

## REVIEW

View Article Online  
View Journal



Cite this: DOI: 10.1039/d4np00031e

# Accelerating enzyme discovery and engineering with high-throughput screening

Eray U. Bozkurt,  Emil C. Ørsted,  Daniel C. Volke  and Pablo I. Nikel \*

Covering: up to August 2024

Enzymes play an essential role in synthesizing value-added chemicals with high specificity and selectivity. Since enzymes utilize substrates derived from renewable resources, biocatalysis offers a pathway to an efficient bioeconomy with reduced environmental footprint. However, enzymes have evolved over millions of years to meet the needs of their host organisms, which often do not align with industrial requirements. As a result, enzymes frequently need to be tailored for specific industrial applications. Combining enzyme engineering with high-throughput screening has emerged as a key approach for developing novel biocatalysts, but several challenges are yet to be addressed. In this review, we explore emergent strategies and methods for isolating, creating, and characterizing enzymes optimized for bioproduction. We discuss fundamental approaches to discovering and generating enzyme variants and identifying those best suited for specific applications. Additionally, we cover techniques for creating libraries using automated systems and highlight innovative high-throughput screening methods that have been successfully employed to develop novel biocatalysts for natural product synthesis.

Received 26th June 2024

DOI: 10.1039/d4np00031e

rsc.li/npr

1. Introduction
2. Sources of natural and engineered enzyme variants
- 2.1 Genome and metagenome mining: unearthing novel biocatalysts
- 2.2 Generation of enzyme mutants *via* gene diversification
3. Tools for high-throughput generation of variant libraries
- 3.1 Automated pipelines for creating gene libraries
- 3.2 Microfluidics platforms for DNA assembly
- 3.3 High-throughput genome engineering protocols
4. Strategies and screening/selection technologies for novel biocatalysts
- 4.1 Strategies for diversification of the catalytic landscape
- 4.2 Enzyme diversification *in vivo*
5. Emergent tools for high-throughput screening/selection of tomorrow's enzymes
- 5.1 Conventional setups are accelerated by integrating liquid handling systems
- 5.2 Single cell-based assays using fluorescence activated cell sorting
- 5.3 Droplet-based technologies greatly enhance screening capabilities for natural products
6. Conclusion

7. Author's contributions
8. Conflicts of interest
9. Acknowledgments
10. References

## 1. Introduction

Enzymes are powerful tools for synthetic chemistry because they can produce (bio)molecules with high selectivity and specificity.<sup>1–4</sup> Biocatalysis, the use of enzymes to perform chemical reactions, offers many advantages over conventional chemical methods, *e.g.*, milder reaction conditions, reduced environmental impact, and access to novel chemical functionalities.<sup>5–8</sup> However, natural enzymes are often not suitable for synthetic applications because they may lack the required activity, stability, or substrate scope.<sup>9</sup> Therefore, enzyme engineering—the modification of enzyme structure and function by rational design or directed evolution—is essential for implementing efficient bioprocesses.<sup>10–13</sup> Such bioprocesses, in turn, are key components of a bioeconomy, *i.e.*, an economic system that relies on renewable biological resources and biotechnological processes to produce goods and services in a sustainable manner.<sup>14–16</sup> Establishing a true bioeconomy has the potential to address global challenges, *e.g.*, climate change, resource depletion, food security, and public health.<sup>17–19</sup> For a global bioeconomy to become a reality, the challenge of

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark. E-mail: erayub@biosustain.dtu.dk; emchr@biosustain.dtu.dk; chdavo@biosustain.dtu.dk; pabnik@biosustain.dtu.dk; Tel: +45 93 5119 18



developing efficient biocatalysts and improving existing ones needs to be urgently addressed.<sup>20,21</sup>

Enzyme engineering enables the discovery or creation of biocatalysts that can mediate new reactions, accept alternative substrates, or operate under different process conditions. The recent developments in computational and experimental techniques have accelerated the process of enzyme engineering and expanded the repertoire of biocatalytic transformations.<sup>22–25</sup> Nowadays, enzyme engineering is a key driver of biocatalysis and bioeconomy, *i.e.*, the sustainable production of chemicals and materials from renewable biological resources.<sup>26–28</sup> Some of the current strategies for the development of high-performance biocatalysts involve screening natural diversity to isolate enzymes with the desired activity, engineering biocatalysts to expand their substrate range, redesigning mechanisms to create novel reactivity, and computationally designing enzymes from scratch. Biobased products and processes require tailored enzymes that fit their specific needs, *e.g.*, activity (rates), stability, sensitivity, and selectivity. The discovery, design, and optimization of enzymes for bioproduction requires a systematic and comprehensive approach.<sup>29–31</sup> Yet, one of the main challenges in enzyme engineering is the screening of enzyme variants, *i.e.*, the selection of the best-performing enzymes from a large and diverse pool of mutants.<sup>32</sup>

In this review, we discuss the growing demand for improved enzymes as biocatalysts for a number of bioprocesses and the increasingly important role of high-throughput screening (HTS) methods in meeting these needs. We explore possible sources of enzyme variants, including state-of-the-art genome mining and gene diversification techniques. Additionally, we examine recent examples of machine learning (ML) and *de novo* enzyme design, highlighting their current limitations. The review also considers the expanding toolbox for HTS library generation and how it can complement *in silico* towards *de novo* biocatalyst

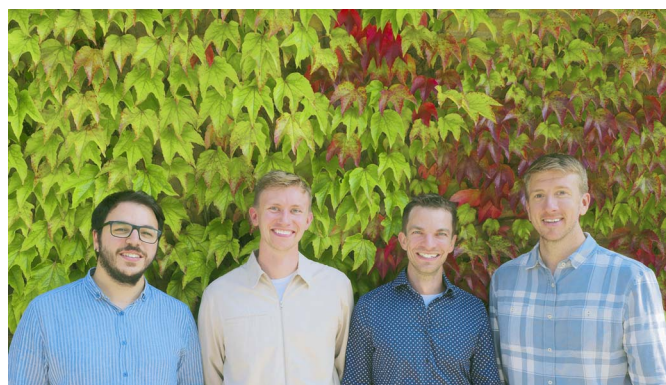
engineering design and selection. We then outline strategies for screening and isolating novel enzymatic activities, focusing on both *in vivo* and *in vitro* HTS approaches. By showcasing recent progress and improvements in these methods, this review provides a roadmap for researchers in natural product synthesis, biocatalysis, and metabolic engineering who are interested in integrating HTS strategies into their workflows.

## 2. Sources of natural and engineered enzyme variants

### 2.1 Genome and metagenome mining: unearthing novel biocatalysts

Nature has already developed myriads of enzymes that can be used as biocatalysts. Instead of creating new catalysts from scratch, the rich diversity of enzymes that exists in nature could be the starting point for selecting biocatalysts.<sup>33–35</sup> This approach takes advantage of the evolutionary optimization that these enzymes have undergone over time, which makes them efficient and specific in their catalytic activities.<sup>36</sup> Traditionally, biocatalyst discovery relied on isolating and cultivating microbes, a laborious and limiting approach. Recent advancements in genome mining offer a powerful alternative to isolation-dependent methodologies.<sup>37</sup>

Genome mining bypasses cultivation by directly analyzing microbial genetic data (metagenomes). This approach unveils a hidden treasure trove of enzymes encoded within hitherto unculturable microbes and hence facilitates the discovery of novel biocatalysts with unique properties. Dedicated software packages (*e.g.*, antiSMASH<sup>38</sup>) help navigating these vast datasets, predicting functionalities based on biosynthetic gene cluster similarities. Additionally, sequence comparison tools (*e.g.*, BLAST) aid in selecting the most promising candidates,<sup>39–43</sup> while EnzymeMiner<sup>44</sup> complements these approaches by automating



From left to right

Eray U. Bozkurt is pursuing his PhD in the Systems Environmental Microbiology (SEM) Group at the Technical University of Denmark (DTU). He is developing high-throughput enzyme engineering platforms, leveraging microfluidics and automation. Eray holds a M.Sc. degree from Bilkent University (Turkey), where he specialized in protein engineering and biosensors. Emil C. Ørsted, a PhD student at the SEM Group, explores synthetic methanol assimilation and lignin valorization by *Pseudomonas putida*. He holds a M.Sc. degree from DTU, focusing on cell factory engineering. He participated in iGEM (2021), developing a CRISPR toolkit for yeast engineering. Daniel C. Volke, Senior Researcher and co-PI at the SEM Group, specializes in expanding the synthetic biology toolbox for reprogramming environmental bacteria. He holds a PhD from RWTH Aachen University and a M.Sc. from the Technical University of Munich (Germany). Pablo I. Nikel is a full Professor at DTU and head of the SEM Group. He earned a PhD in Biotechnology & Molecular Biology in Buenos Aires, and has worked in the USA, France, Austria, and Spain (supported by EMBO and MSCA). His team at DTU is pioneering the implementation of synthetic metabolism, pushing the boundaries of microbial biochemistry to produce new-to-nature chemicals.



the search for soluble enzymes across a wide range of organisms; the latter tool filters candidates based on user-defined criteria (e.g., activity and stability), thus refining the selection process for enzymes that are not only novel but also industrially viable.

While extensive databases like UniProt,<sup>45</sup> Brenda,<sup>46</sup> and RetroBioCat<sup>47</sup> offer a wealth of information on known enzymes, information on how these enzymes interact with non-natural substrates can be limited. This is a key challenge in enzyme engineering because the engineering campaign typically relies on starting with an enzyme that already possesses some activity. Therefore, to explore the vast sequence space found in metagenomic data and find enzymes with the desired activity, bioinformatics and HTS screening can be coupled to overcome the starting activity hurdle and aid in biocatalyst development for industrial applications.<sup>48,49</sup>

AlphaFold2, an AI-driven tool developed by DeepMind<sup>50</sup> that builds upon the foundation of genome mining and bioinformatics, has marked a paradigm shift in our understanding of protein structures and their complex interactions. AlphaFold2 has revolutionized the field by accurately predicting the three-dimensional structures of proteins solely from their amino acid sequences. The implications for the field are relevant because this tool allows for rapid modeling proteins of unknown structures, thereby accelerating the discovery and engineering of novel enzymes and biocatalysts.<sup>51</sup>

The impact of AlphaFold2 extends beyond mere structure prediction and has catalyzed a deeper comprehension of protein dynamics and function. By elucidating the intricate folding patterns that dictate protein activity,<sup>52</sup> this tool has opened new avenues for enzyme engineering, making possible the rational design of proteins with enhanced or novel functionalities. This feature is particularly significant for industrial biocatalysis, where the ability to predict and manipulate protein structures can lead to more efficient and sustainable processes.<sup>53</sup>

The recent introduction of AlphaFold3 has taken these capabilities a step further by incorporating the prediction of protein–ligand interactions into its repertoire.<sup>54</sup> This improvement is a leap forward in computational biology because it enables the prediction of how proteins interact with small molecules, which is crucial for understanding enzyme–substrate relationships. The predictive power of AlphaFold 3 facilitates the exploration of enzyme–substrate interactions with non-natural substrates, and thus provides a powerful tool for enzyme engineers seeking to tailor biocatalysts for specific industrial applications. While many of the tools discussed are effective at predicting protein structure, significant progress is still needed to enhance the predictions on enhancing enzyme activities. Such improvements require further refinement, and potential candidates generated by these tools must undergo rigorous testing to confirm their activity. In this context, a recent review by Gantz *et al.*<sup>55</sup> examines the interplay between HTS systems and ML in protein engineering.

## 2.2 Generation of enzyme mutants *via* gene diversification

Another approach for exploiting the variant sequence space is *in vitro* gene diversification which can be achieved through

random or focused mutagenesis. Random mutagenesis samples the sequence space sparsely to identify hot spots correlated with desired properties without the need for detailed information. The most used method, error-prone PCR (epPCR), relies on low-fidelity DNA polymerases for replication to create mutations, and therefore facilitates rapid engineering of target proteins. By adjusting experimental conditions, e.g., elevating magnesium levels, adding manganese, or using imbalanced concentrations of deoxynucleotide triphosphates (dNTPs), the mutation frequency can be significantly enhanced to reach rates as high as  $8 \times 10^{-3}$  per nucleotide. The mutation rate can be further increased by utilizing mutagenic nucleotide analogues. These analogues have alternate base pairing properties and therefore greatly increase the mutation rate, by up to  $10^{-1}$  per nucleotide, yielding highly mutagenized variants.<sup>56</sup> While this method does not require any structure or function information, it has some limitations. First, *Taq* polymerase favors certain mutations over others and therefore the resulting library is biased.<sup>57</sup> To decrease such bias, another biased polymerase, Mutazyme, has been employed to counterbalance the bias introduced by *Taq* polymerase.<sup>58</sup> A disadvantage of epPCR is the inability of creating adjacent mutations. As a complementary method to epPCR, sequence saturation mutagenesis (SeSaM)<sup>57</sup> can introduce consecutive mutations, providing access to a mutation space that is largely inaccessible by traditional epPCR. Additionally, the creation of huge libraries necessitates the use of HTS techniques. In the absence of such methods, it is necessary to reduce the size of the library.

Focused mutagenesis (also termed as site-directed mutagenesis) is one effective strategy for accomplishing library size reduction. Unlike random mutagenesis, focused mutagenesis refers to diversification of predetermined locations of a gene. Predetermined locations are selected based on the detailed knowledge of the protein's structure and function. Site saturation mutagenesis is the most common type of focused mutagenesis, where all 19 possible amino acids are substituted at targeted positions which results in focused but highly mutagenized, smaller, and smarter libraries. To shrink the library size even further, codon degeneracy can be arranged. For example, by allowing only G or T at the wobble position, the library size can be decreased by 32-fold, making the screening more efficient.<sup>57</sup> Saturation mutagenesis can be performed in iterative rounds, with the best variant from each round used as the template for the next.

A recombination approach can also be used for creating smarter libraries. Homologous recombination is a natural evolutionary mechanism and key to running sequence diversification campaigns by DNA shuffling, which enables the exploration of genetic sequence spaces inaccessible by methods like epPCR and focused mutagenesis. The genes targeted for recombination (often a family of related genes) are first randomly broken down to fragments using DNaseI, and then the fragments of the desired size are selected and purified.<sup>57</sup> Recombination takes place when fragments from different parental genes anneal at regions where there is high similarity in the sequence. It has previously been shown that in many instances DNA shuffling led to the creation of proteins that





exhibit enhanced properties. Since reassembly relies on sequence similarity between DNA fragments, parental genes tend to be reconstructed due to high sequence identity. To facilitate combinatorial DNA shuffling, different techniques have been proposed. Random chimeragenesis on transient template (RACHITT) utilizes single stranded uracil-containing transient templates for library preparation.<sup>59</sup> Donor fragments are first hybridized with transient templates. Heteroduplex strand fragments are stabilized on the transient template in full-length, rather than small overlaps. Next, uracil-DNA-glycosylase treatment, followed by PCR, forms homoduplex double-stranded DNA, resulting in a shuffled DNA library with an increased number of crossovers.<sup>57</sup>

The choice of DNA library preparation method largely depends on the screening technique accessible. An essential distinction between DNA library preparation approaches is the pooling status of the library. Unlike methods utilizing liquid handlers and microfluidics, random mutagenesis generates a pooled library of variants. Consequently, the screening process must be compatible with a pooled library setup. Typically, screening of pooled libraries involves a two-stage process: an initial screening that favors the selection and enrichment of certain variants over others, which is followed by a confirmatory screening to verify the effectiveness of these variants.

In recent years, ML has emerged as a powerful tool in the field of rational design of biocatalysts.<sup>60</sup> The application of ML techniques has significantly accelerated the identification and optimization of enzymes for diverse catalytic reactions and applications in biocomputing.<sup>61</sup> ML algorithms can analyze vast datasets comprising protein structures, sequences, and reaction mechanisms, while extracting meaningful patterns and correlations that can guide the design process.

ML techniques, *e.g.*, deep learning and support vector machines, have been employed to predict enzyme–substrate interactions, substrate specificity, and catalytic activity. These models leverage the wealth of available biological and chemical data to make accurate predictions,<sup>62</sup> streamlining the search for optimal biocatalysts. The integration of ML with experimental approaches has led to more efficient and targeted enzyme engineering, ultimately contributing to the development of tailored biocatalysts for specific applications.

In addition to leveraging ML, *de novo* enzyme design represents a paradigm shift in the rational engineering of biocatalysts.<sup>63–65</sup> Unlike traditional rational enzyme design, where focus is placed on modifying the existing active site or binding pocket of the enzyme to improve its interaction with the substrate or the transition state, *de novo* design involves the creation of entirely new enzymes from scratch. This innovative approach is grounded in a deep understanding of protein structure, function, and the principles governing enzymatic catalysis.<sup>66</sup> The *de novo* enzyme design process typically begins with the identification of a target reaction and the determination of its key catalytic features. Computational tools, *e.g.*, molecular dynamics simulations and quantum mechanical calculations, are then employed to model potential enzyme structures capable of catalyzing the desired reaction. Iterative cycles of design and optimization, guided by both

computational predictions and experimental validation, refine the engineered enzymes for improved efficiency and specificity.

Advances in *de novo* enzyme design have opened avenues for tailoring biocatalysts to meet specific industrial and environmental needs. The ability to custom-design enzymes for non-natural reactions or challenging substrates has broad implications for biotechnology and offers sustainable solutions for various processes ranging from pharmaceutical synthesis to biofuel production. Moreover, integrating ML techniques and principles of *de novo* enzyme design are transforming the biocatalysis landscape to enable the development of highly efficient and tailored catalysts with unprecedented precision and speed. Yeh *et al.*<sup>67</sup> demonstrated these advantages by employing artificial intelligence to guide the creation of novel luciferase enzymes. By optimizing idealized protein scaffolds, they achieved high substrate specificity and catalytic efficiency for oxidative chemiluminescence. This work showcases the potential of deep learning in enzyme design, even without natural templates.

### 3. Tools for high-throughput generation of variant libraries

In the previous section, we described approaches for the systematic discovery and creation of enzymes. With increasing sophistication and speed of *in silico* designs, the construction of these has been tailored to become more efficient. DNA assembly is the foundational step that enables scientists to examine biological systems.<sup>68</sup> However, such assembly most often involves labor intensive, error prone, and repetitive tasks.<sup>69</sup> In fact, DNA assembly can take up more than 50% of the time a researcher spends in the lab. For these very reasons, DNA assembly is an excellent candidate for automation.<sup>70</sup> Developments in technologies, *e.g.*, microfluidics and liquid handling, shaped the future of how biologists construct gene and strain libraries.<sup>71</sup> Another important advancement is the suitability of specific protocols for new technologies, including non-canonical microbial hosts, *e.g.*, *Pseudomonas putida*.<sup>72–76</sup> Modern DNA assembly methods, *e.g.*, Gibson Assembly and Golden Gate Assembly,<sup>77–81</sup> are also promising, particularly because they do not require sequential reagent addition. With increasing use of automated platforms, DNA assembly methods are becoming more advanced and suitable for satisfying the needs of specific automation workflows.<sup>82–84</sup>

#### 3.1 Automated pipelines for creating gene libraries

Molecular cloning lies at the core of construction of any kind of genetic library.<sup>85</sup> Since molecular cloning is such a widespread and highly demanded practice, employment of HTS systems has become essential. These systems are more efficient than manual processes and reduce both human error and the cost of experiments.<sup>70</sup> Notably, liquid handlers minimize the need for direct human intervention, freeing up researchers to focus on more critical tasks. It is therefore crucial that such systems must be accessible to a large proportion of scientists.<sup>82</sup> Low-cost liquid handlers in particular, for example, Opentrons, play an

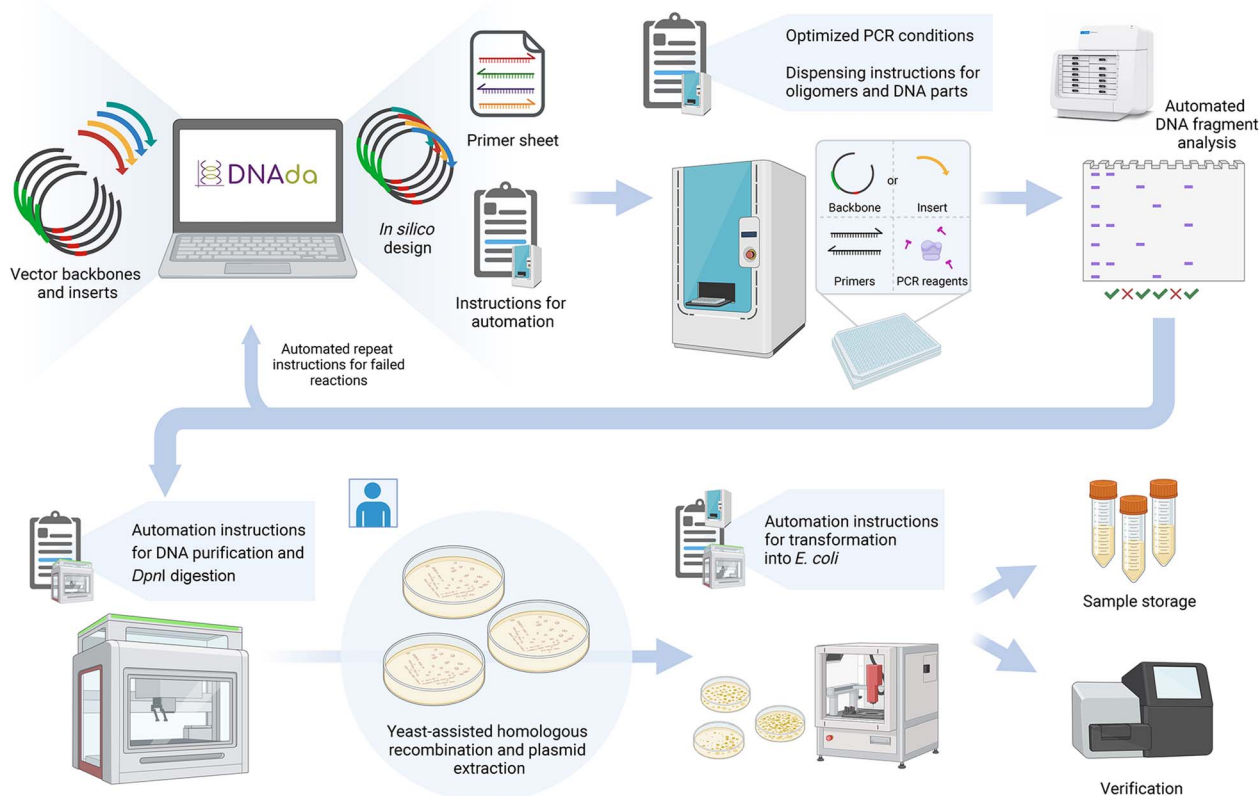


important role in removing the cost obstacle for entry into automated DNA assembly and are becoming prevalent as an accessible choice for workflow standardization.<sup>82</sup>

An excellent example on the use of Opentrons for DNA assembly automation was described in Storch *et al.*<sup>82</sup> The authors reported the implementation of a homology-based cloning method called Biopart Assembly Standard for Idempotent Cloning (BASIC). This assembly method utilizes 21-bp single-stranded overhangs to combine DNA parts. With this method, 88 independent variants, each expressing three genes, were created and tested. Automation decreased the hands-on time from more than 5 hours to 1.5 hours. The authors also calculated that the operating cost was as low as 1.5 \$ per DNA construct, which eliminates the barrier of cost for many practical applications. Despite its many advantages, this method is still not fully integrated and requires operator intervention. Four different scripts need to be initiated at the beginning of each step and the configuration of the plates needs to be arranged. Although introduction of a thermocycler module to OT-2 robots by Opentrons will add to the value of this method and will shorten the hands-on time, alternative methods

provide more flexibility in terms of experimental design. DNA-BOT, despite its low-cost and hence greater accessibility, is yet to be fully integrated with cloning pre-steps, *e.g.*, oligomer design, and post-steps, *e.g.*, sequence verification. Therefore, pre- and post- DNA cloning steps remain as the bottleneck of the workflow.

Comprehensive integration of the procedure is crucial for maintaining high throughput since the overall throughput of the method is going to be evaluated based on the slowest process. Therefore, transition between steps and inclusion of pre- and post-DNA assembly steps should also be considered to achieve more robust workflows. The application developed by Nava *et al.*,<sup>86</sup> termed DNAda, offers a comprehensive platform for DNA assembly (Fig. 1). DeviceEditor,<sup>87</sup> a j5 algorithm-based user-friendly software for DNA assembly, lies at the core of the DNAda platform. The j5 algorithm is a computational tool that creates automated protocols for DNA assembly. Since its introduction in 2012, this algorithm has been one of the key technologies supporting computer-aided design of DNA constructs.<sup>88</sup> DNAda is the first example of integration of an algorithm into automated laboratory workflows.



**Fig. 1** DNAda, an end-to-end, automated DNA assembly platform. The sequence begins with *in silico* combinatorial DNA design using DeviceEditor, which generates primers for PCR amplification of each part and construct. The system also provides instructions for the Echo acoustic liquid handler to arrange oligonucleotides and templates into compatible plates for each design. DNAda can analyze PCR results with an integrated gel electrophoresis system and offers guidelines to redo any unsuccessful amplifications. After successful amplification, the software directs the Echo liquid handler to purify DNA using magnetic beads and coordinates the components for DNA assembly. The process continues with a yeast-assisted homologous recombination protocol, followed by plasmid DNA extraction. DNAda then assists in transforming *Escherichia coli*, selecting colonies, and managing post-procedure steps, including stock preparation and submission for verification of all constructs. Adapted from Nava *et al.*<sup>86</sup>



The DNAda platform utilizes j5-designed DNA assemblies and provides streamlined automation instructions in each step of the DNA assembly process. The workflow starts by designing an *in silico* combinatorial DNA assembly using DeviceEditor, which designs oligonucleotides for PCR amplification for each construct and outputs a purchasing order sheet. Furthermore, the DNAda application provides automation instructions compatible with an Echo Acoustic Liquid Handler (Beckman Coulter). Machine-compatible plates, using appropriate oligonucleotides and templates, can be incorporated to assemble each construct designed *in silico*. Additionally, if amplification results are analyzed using a compatible gel electrophoresis device [Zero Agarose Gel (ZAG) DNA electrophoresis; Agilent Technologies], the DNAda application can interpret the results of PCR and provide instructions to repeat the failed amplifications. The high-level of automation not only decreases hands-on time but also minimizes interactions by the user in terms of sample handling and tracking.

After successful amplification, the program proceeds with Echo liquid handler instructions for DNA purification using magnetic beads and mixing parts for DNA assembly. The user then performs the yeast-assisted homologous recombination protocol and plasmid extraction. After plasmid extraction, the DNAda application assists with transformation to *Escherichia coli*, colony picking and, most importantly, post-steps (e.g., stock preparation and submission for verification).

The DNA assembly methods in the examples above still have flexibility limitations. Restriction based methods, e.g., Golden Gate assembly, require all DNA fragments to be free of “unintended” restriction sites, which places a huge burden on the versatility of the method. Even though homology based cloning methods, e.g., SLIC, SLICE, or Gibson assembly do not have stringent sequence constraints, they are not entirely without limitations.<sup>89</sup> Single-stranded DNA (ssDNA) generated by this method may form secondary structures which significantly decrease the efficiency of certain constructs. The tendency to form secondary structures introduces a degree of bias into these methods. Therefore, in an outstanding example of end-to-end automation at biofoundry scale, Enghiad *et al.*<sup>89,90</sup> developed a novel cloning strategy based on artificial restriction enzymes. The cloning strategy “PfAgo” can create defined ends unlike homology-based methods where exonuclease activity creates >20–40 nt; therefore, “PfAgo” can handle repeated sequences better and theoretically can construct any sequence. This workflow represents a fully automated, end-to-end platform integrating 23 machines. The workflow starts with entering the sequences to be assembled after which the system requires no further human interference. The developed software designs primers and guides for PfAgo cloning. This software is also capable of making smart decisions that overlap with daily wet lab work. For example, if a part is smaller than 80 bp, the software makes the decision to add the part as overhangs by PCR, if feasible. Furthermore, for verification of plasmid assembly, the platform performs confirmatory digestion in which the restriction sites are decided so that the resulting bands are distinct and observable.

The platform can be divided into three main modules: (i) upon receiving the oligo parts, the process from diluting oligonucleotide stocks until PCR purification is performed without human intervention with the help of complex automation equipment, e.g., centrifuges, robotic arms, and liquid handlers; (ii) once the required DNA sequences have been amplified, the cloning module spans from mixing correct sequences to be joined to transformation of assembled products to *E. coli* cells using integrated liquid handlers and thermocyclers; and (iii) the last module consists of isolating single colonies, purification of plasmids and verification of cloning by digestion followed by gel electrophoresis and preparing glycerol stocks for verified plasmids. The workflow has been highly validated by assembling 101 plasmids with different sizes of plasmids, ranging from 5 to 18 kb, with up to 12 DNA parts for six different species. The combination of the three automated modules amount to 20 000 pipetting steps that otherwise would need to be performed by a group of researchers. In this workflow, the operator only has to make decisions based on the experimental designs—and repeat if necessary.

### 3.2 Microfluidics platforms for DNA assembly

Microfluidics technology offers handling of even lower volumes than automation robots, e.g., Echo liquid handler or Opentrons, ranging from microliters to femtoliters.<sup>91</sup> The emergence of Gibson Cloning and Golden Gate Assembly, which do not require sequential reagent addition, have enabled microfluidics applications at DNA level. Developments in digital microfluidics now allow standard biological protocols, e.g., DNA assembly, transformation, and incubation, to become a streamlined and integrated process. Adoption of Lab-On-A-Chip devices is a milestone in biotechnology by minimizing the volumes, cost, effort, and time of the biological processes.<sup>92</sup>

Shih *et al.*<sup>93</sup> proposed a framework on how microfluidics could accelerate the build part of the Design-Build-Test-Learn (DBTL) cycle. The authors showcased a first-case example of using hybrid microfluidics for synthetic biology by adopting three common cloning techniques, i.e., Golden Gate, Gibson Assembly, and yeast assembly. This engineering effort was validated by automatically constructing 32 separate plasmids. The microfluidics device consists of three regions. The first one, the DNA assembly region, is where droplets carrying inserts and vectors are mixed. In this step, necessary reagents for the cloning technique are also added. Next, the droplets are incubated for varying times, depending on the cloning technique. A built-in temperature control ensures the right thermodynamics for DNA assembly. The plasmids are then mixed with cells and the droplets are delivered to the next region where a potential is applied to the droplet containing assembled DNA and cells for electroporation, hence allowing plasmid assembly and electroporation of assembled plasmids on-a-chip. The system shows potential for automating and scaling up DNA assembly and made possible the design of up to 16 unique constructs per run. While effective for assembling plasmids and performing transformations, the ability to handle multiple DNA assemblies simultaneously was not verified. Users need to manage



preparatory and follow-up steps to maintain the throughput offered by the device.

An ideal HTS setup should cover the overall process. When a step is not amenable for automation or needs significant user intervention, it creates a process bottleneck hindering the overall throughput of the workflow. A comprehensive example is provided by Linshiz *et al.*<sup>94</sup> The introduced platform facilitates a seamless integration of design, build, and test phases, beginning with automated *in silico* DNA library design and culminating in evaluation of these libraries. The “DNA Constructor” software was created for optimizing combinatorial DNA library protocols. For DNA assembly, a new method named isothermal hierarchical DNA construction (IHDC) was developed. The advantages of IHDC are that it is independent of temperature cycling and needs reduced process control, making it particularly well-suited for microfluidic environments. In addition, the adaptability of IHDC for DNA cloning was validated using Gibson assembly, Golden Gate assembly, and yeast DNA assembly methods. Heat-shock transformation was performed utilizing a Peltier external temperature controller. Furthermore, Linshiz *et al.*<sup>94</sup> evaluated the DNA library basic assessment methods by combining their setup to an external camera, such that fluorometric and colorimetric assays can be easily monitored.

Overall, the automation era of biology brings standardization to molecular cloning protocols and minimizes the bias introduced by the user. Employment of automation techniques allows higher throughput per round of experiment and significantly decreases hands-on time. Furthermore, the much smaller volumes required per sample enable efficient cost-minimization. Liquid handlers are a vital asset for the next foundational steps of automated biotechnology due to their greater adaptability with current DNA assembly protocols, relative ease of integration with other equipment, and higher throughput compared to microfluidics systems to date. However, there is a limit to miniaturization of liquid handlers due to machine capabilities and evaporation in open-air systems. Microfluidics systems as a closed platform have advantages in terms of down-scaling and therefore cost effectiveness. Both systems have unique advantages and drawbacks, and the future of these technologies may lie at the intersection of microfluidics and robotic systems, with liquid handlers integrated into microfluidics setups to allow manipulation of liquid in microfluidics channels.

### 3.3 High-throughput genome engineering protocols

Numerous products obtained by biocatalysis (both *in vitro* and *in vivo*) have significant commercial importance within pharmaceutical, fine and bulk chemicals, including terpenoids, alcohols, amino acids, organic acids, and antibiotics.<sup>95–98</sup> Many of these compounds are created *via* complex metabolic pathways and are subject to strict regulation in their native context. Furthermore, natural strains often lack the capability to produce these valuable compounds in sufficient quantities.<sup>99</sup> Consequently, extensive genome engineering is needed to reconstruct and debottleneck these metabolic networks to achieve sufficient titers of the products of interest,<sup>96</sup> and to generate new-to-Nature

products.<sup>100–103</sup> High-throughput genome engineering tools are therefore essential for assembling metabolic networks and generating cell-based libraries. In recent years, multiplex automated genome evolution (MAGE)<sup>104</sup> has become a prevalent tool for high-throughput genome engineering.

MAGE is an efficient and rapid tool that allows gene manipulation simultaneously in multiple loci. Initially described for *E. coli*,<sup>105</sup> the technique was quickly adapted for its use in a wide range of prokaryotic and eukaryotic organisms. MAGE utilizes the  $\lambda$ -Red phage proteins to facilitate recombination between short single stranded DNA and the genomic locus of interest. The workflow starts by growing the cells at 30 °C until the culture reaches mid-logarithmic phase. A temperature inducible expression system regulates the expression of  $\lambda$ -Red proteins. Thus, when the cells are transferred to 42 °C, the expression of  $\lambda$ -Red genes is initiated, and the mismatch repair (MMR) system of the host is inactivated.<sup>106</sup> Upon introducing single-stranded, mutation-carrying DNA oligonucleotides *via* electroporation, the oligonucleotides integrate in the lagging strand of the replication fork during DNA replication, yielding mutated genomes.<sup>47</sup> In this way, MAGE allows specific gene manipulations while maintaining a stable genome by using only DNA oligonucleotides and creating genetic diversity that can be tested in a search for better production strains.

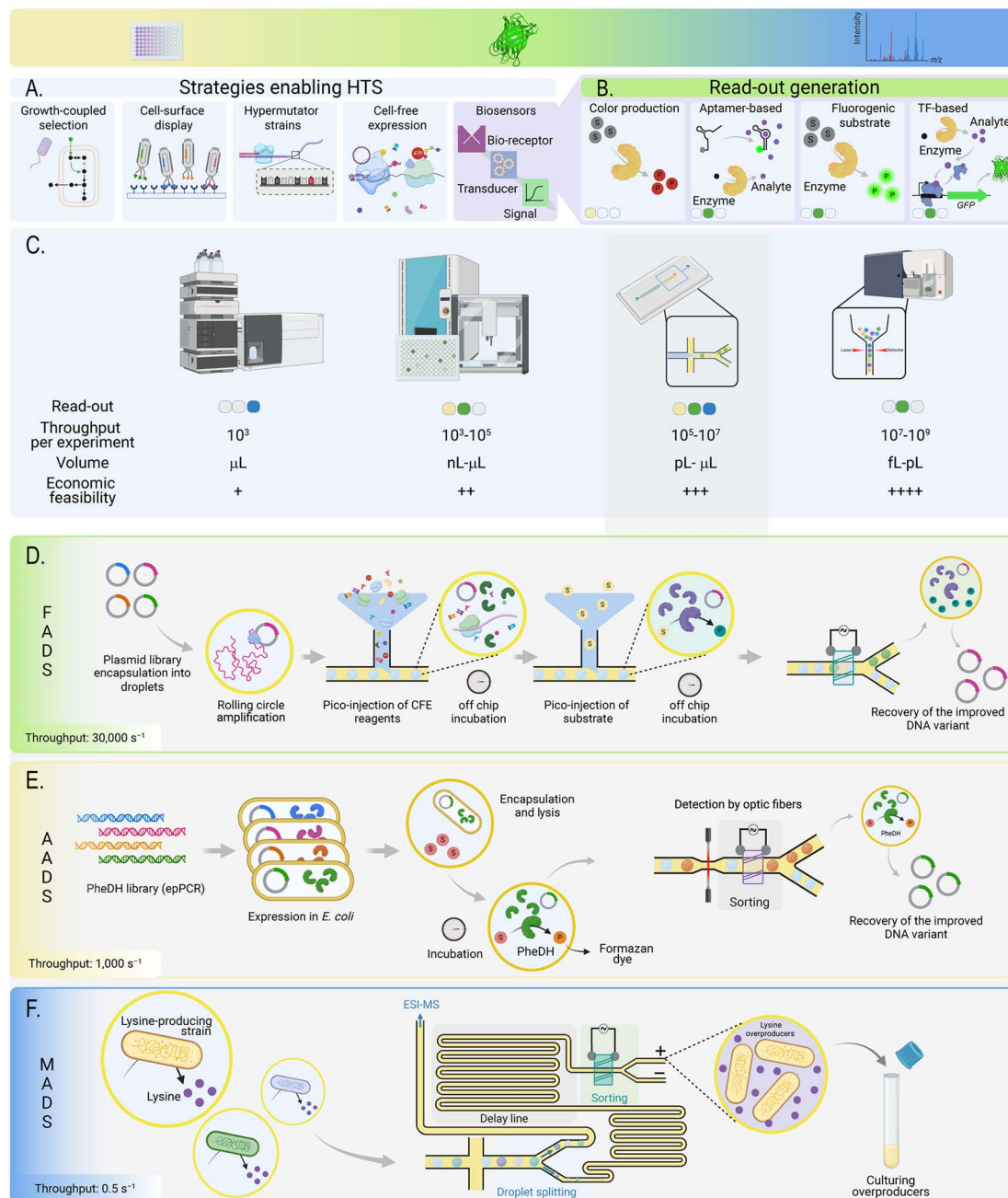
Since the MAGE system was first described in 2009, it has proven useful for many purposes, including (but not limited to) changing the native genetic code of microbial hosts,<sup>107,108</sup> incorporating non-standard amino acids,<sup>109</sup> inserting histidine tags,<sup>110</sup> and optimizing the production of high-value chemicals, *e.g.*, lycopene,<sup>111</sup> curcumin,<sup>112</sup>  $\beta$ -carotene,<sup>113</sup> and L-DOPA, with improvements in the product titer of up to 38.2 fold.<sup>112</sup> For instance, while achieving improved curcumin production, Kang *et al.*<sup>112</sup> constructed an *E. coli* strain library where 5'-untranslated region sequences of six genes playing a role in curcumin pathway were randomized. The strain library was first screened by visual inspection on agar plates for color formation. For selected variants, LC-MS measurements were carried out. Strikingly, the highest performing strains were observed to express significantly less of two of the six enzymes, which led to improved production of curcumin. In another article, L-DOPA production from glucose was improved by tuning the ribosome binding site of 15 genes utilizing MAGE. In yet another example,  $\beta$ -carotene production in *S. cerevisiae*, was improved by creating an *S. cerevisiae* strain library. Here, the promoters, open reading frames, and terminators of four genes in  $\beta$ -carotene production pathway were targeted. The library then showed different levels of  $\beta$ -carotene and diverse colorimetric phenotypes, proving the effectiveness of MAGE.<sup>114</sup>

## 4. Strategies and screening/selection technologies for novel biocatalysts

A major challenge of synthetic biology is the creation of efficient HTS and automated pipelines for the Design-Build-Test-Lean (DBTL) cycle.<sup>115–117</sup> Democratization of such pipelines is another challenge to harness their full potential.<sup>118</sup> In this







**Fig. 2** High-throughput technologies for selecting superior biocatalysts. (A) High-throughput screening (HTS) workflows implemented to facilitate the generation and screening of enzyme variants. (B) Techniques applied for read-out detection in HTS of enzymatic reactions. TF, transcription factor. (C) Comparison of different detection methods that enable versatile HTS workflows. Chromatographic separation, coupled with various detection methods (e.g., LC-MS, HPLC, GC-MS), has a relatively low throughput, but products can be measured and quantified directly. Fluorescent and colorimetric assays in a microtiter plate format can be scaled down to the nL level. Automation is possible when using an acoustic liquid handler. Microfluidics droplet generation is compatible with various manipulation and sorting strategies for different readouts, as shown in panels (D)–(F). FACS enables extremely HTS with tiny reaction volumes. (D) Fluorescence activated droplet sorting (FADS)-based screening platform by Holstein *et al.*<sup>119</sup> A cell-free expression (CFE) system was employed to screen a large ( $10^{14}$ ) protease library. To this end, the DNA encoding the variants was encapsulated together with an isothermal DNA polymerase and PCR components to amplify the template in a completely *in vitro* setup. This approach increases the gene copies in each droplet to 30 000, ensuring adequate protein yield in subsequent CFE processes. CFE reagents were pico-injected into the droplets, which were then incubated for 4 hours to allow protein production. Finally, a fluorogenic casein substrate was pico-injected, and the droplets underwent another incubation period for hydrolytic cleavage to occur, producing a fluorescent signal. The droplets were then sorted based on their fluorescence using FADS. (E) Outline of the absorbance activated droplet sorting (AADS) system by Gielen *et al.*<sup>120</sup> *E. coli* cells carrying a *pheDH* library (obtained via error-prone PCR, epPCR) were encapsulated into droplets and lysed to release the enzymes. Inside these droplets, the exposed enzymes reacted with the substrate to catalyze the formation of formazan dye. Post-incubation, droplets were sorted based on their absorbance. This process identified a *PheDH* variant with an activity increase of up to 2800-fold, achieving read-out rates of 300 droplets per second. (F) Mass-activated droplet sorting (MADS), developed by Holland-Moritz *et al.*<sup>121</sup> and adopted by Payne *et al.*<sup>122</sup> to screen for lysine overproducers. This system analyzes the content of droplets by splitting





section, we describe some selection/screening strategies (Fig. 2A) coupled with HTS and automated pipelines (Fig. 2B) that can be implemented with the technologies described in the previous sections. Such strategies can be generally divided into two categories: selection-based and screening-based. Selection refers to the survival of a functional (strain) variant, while nonfunctional variants were eliminated. In contrast, screening refers to the assessment of as many variants as possible, thus reducing the likelihood of missing a desired trait. The approaches can also be categorized based on several other criteria, including employment of *in vivo* versus *in vitro* approaches, analyte detection method (e.g., colorimetric, fluorometric, or analytical; Fig. 2C), and whether an extracellular or an intracellular compound is detected. The choice of technology varies significantly based on the analyte and the chosen output method. We examine the strategies that can be implemented in HTS setups and how current state-of-the-art technologies can be combined with different screening/selection strategies.

#### 4.1 Strategies for diversification of the catalytic landscape

*In vivo* mutagenesis encompasses a set of techniques used to introduce mutations into the DNA of living organisms. Traditionally, these methods have relied on random mutation of the entire genome, including exposure to mutagens, e.g., UV radiation or mutagenic chemicals, but it is possible nowadays to target and modify specific regions containing genes or entire pathways.<sup>123</sup> Several molecular tools can be implemented to this end, e.g., CRISPR-Cas systems, oligonucleotides, or specialized enzymes,<sup>124–131</sup> which can be designed to recognize and alter specific DNA sequences. By adopting *in vivo* genetic tools to increase enzyme variability, the method can speed up the testing phase of the DBTL cycle and aid in selecting improved variants by coupling the enzymatic activities to growth in HTS setups.<sup>132</sup>

Once *in vivo* mutagenesis has been applied to enhance library complexity, a pool of mutants is conventionally screened by plating on selective agar plates and selecting colonies based on size, fluorescence, or other phenotypical characteristics. One of the most commonly adopted strategies for enzyme selection is the use of biosensors.<sup>133</sup> Biosensors play a crucial role in these systems, serving as molecular sentinels that detect and quantify the presence of specific genetic or phenotypic changes. The use of cell surface display technologies, which present the biosensor on the exterior of the cell, allow for immediate interaction with the extracellular space of the target molecules. The integration of *in vivo* mutagenesis with HTS technologies, e.g., FACS and pico-droplet encapsulation, permits rapid analysis of large numbers of mutants at a scale and speed that were once inconceivable.<sup>134–136</sup> These automated high-throughput methods also significantly reduce the labor and costs associated with traditional mutagenesis techniques, e.g., manual screening of agar plates.<sup>137</sup>

#### 4.2 Enzyme diversification *in vivo*

*In vivo* assessment of enzymes can be achieved (among other strategies) in auxotrophic sensor strains,<sup>132</sup> with systematic growth-coupling designed by modeling,<sup>138</sup> or by using antimetabolite selection strains.<sup>139</sup> Common to all three assessment methods is the generation of a selective pressure that suppresses growth (Fig. 3 and 4), either through disrupting the microorganism's metabolic pathway through targeted deletions of specific genes or through use of metabolic analogs. In this way, growth of the microorganism is restored *via* synthesis of the target molecules.<sup>115</sup> This intentional disruption creates a dependency where the organism's survival and growth under restrictive conditions are solely dependent on the activity of a particular enzyme or metabolic pathway (Fig. 3A). Essentially, the organism can only thrive if the targeted biochemical process is functioning efficiently thereby linking survival to the successful operation of the desired metabolic route.<sup>142</sup> However, the optimization of enzymes through *in vivo* assessment is restricted to those that can be effectively integrated into the host cell metabolism to facilitate growth-coupling, or to those that yield an antimetabolite.

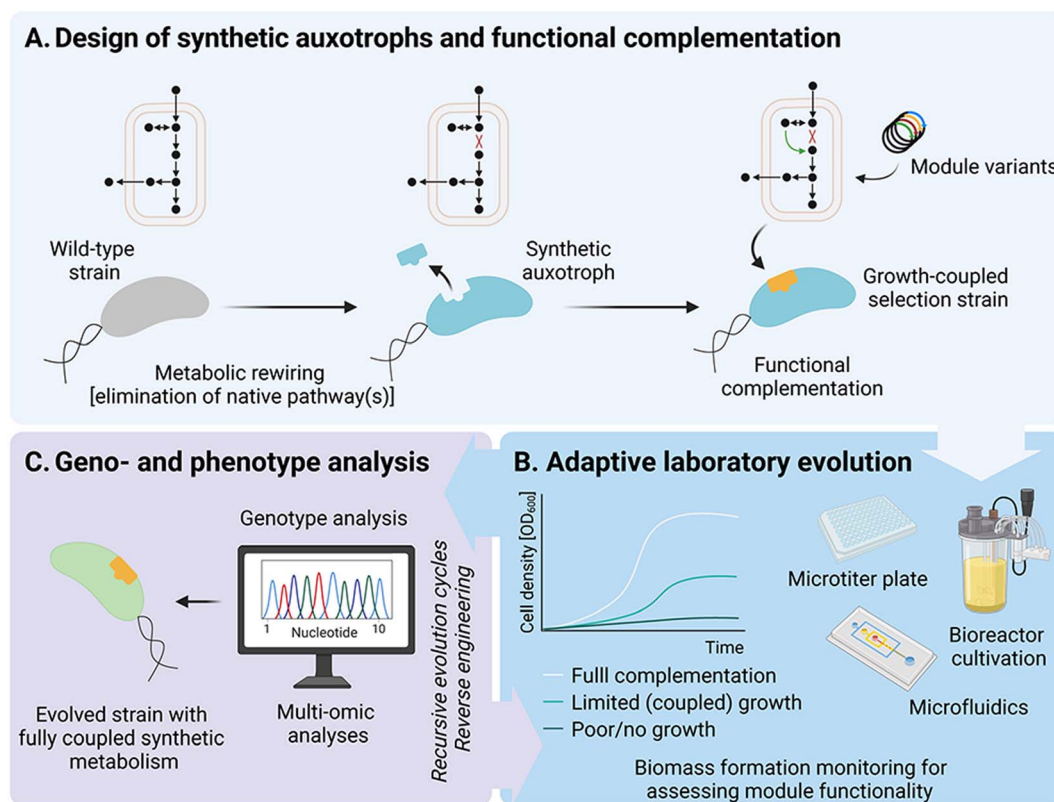
Consistent selective pressure can drive the evolution of auxotrophic selection strains and improve the flux capacity of targeted enzymes. By integrating strategic genetic engineering, growth-coupled selection, and adaptive laboratory evolution (ALE, Fig. 3B), enzyme variants or metabolic pathways can be efficiently identified and refined. Upon genotypic and phenotypic characterization (e.g., aided by multi-omic analyses; Fig. 3C), these methodologies enable harness optimized biological systems for high-throughput, growth-coupled selection for natural product biosynthesis.<sup>115,142</sup>

A recent study by Femmer *et al.*<sup>141</sup> demonstrated the miniaturization of a growth-based, enzyme evolution process where growth behavior correlates with the desired improvements in enzyme performance. Here, the authors engineered an arginine auxotrophic *E. coli* strain for directed evolution of the periplasmic broad-substrate racemase *PpAAR* that catalyzes racemization of D-ornithine (Fig. 4B). The product of *PpAAR*, L-ornithine, is a precursor of arginine synthesis and can therefore complement the auxotrophy. The stringency of the selection can be fine-tuned by including an antimetabolite, D-lysine, which is a competitive inhibitor of *PpAAR*. A HTS platform that leverages alginate droplets with nanoliter reaction volume (NLRs) was implemented to cultivate library clones. By creating conditions where only the bacteria with the most efficient racemase survive, the researchers could identify mutations that enhance enzymatic activity and sort the NLRs through fluorescence-assisted particle sorting.

This method is particularly useful because it enables selection of beneficial traits without the need to directly screen each individual variant, which would be impractical given the large

the droplet population into two fractions: one is sent for ESI-MS analysis and the other is kept in a microfluidic delay line for sorting. Based on the ESI-MS results, droplets are sorted. The utility of the system was demonstrated by employing two *E. coli* strains, i.e., a wild-type strain and a derivative harboring the DapA<sup>E84T</sup> mutation. These strains were encapsulated within emulsions and cultured in droplets to facilitate lysine production. Clones displaying high lysine production levels were subsequently cultured and underwent genotype validation.





**Fig. 3** Growth-coupled selection for metabolic modules. (A) To apply growth-coupled selection, a wild-type microbial strain is rewired through targeted gene deletions that disrupt the native metabolic network, rendering the strain auxotrophic for specific metabolites. These selection strains are then used to test synthetic modules, e.g., enzymatic reactions of interest and their variants, which provide the essential metabolite(s). Successful complementation restores the growth of the synthetic auxotroph. (B) Adaptive laboratory evolution (ALE) can be implemented to improve growth rates and biomass yields. These parameters are a proxy of the synthetic module activity, as they are coupled to each other under selective growth conditions. (C) ALE cycles can be iterated as necessary, and the evolved clones are then subjected to multi-omic characterization and genotyping. This sequential process allows for the identification and reverse-engineering of beneficial mutations in a *naïve* (non-evolved) microbial host. Adapted from Cros *et al.*<sup>140</sup>

number of possibilities. Instead, the growth of the microorganism itself indicates a successful mutation. This connection facilitates rapid screening and identification of enzyme variants with enhanced performance. The results exposed the platform's potential as a simple and effective strategy for *in vivo* enzyme evolution by quantification of residual substrate<sup>143</sup> or enzyme activity, either surface-displayed<sup>144,145</sup> or intracellular.<sup>146</sup> The utility of this method is further highlighted by the fact that it can be easily adapted to most screenings using different antimetabolites.

Selecting improved variants can also be automated using a biosensor. Biosensors are particularly valuable in HTS because they streamline the traditionally labor-intensive and time-consuming process of enzyme selection. Additionally, biosensors facilitate rapid and sensitive detection of specific biological molecules, making them ideal for identifying and selecting enzymes with desired characteristics.<sup>147</sup> In HTS, a large number of variants of enzymes or other proteins can be tested simultaneously for their ability to catalyze a reaction, bind to a substrate, or inhibit a process.<sup>148</sup> By integrating biosensors into microfluidic systems or other automated platforms, thousands of enzyme reactions can be monitored at

once, greatly accelerating the process of identifying the most effective enzymes for industrial processes, therapeutic applications, or research purposes.

Li *et al.*<sup>149</sup> described the utilization of a biosensor to enhance erythritol production in the oleaginous yeast *Yarrowia lipolytica* through a picodroplet-based co-culture system. The methodology revolves around the use of fluorescence-activated droplet sorting (FADS) combined with a transcription factor (TF)-based biosensor. This system facilitates high-throughput isolation of yeast mutants that overproduce erythritol, an important sugar alcohol in the food industry. In the absence of erythritol, the biosensor circuit expresses a repressor protein, inhibiting the expression of reporter gene. In the presence of erythritol, the repressor protein undergoes a conformational change and can no longer bind to the promoter region, allowing expression of the GFP reporter gene. The three-step droplet operation process is illustrated in Fig. 5. Initially, droplets encapsulating yeast mutants were generated and incubated to support erythritol production (Fig. 5A). Next, a fluorescent erythritol-biosensing *E. coli* strain was pico-injected into the droplets (Fig. 5B). Finally, the droplets with the greatest fluorescence, which correlates with the highest erythritol production, were sorted *via* FADS



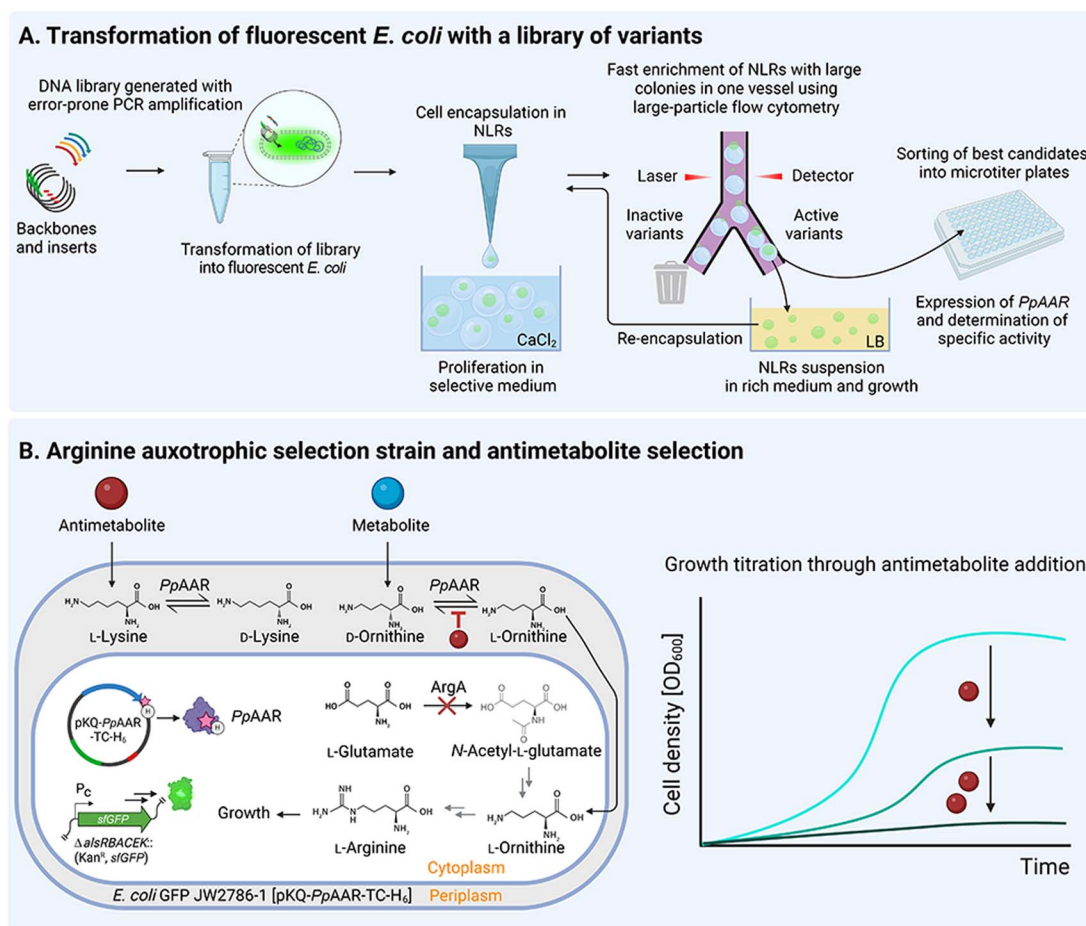


Fig. 4 High-throughput selection of enzyme variants encapsulated in alginate droplets with nanoliter reactors. (A) A DNA library encoding *PpAAR* racemase variants is generated using error-prone PCR amplification, and fluorescent *E. coli* cells expressing a superfolder GFP gene (*sfGFP*) are transformed with the library and encapsulated in nanoliter reactors (NLRs). These NLRs are formed by laminar jet break-up in a calcium chloride (CaCl<sub>2</sub>) solution. The encapsulated cells proliferate in a stringent selective medium, where only those with a functional enzyme can grow. Large-particle flow cytometry is used to rapidly enrich NLRs containing large colonies, effectively separating active and inactive enzyme variants. Clones with active enzyme variants are transferred to rich LB medium for growth before being re-encapsulated for another round of enrichment. The top candidates are sorted into microtiter plates, where the expression of the *PpAAR* racemase variants is analyzed to determine specific activity. (B) Deletion of *argA* creates an arginine auxotrophy that can be rescued by L-ornithine. L-Ornithine can be obtained from externally supplied D-ornithine through *PpAAR* activity. The gene encoding *PpAAR* is expressed from a plasmid that includes tetracycline (TC) and hexahistidine (H) tags, subsequently used for fluorescence-based quantification of *PpAAR* racemase levels in cell-free extracts and for protein purification, respectively. The *PpAAR* variants are transported to the periplasm, where L-lysine and D-ornithine compete for the active site; the selection stringency can be adjusted by titration with the antimetabolite L-lysine. Adapted from Femmer *et al.*<sup>141</sup>

(Fig. 5C). To optimize this process, the temperature and pH within the droplets were controlled, thus separating the erythritol production and detection phases and suppressing background biosensor expression using pH-sensitive erythromycin.

After four rounds of iterative mutagenesis and screening, a high-performing strain, *Yarrowia lipolytica* S4-9, was identified and isolated. This strain had a significant increase in erythritol yield (17%) and production rate (26%) in 5 liters bioreactor cultures compared to the parent strain. The success of the study illustrates the potential of the FADS-TF co-culture system for HTS of extracellular products, paving the way for its application in screening other valuable metabolites as more biosensors become available. The methodology and results collectively highlight the efficiency and effectiveness of the microfluidic-

biosensor co-culture system in enhancing industrial bioproduction and provide valuable information on how background expression of biosensors can be eliminated.

The platform described in the study advances the level of automation for HTS in bioproduction. The use of microfluidic droplets, combined with biosensors and FADS, enables the implementation of a largely automated process that can easily screen 10<sup>6</sup> samples per hour, identifying and isolating high-performing strains with minimal human intervention. However, while the system automates screening and selection of yeast mutants, human oversight is still necessary at certain stages. For instance, setting up the initial parameters, preparing the mutant libraries, and analyzing the final strains for erythritol production require human expertise. Additionally, the





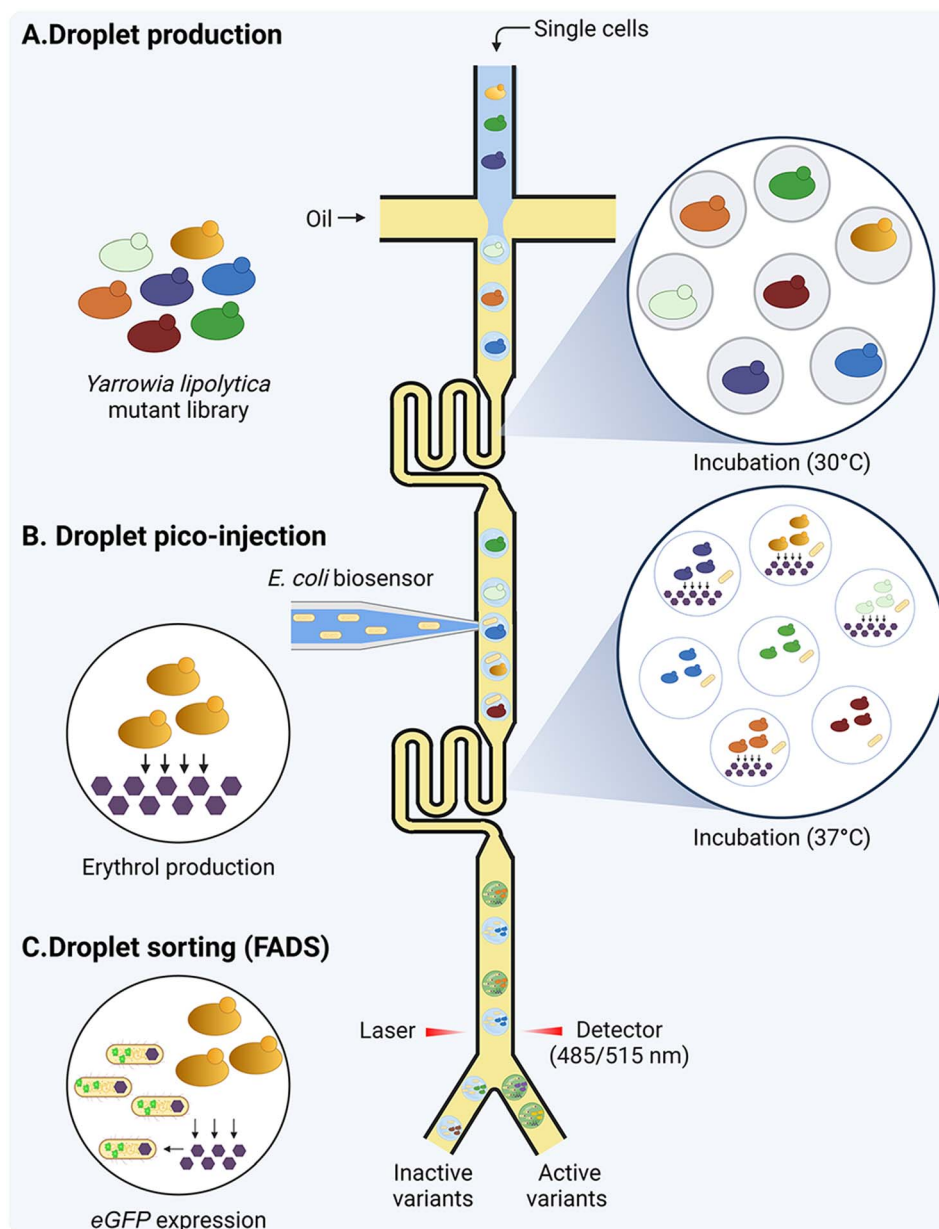


Fig. 5 Fluorescence-activated droplet sorting (FADS) platform for high-throughput screening of *Yarrowia lipolytica* mutants with improved erythritol production. (A) In a first step, single *Y. lipolytica* mutants are encapsulated and collected in a breathable Teflon tube, where they are incubated at 30 °C for 36 hours. During this incubation, the yeast cells proliferate and release erythritol into the droplets. (B) The second step involves injecting an *E. coli* biosensor strain into the droplets. The droplets are then incubated at 37 °C for 12 hours, allowing the *E. coli* biosensor to proliferate and detect erythritol produced by *Y. lipolytica* within the droplets. (C) Droplets with high fluorescence intensity (connected to enhanced GFP, eGFP, which indicates elevated erythritol concentrations), are sorted using FADS (see also Fig. 2D). This process allows for the identification and isolation of high-yield mutant strains. Adapted from Li *et al.*<sup>149</sup>

interpretation of results and decisions for subsequent rounds of mutagenesis and screening also relies on human judgment.

Cell surface display is an advanced biotechnological strategy that enables the presentation of heterologous proteins on the exterior of a microbial cell. This is accomplished by genetic fusion of the protein of interest, referred to as the passenger protein, with a carrier protein that possesses an anchoring domain. This domain facilitates the localization of the fusion protein to the cell membrane. Cell surface display can be

a powerful tool in HTS for enzyme selection due to its ability to link phenotype with genotype for extracellular processes and products. In this context, the phenotype is the observable functional trait of the enzyme, *e.g.*, its ability to catalyze a reaction, while the genotype is the genetic makeup that encodes the enzyme.

In HTS, a vast library of enzyme variants can be displayed on the surface of a cell, typically yeast or bacterium. The screening process involves exposing such enzyme libraries to a substrate



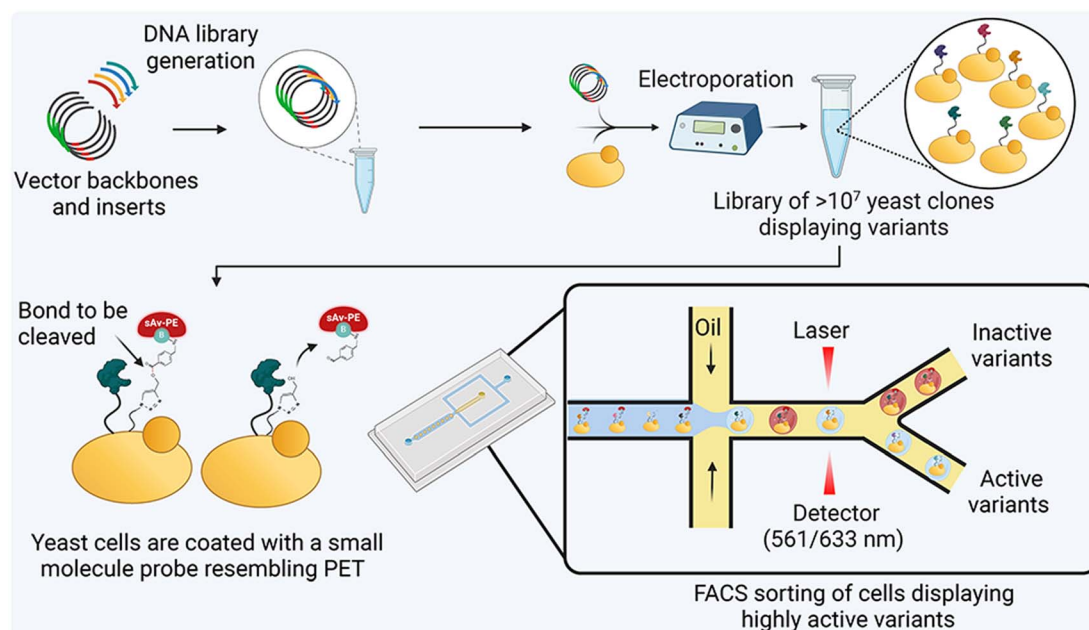
or a reaction condition. The enzymes that can catalyze the desired reaction will bind to or convert the substrate which can be detected by various methods. For instance, if the reaction produces a fluorescent product, the cells with the most active enzymes will glow if excited by light of a specific wavelength. These 'successful' cells can then be isolated using techniques, *e.g.*, FACS. Once isolated, the coding genes of these cells can be sequenced to identify the mutations that led to the improved enzyme function. This method allows researchers to screen millions of enzyme variants simultaneously, making it an ultrahigh-throughput approach. The approach is particularly useful for identifying enzymes with improved properties, *e.g.*, increased activity (rates), stability, or substrate specificity, which are important for industrial applications, *e.g.*, biofuel production, waste treatment, and synthesis of complex chemicals.

Cribari *et al.*<sup>150</sup> have developed a high-throughput yeast cell surface display platform that can rapidly evaluate over 10 million enzyme mutants, as shown in Fig. 6. Their novel methodology was used to enhance the activity of enzymes that degrade synthetic polymers, which is crucial for eco-friendly plastic recycling. The library size for screening of such enzymes was limited to around 10 000 mutants, due to the constraints imposed by polymer degradation assays. In the innovative platform, each yeast cell displays a unique mutant enzyme, and the activity of these enzymes is measured by the change in fluorescence when a synthetic probe, designed to

resemble a polymer of interest, is cleaved. The most active mutants, those showing increased fluorescence, are then isolated using FACS and subsequently sequenced. To demonstrate the efficacy of this platform, the researchers conducted directed evolution experiments on a polyethylene terephthalate (PET)-depolymerizing enzyme, known as leaf and branch compost cutinase (LCC).<sup>151</sup> They discovered mutations that significantly enhanced the enzyme's ability to degrade solid PET films. One mutation, H218Y, was highlighted for its role in improving the enzyme's binding affinity to PET, which was supported by biochemical assays and molecular dynamics simulations. Overall, the study achieved a remarkable increase in the throughput for screening polymer-degrading enzymes by three orders of magnitude.

While mutagenesis takes place concurrently with the selection of the desired phenotypic trait during directed *in vivo* evolution, *in vivo* hypermutators can be used to increase the mutation rate which is naturally between  $10^{-10}$  to  $10^{-9}$  per cell per generation to as high as  $10^{-4}$  within the target gene(s).<sup>152–155</sup> A hypermutator is a synthetic tool used to enhance the mutation rate within, or in proximity to, a gene of interest.<sup>156</sup> Different techniques have been developed for engineering hypermutator strains, as recently reviewed by Molina *et al.*<sup>153</sup>

The key advantage of using hypermutator strains is that they can be used to generate mutational depth quickly in the form of diversified enzyme variants that are highly fit to perform a desired activity without mutating or activating off-target



**Fig. 6** Yeast surface display platform for evolving PET-depolymerizing enzymes. A library of  $>10^7$  yeast clones is prepared in a single test tube, with each yeast cell displaying copies of a polyethylene terephthalate (PET)-depolymerizing enzyme variant on its surface. These cells are coated with a probe molecule resembling PET. For this tagging procedure, yeast cells are first non-specifically coated with an azide-linker construct using NHS labeling. Then, copper click chemistry is used to attach a probe to the linker, which contains an aromatic ester and a terminal biotin. High-activity enzyme mutants cleave the probe more efficiently than low-activity mutants. To detect biotin, yeast cells are stained with a fluorescent streptavidin phycoerythrin conjugate (sAv-PE). Cells displaying low sAv-PE signals are isolated using FACS, based on high levels of probe cleavage and, consequently, enhanced enzyme activity. The sAv-PE signal is normalized to the expression level of the PET-degrading enzyme by immunostaining a myc epitope tag appended to the enzyme. Adapted from Cribari *et al.*<sup>150</sup>

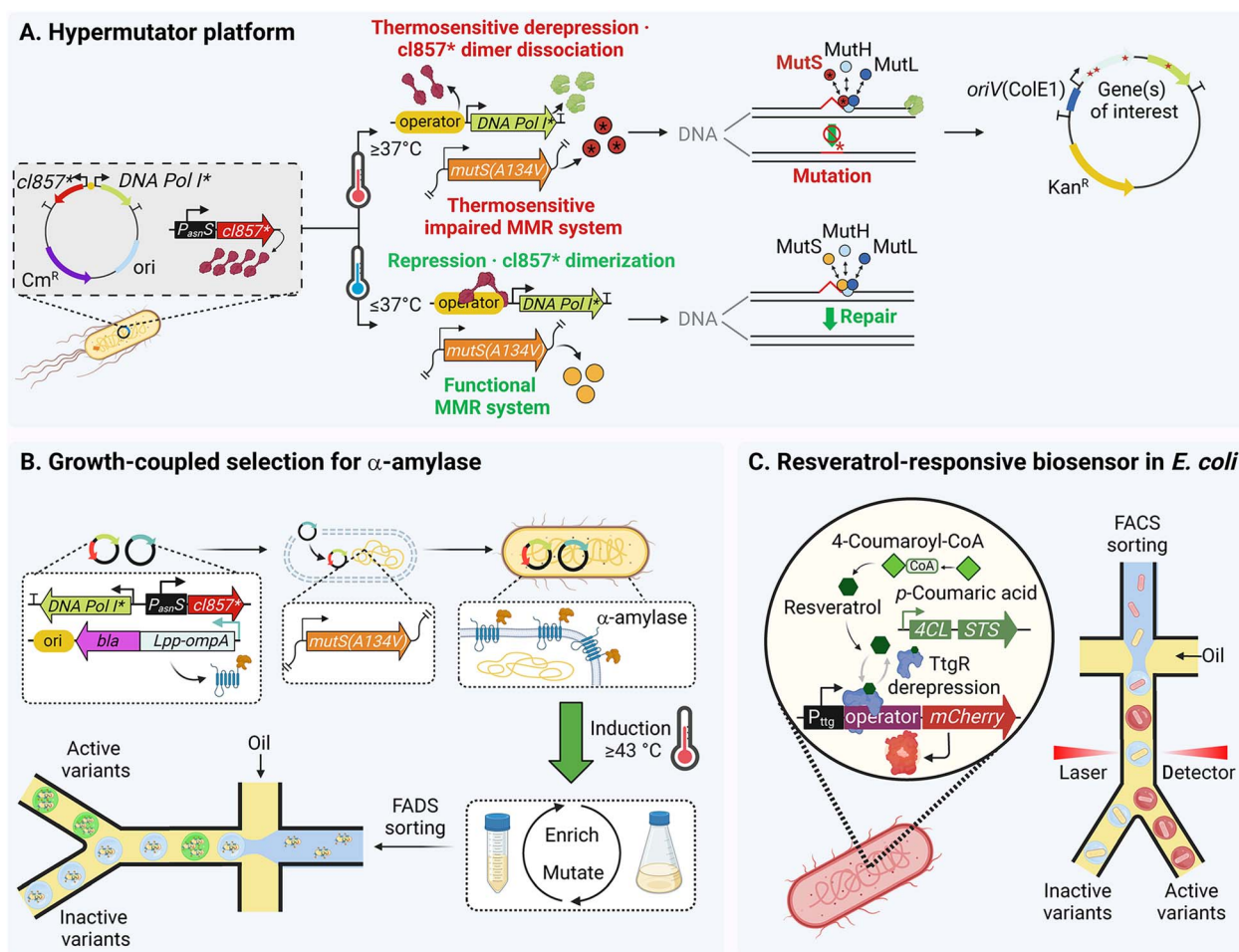


enzymes that might circumvent selection in the chosen strain. The integration of hypermutators with high-throughput capabilities of *in vivo* experimentation opens the door to a multitude of evolutionary pathways, hence enabling researchers to identify the true global maxima within the fitness landscape. Additionally, oscillating the selection pressures has emerged as an alternative technique for traversing different fitness landscapes, thus increasing the chances of reaching a global maximum for the phenotype of interest.<sup>157</sup>

In a recent study, Chen *et al.*<sup>158</sup> developed an innovative ultraHTS-assisted *in vivo* evolution system for enzyme

engineering using a hypermutator allele (Fig. 7A). The continuous *in vivo* evolution utilizes a thermo-inducible error-prone DNA polymerase and a temperature-sensitive *mutS* mutant. This setup facilitated the increased generation and fixation of mutations in an  $\alpha$ -amylase gene cloned in the *E. coli* hypermutator (Fig. 7B). This system had a 600-fold higher mutation rate than the wild-type strain.

Finally, cell-free expression (CFE), also known as *in vitro* protein synthesis, has been increasingly adopted over recent years for high-throughput production and characterization of proteins.<sup>159</sup> CFE refers to protein production without the use of



**Fig. 7** A hypermutator platform for temperature-controlled continuous *in vivo* evolution. (A) The system is based on a plasmid-borne error-prone DNA polymerase I (Pol I\*) gene under control of the temperature-sensitive  $\lambda P_R/P_L$ -*cI857\** expression system. Pol I initiates the replication of ColE1 plasmids and thereby the mutation rate is increased downstream of the ColE1 origin of replication (*ori*) at temperatures above 37 °C. To further enhance the mutation rate, the mismatch repair (MMR) system was rendered temperature-sensitive by incorporating the A134V mutation into *mutS*, which leads to almost wild-type levels of spontaneous mutants at 37 °C, but increases the mutation rate above 43 °C. During replication of the target plasmid at 43 °C, the synergistic action of Pol I\* and *MutS*<sup>A134V</sup> mediates a higher mutation frequency compared to lower temperatures. (B) *In vivo* evolution of  $\alpha$ -amylase (*bla*) for enhanced starch hydrolysis. *Bla* was fused to the carboxyl terminus of the outer membrane chimeric protein Lpp-OmpA, which directs the chimera to the cell surface. An evolved library of *Bla* variants was generated by passing the hypermutator strain from mutagenesis conditions (43 °C, rich medium) to selective conditions (37 °C, minimal medium supplemented with starch as sole carbon source). Single clones of the library were co-embedded with starch in droplets, sorted on the basis of high fluorescence. (C) Using the hypermutator platform for improved resveratrol production using a resveratrol-responsive biosensor. The TtgR repressor blocks *mCherry* transcription in the absence of resveratrol. Upon binding resveratrol, TtgR releases the operator site and *mCherry* is synthesized. A ColE1 plasmid harboring the genes for resveratrol biosynthesis, *i.e.*, stilbene synthase (STS) and *p*-coumarate:CoA ligase (4CL), was introduced into the hypermutator strain. A library of variants was created using iterative mutagenesis passages with consecutive FACS sorting for highly fluorescent cells. Adapted from Chen *et al.*<sup>158</sup>





living cells, circumventing the conventional protein production workflow. While traditional protocols require growing cells and expressing and purifying proteins (typically, consuming several days), CFE can produce proteins within a few hours. Due to the small size of reactions and volume of reagents used, the costs are also significantly decreased. By the same token, CFE increases the capabilities of parallel processing, and CFE can be easily integrated with various detection technologies mentioned in this review. This flexibility opens the possibility of seamless integration of steps and streamlines the pipeline.

## 5. Emergent tools for high-throughput screening/selection of tomorrow's enzymes

One of the main bottlenecks in establishing high-throughput systems is that the read-out needs to be converted into a trackable signal, either colorimetric or fluorometric, *via* product formation or substrate depletion, thus establishing an indirect link to monitor catalytic activity,<sup>160</sup> as shown in Fig. 2B. In this section, we explore methodologies and considerations of HTS and selection tools and related techniques, and how these methodologies are integrated into screening and selection workflows for enzyme discovery.

In colorimetric assays, enzymes in general produce a colored compound which can be detected through absorption. Spatial separation of library members and genotype-phenotype linkage is achieved using mostly microtiter plates. Employment of liquid handlers and other robotic platforms can increase the throughput to  $10^4$  variants per experiment.<sup>161</sup> Fluorometric detection is arguably the most widespread method for evaluating the activity of enzymes. Enzyme activity needs to be linked with the production of a fluorescent molecule to track the performance of the enzyme. There are several well-established fluorescent coupling strategies in the literature that constitute the basis for any HTS setup.<sup>162</sup> A basic and commonly used method to establish the linkage is through utilization of fluorogenic substrates.<sup>163</sup> These have been utilized in several publications to evaluate the activity of enzymes, *e.g.*, oxidoreductases and lyases, or in the identification of novel enzymes.<sup>119,163,164</sup> If no direct colorimetric and fluorometric assays are available, the product of the enzyme of interest can be processed further with enzymes that use the non-fluorescent metabolite of interest and convert it to a fluorescent product. As an alternative, biosensors can also be employed to speed up screening processes. The non-fluorescent metabolite formed as the result of the enzyme activity can be detected at high specificity and sensitivity through a transcription-based, RNA-based, or DNA-based biosensor. So far, applications with aptamers, transcription factors, and riboswitches have been successfully demonstrated. As a response to the metabolite binding, the biosensor can generate fluorescent products that correspond to the enzyme activity.<sup>162</sup> Although indirect measurements are not very sensitive and usually require confirmation after initial screening, they still provide valuable information and allow faster screening compared to direct selection methods. In the

absence of colorimetric and fluorometric assays, however, the samples need to be processed label-free using analytical methods, *e.g.*, GC-MS, LC-MS, and HPLC, which hinders the overall throughput.<sup>160</sup>

### 5.1 Conventional setups are accelerated by integrating liquid handling systems

Liquid manipulation is a practice at the core of research in life sciences. Utilization of liquid handling systems has expanded significantly in the recent years,<sup>165</sup> including tools for automated analysis of growth parameters.<sup>166</sup> Conventional workflows utilizing 96-, 384- and 1536-well plates have proven themselves useful for a broad range of applications due to ensuring physical separation and parallel processing. The library size that can be effectively screened using conventional workflows is limited to  $10^4$  samples.<sup>167</sup> However, due to the reliability of the workflow, it is very well suited to experiments with library size  $<10^4$ . A successful application was reported by Zhou *et al.*<sup>168</sup> These authors tuned the expression of genes involved in the (2S)-naringenin synthesis pathway by determining and overcoming the bottleneck step *via* HTS.<sup>168</sup> They utilized an iterative high-throughput balancing strategy to optimize the expression of each gene in the pathway to ensure the highest production. In their pipeline, promoters with different strength were randomly introduced upstream of the pathway genes, and the diversified pathways were cloned into *E. coli*. The resulting 4800 strain variants were screened for (2S)-naringenin and *p*-coumarate production by assessing fluorescence and absorbance based on scanning excitation/emission data acquired using standards of each compound. Iterative screening cycles isolated the best candidate strain, which reached 192 mg per L (2S)-naringenin while also producing 29 mg per L *p*-coumarate. Chalcone synthase was identified as the rate-limiting step and a target for balancing production fluxes.

Microtiter plate-based setups are still used in most of the label-free detection approaches. A recent example of high-throughput analyte detection is reported by Alborno *et al.*<sup>169</sup> In the article, the authors focused on improving surfactin yield in *Bacillus* and achieved 160% improvement by refining the media formulation. Surfactin is a promising biosurfactant due to its thermostability and has applications in wide range of industries. The pipeline includes extensive microtiter plate experiments where *Bacillus* cells were grown with combinations of media formulations. For yield measurements, they developed a fast method to perform metabolomics analysis of the samples using flow-injection mass-spectrometry. The method can analyze a sample in 3 minutes, which allows completion of a 96-well plate analysis in 5 hours. To assess the yield and design the next round of experiments, they used iterative cycles of active-learning algorithm directed media formulations for maximum production.

Reliable liquid handling systems are in high demand due to the need to work with reduced volumes and parallel sample processing for HTS. Downsizing samples and precise liquid manipulation were made possible by integrating liquid



handling devices with microtiter plate operations. In a recent example, Ahsan *et al.*<sup>170</sup> demonstrated that a combination of genome-mining, CFE, and liquid handlers could efficiently screen polyester polyurethane degrading enzymes. The authors used biofilm bioinformatics tools to identify organisms growing on the metal surface of aircraft and trucks. Next, in an example of use of genome mining tools, they used a colorogenic esterase probe, 4-nitrophenyl hexanoate (4-NPH), to detect the activity of putative enzymes at 405 nm. The putative enzymes were then produced in CFE reactions. Liquid handlers automated the mixing of CFE reactions, production of enzymes, and addition of 4-NPH. The absorbance measured at 405 revealed the top performing enzymes with the highest esterase activity.

Due to the versatility of liquid handling systems, they can be employed to process samples with any read-out type, whether colorimetric, fluorometric, or analytical. In another example, Beeman *et al.*<sup>171</sup> utilized liquid handling systems to prepare a pipeline for high-throughput integrated MALDI-MS analysis for c-MET kinase assay in a 1536-well assay setup. Although label-free detection methods are known to be time-consuming, with the advancement of hardware systems it has become possible to create an HT label-free detection method. A Rapiflex MALDI Pharma Pulse system was employed to attain a throughput of up to 1 million variants analyzed per week. The assay principle is based upon phosphorylation of the peptide substrate, SRCtide, which leads to an 80 Da shift in *m/z* spectrum. Upon phosphorylation, both SRC and *p*-SRC can be quantified using the MALDI-MS system to provide insights into the behavior and the activity of the enzymes tested. Due to working with such small volumes, it is important to highlight the speed of the liquid handling system and possible evaporation. Acoustic liquid handling systems are overall faster than pipettor-based systems. Due to evaporation at such small scales, the authors reported that utilization of acoustic liquid handler does not interfere with the reliability of their workflow.

## 5.2 Single cell-based assays using fluorescence activated cell sorting

FACS is a highly advanced, commercialized microfluidics technique that allows rapid evaluation and sorting at single cell level. The technology possesses by far the greatest processing speed,<sup>172</sup> at up to  $10^5$  cells per s, allowing the screening of a library size of  $\sim 10^9$  clones per experiment.<sup>167</sup> FACS has been utilized in numerous examples of screening extensive libraries of enzymes,<sup>173–177</sup> including hydrolases,<sup>178</sup> oxidoreductases,<sup>179</sup> transferases,<sup>180</sup> and isomerases.<sup>141,181</sup> The cellular membrane provides a natural compartmentalization of single cells, trapping the assay and the output, and linking genotype to phenotype.<sup>167,172</sup> Tan *et al.*<sup>182</sup> adopted the high-throughput power of FACS to improve the activity of a fucosyltransferase (FutA). They developed several substrate derivatives labeled with fluorescent tags for the fucosylation reaction, which enabled FutA variants to process these substrates and emit a fluorescent signal. The screening process began by diversifying *futA* through error-prone PCR to create a library of approximately 4 million colonies. The cells containing this library were then cultured in LB

medium, and gene expression was induced. After harvesting, the cells were treated with fucose for 30 minutes. Subsequently, the cells were sorted based on their fluorescence intensity with high stringency. From positive sorted cells, the *futA* gene was amplified to generate a library for subsequent rounds of enrichment.

## 5.3 Droplet-based technologies greatly enhance screening capabilities for natural products

Despite many successful applications and HTS performance, applications of commercialized FACS systems are mainly limited to intracellular reactions. FACS systems are thus unable to examine secreted enzymes and extracellular metabolites, which hampers versatile experiment design. Commercialized FACS on *in vitro* encapsulation (based on water–oil–water emulsions) have therefore been conducted to emulate and study extracellular space. The main problem is that water-in-oil droplets cannot be sorted with FACS systems because FACS systems require an aqueous carrier phase. Although water–oil–water emulsions bring more functionality, working with double emulsions requires another round of encapsulation and these emulsions are less durable compared to water-in-oil emulsions.<sup>183</sup> Water-in-oil droplet-based assays can compensate for some of the weaknesses of FACS-based assays. Firstly, such assays are suitable to study extracellular metabolites or proteins. Secondly, manipulations in the content of reactions are made possible through pico-injection and droplet merging technologies. Thirdly, cells can also be co-cultured in the droplet environment to study relationships between genetically engineered or natural strains.<sup>184–188</sup> Fourthly, FACS is a strictly fluorescent-coupled technique. With the technical advancements in microfluidics, droplet microfluidics have been applied to sorting setups based on colorimetric or label-free detection, which extends analytical capacity beyond the capabilities of FACS. Additionally, further reduction in volume decreases consumption of reagents and costs. In a standard workflow, individual cells each carrying DNA variants are encapsulated in monodispersed droplets. Cells are incubated for varying times to allow the key reaction to take place and produce the output metabolite, usually fluorescent molecules. Next, the droplets are sorted according to the output signal. Selected droplets can be used for downstream analysis or for another round of enrichment until the desired functionalities are met. Debon *et al.*<sup>189</sup> used a similar workflow to engineer catalytically superior amine oxidases that can accept an industrially relevant non-natural substrate. They used *E. coli* cells to express a gene library and encapsulated single cells with the target substrate, lysis reagent, and reporter cascade. In each droplet, cells are lysed, liberating amino oxidases and allowing a reaction to create a fluorescent readout. The droplets were sorted three times to ensure the highest performers were selected with high stringency. As the result of their screening, they reported a 960-fold increase in catalytic efficiency compared to the wild-type enzyme.

CFE systems promise an opportunity to further accelerate the screening process by eliminating plasmid assembly, transformation, and cell-culturing steps, and hence minimizing



reagent consumption.<sup>190,191</sup> In a clever design, Holstein *et al.*<sup>119</sup> used a CFE system to screen a protease library of  $10^{14}$  variants in droplets (Fig. 2D), which is not possible *in vivo* due to transformation efficiency limitations. They encapsulated their DNA variants with isothermal DNA polymerase and PCR reagents to achieve DNA variant enrichment in their entirely *in vitro* setup. This step accomplished the enrichment of DNA variants in the droplets of up to 30 000 copies, which ensures sufficient yields of protein in CFE. The CFE reagents were then pico-injected into droplets which were incubated for 4 hours for protein production to take place. After production, the fluorogenic casein substrate was pico-injected and the droplets were again incubated for hydrolytic cleavage, which emits a fluorescent signal. Droplets were then sorted using a fluorescence-assisted droplet sorting device, the higher performers were selected, and a few selected variants were subjected to confirmatory assays in microtiter plates to identify a 5-fold better performing variant.

Droplet microfluidics systems are more versatile than FACS-systems because they can integrate with other devices to allow different readout types, and manipulations can be performed on the content of droplets. So far, FADS has been integrated and modified to work with colorimetric assays (absorbance-activated droplet sorting; AADS) and label-free detection (mass activated droplet sorting; MADS). A sophisticated example of absorbance-based detection was reported in Gielen *et al.*<sup>120</sup> As outlined in Fig. 2E, the authors improved the yields of phenylalanine dehydrogenase (PheDH), which catalyzes the  $\text{NAD}^+$ -dependent deamination of amino acids. PheDH converts L-phenylalanine (L-Phe) to phenylpyruvate, which is coupled to the oxidation of WST-1 to the dye WST-1 formazan. For the screening setup, two optical fibers were placed across a microfluidic channel to measure any decrease in transmittance when a droplet passes through, and hence created a novel “absorbance-activated droplet sorting” (AADS). After validating the screening setup, the *E. coli* library harboring PheDH variants was encapsulated, and the cells were lysed inside the droplet. Enzyme variants were exposed to the substrate inside the droplets and, depending on the activity of the variant, catalyzed the formation of formazan dye. After a period of incubation, droplets with varying levels of formazan dye passed through the microfluidic channel, where absorbance of each droplet was measured. The authors successfully demonstrated enrichment of PheDH variants of up to 2800-fold in activity, at a speed of up to 300 droplets per second. Medcalf *et al.*<sup>192</sup> then utilized the same target and the same setup with modifications to show increased speed of sorting, achieving 1000 droplets per second. Although sorting is slower than similar setups with fluorescence read-out, this is still a major advancement over the throughput of widely used microtiter plates for colorimetric assays.

Another very promising setup, illustrated in Fig. 2F, has been published by Holland-Moritz *et al.*<sup>121</sup> These authors designed a mass spectrometry integrated droplet microfluidics-based setup that can gather information about the unlabeled content of droplets. In their setup, they integrated electrospray ionization mass spectrometry (ESI-MS) for analysis of droplet content and demonstrated the utility of their setup using

a transaminase that converts 1-(imidazo[2,1-*b*]thiazol-6-yl)propan-2-amine into 1-(imidazo[2,1-*b*]thiazol-6-yl)propan-2-one. Transaminases were expressed *in vitro* and encapsulated, and the authors tracked the decrease of the substrate concentration to identify enzymes displaying high activity. Since mass spectrometry analysis destroys the sample, Holland-Moritz *et al.*<sup>121</sup> split the droplet in two prior to analysis—one droplet for analysis and the other for retaining the information. While a major part of the sample was sent for MS analysis, a small part was retained in microfluidic channels until the analysis is complete and the decision is made regarding the fate of the droplet. Although the throughput of the system droplet per second as well as sufficient product needs to be present to match the detection limit of MS, the method is very valuable for gathering information about almost any reaction that takes place 100-fold faster than standard HPLC/MS measurements.<sup>192</sup> The system was improved later at the same laboratory and the authors demonstrated the utility of the strategy for sorting lysine overproducer strains at increased sorting speed.<sup>122</sup> In this work, two different *E. coli* strains were utilized—wild-type and a derivative carrying DapA<sup>E84T</sup>, which is an improved variant for lysine production. The strains were encapsulated in emulsions and cultured in droplets for growth and lysine production. Similarly to the workflow of Holland-Moritz *et al.*,<sup>121</sup> the droplets were split in two and positively-sorted droplets were cultured and sent for validation of mutant genotype.

## 6. Conclusion

The evolution and engineering of enzymes continues to be instrumental to advance biotechnology. Over the last decades, enzymes have been instrumental in addressing a key challenge: transformation of industries to minimize environmental footprints by using renewable resources and producing less waste. However, despite the progress made, significant challenges remain to be addressed, particularly in the build and test phases of the DBTL cycle, which are often labor-intensive and time-consuming. Moving forward, key developments can help overcome these challenges. Firstly, the development of more sophisticated, smarter designs and the integration of advanced prediction tools driven by ML will probably replace the need for extensive wet lab experimentation and enable the creation of smaller libraries with efficient testing pipelines. Furthermore, there is a need for the continued advancement and adoption of end-to-end automated pipelines and high-throughput assays, which have already shown promise in reducing workloads and accelerating enzyme engineering efforts. The future should see these systems becoming more robust, widely accessible and seamlessly integrated into daily laboratory routines, minimizing human intervention and standardizing procedures. Moreover, innovations in droplet microfluidics and CFE systems represent the next frontier in enzyme screening, and further refinement of these technologies will significantly enhance throughput and accuracy. As droplet microfluidics systems continue to be refined, they are poised to become key tools for processing large libraries of enzyme candidates with unprecedented speed and precision. These developments have





expanded the capabilities of enzyme engineering and will be vital for a future transition to a sustainable bioeconomy.

## 7. Author's contributions

E. U. B., E. C. Ø., D. C. V., and P. I. N.: conceptualization, visualization, writing – original draft; P. I. N.: resources, funding acquisition, supervision, project administration, writing – review & editing.

## 8. Conflicts of interest

None declared.

## 9. Acknowledgments

The authors would like to acknowledge the work by many researchers in the field of biocatalysis, directed evolution, and high-throughput methodologies who have made authoritative contributions to the field, the work of whom could not always be cited because of space reasons. The financial support from The Novo Nordisk Foundation (NNF10CC1016517 and NNF18CC0033664) and from the European Union's Horizon2020 Research and Innovation Program under grant agreement No. 814418 (*SinFonia*) to P. I. N. is gratefully acknowledged. The responsibility of this article lies with the authors. The NNF and the European Union are not responsible for any use that may be made of the information contained therein. The figures were created with <https://www.biorender.com/>.

## 10. References

- U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, *Nature*, 2012, **485**, 185–194.
- R. Buller, S. Lutz, R. J. Kazlauskas, R. Snajdrova, J. C. Moore and U. T. Bornscheuer, *Science*, 2023, **382**, eadh8615.
- Y. Zhou, S. Wu and U. T. Bornscheuer, *Chem. Commun.*, 2021, **57**, 10661–10674.
- B. Hauer, *ACS Catal.*, 2020, **10**, 8418–8427.
- K. Chen and F. H. Arnold, *Nat. Catal.*, 2020, **3**, 203–213.
- C. P. S. Badenhurst and U. T. Bornscheuer, *Trends Biochem. Sci.*, 2018, **43**, 180–198.
- S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius and U. T. Bornscheuer, *Angew Chem. Int. Ed. Engl.*, 2021, **60**, 88–119.
- D. C. Miller, S. V. Athavale and F. H. Arnold, *Nat. Synth.*, 2022, **1**, 18–23.
- S. C. Hammer, A. M. Knight and F. H. Arnold, *Curr. Opin. Green Sustainable Chem.*, 2017, **7**, 23–30.
- E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner and S. L. Flitsch, *Nat. Rev. Methods Primers*, 2021, **1**, 46.
- C. Paul, U. Hanefeld, F. Hollmann, G. Qu, B. Yuan and Z. Sun, *Mol. Catal.*, 2024, **555**, 113874.
- E. Orsi, L. Schada von Borzyskowski, S. Noack, P. I. Nikel and S. N. Lindner, *Nat. Commun.*, 2024, **15**, 3447.
- D. C. Volke, A. Cros, D. H. Welner and P. I. Nikel, *Chem Catal.*, 2022, **2**, 2400–2402.
- K. Pandey, M. Pandey, V. Kumar, U. Aggarwal and B. Singhal, *Syst. Microbiol. Biomanuf.*, 2024, **4**, 407–424.
- S. M. Ioannidou, C. Pateraki, D. Ladakis, H. Papapostolou, M. Tsakona, A. Vlysidis, I. K. Kookos and A. Koutinas, *Bioresour. Technol.*, 2020, **307**, 123093.
- A. P. Zeng, *Biotechnol. Adv.*, 2019, **37**, 508–518.
- F. Federici, E. Orsi and P. I. Nikel, *ChemCatChem*, 2023, **15**, e202300746.
- H. Stichnothe, *Adv. Biochem. Eng. Biotechnol.*, 2019, **166**, 519–539.
- M. Guo and W. Song, *New Biotechnol.*, 2019, **49**, 48–57.
- D. Wield, R. Hanlin, J. Mittra and J. Smith, *Sci. Public Pol.*, 2013, **40**, 17–24.
- J. Nielsen, C. B. Tillegreen and D. Petranovic, *Trends Biotechnol.*, 2022, **40**, 1160–1172.
- T. Dinmukhamed, Z. Huang, Y. Liu, X. Lv, J. Li, G. Du and L. Liu, *Syst. Microbiol. Biomanuf.*, 2021, **1**, 15–23.
- C. Zeymer and D. Hilvert, *Annu. Rev. Biochem.*, 2018, **87**, 131–157.
- J. D. Winkler and K. C. Kao, *Genomics*, 2014, **104**, 406–411.
- S. Mazurenko, Z. Prokop and J. Damborský, *ACS Catal.*, 2020, **10**, 1210–1223.
- R. Riaz, M. Ashraf, N. Hussain, Z. Baqar, M. Bilal and H. M. N. Iqbal, *Catal. Lett.*, 2023, **153**, 1587–1601.
- J. S. Cho, G. B. Kim, H. Eun, C. W. Moon and S. Y. Lee, *JACS Au*, 2022, **2**, 1781–1799.
- J. Nielsen and J. D. Keasling, *Cell*, 2016, **164**, 1185–1197.
- I. Victorino da Silva Amatto, N. Gonsales da Rosa-Garzon, F. Antônio de Oliveira Simões, F. Santiago, N. Pereira da Silva Leite, J. Raspante Martins and H. Cabral, *Biotechnol. Appl. Biochem.*, 2022, **69**, 389–409.
- D. T. Monterrey, I. Ayuso-Fernández, I. Oroz-Guinea and E. García-Junceda, *Biotechnol. Adv.*, 2022, **60**, 108016.
- O. Vavra, J. Damborský and D. Bednar, *Biotechnol. Adv.*, 2022, **60**, 108009.
- R. Vanella, G. Kovacevic, V. Doffini, J. Fernández de Santaella and M. A. Nash, *Chem. Commun.*, 2022, **58**, 2455–2467.
- M. Vasina, P. Vanacek, J. Damborský and Z. Prokop, *Methods Enzymol.*, 2020, **643**, 51–85.
- E. M. Lammens, P. I. Nikel and R. Lavigne, *Nat. Commun.*, 2020, **11**, 5294.
- A. Sánchez-Pascuala, V. de Lorenzo and P. I. Nikel, *ACS Synth. Biol.*, 2017, **6**, 793–805.
- M. H. Medema, T. de Rond and B. S. Moore, *Nat. Rev. Genet.*, 2021, **22**, 553–571.
- N. Ziemert, M. Alanjary and T. Weber, *Nat. Prod. Rep.*, 2016, **33**, 988–1005.
- M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339–W346.
- T. Weber, *Int. J. Med. Microbiol.*, 2014, **304**, 230–235.



- 40 M. H. Medema and M. A. Fischbach, *Nat. Chem. Biol.*, 2015, **11**, 639–648.
- 41 S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403–410.
- 42 T. Weber and H. U. Kim, *Synth. Syst. Biotechnol.*, 2016, **1**, 69–79.
- 43 R. D. King, P. H. Wise and A. Clare, *Bioinformatics*, 2004, **20**, 1110–1118.
- 44 J. Hon, S. Borko, J. Stourac, Z. Prokop, J. Zendulka, D. Bednar, T. Martinek and J. Damborský, *Nucleic Acids Res.*, 2020, **48**, W104–W109.
- 45 E. Coudert, S. Gehant, E. de Castro, M. Pozzato, D. Baratin, T. Neto, C. J. A. Sigrist, N. Redaschi and A. Bridge, *Bioinformatics*, 2023, **39**, btac793.
- 46 L. Jeske, S. Placzek, I. Schomburg, A. Chang and D. Schomburg, *Nucleic Acids Res.*, 2019, **47**, D542–D549.
- 47 W. Finnigan, L. J. Hepworth, S. L. Flitsch and N. J. Turner, *Nat. Catal.*, 2021, **4**, 98–104.
- 48 B. N. Hogg, C. Schnepel, J. D. Finnigan, S. J. Charnock, M. A. Hayes and N. J. Turner, *Angew Chem. Int. Ed. Engl.*, 2024, **63**, e202402316.
- 49 F. Berini, C. Casciello, G. L. Marcone and F. Marinelli, *FEMS Microbiol. Lett.*, 2017, **364**, fnx211.
- 50 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583–589.
- 51 D. Meijer, M. A. Benididir, C. W. Coley, Y. M. Mejri, M. Öztürk, J. J. J. van der Hooft, M. H. Medema and A. Skiredj, *Nat. Prod. Rep.*, 2024, DOI: [10.1039/d1034np00008k](https://doi.org/10.1039/d1034np00008k).
- 52 J. Wang, J. L. Watson and S. L. Lisanza, *Cold Spring Harbor Perspect. Biol.*, 2024, **16**, a041472.
- 53 E. N. Kissman, M. B. Sosa, D. C. Millar, E. J. Koleski, K. Thevasundaram and M. C. Y. Chang, *Nature*, 2024, **631**, 37–48.
- 54 J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Židek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis and J. M. Jumper, *Nature*, 2024, **630**, 493–500.
- 55 M. Gantz, S. V. Mathis, F. E. H. Nintzel, P. Lio and F. Hollfelder, *Faraday Discuss.*, 2024, **252**, 127–156.
- 56 M. Zaccolo, D. M. Williams, D. M. Brown and E. Gherardi, *J. Mol. Biol.*, 1996, **255**, 589–603.
- 57 Y. Wang, P. Xue, M. Cao, T. Yu, S. T. Lane and H. Zhao, *Chem. Rev.*, 2021, **121**, 12384–12444.
- 58 T. Vanhercke, C. Ampe, L. Tirry and P. Denolf, *Anal. Biochem.*, 2005, **339**, 9–14.
- 59 W. M. Coco, *Methods Mol. Biol.*, 2003, **231**, 111–127.
- 60 C. E. Lawson, J. M. Martí, T. Radivojević, S. V. R. Jonnalagadda, R. Gentz, N. J. Hillson, S. Peisert, J. Kim, B. A. Simmons, C. J. Petzold, S. W. Singer, A. Mukhopadhyay, D. Tanjore, J. G. Dunn and H. García-Martín, *Metab. Eng.*, 2021, **63**, 34–60.
- 61 A. Goñi-Moreno and P. I. Nikel, *Front. Bioeng. Biotechnol.*, 2019, **7**, 40.
- 62 D. Listov, C. A. Goverde, B. E. Correia and S. J. Fleishman, *Nat. Rev. Mol. Cell Biol.*, 2024, **25**, 639–653.
- 63 B. Markus, G. C. C. K. Andreas, K. Arkadij, L. Stefan, O. Gustav, S. Elina and S. Radka, *ACS Catal.*, 2023, **13**, 14454–14469.
- 64 L. Chen, Q. Li, K. F. A. Nasif, Y. Xie, B. Deng, S. Niu, S. Pouriyeh, Z. Dai, J. Chen and C. Y. Xie, *Int. J. Mol. Sci.*, 2024, **25**, 8426.
- 65 M. Fuxreiter, *Curr. Opin. Struct. Biol.*, 2024, **87**, 102834.
- 66 J. Zhou and M. Huang, *Chem. Soc. Rev.*, 2024, **53**, 8202–8239.
- 67 A. H. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. DeWitt, L. Carter, K. N. Houk and D. Baker, *Nature*, 2023, **614**, 774–780.
- 68 R. A. Hughes and A. D. Ellington, *Cold Spring Harbor Perspect. Biol.*, 2017, **9**, a023812.
- 69 N. Tenhaef, R. Stella, J. Frunzke and S. Noack, *ACS Synth. Biol.*, 2021, **10**, 589–599.
- 70 R. Chao, S. Mishra, T. Si and H. Zhao, *Metab. Eng.*, 2017, **42**, 98–108.
- 71 M. Conejero-Muriel, I. Rodríguez-Ruiz, S. Martínez-Rodríguez, A. Llobera and J. A. Gavira, *Lab Chip*, 2015, **15**, 4083–4089.
- 72 V. de Lorenzo, D. Pérez-Pantoja and P. I. Nikel, *J. Bacteriol.*, 2024, **206**, e0013624.
- 73 D. C. Volke, P. Calero and P. I. Nikel, *Trends Microbiol.*, 2020, **28**, 512–513.
- 74 P. I. Nikel and V. de Lorenzo, *Metab. Eng.*, 2018, **50**, 142–155.
- 75 A. Weimer, M. Kohlstedt, D. C. Volke, P. I. Nikel and C. Wittmann, *Appl. Microbiol. Biotechnol.*, 2020, **104**, 7745–7766.
- 76 N. T. Wirth and P. I. Nikel, *Chem Catal.*, 2021, **1**, 1234–1259.
- 77 L. Ortiz, M. Paván, L. McCarthy, J. Timmons and D. M. Densmore, *J. Visualized Exp.*, 2017, **130**, e54703.
- 78 R. Chao, J. Liang, I. Tasan, T. Si, L. Ju and H. Zhao, *ACS Synth. Biol.*, 2017, **6**, 678–685.
- 79 M. Storch, A. Casini, B. Mackrow, T. Fleming, H. Trewitt, T. Ellis and G. S. Baldwin, *ACS Synth. Biol.*, 2015, **4**, 781–787.
- 80 D. C. Volke, R. A. Martino, E. Kozaeva, A. M. Smania and P. I. Nikel, *Nat. Commun.*, 2022, **13**, 3026.
- 81 E. Kozaeva, Z. S. Nielsen, M. Nieto-Domínguez and P. I. Nikel, *Nucleic Acids Res.*, 2024, **52**, e19.



- 82 M. Storch, M. C. Haines and G. S. Baldwin, *Synth. Biol.*, 2020, **5**, ysaa010.
- 83 N. Gurdo, D. C. Volke, D. McCloskey and P. I. Nikel, *New Biotechnol.*, 2023, **74**, 1–15.
- 84 N. Gurdo, D. C. Volke and P. I. Nikel, *Trends Biotechnol.*, 2022, **40**, 1148–1159.
- 85 T. Ellis, T. Adie and G. S. Baldwin, *Integr. Biol.*, 2011, **3**, 109–118.
- 86 A. A. Nava, A. L. Fear, N. Lee, P. Mellinger, G. Lan, J. McCauley, S. Tan, N. Kaplan, G. Goyal, R. C. Coates, J. Roberts, Z. Johnson, R. Hu, B. Wu, J. Ahn, W. E. Kim, Y. Wan, K. Yin, N. Hillson, R. W. Haushalter and J. D. Keasling, *ACS Synth. Biol.*, 2023, **12**, 3506–3513.
- 87 J. Chen, D. Densmore, T. S. Ham, J. D. Keasling and N. J. Hillson, *J. Biol. Eng.*, 2012, **6**, 1.
- 88 N. J. Hillson, R. D. Rosengarten and J. D. Keasling, *ACS Synth. Biol.*, 2012, **1**, 14–21.
- 89 B. Enghiad, P. Xue, N. Singh, A. G. Boob, C. Shi, V. A. Petrov, R. Liu, S. S. Peri, S. T. Lane, E. D. Gaither and H. Zhao, *Nat. Commun.*, 2022, **13**, 2697.
- 90 B. Enghiad and H. Zhao, *ACS Synth. Biol.*, 2017, **6**, 752–757.
- 91 Y. Ding, P. D. Howes and A. J. de Mello, *Anal. Chem.*, 2020, **92**, 132–149.
- 92 P. C. Gach, K. Iwai, P. W. Kim, N. J. Hillson and A. K. Singh, *Lab Chip*, 2017, **17**, 3388–3400.
- 93 S. C. Shih, G. Goyal, P. W. Kim, N. Koutsoubelis, J. D. Keasling, P. D. Adams, N. J. Hillson and A. K. Singh, *ACS Synth. Biol.*, 2015, **4**, 1151–1164.
- 94 G. Linshiz, E. Jensen, N. Stawski, C. Bi, N. Elsbree, H. Jiao, J. Kim, R. Mathies, J. D. Keasling and N. J. Hillson, *J. Biol. Eng.*, 2016, **10**, 3.
- 95 I. Mougiakos, E. F. Bosma, J. Ganguly, J. van der Oost and R. van Kranenburg, *Curr. Opin. Biotechnol.*, 2018, **50**, 146–157.
- 96 Y. Vervoort, A. G. Linares, M. Roncoroni, C. Liu, J. Steensels and K. J. Verstrepen, *Curr. Opin. Biotechnol.*, 2017, **46**, 120–125.
- 97 E. Orsi, N. J. Claassens, P. I. Nikel and S. N. Lindner, *Biotechnol. Adv.*, 2022, **60**, 108035.
- 98 N. L. Bitzenhofer, L. Kruse, S. Thies, B. Wynands, T. Lechtenberg, J. Rönitz, E. Kozaeva, N. T. Wirth, C. Eberlein, K. E. Jaeger, P. I. Nikel, H. J. Heipieper, N. Wierckx and A. Loeschcke, *Essays Biochem.*, 2021, **65**, 319–336.
- 99 V. Singh and D. Braddick, *Syst. Synth. Biol.*, 2015, **9**, 1–9.
- 100 I. Pardo, D. Bednar, P. Calero, D. C. Volke, J. Damborský and P. I. Nikel, *ACS Catal.*, 2022, **12**, 6570–6577.
- 101 T. Kittilä, P. Calero, F. Fredslund, P. T. Lowe, D. Tezé, M. Nieto-Domínguez, D. O'Hagan, P. I. Nikel and D. H. Welner, *Microb. Biotechnol.*, 2022, **15**, 1622–1632.
- 102 P. Calero, D. C. Volke, P. T. Lowe, C. H. Gottfredsen, D. O'Hagan and P. I. Nikel, *Nat. Commun.*, 2020, **11**, 5045.
- 103 L. Martinelli and P. I. Nikel, *Microb. Biotechnol.*, 2019, **12**, 187–190.
- 104 J. van der Oost and C. Patinios, *Trends Biotechnol.*, 2023, **41**, 396–409.
- 105 T. M. Wannier, P. N. Ciaccia, A. D. Ellington, G. T. Filsinger, F. J. Isaacs, K. Javanmardi, M. A. Jones, A. M. Kunjapur, Á. Nyerges, C. Pal, M. G. Schubert and G. M. Church, *Nat. Rev. Methods Primers*, 2021, **1**, 7.
- 106 L. Fernández-Cabezón, A. Cros and P. I. Nikel, *ACS Synth. Biol.*, 2021, **10**, 1214–1226.
- 107 F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson and G. M. Church, *Science*, 2011, **333**, 348–353.
- 108 P. I. Nikel, *Trends Biotechnol.*, 2019, **37**, 1036–1038.
- 109 M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church and F. J. Isaacs, *Science*, 2013, **342**, 357–360.
- 110 H. H. Wang, P. Y. Huang, G. Xu, W. Haas, A. Marblestone, J. Li, S. P. Gygi, A. C. Forster, M. C. Jewett and G. M. Church, *ACS Synth. Biol.*, 2012, **1**, 43–52.
- 111 H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest and G. M. Church, *Nature*, 2009, **460**, 894–898.
- 112 S. Y. Kang, K. T. Heo and Y. S. Hong, *ACS Synth. Biol.*, 2018, **7**, 2054–2062.
- 113 E. M. Barbieri, P. Muir, B. O. Akhuetie-Oni, C. M. Yellman and F. J. Isaacs, *Cell*, 2017, **171**, 1453–1467.e1413.
- 114 Z. Shi and G. Wang, *Eng. Biol. Microb. Biosynth.*, 2024, vol. 15, pp. 315–343.
- 115 E. Orsi, N. J. Claassens, P. I. Nikel and S. N. Lindner, *Nat. Commun.*, 2021, **12**, 5295.
- 116 P. Calero and P. I. Nikel, *Microb. Biotechnol.*, 2019, **12**, 98–124.
- 117 J. Foldi, J. A. Connolly, E. Takano and R. Breitling, *ACS Synth. Biol.*, 2024, **13**, 2684–2692.
- 118 A. S. Karim, D. M. Brown, C. M. Archuleta, S. Grannan, L. Aristilde, Y. Goyal, J. N. Leonard, N. M. Mangan, A. Prindle, G. J. Rocklin, K. J. Tyo, L. Zoloth, M. C. Jewett, S. Calkins, N. P. Kamat, D. Tullman-Ercek and J. B. Lucks, *Nat. Commun.*, 2024, **15**, 5425.
- 119 J. M. Holstein, C. Gylstorff and F. Hoffelder, *ACS Synth. Biol.*, 2021, **10**, 252–257.
- 120 F. Gielen, R. Hours, S. Emond, M. Fischlechner, U. Schell and F. Hoffelder, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E7383–E7389.
- 121 D. A. Holland-Moritz, M. K. Wismer, B. F. Mann, I. Farasat, P. Devine, E. D. Guetschow, I. Mangion, C. J. Welch, J. C. Moore, S. Sun and R. T. Kennedy, *Angew. Chem. Int. Ed. Engl.*, 2020, **59**, 4470–4477.
- 122 E. M. Payne, B. E. Murray, L. I. Penabad, E. Abbate and R. T. Kennedy, *Anal. Chem.*, 2023, **95**, 15716–15724.
- 123 L. Fernández-Cabezón, A. Cros and P. I. Nikel, *Biotechnol. J.*, 2019, **14**, 1800439.
- 124 M. S. Morrison, C. J. Podracky and D. R. Liu, *Nat. Chem. Biol.*, 2020, **16**, 610–619.
- 125 A. Zimmermann, J. E. Prieto-Vivas, K. Voordeckers, C. Bi and K. J. Verstrepen, *Trends Microbiol.*, 2024, **32**, 884–901.



- 126 D. C. Volke, E. Orsi and P. I. Nikel, *Curr. Opin. Microbiol.*, 2023, **75**, 102353.
- 127 N. T. Wirth, E. Kozaeva and P. I. Nikel, *Microb. Biotechnol.*, 2020, **13**, 233–249.
- 128 R. Silva-Rocha, E. Martínez-García, B. Calles, M. Chavarria, A. Arce-Rodríguez, A. de las Heras, A. D. Páez-Espino, G. Durante-Rodríguez, J. Kim, P. I. Nikel, R. Platero and V. de Lorenzo, *Nucleic Acids Res.*, 2013, **41**, D666–D675.
- 129 J. Turlin, B. Dronsella, A. De Maria, S. N. Lindner and P. I. Nikel, *Metab. Eng.*, 2022, **74**, 191–205.
- 130 D. C. Volke, J. Turlin, V. Mol and P. I. Nikel, *Microb. Biotechnol.*, 2020, **13**, 222–232.
- 131 N. Krink, P. I. Nikel and C. L. Beisel, *EMBO Rep.*, 2024, **25**, 1694–1699.
- 132 J. M. Hernández-Sancho, A. Boudigou, M. V. G. Alván-Vargas, D. Freund, J. A. Bååth, P. Westh, K. Jensen, L. N. García, D. C. Volke and P. I. Nikel, *Nat. Commun.*, 2024, In press.
- 133 J. A. Kaczmarek and K. L. J. Prather, *J. Ind. Microbiol. Biotechnol.*, 2021, **48**, kuab049.
- 134 D. Josephides, S. Davoli, W. Whitley, R. Ruis, R. Salter, S. Gokkaya, M. Vallet, D. Matthews, G. Benazzi, E. Shvets, F. Gesellchen, D. Geere, X. Liu, X. Li, B. Mackworth, W. Young, Z. Owen, C. Smith, D. Starkie, J. White, B. Sweeney, M. Hinchliffe, S. Tickle, D. J. Lightwood, M. Rehak, F. F. Craig and D. Holmes, *SLAS Technol.*, 2020, **25**, 177–189.
- 135 M. Kortmann, C. Mack, M. Baumgart and M. Bott, *ACS Synth. Biol.*, 2019, **8**, 274–281.
- 136 A. Jain, S. Stavarakis and A. de Mello, *Curr. Opin. Biotechnol.*, 2024, **87**, 103097.
- 137 L. Calzadiaz-Ramírez, C. Calvó-Tusell, G. M. M. Stoffel, S. N. Lindner, S. Osuna, T. J. Erb, M. Garcia-Borràs, A. Bar-Even and C. G. Acevedo-Rocha, *ACS Catal.*, 2020, **10**, 7512–7525.
- 138 A. von Kamp and S. Klamt, *Nat. Commun.*, 2017, **8**, 15956.
- 139 J. Buerger, L. S. Gronenberg, H. J. Genée and M. O. A. Sommer, *Curr. Opin. Biotechnol.*, 2019, **59**, 85–92.
- 140 A. Cros, G. Alfaro-Espinoza, A. de Maria, N. T. Wirth and P. I. Nikel, *Curr. Opin. Biotechnol.*, 2022, **74**, 180–193.
- 141 C. Femmer, M. Bechtold, M. Held and S. Panke, *Metab. Eng.*, 2020, **59**, 15–23.
- 142 S. Wenk, O. Yishai, S. N. Lindner and A. Bar-Even, *Methods Enzymol.*, 2018, **608**, 329–367.
- 143 B. L. Wang, A. Ghaderi, H. Zhou, J. Agresti, D. A. Weitz, G. R. Fink and G. Stephanopoulos, *Nat. Biotechnol.*, 2014, **32**, 473–478.
- 144 J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J. C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths and D. A. Weitz, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 4004–4009.
- 145 R. Ostafe, R. Prodanovic, W. Lloyd Ung, D. A. Weitz and R. Fischer, *Biomicrofluidics*, 2014, **8**, 041102.
- 146 T. Beneyton, F. Coldren, J. C. Baret, A. D. Griffiths and V. Taly, *Analyst*, 2014, **139**, 3314–3323.
- 147 H. G. Hwang, D. Y. Ye and G. Y. Jung, *Biotechnol. Adv.*, 2023, **69**, 108251.
- 148 J. S. Andon, B. Lee and T. Wang, *Org. Biomol. Chem.*, 2022, **20**, 5891–5906.
- 149 S. Li, Y. Zhang, L. Li, Y. Yuan, H. Sun, X. H. Xing, X. Wang and C. Zhang, *Biochem. Eng. J.*, 2023, **198**, 109036.
- 150 M. A. Cribari, M. J. Unger, I. C. Unarta, A. N. Ogorek, X. Huang and J. D. Martell, *J. Am. Chem. Soc.*, 2023, **145**, 27380–27389.
- 151 S. Sulaiman, S. Yamato, E. Kanaya, J. J. Kim, Y. Koga, K. Takano and S. Kanaya, *Appl. Environ. Microbiol.*, 2012, **78**, 1556–1562.
- 152 M. Imhof and C. Schlotterer, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 1113–1117.
- 153 R. S. Molina, G. Rix, A. A. Mengiste, B. Álvarez, D. Seo, H. Chen, J. E. Hurtado, Q. Zhang, J. D. García-García, Z. J. Heins, P. J. Almhjell, F. H. Arnold, A. S. Khalil, A. D. Hanson, J. E. Dueber, D. V. Schaffer, F. Chen, S. Kim, L. Á. Fernández, M. D. Shoulders and C. C. Liu, *Nat. Rev. Methods Primers*, 2022, **2**, 36.
- 154 G. Rix and C. C. Liu, *Curr. Opin. Chem. Biol.*, 2021, **64**, 20–26.
- 155 T. J. Overbeck, D. L. Welker, J. E. Hughes, J. L. Steele and J. R. Broadbent, *Appl. Environ. Microbiol.*, 2017, **83**, 011200.
- 156 T. W. Wang, H. Zhu, X. Y. Ma, T. Zhang, Y. S. Ma and D. Z. Wei, *Mol. Biotechnol.*, 2006, **34**, 55–68.
- 157 A. C. Carpenter, A. M. Feist, F. S. M. Harrison, I. T. Paulsen and T. C. Williams, *Metab. Eng. Commun.*, 2023, **17**, e00227.
- 158 S. Chen, Z. Yang, Z. Zhong, S. Yu, J. Zhou, J. Li, G. Du and G. Zhang, *Biotechnol. Biofuels Bioprod.*, 2024, **17**, 9.
- 159 Y. C. Kwon and M. C. Jewett, *Sci. Rep.*, 2015, **5**, 8663.
- 160 Y. V. Sheludko and W. D. Fessner, *Curr. Opin. Struct. Biol.*, 2020, **63**, 123–133.
- 161 L. M. Mayr and D. Bojanic, *Curr. Opin. Pharmacol.*, 2009, **9**, 580–588.
- 162 J. Jiang, G. Yang and F. Ma, *Biotechnol. Adv.*, 2023, **66**, 108173.
- 163 J. D. Schnettler, O. J. Klein, T. S. Kaminski, P. Y. Colin and F. Hollfelder, *J. Am. Chem. Soc.*, 2023, **145**, 1083–1096.
- 164 H. Yuan, R. Tu, X. Tong, Y. Lin, Y. Zhang and Q. Wang, *J. Ind. Microbiol. Biotechnol.*, 2022, **49**, kuac007.
- 165 T. C. Silva, M. Eppink and M. Ottens, *J. Chem. Technol. Biotechnol.*, 2022, **97**, 2365–2375.
- 166 N. T. Wirth, J. Funk, S. Donati and P. I. Nikel, *Nat. Protoc.*, 2023, **18**, 2401–2403.
- 167 K. K. Kwon, J. Lee, H. Kim, D. H. Lee and S. G. Lee, *Curr. Opin. Syst. Biol.*, 2024, **37**, 100487.
- 168 S. Zhou, Y. Lyu, H. Li, M. A. G. Koffas and J. Zhou, *Biotechnol. Bioeng.*, 2019, **116**, 1392–1404.
- 169 R. V. Alborno, D. Oyarzún and K. Burgess, *Comput. Struct. Biotechnol. J.*, 2024, **23**, 1226–1233.
- 170 A. Ahsan, D. Wagner, V. A. Varaljay, V. Roman, N. Kelley-Loughnane and N. F. Reuel, *Synth. Biol.*, 2024, **9**, ysae005.
- 171 K. Beeman, J. Baumgärtner, M. Laubenheimer, K. Hergesell, M. Hoffmann, U. Pehl, F. Fischer and J. C. Pieck, *SLAS Discovery*, 2017, **22**, 1203–1210.
- 172 T. Kim, M. Ko, E. Rha, H. Kim and H. Lee, *Biotechnol. Bioprocess Eng.*, 2024, **29**, 25–33.





- 173 H. Xiao, Z. Bao and H. Zhao, *Ind. Eng. Chem. Res.*, 2015, **54**, 4011–4020.
- 174 J. K. Michener and C. D. Smolke, *Metab. Eng.*, 2012, **14**, 306–316.
- 175 S. Becker, H. U. Schmoldt, T. M. Adams, S. Wilhelm and H. Kolmar, *Curr. Opin. Biotechnol.*, 2004, **15**, 323–329.
- 176 Y. S. Kim, H. C. Jung and J. G. Pan, *Appl. Environ. Microbiol.*, 2000, **66**, 788–793.
- 177 L. Yi, M. C. Gebhard, Q. Li, J. M. Taft, G. Georgiou and B. L. Iverson, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7229–7234.
- 178 E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly and A. D. Griffiths, *Chem. Biol.*, 2005, **12**, 1291–1300.
- 179 C. Pitzler, G. Wirtz, L. Vojcic, S. Hiltl, A. Böker, R. Martinez and U. Schwaneberg, *Chem. Biol.*, 2014, **21**, 1733–1742.
- 180 R. A. Scheele, L. H. Lindenburg, M. Petek, M. Schober, K. N. Dalby and F. Hollfelder, *Nat. Commun.*, 2022, **13**, 844.
- 181 C. Femmer, M. Bechtold and S. Panke, *Biotechnol. Bioeng.*, 2020, **117**, 2683–2693.
- 182 Y. Tan, Y. Zhang, Y. Han, H. Liu, H. Chen, F. Ma, S. G. Withers, Y. Feng and G. Yang, *Sci. Adv.*, 2019, **5**, eaaw8451.
- 183 M. Gantz, S. Neun, E. J. Medcalf, L. D. van Vliet and F. Hollfelder, *Chem. Rev.*, 2023, **123**, 5571–5611.
- 184 H. Lee, J. I. Baek, S. J. Kim, K. K. Kwon, E. Rha, S. J. Yeom, H. Kim, D. H. Lee, D. M. Kim and S. G. Lee, *Front. Bioeng. Biotechnol.*, 2020, **8**, 358.
- 185 G. Sunantha and N. Vasudevan, *Sci. Total Environ.*, 2021, **759**, 143544.
- 186 S. Kim, S. H. Jin, H. G. Lim, B. Lee, J. Kim, J. Yang, S. W. Seo, C. S. Lee and G. Y. Jung, *Lab Chip*, 2021, **21**, 4455–4463.
- 187 Y. Wang, S. Li, N. Xue, L. Wang, X. Zhang, L. Zhao, Y. Guo, Y. Zhang and M. Wang, *ACS Synth. Biol.*, 2023, **12**, 1761–1771.
- 188 E. Hua, Y. Zhang, K. Yun, W. Pan, Y. Liu, S. Li, Y. Wang, R. Tu and M. Wang, *ACS Synth. Biol.*, 2022, **11**, 2697–2708.
- 189 A. Debon, M. Pott, R. Obexer, A. P. Green, L. Friedrich, A. D. Griffiths and D. Hilvert, *Nat. Catal.*, 2019, **2**, 740–747.
- 190 L. Jiang, J. Zhao, J. Lian and Z. Xu, *Synth. Syst. Biotechnol.*, 2018, **3**, 90–96.
- 191 B. J. Rasor, B. Vögeli, G. M. Landwehr, J. W. Bogart, A. S. Karim and M. C. Jewett, *Curr. Opin. Biotechnol.*, 2021, **69**, 136–144.
- 192 E. J. Medcalf, M. Gantz, T. S. Kaminski and F. Hollfelder, *Anal. Chem.*, 2023, **95**, 4597–4604.

