarboxylase; Car1: arginase; speB: agmatine ureohydrolase; Spe1: ornithine decarboxylase; Fms1: polyamine oxidase; AbPMT1 and DsPMT1: *N*-methyltransferases; DmMPO1^{AC-PTS1}: engineered *N*-methylputrescine oxidase; AbPYKS: pyrrolidine ketide synthase; AbCYP82M3: tropinone synthase; AtATR1: cytochrome P450 reductase; DsTR1: tropinone reductase 1; Aro8 and Aro9: aromatic aminotransferases; WPPR: phenylpyruvate reductase; AbUGT: UDP-glucosyltransferase; DsRed-AbLS: littorine synthase; AbCYP80F1: littorine mutase; DsHDH: hyoscyamine dehydrogenase; DsH6H: hyoscyamine 6 β -hydroxylase/dioxygenase. Starting materials are highlighted in orange. The products are highlighted in blue. Key enzymes in this synthetic route are highlighted in purple.

Avena sativa (AsADC) and agmatine ureohydrolase from *E. coli* (speB) which constitute a pathway to decarboxylate arginine, the yield of cadaverine was effectively enhanced. After entering the tropine biosynthesis module, one terminal amino group of putrescine is methylated by putrescine *N*-methyltransferases (AbPMT1 and DsPMT1), and the other one is deaminated into the aldehyde by an engineered *N*-methylputrescine oxidase (DmMPO1). The formed 4-methylaminobutanal self-cyclizes into *N*-methylpyrrolinium. The newly discovered key enzymes pyrrolidine ketide synthase (AbPYKS), tropinone synthase (AbCYP82M3) and cytochrome P450 reductase (AtATR1) catalyse the sequential malonylation and cyclisation of *N*-methylpyrrolinium to generate tropinone, in which the carbonyl group is subsequently reduced to a hydroxyl group by tropinone reductase 1 (DsTR1) to obtain tropine. In the PLA glucoside module, phenylalanine is deaminated and reduced to PLA by aromatic aminotransferases (Aro8 and Aro9) and phenylpyruvate reductase (PPR), respectively. Then a newly identified PLA UDP-glucosyltransferase (UGT84A27) achieves the glucosylation of PLA. Littorine is produced in the TA scaffold module through acyl-transfer catalysed by littorine synthase (AbLS). In the last TA module, littorine is converted to hyoscyamine by littorine mutase (AbCYP80F1) and hyoscyamine dehydrogenase (HDH), and further to scopolamine by hyoscyamine 6 β -hydroxylase/dioxygenase (H6H).

The design of the heterologous biosynthetic pathway of hyoscyamine and scopolamine basically refers to the natural plant synthesis pathway. Through the mining of plant transcriptome data, the successful identification of the key enzymes AbPYKS, AbCYP82M3, AbUGT, AbLS, AbCYP80F1, DsHDH and DsH6H was the prerequisite for constructing a complete synthetic pathway. In addition, screening orthologues and introducing microorganism-derived enzymes are able to improve the efficiency of the single-step reaction and product yield. However, the enzymes involved in the original synthesis pathway are expressed in different tissue cells in plants and various subcellular structures in plant cells. Due to lacking the corresponding differentiated cells and subcellular structures, it is a great challenge to construct such complex enzymatic cascades in most heterologous microbial cells. In this influential biosynthesis example, enzymes are expressed in the cytoplasm, mitochondria, peroxisome, vacuole, and membrane of vacuole and endoplasmic reticulum, respectively, according to their protein transmembrane structure, cofactor regeneration and electron transport chain required for activity. This strategy of 'regionalised' expression not only ensures the activity and function of the enzyme, but also brings convenience to the regulation of synthetic pathways. The final product titres were reported as 30 to 80 $\mu\text{g L}^{-1}$, hence extensive further optimisation is needed to reach grams per litre. Nevertheless, this





Scheme 16 Light-induced natural and promiscuous enzymatic transformations. Fatty acid photodecarboxylase (A) is the only natural photoenzyme employed for biotransformations. The other examples (B–G) utilize the photophysical capabilities of nicotinamide and flavin cofactors to generate radical species (key intermediates are shown in the green boxes) as reactive intermediates.

above reactions, inclusion of photoorganoredox catalysts that have strong binding to proteins can unlock further promiscuous reactions: the rose bengal radical anion, which can be formed outside of the protein by light illumination, can strongly associate with a double bond reductase and thus inject its electron into an α -acetoxyketone located in the active site (Scheme 16G).⁵⁰⁰ Here, the role of the protein is to attenuate the redox potential of the ketone substrate by a hydrogen bonding interaction to its carbonyl group. This ensures that the single electron transfer from the photocatalyst only occurs to the enzyme-bound ketone substrate and not in the bulk reaction solution, and thus facilitates the exclusive asymmetric H-atom delivery from the enzyme-bound nicotinamide cofactor. The obtained α -chiral ketones or amides can serve as valuable synthons suitable for further functionalization.

For the construction of more complex molecules, it is a plausible next step to combine photocatalytic reactions with further enzymatic conversions in a one-pot sequential fashion or even as one-pot concurrent cascade. The feasibility of this approach has been demonstrated very recently,⁵⁰¹ and a couple of examples are available from the last two years. In most cases, the photocatalytic step generates a (reactive) intermediate that is then converted by an enzyme in the second step.^{487,488} Exploited photoreactions include *e.g.*, oxyfunctionalisation of alkanes⁵⁰² or alcohols⁵⁰³ by the water-soluble photocatalyst sodium anthraquinone sulfate to yield ketones or aldehydes. Many different enzymatic reactions can then be coupled as the second step, yielding amines and cyanohydrins, amongst others. However, the one-pot two-step protocol was more



