

Automated in vivo enzyme engineering accelerates biocatalyst optimization

Received: 21 December 2023

Accepted: 4 March 2024

Published online: 24 April 2024

Enrico Orsi¹, Lennart Schada von Borzyskowski², Stephan Noack³,
Pablo I. Nikel¹ & Steffen N. Lindner^{4,5}✉

Achieving cost-competitive bio-based processes requires development of stable and selective biocatalysts. Their realization through in vitro enzyme characterization and engineering is mostly low throughput and labor-intensive. Therefore, strategies for increasing throughput while diminishing manual labor are gaining momentum, such as in vivo screening and evolution campaigns. Computational tools like machine learning further support enzyme engineering efforts by widening the explorable design space. Here, we propose an integrated solution to enzyme engineering challenges whereby ML-guided, automated workflows (including library generation, implementation of hypermutation systems, adapted laboratory evolution, and in vivo growth-coupled selection) could be realized to accelerate pipelines towards superior biocatalysts.

The development of tailored and efficient bio-based processes is essential for applications as diverse as biopharmaceutical production, industrial biotechnology, food technology, crop improvement, and bioremediation. To establish such profitable bio-based processes, biocatalysts that can perform substrate-to-product conversions with high volumetric productivities ($g_{\text{product}} L^{-1} h^{-1}$), yields ($g_{\text{product}} g_{\text{substrate}}^{-1}$), and selectivities (enantiomeric excess) are essential¹. To reach improvements in these performance indicators and optimize chemical conversions, enzyme engineering has been developed as one of the pillars of synthetic biology², realizing enzyme optimization and development from the single reaction step to entire metabolic pathways^{2,3}.

Current efforts in bioengineering aim at designing biological systems that provide enzymatic activities beyond what has been developed and optimized by nature^{2,4,5}. Implementing these innovations can further develop a bio-based economy^{6–8}. Hence, it is desirable to design novel, new-to-nature enzymatic activities as the key parts needed to assemble complete synthetic pathways^{2,9} or used in enzyme-driven catalysis applications directly in synthetic processes (e.g., in the striking case of in vitro conversion of CO_2/H_2 or methanol into starch¹⁰). However, creating and optimizing such new-to-nature

reactions is a challenging task, for which the use of rational protein design accompanied by in vitro enzyme activity measurements or adaptive laboratory evolution (ALE) might not be sufficient^{11,12}.

At this point, directed evolution comes in handy, as it allows to perform Darwinian evolution in a test tube by increasing mutation and recombination rates within a target gene^{13,14}. Two types of directed evolution approaches are possible and differ in the environment where the evolution takes place. In vitro-directed evolution occurs outside a living organism, whereas in vivo evolution takes place within living systems. Both strategies have pros- and cons- which have been discussed elsewhere^{15,16}. In recent years, in vivo-directed evolution approaches have emerged as promising tools to use in protein engineering campaigns^{11,16}. The use of these approaches combined with growth-coupled selection (meaning coupling the enzymatic activity of interest to microbial fitness) has been applied for different optimization strategies^{17,18}. At the same time, automated biofoundries are becoming pivotal in supporting high-throughput efforts for engineering biology^{19–22}. Hence, the use of these infrastructures for protein engineering is gaining momentum²³. Moreover, the use of artificial intelligence (AI) and machine learning (ML) is aiding important endeavors in the design of new biological systems, from protein- to

¹The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark. ²Institute of Biology Leiden, Leiden University, 2333 BE Leiden, The Netherlands. ³Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany. ⁴Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany. ⁵Department of Biochemistry, Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität, 10117 Berlin, Germany. ✉e-mail: steffen.lindner@charite.de

organism level^{24–27}. Therefore, we are witnessing a paradigm shift in our ability and capacity to engineer biological systems. A combination of these technologies might result in the establishment of self-driving labs and workflows, which potentially accelerate scientific discoveries and innovation while reducing human errors^{28–30}. In this article, we discuss how the integration of ML, in vivo continuous evolution, and the use of automated biofoundries will accelerate the generation of new and competitive biocatalysts capable of supporting the transition towards a circular bio-based economy. Definitions of the most important technical terms used within the text are described in the Box 1.

An integrated workflow for accelerating in vivo enzyme engineering

To enable the workflow proposed in this article, different fields of expertise need to be integrated (Fig. 1). In brief, ML is used as input for (i) predicting the modifications required for engineering the target enzyme(s) and (ii) supporting the design of auxotroph selection strains by suggesting target genes to delete. Then, these selection strains are created through gene deletions. Subsequently, in vivo hypermutators can be exploited to increase the mutation rate within the target gene. Following the principle of growth-coupled selection, high-throughput and continuous cultivation platforms can be used for enriching the microbial population with clones containing the evolved target enzyme(s). Finally, sequencing of the enriched clones can inform the success of the evolution campaign. This step can also benefit from the use of ML-guided computational tools. If necessary, this pipeline can be iterated through several rounds.

The abovementioned steps should be intended as stand-alone workflows, which can be integrated by mobile robot units (Fig. 1). In the following sections of the manuscript, we dive more in-depth into the different aspects of this pipeline, eventually suggesting their integration using state-of-the-art automated workflows. Finally, we discuss caveats and limitations of this concept.

Enzyme engineering generates improved biocatalysts

Mastering enzyme engineering is vital to enable the optimization of existing bioprocesses or the exploration of new ones. Here, we cannot give a comprehensive overview of this large and rapidly developing area of research, but rather present key approaches and concepts. For a more detailed summary, the reader is referred to the existing literature^{4,31–33}.

The targets for enzyme engineering are highly diverse. To broaden the substrate range, the active site must be opened and remodeled; to improve the substrate specificity or enantioselectivity, the active site must be altered to only accommodate one type of desired substrate or intermediate³³; to develop novel catalytic functions, general principles of catalysis and transition state stabilization have to be applied to modify a suitable scaffold enzyme. To increase the thermal stability of enzymes and allow the catalysis of industrial processes at high temperatures, enzymes must be modified by introducing additional hydrogen bonds and salt bridges, rigidizing flexible residues, creating a more compact core region, or decreasing surface area hydrophobicity^{34,35}. Computational tools and predictions can help in identifying relevant amino acid residues and regions of interest to be mutated^{36,37}. Prediction of relevant residues to be mutated can be supported by experimentally determined protein structures, or—in their absence—by the data-driven protein structure prediction tool AlphaFold³⁸, which has made high-quality protein structure models easily available for the global research community since 2021.

Previously established techniques of enzyme engineering, such as rational mutagenesis based on in silico predictions or structure analysis, semi-rational mutagenesis (i.e., the combination of site-directed mutagenesis with random mutagenesis or directed evolution³⁹), and directed evolution of a suitable parent enzyme^{13,40}, can provide the starting points for screenings of improved biocatalysts and iterative

BOX 1

Definitions

In vivo engineering: optimization of enzymes or pathways within a living organism. It allows for direct application of in vivo genetic tools to increase enzyme variance and test them within their natural cellular context. When enzymatic activities can be coupled to growth the technique can strongly enhance the testing phase of DBTL cycles and allow selection of improved variants.

DBTL cycle: iterative workflow in bioengineering that involves the design of e.g., enzyme variants, building the according enzyme library, testing functionality of the library, and learning from the results obtained as the foundation for the next iteration. Ideally, this cyclical process allows the development of desired functionalities that can be further optimized by using several cycles.

Selection strains: mutant strains obtained through gene deletions that display an auxotrophy towards one or more essential metabolites (i.e., a biomass precursor) or an impaired general metabolic task (e.g., redox cofactor regeneration).

Hypermutator: synthetic biology tools used to enhance the rate of mutations within a gene of interest. In the context of this manuscript, we refer to in vivo hypermutators to describe tools capable of increasing the mutation rates in the in vivo context of a living organism.

Biofoundry: infrastructure capable of performing molecular biology operations and phenotyping through an automated Design-Build-Test-Learn pipeline. It includes robotic liquid handling, high-throughput analytics systems, and software for the analysis of the large amount of data generated.

Machine learning: a subset of artificial intelligence that self-develops algorithms for, e.g., enzyme engineering by analyzing data from previous experiments for its training. This training enables it to improve performance on a specific task over time without programming, allowing the systems to learn from data and make informed predictions or decisions.

cycles of enzyme development (Fig. 2). However, efficiently enhancing catalytic properties of enzymes requires maneuvering through complex and rugged fitness landscapes, where the relationship between enzyme sequence (genotype) and functional characteristics (phenotype) is difficult to predict. This means that optimization trajectories frequently result in diminishing returns (i.e., additional mutations only result in minimal improvements of an enzyme) and undesired tradeoff effects (e.g., substrate specificity is improved, but enzyme turnover number is strongly decreased)^{41,42}. When the abovementioned methods no longer yield enhancements, it often remains uncertain if an enzyme has already reached its maximum catalytic efficiency, or if there are other possible combinations of mutations that could generate further improvements⁴³. Therefore, the construction of large combinatorial enzyme libraries is a key approach in enzyme engineering (Fig. 2). By introducing diversity into enzyme sequences, libraries can be screened or selected for desired properties. High-throughput screening methods, including droplet-based microfluidics and fluorescence-activated cell sorting (FACS), enable the isolation and identification of enzymes with improved features, as long as a suitable readout is available (Fig. 3).

In addition to the previously discussed methods, ML approaches are now being more commonly employed to identify data patterns that aid in forecasting protein structures, enhancing enzyme stability, solubility, and function, predicting substrate specificity, and facilitating rational de novo protein design^{27,44,45}.

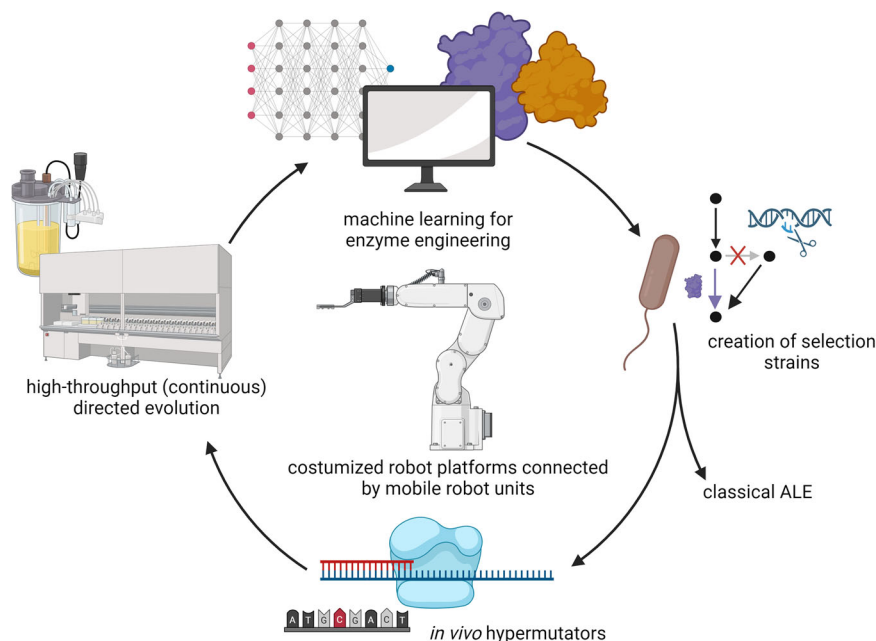


Fig. 1 | Conceptual overview of the pipeline proposed in this article. This workflow aims at generating superior catalysts by combining the use of machine learning, growth-coupled selection, in vivo hypermutators, and high-throughput

cultivations. Each part can be envisioned as a stand-alone module, which can, in principle, be connected using mobile robot units. Created with BioRender.com.

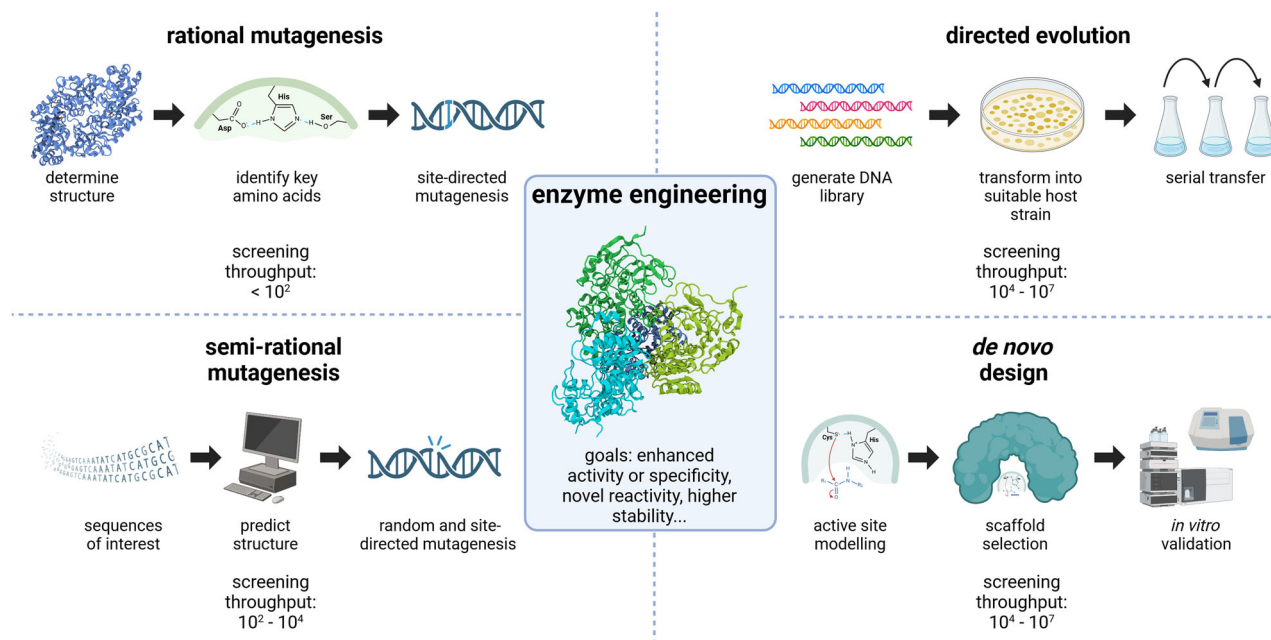


Fig. 2 | Enzyme engineering is applied to improve the properties of biocatalysts in a desired way. This includes increasing enzyme activity, substrate specificity, or enantioselectivity, introducing novel reactivities, or improving protein stability, among other goals. Methods that are applied for this purpose range from rational

mutagenesis of key amino acids to semi-rational approaches and directed evolution of gene sequences from DNA libraries. De novo enzyme design can be applied to generate biocatalysts that are free from the constraints of existing enzymes. Created with BioRender.com.

De novo enzyme design complements enzyme engineering

The aim of de novo enzyme design is the development of novel enzymes with desired functions from scratch, without relying on naturally occurring enzymes as starting points. Various protein design algorithms, such as Rosetta^{46,47} (<https://www.rosettacommons.org/>), have been developed to predict and optimize enzyme sequences based on desired functions. Rosetta relies on the mechanistic modelling of proteins using energy fields to guide the design process and

explore the vast sequence space for enzyme engineering³¹. Already 15 years ago, the first high-profile designer enzymes, e.g., for the catalysis of retro-aldol reactions, were reported⁴⁸. Some current highlights in this rapidly growing field include the de novo design of an eight stranded β -barrel protein that functions as a retro-aldolase, whose activity and stereoselectivity were further improved using directed evolution⁴⁹, and the creation of artificial luciferase enzymes from scratch, whose catalytic efficiency is comparable to that of natural

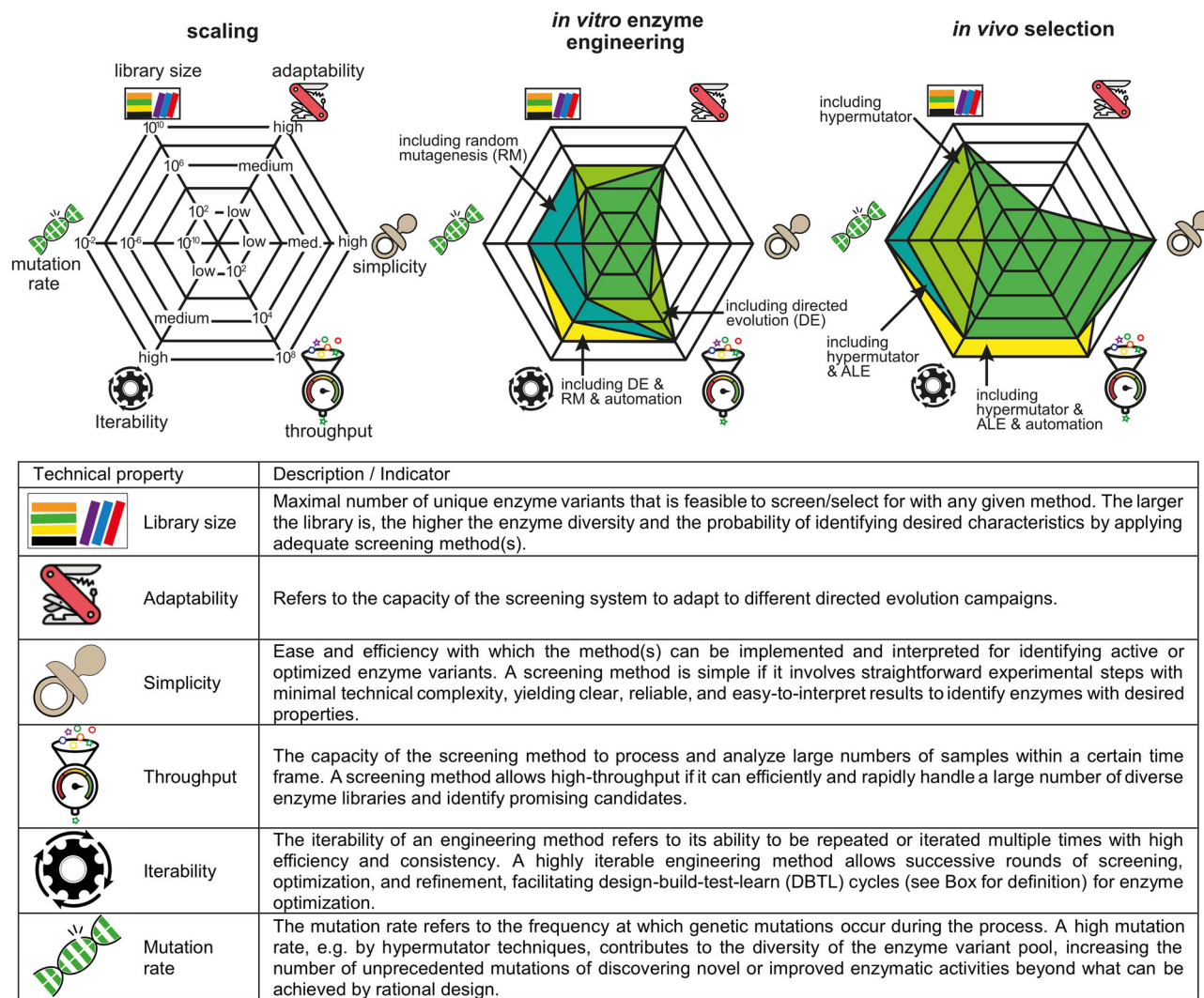


Fig. 3 | Comparison of *in vitro* and *in vivo* approaches for enzyme optimization.

The figure shows and scores the properties of the two methods for several key performance indicators (KPIs). Moreover, the use of additional technical

improvements to the approaches is included, such as random mutagenesis (RM), directed evolution (DE), hypermutators and adaptive laboratory evolution (ALE). The impact of these improvements on the KPIs is also depicted in the figure.

luciferases, while having a much higher substrate specificity and very high thermostability⁵⁰. The *de novo* design of enzymes that bind complex cofactors, such as heme irons, is also not an obstacle anymore: recently, the creation of a heme enzyme with a tunable substrate-binding pocket and its further engineering into an efficient carbene transferase was reported⁵¹. The combination of protein design with iterative mutagenesis for efficient enzyme engineering should not be underestimated. A good example is the conversion of a designed enzyme with modest activity for carbon-carbon bond formation between aldehydes and enones⁵² into a highly efficient biocatalyst via fourteen rounds of both local and global mutagenesis, coupled to high-throughput spectrophotometric assays as time-efficient readout⁵³.

It seems likely that *de novo* protein design will be feasible and widely used for all types of enzymes in the near future. The fine-tuning of the deep learning neural network RoseTTAFold⁵⁴ on protein structure denoising tasks resulted in RFdiffusion, a generative model of protein backbones with outstanding performance on protein monomer design, enzyme active site scaffolding (Fig. 2), and metal-binding protein design, which only requires simple molecular specifications as input⁵⁵.

Computational enzyme design is especially valuable to realize a novel metabolic pathway in which a natural enzyme for one reaction

step is lacking⁵⁶. Here, *de novo* design can supply enzyme candidates that catalyze the desired conversion, often only with initially low activities. These enzymes can subsequently be improved by mutagenesis and directed evolution, making it possible to implement efficient new-to-nature bioconversion routes.

ML-supported pathway design increases the engineering design space

The design and implementation of non-natural metabolic pathways is a complex and highly time-consuming task. ML can alleviate this challenge by automating several stages of the pathway design process²⁴. Specifically, ML algorithms can efficiently predict and analyze metabolic reactions, aiding in retrosynthesis approaches (i.e., the identification of potential pathways for the production of specific, desired compounds^{57–59}). Furthermore, ML is a powerful tool that can efficiently detect patterns in large sets of data. It has been extensively employed for analyzing datasets obtained through high-throughput technologies in order to create data-based models for intricate bio-processes. The integration of ML with the Design-Build-Test-Learn cycle commonly applied in synthetic biology can accelerate the development process²⁶. It can also assist in optimizing the metabolic engineering process, by intelligently exploring and designing different

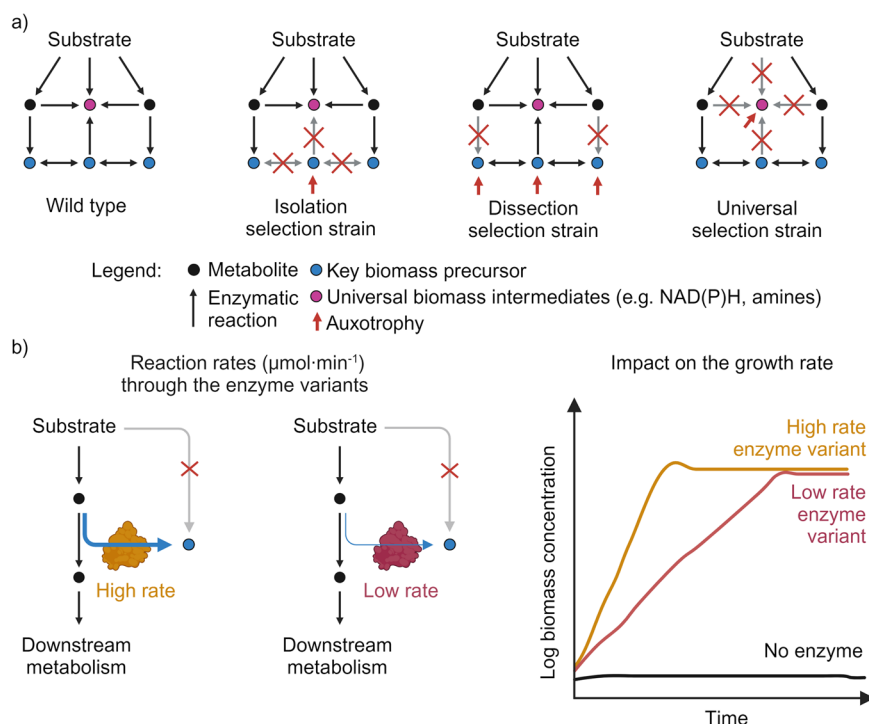


Fig. 4 | Adopting dedicated selection strains supports efficient enzyme screening. **a** Different types of auxotrophic selection strains can be used for in vivo enzyme evolution. The first group includes isolation and dissection strains; both strains are incapable of synthesizing essential biomass precursor(s) by blocking or isolating specific metabolic nodes. The second group is represented by strains unable to produce universal biomass intermediates (e.g., synthesis of NAD(P)H or amines). Red arrows indicate the auxotrophy generated. **b** Effect of reaction rates

on selection strain's growth profile. A selection strain produces two different enzyme variants capable of replenishing synthesis of a key biomass precursor through their enzymatic activity. The two enzyme variants display different reaction rates ($\mu\text{mol}\cdot\text{min}^{-1}$). These different reaction rates will impact the in vivo selection, as the two selection strains can be distinguished by their different growth rates (h^{-1}). Created with BioRender.com.

combinations of enzymes and genetic modifications to enhance pathway efficiency and yield.

A good example of this is METIS, a flexible active ML workflow that enables the efficient optimization of biological targets with minimal experiments⁶⁰. The effectiveness of this approach was demonstrated across a range of applications, such as cell-free transcription and translation, genetic circuits, and a synthetic carbon dioxide fixation cycle with 27 variables. The performance of these systems was enhanced by one to two orders of magnitude. Moreover, the workflow identified the relative importance of individual factors in system performance, uncovering previously unknown interactions and bottlenecks. It can be expected that similar workflows will realize the easy optimization and prototyping of diverse genetic and metabolic networks by a broad user base in the near future.

Since characterization, structure prediction, and de novo design of enzyme function as well as drafting and prototyping of metabolic pathways, largely benefit from the plethora of innovative methods summarized above, the possible design space of biological engineers vastly increases in size. Therefore, in vitro testing and screening of enzymes and metabolic pathways might be limited in capacity with consequent constraints in the optimization process of engineered biological systems (Fig. 3).

Selection strains allow high-throughput in vivo enzyme screening

As discussed above, an alternative to in vitro testing for enzyme or pathway screening is represented by in vivo assessment using auxotroph sensor strains (henceforth referred to as selection strains), systematic growth-coupling designed by modeling⁶¹, or using antimetabolite selection strains⁶². These rely on the selective pressure generated by metabolite analogs which inhibit growth. As a

consequence, growth restoration is possible via enhanced enzyme production or synthesis of the target molecules⁶². However, it is important to note at this stage that in vivo selection might not be possible to exploit if the enzyme to be optimized cannot be linked to the metabolism of the host cell in a suitable way to enable growth-coupling, or when it does not produce an antimetabolite.

In the context of this article, we focus on auxotroph selection strains as a platform for enabling in vivo enzyme screening and evolution. In general terms, these selection strains are obtained through gene deletions which interrupt the host's metabolic network. In other words, in such strains, the biosynthesis of key biomass precursors or essential metabolic functions is blocked^{63,64}. Growth of these strains can be restored when supplying the "missing" biomass building blocks or when introducing metabolic modules (i.e., enzymatic reactions of interest) that reestablish the biosynthesis of essential metabolites. Hence, growth becomes a straightforward readout of the module's activity^{63,64}. Multiple selection strains can be generated for the same auxotrophy so that such a demand can cover different ranges of sensitivity (i.e., pulling force of the selection)⁶⁵. This feature exhibits the advantage of creating different intensities of selective pressure, which can be exploited for screening purposes. In other words, selection strains are convenient platforms to explore for enzyme evolution purposes^{17,18}, and their throughput is limited only by the transformation efficiency^{66,67}.

Selection strains can be categorized into two main groups, depending on how the auxotrophy is designed (Fig. 4a): to the first group belongs to strains presenting metabolic "isolation" or "dissection", whereas the second one includes strains deficient in a universal metabolic task (i.e., cofactor regeneration or provision of amino groups). The first group includes strains that cannot generate an essential biomass precursor molecule or an intermediate metabolite

responsible for the synthesis of a biomass precursor molecule (isolation strains). This type of strain was crucial for, e.g., the stepwise implementation of the different modules of the reductive glycine pathway prior to the demonstration of full formatotrophy^{68,69}. Similarly, “dissection” strains are incapable of synthesizing a key biomass component or one of its precursors. Moreover, in this case, the segmentation of the metabolic network is not limited to a single key metabolite but rather to a whole metabolic region (including several biomass precursor molecules). Such a broader selection range requires a higher enzymatic activity to support growth. Several studies are based on the use of dissection strains, and include, e.g., the generation of a hemi-autotrophic and an autotrophic *E. coli* growing through the Calvin-Benson-Bassham cycle^{70,71}, full formatotrophic growth via the reductive glycine pathway⁷², test of shunts for the ribulose monophosphate⁷³ or the Gnd-Entner-Doudoroff⁷⁴ cycles. Another striking example of this type of selection strains is a 3-phosphoglycerate sensor that can respond to several orders of magnitude of 3-phosphoglycerate concentrations⁶⁵.

The second group of selection strains includes mutants unable to perform a metabolic function common to multiple biochemical blocks (Fig. 4a). Examples of this sort are mutants deficient in cofactor regeneration, either in the form of NADH⁷⁵ or NADPH⁷⁶. A plethora of growth-coupled selection strategies have been developed using this type of auxotrophy, both for enzyme screening and for directed evolution campaigns at different throughput levels^{66,77–83}. Strains that lack the ability to fix ammonium to make amino acids and other essential amine metabolites also belong to this category. These can be used to select for a broad range of amine-generating reactions, e.g., for the exploration of alternative amination routes⁸⁴ or for supporting the directed evolution of amine-related enzymes⁸⁵.

Once the selection strains are equipped with the module of interest, growth restoration works as a proxy for the module's enzymatic activity^{63,64}. In particular, the growth rate μ (h^{-1}) can be used as a coarse-grained proxy for the reaction rate ($\mu\text{mol}\cdot\text{min}^{-1}$) of the target enzyme in vivo, i.e., in the context of a dedicated selection strain (Fig. 4b). For example, a selection strain expressing two different gene variants encoding for the same enzymatic activity (e.g., unevolved and evolved) might present a different growth rate for the two clones as a consequence of different reaction rates through the target enzymes (Fig. 4b). Moreover, changes in expression levels of the gene of interest (also as a consequence of evolution) might result in an improved growth rate. In summary, using growth-coupled selection strategies represents a cheap and resourceful approach for determining enzymatic activities in vivo.

Combining growth-coupling to directed evolution for new phenotypes

When exploring new-to-nature enzymatic reactions (e.g., formylphosphate reductase⁸⁶ or glycolyl-CoA carboxylase⁸⁷), it can be useful to expand the solution space of mutations which can be screened. In this situation, directed evolution becomes a useful tool as it increases genetic diversity within a sequence of interest, provided that a high-throughput per experiment can be achieved. Several directed evolution strategies have been developed through the years. They are divided mainly into two groups, based on where the diversification of the starting genetic sequence occurs: in vitro and in vivo mutagenesis. Both approaches have been extensively reviewed in literature^{11,13,14,16,17,23,40,67,88,89}.

In vitro mutagenesis approaches generate a library of gene variants in a test tube, which is then transformed into an adequate strain and screened for a readout of interest. Hence, transformation efficiency becomes the bottleneck for the number of gene variants one could recover and screen. In aid to this limitation, microfluidics solutions for high-throughput electroporation are becoming available which circumvent these shortcomings⁹⁰. The most common in vitro

techniques include (but are not limited to) error-prone PCR, site saturation mutagenesis and recombination-based DNA shuffling. In vitro-directed evolution can also be combined with the use of selection strains for in vivo screening of the evolved enzyme activity, as in the case of a formate dehydrogenase with improved specificity toward NADP⁸⁰. We refer to excellent reviews on the topic for more in-depth comparisons of the in vitro techniques available^{3,14,23}.

The use of in vivo mutagenesis strategies allows to bypass the bottleneck of transformation efficiency and perform gene diversification within the cell. Multiplex automated genome engineering (MAGE) as well as CRISPR-Cas technologies⁹¹ or zinc finger nucleases⁹² mediated tools are examples of in vivo directed mutagenesis based on mediated allelic replacement⁹³. In the case of MAGE, a pool of single-stranded DNA oligos with degenerated sequences is transformed into cells, which generates a variety of genetic modifications in vivo. By iterating this transformation step, it is possible to enhance library complexity and generate a pool of mutants which can be screened once plated on e.g., selective agar plates⁹³. The use of the abovementioned methods has been extended to multiple species beyond model laboratory strains⁹⁴. Altogether, the creation of genome-edited library strains instead of plasmid-based ones enables rapid adjustment of the strategy depending on the results of the preceding iteration.

An additional benefit of in vivo enzyme library generation is the ability to combine library generation with techniques that significantly elevate the mutation rate of the target gene. When employing selection strains for enzyme development, mutagenesis takes place concurrently with the selection of the desired phenotypic trait. These hypermutation methods facilitate rapid introduction of mutations into a gene, increasing the mutation rate (naturally between 10^{-10} and 10^{-9} to as high as 10^{-4})¹¹. Thus, these methods surpass the typical mutation rates achieved through ALE experiments, enabling the quick generation of diversified enzyme variants. Moreover, they significantly reduce the mutation or activation of off-target enzymes that might circumvent selection in the chosen strain, e.g., by activating silent genes or by mutating an enzyme to enhance its promiscuous activity. Hence, these techniques facilitate a more thorough exploration of the fitness landscape, aiding in the creation of enzyme variants that surpass local fitness maxima.

Several hypermutation techniques have been developed, as extensively reviewed recently¹¹. Most of these methods are based on error-prone DNA polymerase (OrthoRep)^{95,96}, nCas9-mediated DNA nicking combined with error-prone DNA polymerase (EvolVR)⁹⁷, or nucleobase deaminase/T7 RNA polymerase (MutaT7)⁹⁸. These techniques have been continually refined since their inception, with ongoing development focused on enhancing mutation rates and profiles. Derivatives of MutaT7 technology include e.g., extension of this technology to *S. cerevisiae*⁹⁹, improvement of its mutation rate¹⁰⁰, and fusion of a new deaminase combined with the introduction of dCas9 to obtain more control over T7 RNA polymerase¹⁰¹. Further utilization of OrthoRep allowed, e.g., to evolve custom antibodies to display on yeast's surface¹⁰² or an improved version of tryptophan synthase for synthesizing L-tryptophan from indole and L-serine¹⁰³. Another recent update of the OrthoRep system claims an improved rate of in vivo substitution per base ($>10^{-4}$)¹⁰⁴. Also, during the revision of this manuscript, a new technique was published that relies on an orthogonal DNA polymerase¹⁰⁵. In this system, user-defined DNA is introduced into an *E. coli* cell in such a way that it is selectively copied and mutated by a distinct replication machinery which is independent from the one responsible of duplicating the strain's genome. This approach resulted in the enhancement of the mutation rate in the target replicon between two to four orders of magnitude¹⁰⁵.

In addition to enzyme-based hypermutator tools, phages have been utilized as vectors to introduce variations into a target gene. This approach, known as phage-assisted continuous evolution (PACE), has

caught significant interest¹⁰⁶. In PACE, engineered phages are employed to introduce sequence variations. Leveraging the remarkably short lifecycles of phages, this method accelerates evolution cycles and enhances mutation rates in a gene of interest in the host bacterium¹⁰⁷.

In conclusion, state-of-the-art directed evolution techniques are available to develop enzymatic reactions in vivo, simplifying optimization workflows. Additionally, an expansion of the solution space beyond prediction, achievable with targeted hypermutator tools, introduces the necessary genetic diversity. The combination of these techniques^{108,109} maximizes the diversity of the library, with its size theoretically constrained only by the number of cells in the culture. Finally, the combination of the workflow with ALE promotes the enrichment of more optimal variants (Fig. 3).

ALE further enhances emerging phenotypes

Once the round(s) of directed mutagenesis enable the emergence of the activity of interest, it is possible to exploit the power of ALE to further enhance the target reaction rate. Many excellent reviews discuss the set of techniques associated with this approach, and we refer to them for a more thorough read; see for example^{12,110–113}. In the context of enzyme evolution, the use of ALE in combination with selection strains has also been described^{114–116}.

To achieve an improved phenotype, 100–500 generations are generally sufficient¹¹². These can be obtained using mainly three different experimental approaches: (i) serial batch dilutions or continuous cultivation either as (ii) chemostat or (iii) turbidostat¹¹¹. In a serial batch, a growing microbial population is propagated by serial dilutions over time while the stress factor is kept constant or increased. In this setup, the growth conditions are dynamic throughout the growth, and the moment of growth chosen for dilution has an impact on the phenotype that is being selected for. Instead, in a chemostat, the culture conditions are kept constant throughout the cultivation; influx and efflux of medium are equal, and the dilution rate sets the specific growth rate of the microbial population. A steady exponential growth is imposed on growing cells while a limiting essential nutrient determines the selective pressure. Subpopulations slower at consuming the limiting nutrient will be washed out from the cultivation and removed from the bioreactor. In this cultivation setup, both the concentration of the limiting nutrient in the feeding and the dilution rate can be controlled by the user. An overview on the basis of ALE using chemostats can be found in literature^{117,118}. A turbidostat differs from a chemostat as its dilution rate is controlled by the turbidity of the culture. Here, the goal is to maintain the turbidity constant. This system allows to select for a population of cells capable of growing at $\mu_{\max}$ and does not require the introduction of a limiting nutrient. The use of turbidostat in studying enzyme evolution has also been reported in literature¹¹⁹. Hence, depending on the phenotype one wants to select for, these different cultivation conditions can be used to support ALE efforts.

One common characteristic of the abovementioned ALE approaches is the constant selective pressure that is imposed on the system. An emerging alternative consists of the use of oscillating pressures for traversing different fitness landscapes and increasing the chances of reaching a global maximum for the phenotype of interest¹²⁰. In particular, the use of this strategy allows the exploration of mutations that would be otherwise deleterious during constant pressure. This approach allowed e.g., a change in cofactor specificity when an NADPH-auxotrophy was imposed in *E. coli*¹¹⁶. Therefore, the use of such oscillation in combination with directed evolution might allow to evolve enzyme activities through changing rugged fitness landscapes¹²⁰. In summary, we posit that ALE should be regarded as a complementary approach supporting directed evolution for the emergence of novel enzymatic reactions in biocatalysts (Fig. 3).

In the quest for optimal microbial hosts for in vivo enzyme engineering

While *E. coli* and *S. cerevisiae* have been historically used as model microbial platforms for growth-coupled selection of enzymes and synthetic pathways, non-canonical hosts have increasingly gained attention as alternatives. Among bacterial species, *E. coli* continues to be a preferred option, and the principle of increased fitness over time in the presence of selective pressure has been exploited extensively—epitomized by the classical long-term evolution experiment (LTEE), where cells evolved to optimize carbon utilization pathways towards maximizing growth over 50,000 generations¹²¹. Building on this notion, and just to mention some key studies over the last 5 years, *E. coli* has been used for the selection and evolution of the activity of several enzymes (e.g., proteases¹²², deaminases¹²³ and formate dehydrogenases⁸⁰) and enzymes displaying emergent properties (either natural or engineered, e.g., using non-canonical redox cofactors^{124,125}). *S. cerevisiae* has been likewise used to evolve bacterial enzymes, e.g., an efficient tryptophan synthase from *Thermotoga maritima* using OrthoRep¹⁰³.

While these examples illustrate the value of well-established microbial hosts, there are enzymes and pathways involving reaction substrates, intermediates and products that require a more robust host organism for in vivo engineering. Therefore, environmental bacteria thriving in habitats characterized by changing physicochemical conditions, with multiple abiotic and biotic factors (e.g., presence of stressors, salinity levels, pH values and interaction with other microbes) that play a role in shaping their physiology and metabolism, might be suitable hosts for future in vivo engineering projects. *Pseudomonas putida*, a non-pathogenic, Gram-negative soil bacterium¹²⁶, constitutes an archetypal example of a microbe displaying ‘built-in’ robustness, derived from the extreme environments it can colonize. *P. putida* has been used for multiple applications in metabolic engineering, especially towards bioprocesses that require the use of solvents or toxic substrates and products¹²⁷. Although selection schemes based on growth-coupling strategies have been implemented in *P. putida*^{128,129}, adopting this bacterium as the host for in vivo evolution of enzymes remains a relatively unexplored endeavor. *P. putida* could be an attractive option for the evolution of enzymes generating aromatic aldehydes¹³⁰ and other, similarly reactive intermediates, since such chemical species are part of its native biochemistry, e.g., as metabolites within degradation pathways for aromatic xenobiotics. Moreover, the native metabolic architecture in *P. putida* KT2440 is geared towards catabolic overproduction of reducing power in the form of NADPH¹³¹, which could further support evolving reactions that require large amounts of redox currency.

Similarly, other strains with properties that are relevant for an automated in vivo engineering process, but not present in *E. coli* or *S. cerevisiae*, could be exploited. The marine bacterium *Vibrio natriegens* is the fastest-growing microbe described so far. A doubling time of less than 10 minutes on rich medium¹³² might enable a faster automated in vivo enzyme engineering process, compared to currently used model species. Since many genetic tools, including plasmids with diverse promoters, ribosome binding sites, and resistance markers¹³³, regulatory parts¹³⁴, and a system for multiplex genome editing by natural transformation¹³⁵, are already available for this bacterium, it is likely that it will be harnessed as a chassis for in vivo enzyme engineering in the near future.

Given the expanding wealth of synthetic biology tools available for strain domestication^{136–139}, it is not unthinkable that it will become possible to choose any bacterium of interest that is naturally suitable to handle the reaction(s) to be improved or evolved, and use it as a chassis for automated in vivo enzyme engineering. This approach can be extended by using a given enzyme engineering host also directly as a production strain; e.g., the halophilic bacterium *Halomonas bluephagenesis*⁵ could be used to generate an improved enzyme that is

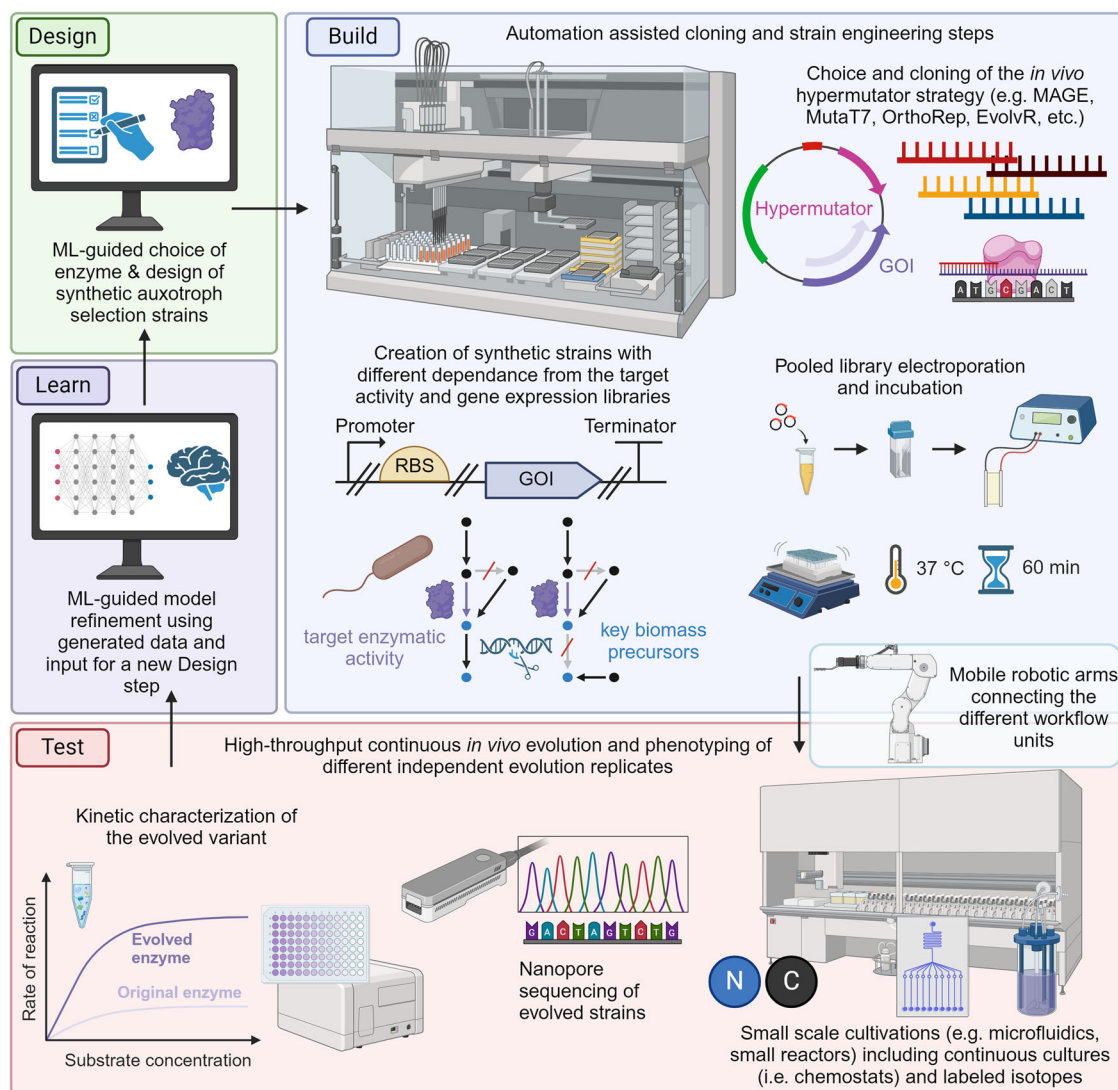


Fig. 5 | Proposed pipeline for the combined use of machine learning, automation, *in vivo* mutagenesis, and growth-coupled selection for the directed evolution of enzymes using a biofoundry. Sequential numbers indicate the steps to be taken following the Design-Build-Test-Learn paradigm. ML machine learning,

RBS ribosome binding site, GOI gene of interest. For the definition of the hypermutators MAGE, MutaT7, OrthoRep, and EvolvR we refer the reader to the main text. Created with BioRender.com.

subsequently applied in the high salt medium for the conversion of algal biomass into a desired value-added product. Similarly, use of thermophilic bacteria could be exploited for evolving thermostable enzyme variants^{140–142}. Moreover, it might be beneficial to use bacteria that can naturally produce cofactors which are required for an enzyme of interest for the *in vivo* enzyme engineering procedure. Relevant examples include cofactors such as pyrroloquinoline quinone (PQQ; redox coenzyme in dehydrogenases) or heme (prosthetic group for oxygen-carrying or electron transfer). PQQ is a common coenzyme for alcohol dehydrogenases in *P. putida*¹⁴³ or *Methylobacterium extorquens*¹⁴⁴; and while the heme biosynthetic pathway is present in *E. coli*¹⁴⁵, other bacteria, such as the metal-reducing *Shewanella oneidensis*, have many more enzymes that require this prosthetic group^{146–148}.

Towards the automated generation of optimal biocatalysts

The workflow for *in vivo*-directed evolution of enzymes can be executed through automated setups in a biofoundry. In fact, starting the proposed *in vivo* enzyme engineering workflow with a specific sensor strain, the task of integrating an efficient module to rescue and enhance cell growth may lead to the requirement of testing

different DNA parts in a combinatorial setup. In particular, when the module contains two or more enzymes in a reaction sequence, fine-tuned expression of the underlying genes is required to enable a balanced high carbon flux to maximize growth rate. The latter depends upon the right combinations of multiple DNA parts (i.e., promoter, ribosomal binding site, gene of interest, terminator) in functional transcription units¹⁴⁹, and the number of strain constructs to be tested can increase rapidly. To meet this challenge with reasonable personnel, material and time expenditure, the standardization, miniaturization and automation of strain engineering workflows is essential.

Emerging biofoundries around the globe are providing automation capabilities for setting up such workflows^{19,20,150} by transferring and combining available methods for modular DNA assembly, highly parallel transformation and incubation, image-based colony identification and multi-pin picking, as well as plasmid library preparation using canonical hosts such as *E. coli*^{151,152} or *S. cerevisiae*¹⁵³ as a basis. Recently, robotics-assisted modular cloning^{154–158}, high-throughput transformation¹⁵⁹ and monoclonal colony cultivation and picking¹⁶⁰ have been introduced for other industrially relevant organisms such as *P. putida*

or *Corynebacterium glutamicum*. The correctness of assembled and cloned plasmids can be verified easily with high-throughput using colony PCR or Oxford Nanopore sequencing¹⁶¹.

In the next step, the resulting first-generation selection strains will be used for targeted diversity generation with MAGE or other in vivo hypermutators, where the same standardized modules can be integrated into automated workflows. Most importantly, ML methods can be employed to enable autonomous exploration of the enzyme fitness landscape of combinatorial mutagenesis libraries¹⁶². In the same vein, reliable and autonomous growth phenotyping of resulting second-generation sensor strains has become possible by combining automated microbioreactor platforms¹⁶³ with appropriate data processing tools¹⁶⁴.

As mentioned above, ALE is another important tool to fully exploit the genetic diversity of sensor strains and enrich the best performing variants. Depending on the required scale, throughput and additional selection pressure¹²⁰, ALE technologies are available for automated operation at laboratory scale¹⁶⁵, small scale¹⁶⁶ and single cell level¹⁶⁷. Genome-wide identification of resulting beneficial mutations or awakened latent enzyme activities is also enhanced by automation of RNASeq technology¹⁶⁸. Finally, to confirm module activity and identify competing routes that should be inactivated, miniaturized and automated ¹³C-/¹⁵N-labeling experiments¹⁶⁹ can be performed in combination with highly informative and accurate LC-QToF mass spectrometry¹⁷⁰.

The abovementioned technologies can be combined in an automated, ML-guided pipeline. We envision that the combination of in vivo mutagenesis with the screening power of growth-coupled selection will enable to enhance the throughput of biofoundries for enzyme engineering campaigns (Fig. 5). From a technical point of view, the realization of the depicted pipeline is certainly not feasible on the basis of a large solitary platform but requires the combination of a number of customized robot platforms, ultimately connected by mobile robot units, enabling distributed workflows and complex scheduling. Moreover, there are still many pitfalls in integrating specific devices (with different interfaces) into liquid-handling stations, setting up functional and resource-efficient cloning workflows, and implementing a flexible and user-friendly digital infrastructure for running automated experiments, including real-time data processing for loop closure.

Despite these technical challenges, discussions on the perspective of self-driving labs have appeared in the scientific literature^{29,30}. Moreover, there is no evidence of their concrete implementation in fully automated experimental setups²⁸. We, therefore, expect the possibility of extending such automated workflows also to the in vivo engineering of biocatalysts in the coming future.

Outlook and final remarks

In this manuscript, we reasoned on the benefits of combining in vivo mutagenesis with growth-coupled selection strategies. As mentioned above, we believe that their use, combined with ML-guided automation, will accelerate enzyme engineering campaigns in the future. However, despite being a promising approach, there are caveats associated with these approaches, which are important to consider and are addressed in this final paragraph.

Relying on growth-coupled selection for in vivo enzyme screening displays an inherent limit in the detection threshold. This is dictated by the minimum enzymatic activity required to replenish the metabolite pool associated with the auxotrophy (i.e., a biomass precursor or a generalist metabolic function). Therefore, if the enzymatic activity is present but at too low level, growth complementation will not occur, and thus, it will not be detected. Therefore, alternative methods can be used, such as transcription factor-based biosensors^{171,172}. In principle, if the target enzyme activity can be coupled to transcriptional

activation, these types of biosensors could be used as initial step of high-throughput screening and evolution¹⁷³ and detect enzyme activity realizing, e.g., synthesis of a fluorescent protein. Use of such strains might require some adjustments to the automated pipeline presented above, such as the application of FACS to identify and isolate promising candidates. Moreover, the selective pressure imposed on the strains during the in vivo selection can induce the emergence of an underground metabolism or the generation of mutations which bypass the selection. These activities, although scientifically interesting^{174,175}, provide a risk of experimental failure, which can be avoided through some measures. These include, i.e., a physiological characterization of the selection strain after its engineering; the use of strains with a quantitatively different dependence on the activity of interest (in terms of mmol essential metabolite g_{CDW}⁻¹); use of RNAseq or proteomics data to identify possible targets responsible for breaking the selection. Moreover, prior to using the selection strains for growth-coupled experiments, it is recommended to undergo an ALE experiment under selective conditions to identify possible moonlight reactions which can hamper evolutionary campaigns¹¹⁶. These pieces of information can then instruct the ML pipeline to curate predictions for gene deletions for the construction of new selection strains.

Besides, some engineering or evolution campaigns might involve enantioselective enzymes or simply enzymes whose activity cannot be coupled to growth. In these cases, the use of growth-coupled selection is not possible, and the approach described in this manuscript would not lead to the attainment of improved biocatalysts.

Another important caveat is related to the use of in vivo hypermutators. Despite all the benefits mentioned above, some techniques display an inherent bias towards a certain type of mutation. This can create diversification in the evolutionary landscape with consequent constrained capacity for long-term sequence space exploration¹⁷⁶.

Finally, it is important to note that the optimized enzyme obtained at the end of the workflow must be tested in the context of its final purpose. Therefore, other iterative Design-Build-Test-Learn cycles might be required, e.g., in the context of a production strain to assess the effectiveness of the evolution campaign for biomanufacturing, as previously suggested in the use of growth-coupled selection for cell factories optimization⁶⁴.

References

- Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
- Erb, T. J., Jones, P. R. & Bar-Even, A. Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol.* **37**, 56–62 (2017). **The idea of synthetic metabolism is defined as a level of metabolic engineering with an increased design and solution space beyond what nature has evolved for.**
- Volk, M. J. et al. Metabolic engineering: methodologies and applications. *Chem. Rev.* **9**, 5521–5570 (2022).
- Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat. Catal.* **3**, 203–213 (2020).
- Tan, D., Xue, Y. S., Aibaidula, G. & Chen, G. Q. Unsterile and continuous production of polyhydroxybutyrate by *Halomonas* TD01. *Bioresour. Technol.* **102**, 8130–TD8136 (2011).
- Clomburg, J. M., Crumbley, A. M. & Gonzalez, R. Industrial biomanufacturing: the future of chemical production. *Science* **355**, aag0804 (2017).
- Zhang, Y. H. P., Sun, J. & Ma, Y. Biomanufacturing: history and perspective. *J. Ind. Microbiol. Biotechnol.* **44**, 773–784 (2017).
- Buller, R. et al. From nature to industry: harnessing enzymes for biocatalysis. *Science* **382**, eadh8615 (2023).
- Arnold, F. H. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed. Engl.* **57**, 4143–4148 (2018).

10. Cai, T. et al. Cell-free chemoenzymatic starch synthesis from carbon dioxide. *Science* **373**, 1523–1527 (2021).
11. Molina, R. S. et al. In vivo hypermutation and continuous evolution. *Nat. Rev. Methods Primers* **2**, 37 (2022). **Extensive review on the use of in vivo hypermutator techniques for continuous evolution.**
12. Wu, Y., Jameel, A., Xing, X. H. & Zhang, C. Advanced strategies and tools to facilitate and streamline microbial adaptive laboratory evolution. *Trends Biotechnol.* **40**, 38–59 (2022).
13. Wang, Y. et al. Directed evolution: methodologies and applications. *Chem. Rev.* **121**, 12384–12444 (2021).
14. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
15. Golynskiy, M. V., Haugner, J. C. 3rd, Morelli, A., Morrone, D. & Seelig, B. In vitro evolution of enzymes. *Methods Mol. Biol.* **978**, 73–92 (2013).
16. Badran, A. H. & Liu, D. R. In vivo continuous directed evolution. *Curr. Opin. Chem. Biol.* **24**, 1–10 (2015).
17. Li, Z., Deng, Y. & Yang, G. Y. Growth-coupled high throughput selection for directed enzyme evolution. *Biotechnol. Adv.* **68**, 108238 (2023).
18. Chen, J., Wang, Y., Zheng, P. & Sun, J. Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends Biotechnol.* **40**, 773–776 (2022).
19. Gurdo, N., Volke, D. C., McCloskey, D. & Nikel, P. I. Automating the design-build-test-learn cycle towards next-generation bacterial cell factories. *N. Biotechnol.* **74**, 1–15 (2023). **This review discusses in detail recent advances in automating the design-build-test-learn pipeline.**
20. Tellechea-Luzardo, J., Otero-Muras, I., Goni-Moreno, A. & Carbone, P. Fast biofoundries: coping with the challenges of biomanufacturing. *Trends Biotechnol.* **40**, 831–842 (2022).
21. Chao, R., Mishra, S., Si, T. & Zhao, H. Engineering biological systems using automated biofoundries. *Metab. Eng.* **42**, 98–108 (2017).
22. Hillson, N. et al. Building a global alliance of biofoundries. *Nat. Commun.* **10**, 2040 (2019).
23. Yu, T., Boob, A. G., Singh, N., Su, Y. & Zhao, H. In vitro continuous protein evolution empowered by machine learning and automation. *Cell Syst.* **14**, 633–644 (2023). **Review on advancements in machine learning and lab automation for rapid protein engineering through directed evolution.**
24. Lawson, C. E. et al. Machine learning for metabolic engineering: a review. *Metab. Eng.* **63**, 34–60 (2021).
25. Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Machine learning applications in systems metabolic engineering. *Curr. Opin. Biotechnol.* **64**, 1–9 (2020).
26. Cheng, Y. et al. Machine learning for metabolic pathway optimization: a review. *Comput. Struct. Biotechnol. J.* **21**, 2381–2393 (2023).
27. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
28. Rapp, J. T., Bremer, B. J. & Romero, P. A. Self-driving laboratories to autonomously navigate the protein fitness landscape. *Nat. Chem. Eng.* **1**, 97–107 (2024). **Demonstration of a self-driving automated robotic system that designs, tests, and provides feedbacks on a protein engineering pipeline.**
29. Häse, F., Roch, L. M. & Aspuru-Guzik, A. Next-generation experimentation with self-driving laboratories. *Trends Chem.* **1**, 282–291 (2019).
30. Martin, H. G. et al. Perspectives for self-driving labs in synthetic biology. *Curr. Opin. Biotechnol.* **79**, 102881 (2023).
31. Lovelock, S. L. et al. The road to fully programmable protein catalysis. *Nature* **606**, 49–58 (2022).
32. R., C & Maranas, C. D. From directed evolution to computational enzyme engineering—a review. *AIChE J.* **66** <https://par.nsf.gov/servlets/purl/10170897> (2020).
33. Qu, G., Li, A., Acevedo-Rocha, C. G., Sun, Z. & Reetz, M. T. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed. Engl.* **59**, 13204–13231 (2020).
34. Nezhad, N. G. et al. Thermostability engineering of industrial enzymes through structure modification. *Appl. Microbiol. Biotechnol.* **106**, 4845–4866 (2022).
35. Sun, Z., Liu, Q., Qu, G., Feng, Y. & Reetz, M. T. Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem. Rev.* **119**, 1626–1665 (2019).
36. Planas-Iglesias, J. et al. Computational design of enzymes for biotechnological applications. *Biotechnol. Adv.* **47**, 107696 (2021).
37. Sumbalova, L., Stourac, J., Martinek, T., Bednar, D. & Damborsky, J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res.* **46**, W356–W362 (2018).
38. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
39. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16**, 378–384 (2005).
40. McLure, R. J., Radford, S. E. & Brockwell, D. J. High-throughput directed evolution: a golden era for protein science. *Trends Chem.* **4**, 278–291 (2022).
41. Tokuriki, N. et al. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* **3**, 1257 (2012).
42. Tawfik, D. S. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency. *Curr. Opin. Chem. Biol.* **21**, 73–80 (2014).
43. Goldsmith, M. & Tawfik, D. S. Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol.* **47**, 140–150 (2017).
44. Mazurenko, S., Prokop, Z. & Damborsky, J. Machine learning in enzyme engineering. *ACS Catal.* **10**, 1210–1223 (2020).
45. Kouba, P. et al. Machine learning-guided protein engineering. *ACS Catal.* **13**, 13863–13895 (2023).
46. Alford, R. F. et al. The rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
47. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987–2998 (2010).
48. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
49. Kipnis, Y. et al. Design and optimization of enzymatic activity in a de novo beta-barrel scaffold. *Protein Sci.* **31**, e4405 (2022).
50. Yeh, A. H. et al. De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).
51. Kalvet, I. et al. Design of heme enzymes with a tunable substrate binding pocket adjacent to an open metal coordination site. *J. Am. Chem. Soc.* **145**, 14307–14315 (2023).
52. Bjelic, S. et al. Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-Hillman reaction. *ACS Chem. Biol.* **8**, 749–757 (2013).
53. Crawshaw, R. et al. Engineering an efficient and enantioselective enzyme for the Morita-Baylis-Hillman reaction. *Nat. Chem.* **14**, 313–320 (2022).

54. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
55. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
56. Siegel, J. B. et al. Computational protein design enables a novel one-carbon assimilation pathway. *Proc. Natl. Acad. Sci.* **112**, 3704–3709 (2015).
57. Koch, M., Duigou, T. & Faulon, J.-L. Reinforcement learning for bioretrosynthesis. *ACS Synth. Biol.* **9**, 157–168 (2020).
58. Yu, T. et al. Machine learning-enabled retrobiosynthesis of molecules. *Nat. Catal.* **6**, 137–151 (2023).
59. Zheng, S. et al. Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP. *Nat. Commun.* **13**, 3342 (2022).
60. Pandi, A. et al. A versatile active learning workflow for optimization of genetic and metabolic networks. *Nat. Commun.* **13**, 3876 (2022).
61. Von Kamp, A. & Klamt, S. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat. Commun.* **8**, 15956 (2017). **Study confirming the feasibility of coupling growth to production across diverse organisms, underscoring its importance for rational metabolic engineering.**
62. Buerger, J., Gronenberg, L. S., Genee, H. J. & Sommer, M. O. A. Wiring cell growth to product formation. *Curr. Opin. Biotechnol.* **59**, 85–92 (2019).
63. Wenk, S., Yishai, O., Lindner, S. N. & Bar-Even, A. An engineering approach for rewiring microbial metabolism. *Methods Enzymol.* **608**, 329–367 (2018).
64. Orsi, E., Claassens, N. J., Nikel, P. I. & Lindner, S. N. Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nat. Commun.* **12**, 5295 (2021). **Comment article describing the use of growth-coupling within the context of the design-build-test-learn cycle.**
65. Aslan, S., Noor, E., Benito Vaquerizo, S., Lindner, S. N. & Bar-Even, A. Design and engineering of *E. coli* metabolic sensor strains with a wide sensitivity range for glycerate. *Metab. Eng.* **57**, 96–109 (2020).
66. Nielsen, J. R., Weusthuis, R. A. & Huang, W. E. Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnol. Adv.* **63**, 108102 (2023).
67. Xiao, H., Bao, Z. & Zhao, H. High throughput screening and selection methods for directed enzyme evolution. *Ind. Eng. Chem. Res.* **54**, 4011–4020 (2015).
68. Yishai, O., Bouzon, M., Döring, V. & Bar-Even, A. In vivo assimilation of one-carbon via a synthetic reductive glycine pathway in *Escherichia coli*. *ACS Synth. Biol.* **7**, 2023–2028 (2018).
69. Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S. N. & Bar-Even, A. Engineered assimilation of exogenous and endogenous formate in *Escherichia coli*. *ACS Synth. Biol.* **6**, 1722–1731 (2017).
70. Gleizer, S. et al. Conversion of *Escherichia coli* to generate all biomass carbon from CO₂. *Cell* **179**, 1255–1263 (2019).
71. Antonovsky, N. et al. Sugar synthesis from CO₂ in *Escherichia coli*. *Cell* **166**, 115–125 (2016).
72. Kim, S. et al. Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* **16**, 538–545 (2020).
73. He, H., Edlich-Muth, C., Lindner, S. N. & Bar-Even, A. Ribulose monophosphate shunt provides nearly all biomass and energy required for growth of *E. coli*. *ACS Synth. Biol.* **7**, 1601–1611 (2018).
74. Satanowski, A. et al. Awakening a latent carbon fixation cycle in *Escherichia coli*. *Nat. Commun.* **11**, 5812 (2020).
75. Wenk, S. et al. An “energy-auxotroph” *Escherichia coli* provides an in vivo platform for assessing NADH regeneration systems. *Biotechnol. Bioeng.* **117**, 3422–3434 (2020).
76. Lindner, S. N. et al. NADPH-auxotrophic *E. coli*: a sensor strain for testing in vivo regeneration of NADPH. *ACS Synth. Biol.* **7**, 2742–2749 (2018).
77. Trinh, C. T., Liu, Y. & Conner, D. J. Rational design of efficient modular cells. *Metab. Eng.* **32**, 220–231 (2015).
78. Zhang, L., King, E., Luo, R. & Li, H. Development of a high-throughput, in vivo selection platform for NADPH-dependent reactions based on redox balance principles. *ACS Synth. Biol.* **7**, 1715–1721 (2018).
79. Kramer, L. et al. Engineering carboxylic acid reductase (CAR) through a whole-cell growth-coupled NADPH recycling strategy. *ACS Synth. Biol.* **9**, 1632–1637 (2020).
80. Calzadiaz-Ramirez, L. et al. In vivo selection for formate dehydrogenases with high efficiency and specificity toward NADP⁺. *ACS Catal.* **10**, 7512–7525 (2020).
81. Maxel, S. et al. A growth-based, high-throughput selection platform enables remodeling of 4-hydroxybenzoate hydroxylase active site. *ACS Catal.* **10**, 6969–6974 (2020).
82. Maxel, S. et al. Growth-based, high-throughput selection for NADH preference in an oxygen-dependent biocatalyst. *ACS Synth. Biol.* **10**, 2359–2370 (2021).
83. Maxel, S. et al. In vivo, high-throughput selection of thermostable cyclohexanone monooxygenase (CHMO). *Catalysts* **10**, 935 (2020).
84. Schulz-Mirbach, H. et al. On the flexibility of the cellular amination network in *E. coli*. *Elife* **11**, e77492 (2022).
85. Wu, S. et al. A growth selection system for the directed evolution of amine-forming or converting enzymes. *Nat. Commun.* **13**, 7458 (2022).
86. Nattermann, M. et al. Engineering a new-to-nature cascade for phosphate-dependent formate to formaldehyde conversion in vitro and in vivo. *Nat. Commun.* **14**, 2682 (2023).
87. Marchal, D. G. et al. Machine learning-supported enzyme engineering toward improved CO(2)-fixation of glycolyl-CoA carboxylase. *ACS Synth. Biol.* **12**, 3521–3530 (2023).
88. d’Oelsnitz, S. & Ellington, A. Continuous directed evolution for strain and protein engineering. *Curr. Opin. Biotechnol.* **53**, 158–163 (2018).
89. Rix, G. & Liu, C. C. Systems for in vivo hypermutation: a quest for scale and depth in directed evolution. *Curr. Opin. Chem. Biol.* **64**, 20–26 (2021).
90. Iwai, K. et al. Scalable and automated CRISPR-based strain engineering using droplet microfluidics. *Microsyst. Nanoeng.* **8**, 31 (2022).
91. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
92. Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764 (2003).
93. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
94. Nyerges, A. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl. Acad. Sci. USA* **113**, 2502–2507 (2016).
95. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* **175**, 1946–1957.e1913 (2018).
96. Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
97. Halperin, S. O. et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).

98. Moore, C. L., Papa, L. J. 3rd & Shoulders, M. D. A processive protein chimera introduces mutations across defined DNA regions in vivo. *J. Am. Chem. Soc.* **140**, 11560–11564 (2018).
99. Cravens, A., Jamil, O. K., Kong, D., Sockolosky, J. T. & Smolke, C. D. Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering. *Nat. Commun.* **12**, 1579 (2021).
100. Park, H. & Kim, S. Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Res.* **49**, e32 (2021).
101. Alvarez, B., Mencia, M., de Lorenzo, V. & Fernandez, L. A. In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.* **11**, 6436 (2020).
102. Wellner, A. et al. Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat. Chem. Biol.* **17**, 1057–1064 (2021).
103. Rix, G. et al. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **11**, 5644 (2020).
104. Rix, G. et al. Continuous evolution of user-defined genes at 1-million-times the genomic mutation rate (Cold Spring Harbor Laboratory, 2023).
105. Tian, R. et al. Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. coli*. *Science* **383**, 421–426 (2024). **Method to rapidly mutate defined DNA in *E. coli*, speeding up evolution without harmful off-target effects.**
106. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
107. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
108. Zhong, Z. et al. Automated continuous evolution of proteins in vivo. *ACS Synth. Biol.* **9**, 1270–1276 (2020).
109. Huang, T. P. et al. High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. *Nat. Biotechnol.* **41**, 96–107 (2023).
110. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution - principles and applications for biotechnology. *Micro. Cell Fact.* **12**, 64 (2013).
111. Mavrommati, M., Daskalaki, A., Papanikolaou, S. & Aggelis, G. Adaptive laboratory evolution principles and applications in industrial biotechnology. *Biotechnol. Adv.* **54**, 107795 (2022).
112. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56**, 1–16 (2019).
113. Wang, G. et al. Recent progress in adaptive laboratory evolution of industrial microorganisms. *J. Ind. Microbiol. Biotechnol.* **50**, kuac023 (2023).
114. Zelle, R. M., Harrison, J. C., Pronk, J. T. & Van Maris, A. J. A. Anaerobic role for cytosolic malic enzyme in engineered *Saccharomyces cerevisiae* strains. *Appl. Environ. Microbiol.* **77**, 732–738 (2011).
115. Luo, H. et al. Coupling S-adenosylmethionine-dependent methylation to growth: design and uses. *PLoS Biol.* **17**, e2007050 (2019).
116. Bouzon, M. et al. Change in cofactor specificity of oxidoreductases by adaptive evolution of an *Escherichia coli* NADPH-auxotrophic strain. *mBio* **12**, e0032921 (2021).
117. Wortel, M. T., Bosdriesz, E., Teusink, B. & Bruggeman, F. J. Evolutionary pressures on microbial metabolic strategies in the chemostat. *Sci. Rep.* **6**, 29503 (2016).
118. Gresham, D. & Hong, J. The functional basis of adaptive evolution in chemostats. *FEMS Microbiol. Rev.* **39**, 2–16 (2015).
119. Counago, R., Chen, S. & Shamoo, Y. In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol. Cell* **22**, 441–449 (2006).
120. Carpenter, A. C., Feist, A. M., Harrison, F. S. M., Paulsen, I. T. & Williams, T. C. Have you tried turning it off and on again? Oscillating selection to enhance fitness-landscape traversal in adaptive laboratory evolution experiments. *Metab. Eng. Commun.* **17**, e00227 (2023). **Perspective article discussing oscillating selection pressures as a tool to enable fitness valley crossing and optimum peak shifting.**
121. Wiser, M. J., Ribeck, N. & Lenski, R. E. Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367 (2013).
122. Kross, C. et al. PROFICS: a bacterial selection system for directed evolution of proteases. *J. Biol. Chem.* **297**, 101095 (2021).
123. Long, M. et al. Directed evolution of ornithine cyclodeaminase using an EvolvR-based growth-coupling strategy for efficient biosynthesis of L-proline. *ACS Synth. Biol.* **9**, 1855–1863 (2020).
124. King, E. et al. Orthogonal glycolytic pathway enables directed evolution of noncanonical cofactor oxidase. *Nat. Commun.* **13**, 7282 (2022).
125. Zhang, L. et al. Directed evolution of phosphite dehydrogenase to cycle noncanonical redox cofactors via universal growth selection platform. *Nat. Commun.* **13**, 5021 (2022).
126. Belda, E. et al. The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. *Environ. Microbiol.* **18**, 3403–3424 (2016).
127. Weimer, A., Kohlstedt, M., Volke, D. C., Nikel, P. I. & Wittmann, C. Industrial biotechnology of *Pseudomonas putida*: advances and prospects. *Appl. Microbiol. Biotechnol.* **104**, 7745–7766 (2020).
128. Wirth, N. T. et al. A synthetic C2 auxotroph of *Pseudomonas putida* for evolutionary engineering of alternative sugar catabolic routes. *Metab. Eng.* **74**, 83–97 (2022).
129. Eng, T. et al. Maximizing microbial bioproduction from sustainable carbon sources using iterative systems engineering. *Cell Rep.* **42**, 113087 (2023).
130. Yuan, Z., Liao, J., Jiang, H., Cao, P. & Li, Y. Aldehyde catalysis - from simple aldehydes to artificial enzymes. *RSC Adv.* **10**, 35433–35448 (2020).
131. Nikel, P. I. et al. Reconfiguration of metabolic fluxes in *Pseudomonas putida* as a response to sub-lethal oxidative stress. *ISME J.* **15**, 1751–1766 (2021).
132. Eagon, R. G. *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 minutes. *J. Bacteriol.* **83**, 736–737 (1962).
133. Tschirhart, T. et al. Synthetic biology tools for the fast-growing marine bacterium *Vibrio natriegens*. *ACS Synth. Biol.* **8**, 2069–2079 (2019).
134. Wu, F. et al. Design and reconstruction of regulatory parts for fast-growing *Vibrio natriegens* synthetic biology. *ACS Synth. Biol.* **9**, 2399–2409 (2020).
135. Dalia, T. N. et al. Multiplex genome editing by natural transformation (MuGENT) for synthetic biology in *Vibrio natriegens*. *ACS Synth. Biol.* **6**, 1650–1655 (2017).
136. Nikel, P. I., Martinez-Garcia, E. & de Lorenzo, V. Biotechnological domestication of pseudomonads using synthetic biology. *Nat. Rev. Microbiol.* **12**, 368–379 (2014).
137. Schada von Borzyskowski, L. Taking synthetic biology to the seas: from blue chassis organisms to marine aquaforming. *ChemBioChem* **24**, e202200786 (2023).
138. Riley, L. A. & Guss, A. M. Approaches to genetic tool development for rapid domestication of non-model microorganisms. *Bio-technol. Biofuels* **14**, 30 (2021).

139. Volke, D. C., Orsi, E. & Nikel, P. I. Emergent CRISPR-Cas-based technologies for engineering non-model bacteria. *Curr. Opin. Microbiol.* **75**, 102353 (2023).
140. Rigoldi, F., Donini, S., Redaelli, A., Parisini, E. & Gautieri, A. Review: engineering of thermostable enzymes for industrial applications. *APL Bioeng.* **2**, 011501 (2018).
141. Atalah, J., Caceres-Moreno, P., Espina, G. & Blamey, J. M. Thermophiles and the applications of their enzymes as new biocatalysts. *Bioresour. Technol.* **280**, 478–488 (2019).
142. Han, H. et al. Improvements of thermophilic enzymes: from genetic modifications to applications. *Bioresour. Technol.* **279**, 350–361 (2019).
143. Turlin, J., Puiggene, O., Donati, S., Wirth, N. T. & Nikel, P. I. Core and auxiliary functions of one-carbon metabolism in *Pseudomonas putida* exposed by a systems-level analysis of transcriptional and physiological responses. *mSystems* **8**, e0000423 (2023).
144. Ghosh, M., Avezoux, A., Anthony, C., Harlos, K. & Blake, C. C. X-ray structure of PQQ-dependent methanol dehydrogenase. *EXS* **71**, 251–260 (1994).
145. Jordan, P. M., Mgbeje, B. I., Thomas, S. D. & Alwan, A. F. Nucleotide sequence for the *hemD* gene of *Escherichia coli* encoding uroporphyrinogen III synthase and initial evidence for a hem operon. *Biochem. J.* **249**, 613–616 (1988).
146. Mowat, C. G. et al. Octaheme tetrathionate reductase is a respiratory enzyme with novel heme ligation. *Nat. Struct. Mol. Biol.* **11**, 1023–1024 (2004).
147. Pitts, K. E. et al. Characterization of the *Shewanella oneidensis* MR-1 decaheme cytochrome MtrA: expression in *Escherichia coli* confers the ability to reduce soluble Fe(III) chelates. *J. Biol. Chem.* **278**, 27758–27765 (2003).
148. Schwalb, C., Chapman, S. K. & Reid, G. A. The membrane-bound tetrahaem c-type cytochrome CymA interacts directly with the soluble fumarate reductase in *Shewanella*. *Biochem. Soc. Trans.* **30**, 658–662 (2002).
149. Zhang, M., Holowko, M. B., Hayman Zumpe, H. & Ong, C. S. Machine learning guided batched design of a bacterial ribosome binding site. *ACS Synth. Biol.* **11**, 2314–2326 (2022).
150. Ko, S. C., Cho, M., Lee, H. J. & Woo, H. M. Biofoundry palette: planning-assistant software for liquid handler-based experimentation and operation in the biofoundry workflow. *ACS Synth. Biol.* **11**, 3538–3543 (2022).
151. Iverson, S. V., Haddock, T. L., Beal, J. & Densmore, D. M. CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth. Biol.* **5**, 99–103 (2016).
152. Bryant, J. A. Jr, Kellinger, M., Longmire, C., Miller, R. & Wright, R. C. AssemblyTron: flexible automation of DNA assembly with open-trons OT-2 lab robots. *Synth. Biol.* **8**, ysac032 (2023).
153. Malci, K. et al. Standardization of synthetic biology tools and assembly methods for *saccharomyces cerevisiae* and emerging yeast species. *ACS Synth. Biol.* **11**, 2527–2547 (2022).
154. Martinez-Garcia, E. et al. SEVA 4.0: an update of the Standard European Vector Architecture database for advanced analysis and programming of bacterial phenotypes. *Nucleic Acids Res.* **51**, D1558–D1567 (2023).
155. Keating, K. W. & Young, E. M. Systematic part transfer by extending a modular toolkit to diverse bacteria. *ACS Synth. Biol.* **12**, 2061–2072 (2023).
156. Blazquez, B. et al. Golden standard: a complete standard, portable, and interoperable MoClo tool for model and non-model proteobacteria. *Nucleic Acids Res.* **51**, e98 (2023).
157. Kang, D. H., Ko, S. C., Heo, Y. B., Lee, H. J. & Woo, H. M. Robo-MoClo: a robotics-assisted modular cloning framework for multiple gene assembly in biofoundry. *ACS Synth. Biol.* **11**, 1336–1348 (2022).
158. Nava, A. A. et al. Automated platform for the plasmid construction process. *ACS Synth. Biol.* **12**, 3506–3513 (2023).
159. Tenhaef, N., Stella, R., Frunzke, J. & Noack, S. Automated rational strain construction based on high-throughput conjugation. *ACS Synth. Biol.* **10**, 589–599 (2021).
160. Jian, X. et al. Single-cell microliter-droplet screening system (MISS Cell): an integrated platform for automated high-throughput microbial monoclonal cultivation and picking. *Biotechnol. Bioeng.* **120**, 778–792 (2023).
161. Vegh, P., Donovan, S., Rosser, S., Stracquadanio, G., & Fragkoudis, R. Biofoundry-scale DNA assembly validation using cost-effective high-throughput long read sequencing. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2023.09.19.558498v1> (2023).
162. Hu, R. et al. Protein engineering via Bayesian optimization-guided evolutionary algorithm and robotic experiments. *Brief. Bioinform.* **24**, bbac570 (2023).
163. Helleckes, L. M. et al. From frozen cell bank to product assay: high-throughput strain characterisation for autonomous design-build-test-learn cycles. *Micro. Cell Fact.* **22**, 130 (2023).
164. Helleckes, L. M., Osthege, M., Wiechert, W., von Lieres, E. & Oldiges, M. Bayesian calibration, process modeling and uncertainty quantification in biotechnology. *PLoS Comput. Biol.* **18**, e1009223 (2022).
165. Bromig, L. & Weuster-Botz, D. Accelerated adaptive laboratory evolution by automated repeated batch processes in parallelized bioreactors. *Microorganisms* **11**, 275 (2023).
166. Halle, L. et al. Robotic workflows for automated long-term adaptive laboratory evolution: improving ethanol utilization by *Corynebacterium glutamicum*. *Micro. Cell Fact.* **22**, 175 (2023).
167. Rosenthal, R. G., Diana Zhang, X., Durdic, K. I., Collins, J. J. & Weitz, D. A. Controlled continuous evolution of enzymatic activity screened at ultrahigh throughput using drop-based microfluidics. *Angew. Chem. Int. Ed. Engl.* **62**, e202303112 (2023).
168. Garcia, B. J. et al. A toolkit for enhanced reproducibility of RNASeq analysis for synthetic biologists. *Synth. Biol.* **7**, ysac012 (2022).
169. Niesser, J., Muller, M. F., Kappelmann, J., Wiechert, W. & Noack, S. Hot isopropanol quenching procedure for automated microtiter plate scale (¹³C)-labeling experiments. *Microb. Cell Fact.* **21**, 78 (2022).
170. Kappelmann, J., Beyss, M., Noh, K. & Noack, S. Separation of ¹³C- and ¹⁵N-isotopologues of amino acids with a primary amine without mass resolution by means of o-phthalaldehyde derivatization and collision induced dissociation. *Anal. Chem.* **91**, 13407–13417 (2019).
171. Li, J. W., Zhang, X. Y., Wu, H. & Bai, Y. P. Transcription factor engineering for high-throughput strain evolution and organic acid bioproduction: a review. *Front. Bioeng. Biotechnol.* **8**, 98 (2020).
172. Mitchler, M. M., Garcia, J. M., Montero, N. E. & Williams, G. J. Transcription factor-based biosensors: a molecular-guided approach for natural product engineering. *Curr. Opin. Biotechnol.* **69**, 172–181 (2021).
173. Cheng, F., Tang, X. L. & Kardashliev, T. Transcription factor-based biosensors in high-throughput screening: advances and applications. *Biotechnol. J.* **13**, e1700648 (2018).
174. Notebaart, R. A., Kintsjes, B., Feist, A. M. & Papp, B. Underground metabolism: network-level perspective and biotechnological potential. *Curr. Opin. Biotechnol.* **49**, 108–114 (2018).
175. Rosenberg, J. & Commichau, F. M. Harnessing underground metabolism for pathway development. *Trends Biotechnol.* **37**, 29–37 (2019).
176. Napiorkowska, M., et al. YeastIT: reducing mutational bias for in vivo directed evolution using a novel yeast mutator strain based on dual adenine-/cytosine-targeting and error-prone DNA repair. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2023.11.20.567881v1> (2023).

Acknowledgements

The authors thank Carlos Acevedo-Rocha and Ari Satanowski for critical reading of the manuscript. E.O. was supported by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 101065339 (ROAD). S.N. acknowledges funding from the German Federal Ministry of Education and Research (BMBF) (grant number 031B1134A) as part of the "Auto-BioTech" innovation lab. P.I.N. acknowledges funding from the Nordisk Foundation (grants NNF10CC1016517 and NNF18CC0033664). S.N.L. acknowledges funding from the BMBF grants MaxKat (031B1028) and ForceYield 2.2 (031B1337B).

Author contributions

S.N.L. and E.O. structured and composed an outline of the manuscript. L.S.v.B., P.I.N. and S.N. provided extensive feedback on this outline. All authors wrote sections of the manuscript and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Steffen N. Lindner.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024