Automating the design-build-test-learn cycle towards next-generation bacterial cell factories

Gurdo, Nicolás; Volke, Daniel C.; McCloskey, Douglas; Nikel, Pablo Iván

Published in:
New Biotechnology

Link to article, DOI:
10.1016/j.nbt.2023.01.002

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

New BIOTECHNOLOGY 74 (2023) 1–15

Contents lists available at ScienceDirect

New BIOTECHNOLOGY

journal homepage: www.elsevier.com/locate/nbt

Automating the design-build-test-learn cycle towards next-generation bacterial cell factories

Nicolás Gurdo, Daniel C. Volke, Douglas McCloskey, Pablo Iván Nikel *

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens, Lyngby, Denmark

ARTICLE INFO

Keywords:
Synthetic biology
Biofoundry
DBTL cycle
Automation
Machine learning
Metabolic engineering
Synthetic metabolism
Bacteria

ABSTRACT

Automation is playing an increasingly significant role in synthetic biology. Groundbreaking technologies, developed over the past 20 years, have enormously accelerated the construction of efficient microbial cell factories. Integrating state-of-the-art tools (e.g. for genome engineering and analytical techniques) into the design-build-test-learn cycle (DBTLc) will shift the metabolic engineering paradigm from an almost artisanal labor towards a fully automated workflow. Here, we provide a perspective on how a fully automated DBTLc could be harnessed to construct the next-generation bacterial cell factories in a fast, high-throughput fashion. Innovative toolsets and approaches that pushed the boundaries in each segment of the cycle are reviewed to this end. We also present the most recent efforts on automation of the DBTLc, which heralds a fully autonomous pipeline for synthetic biology in the near future.

1. Introduction

Supported by synthetic biology (SynBio), metabolic engineering has shifted from the traditional, trial-and-error approaches in the late 1990 s and early 2000 s towards a truly rational effort. This transition has been promoted by novel tools, e.g. advanced genome editing protocols, multipart gene and genome assembly, genome-scale metabolic reconstructions and high-throughput phenotype analysis. Virtually all of these techniques are still implemented step-by-step, performed in a manual and iterative — almost artisanal — fashion. Manual labor is a major source of non-systematic errors, leading to disproportionate resource consumption and considerable production of wastes, imprecise designs, non-scalable, laboratory-specific techniques and selective data recording [1]. Automated processes, rapidly emerging across fields, became an alternative to overcome these limitations. Automation routines in biotechnology are supported by robotics, DNA sequencing, data processing and artificial intelligence (AI) — as well as standardization of biological parts. These developments will not only speed up workflows and increase reproducibility, but they will also enable new applications of metabolic engineering beyond the customary handful of target molecules [2]. When systematically adopted, these approaches can help to bridge the gap between (i) design and construction of genetic circuits encoding defined metabolic modules, (ii) combinatorial, multipart DNA assembly and (iii) performance analysis, for the time being, carried out individually. Automation will also accelerate re-design of genetic circuits, adopting alternative genetic parts to enhance performance.

In the broad engineering field, computer-aided design and analysis is known as *design automation*. This concept has been extended to SynBio as *bio-design automation* (BDA) [3]. Here, a particular input (e.g. blueprint of a metabolic pathway, plasmid or synthetic construct) is transformed into a physical entity [e.g. a biological *chassis* (see Fig. 1 for a glossary of relevant terms) equipped with the information needed to produce a protein or metabolite of interest]. BDA has been implemented at each step of the *design-build-testing-learn cycle* (DBTLc) to enable fully

Abbreviations: DBTLc, design-build-test-learn cycle; MFA, metabolic flux analysis; SynBio, synthetic biology; AI, artificial intelligence; BDA, biodesign automation; ML, machine learning; RBS, ribosome binding site; SBOL, SynBio open language; MAGE, multiplex automated genome engineering; USER, uracil-specific excision reagent; LCR, ligase chain reaction; SRM/MRM, selected- and multiple-reaction monitoring; DDA, data dependent analysis; DIA, data independent analysis; FIA, flow-injection analysis; SWATH-MS, sequential window acquisition of all theoretical mass spectra; HRMS, high resolution mass spectrometry; FBA, flux balance analysis; GSMM, genome-scale metabolic model; COBRA, constraint-based reconstruction and analysis; tFBA, thermodynamics-based FBA; FVA, flux variability analysis; pFBA, parsimonious FBA; MDVs, mass distribution vectors; EMU, elementary metabolite units; DL, deep learning; SVMs, support vector machines; VAE, variational autoencoder; GAN, generative adversarial network; GNNs, graph neural networks; PINNs, physics-informed neural networks; TPOT, tree-based pipeline optimization tool.

* Correspondence to: The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark.
E-mail address: pabnik@biosustain.dtu.dk (P.I. Nikel).

https://doi.org/10.1016/j.nbt.2023.01.002
Received 4 December 2022; Received in revised form 15 January 2023; Accepted 22 January 2023
Available online 1 February 2023

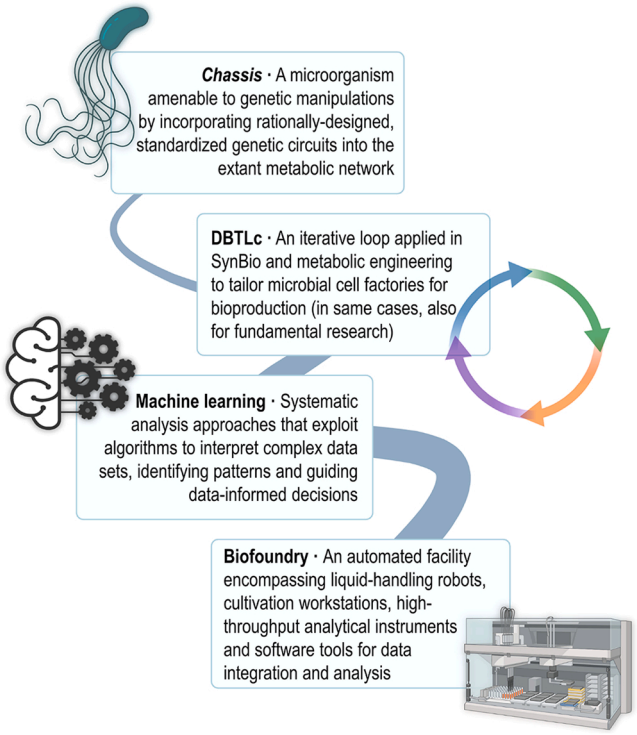

Fig. 1. Key definitions used in this article in the context of automating the design-build-test-learn cycle (DBTLc) of Synthetic Biology. Each of these terms— from top to bottom—is linked to the stepwise process of rationally constructing microbial cell factories. The sequence starts with the selection of an adequate *chassis* (microbial host), and continues through DBTLc iterations, fueled by machine learning algorithms. Embedding these features in the setting of a biofoundry could help delivering the next generation of microbial cell factories.

engineered biological hosts. BDA broadened the scope of these efforts, e.g. towards preparing and testing DNA libraries encompassing several thousand building-blocks or standardizing complex genetic architectures. These developments enabled multi-part, high-throughput assembly on a scale that would not have been possible otherwise. Analytical techniques are also available to assay metabolites, proteins and compounds of interest in the engineered hosts [4] carrying natural or synthetic pathways [5] to test the performance of genetic constructs [6]. Finally, results gathered at each of the stages above are processed with machine learning (ML) algorithms deployed to decide on the subsequent DBTLc rounds. DBA strategies (and, more recently, ML tools), complemented by novel SynBio approaches in each DBTL module, are supporting a transition from an artisanal exercise towards a fully standardized, iterative workflow. Such shift will trigger a significant reduction in the costs associated with each operational stage, and will improve productivity, reproducibility and precision [7]. These have been major difficulties hampering a true bioeconomy, whereby goods are sustainably produced with microbial cell factories from renewable feedstocks.

Against this background, in this review major breakthroughs over the last two decades towards bringing the DBTLc to a fully automated routine are discussed, describing key milestones at each segment of the cycle. State-of-the-art studies that incorporate automation and ML methodologies are discussed in the context of metabolic engineering. We conclude by outlining current and future challenges in this ambition, and avenues whereby experimental procedures will become part of fully automated workflows are proposed.

2. Paving the way towards automation in SynBio and metabolic engineering

Most laboratory work is performed manually, with minimal incorporation of automation strategies. Despite the massive expansion of SynBio, metabolic engineering and systems biology, the transition from hand-work to high-throughput and robust automated procedures is still a meandering path. Breakthrough methodologies have been progressively implemented into the DBTLc, contributing to the continuous improvement of biofoundries. Software tools, high-throughput DNA sequencing, omics technologies and ML approaches have pushed the boundaries for automation (Fig. 2). In the sections below, the key steps towards these goals are covered, from *in silico* design to in vivo implementation of genetic circuits in bacterial hosts, and the role of automation is illustrated with recent examples in the SynBio domain.

3. Shaping and exploring metabolic networks in silico

Computational design tools help drafting metabolic pathway designs *de novo* [8]. Repository databases can be used to select and assemble the pathway(s) of interest. Here, the *Kyoto Encyclopedia of Genes and Genomes* (KEGG) is among the first knowledge-bases for systematic analysis of gene functions, cellular processes, chemical compounds and enzymes [9]. The *Braunschweig Enzyme Database* (BRENDA) provides enzyme/-ligand information, and facilitates searches of functional and molecular parameters of enzymatic reactions [10]. MetaCyc joined only two years later as a catalog of metabolic reactions and enzymes in different microorganisms [11]. These databases are continuously updated and refined, making them a first-option in selecting activities and routes to implement in pathway design.

Software packages, designed to harvest information from these databases and to identify feasible designs, contributed to rationally-designed metabolic architectures. *OptKnock* was among the first platforms for gene knock-out strategies towards efficient bioproduction [12], usually by deleting genes encoding competing reactions and through manipulations that couple biomass formation with production [13,14]. Multiple *in silico* constraint-based strain design strategies and algorithms have been developed since [15]. Depending on the native metabolic complexity, identifying and blocking all potential competing routes could be challenging. Nearly complete cut sets have been implemented for metabolic engineering of *Escherichia coli* [16] and *Pseudomonas putida* [17,18]. The combinatorial space to connect an existing metabolic network with a desired product can also be navigated with *RetroPath*. This open-source and modular command line rationalizes pathway choice by exploring all possible connections [19]. Next to this, candidate enzymes can be selected with *Selenzyme*, based on existing databases that evaluate sequence similarity and catalyzed reactions, among other parameters. Selenzyme has been already adopted in some automated biofoundry workflows [20]. Finally, and complementary to these developments, unmapped enzyme sequences for biocatalysis can be obtained from *EnzymeMiner*. This easy-to-use computational tool ranks sequences based on likelihood of catalytic activity and the possibility of producing the corresponding polypeptide as a soluble protein in *E. coli* [21].

After choosing the parts for a given metabolic design, the DNA encoding them has to be drafted and synthesized. *GeneDesigner* is among the first software packages for fast design of synthetic DNA. The addition, edition and blending of structural and regulatory elements (e.g. promoters, open reading frames and DNA parts) is facilitated through an intuitive interface, displaying a hierarchical DNA/protein map. Codon optimization and real-time calculation of oligonucleotide annealing temperatures, sequencing primer generator, inclusion of restriction sites and sequence-identity optimization complete the software features [22]. Some years later, software emerged for using formalized parts in scarless assembly techniques, e.g. *GenoCAD* [23,24]. Standardization has been key to these developments. *BioBricks*, for instance, are a set of reusable,

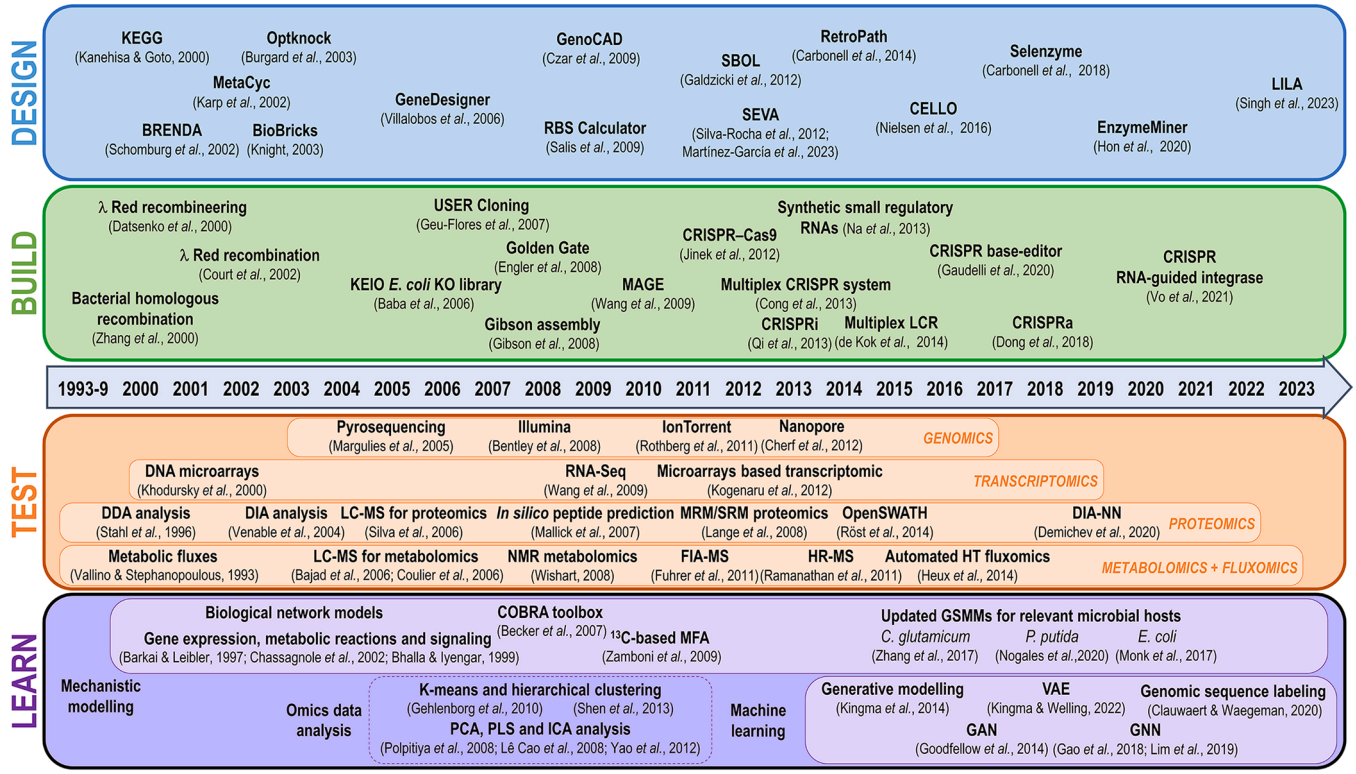

Fig. 2. Timeline showing selected enabling technologies and approaches developed in the design-build-test-learn cycle of Synthetic Biology over the past 30 years (from left to right). The diagram illustrates some key breakthroughs in each stage of the design-build-test-learn cycle (DBTLc): *Design* (blue), *Build* (green), *Test* (orange) and *Learn* (purple). Each methodology is referred to (and explained in detail) in the text. Note that the list of examples is non-exhaustive due to space constraints; abbreviations are provided in the text.

standard DNA parts containing elementary functions that can be combined in compatible vectors in a modular fashion [25]. Similarly, plasmid structures have been standardized through the *Standard European Vector Architecture* (SEVA) that enables exchangeability of multiple DNA modules (e.g. antibiotic selection markers and origins of replication) and constantly updated [26,27]. *RBS Calculator* and other tools can be used in these designs to predict the translation-initiation rate of each *START* codon and to optimize synthetic ribosome binding sites—thereby refining control over protein production [28].

The increased complexity of genetic circuits called for a readily-accessible biological language to represent and visualize *in silico* SynBio design, e.g. the *Synthetic Biology Open Language* (SBOL) [29]. This framework evolved in different versions, culminating in the updated SBOL2.3 [30]. Efforts to integrate more complex functions, blending fully programmable genetic circuits, gave rise to *CELLO* [31], a hardware description language that builds on principles of electronic design. *CELLO* exploits the Verilog language, parsed through algorithms that create circuit diagrams, assign Boolean logic gates, balance constraints to build synthetic DNA and simulate circuit performance. The pipeline was applied to design 60 circuits in *E. coli*, and the cognate DNA fragments (880,000 bp) were built as specified with no additional tuning required. Out of these 60 designs, 45 circuits performed correctly in every output state (up to 10 regulators and 55 independent functional parts), indicating that 92 % of the 412 output states functioned as predicted. An overview of the available *in silico* tools in this section is presented in Fig. 2 (Design).

4. Building biological *chassis* by harnessing advanced SynBio tools

Automated assembly pipelines require efficiency and versatility towards incorporation of novel functions into the host of choice. Toolboxes have been developed during the last years to meet this general criterion. Genome-wide modifications combine fast multi-part DNA assembly techniques and genome engineering methodologies [32,33]. A simple and efficient way to disrupt genes in *E. coli* was developed some 20 years ago by making use of the λ *Red* recombinase functions. PCR products, containing an antibiotic marker and homology regions to the target gene or locus, are genomically integrated by the phage recombinase, interrupting the region to be eliminated [34]. Almost simultaneously, protocols to insert DNA fragments into the *E. coli* chromosome were developed based on homologous recombination [35]. An *E. coli* knock-out library (i.e. the KEIO collection) of strains harboring single deletions of all non-essential genes was created by combining these techniques [36,37]. The KEIO collection comprises *E. coli* mutants for 303 genes, including 37 genes of unknown function. This resource enabled studying loss-of-function phenotypes to a scale never attempted before, providing key information when engineering *E. coli* strains for chemical production. λ *Red*-based recombination was tailored to use oligonucleotides instead of double-stranded DNA, which increased recombination efficiency and broadened the application spectrum [38]; *multiplex automated genome engineering* (MAGE) relies on these efforts [39].

A breakthrough in molecular biology has been the discovery of *clustered regularly interspaced short palindromic repeats* (CRISPR) and associated Cas proteins, repurposed for gene and genome editing protocols [40]. To expand the breadth of these applications, multiplex-editing techniques were implemented to engineer several sites in the eukaryotic genome simultaneously [41]. The same principles were combined with recombineering or homologous recombination to adapt CRISPR/Cas methodologies in prokaryotes [42], lacking non-homologous end joining [43–45]. The number of microbial species that can be genetically accessed with these toolsets continues to increase [46–49], incorporating non-traditional hosts to the list of *chassis* for

metabolic engineering. Along the same line, base-editors were recently developed based on CRISPR/Cas technologies, enabling targeted and precise manipulations at single-base resolution [50–52]. Other tools have been engineered to control gene expression without altering chromosomal sequences. Synthetic small regulatory RNAs, for instance, lower expression by inhibiting translation [53], although multiplexing is to be demonstrated. Similarly, CRISPR interference (CRISPRi) decreases expression by blocking gene transcription [54–56]. CRISPR activation (CRISPRa), in contrast, boosts gene expression [57]. Upgraded CRISPR/Cas protocols are reported virtually on a weekly basis, including strategies for handling increasingly long DNA fragments (CRISPR RNA-guided integrases) [58], and it can be anticipated that these methodologies will become daily laboratory procedures in the near future.

All of the protocols listed here rely on DNA assembly, and cloning strategies have been developed to assemble complex and large genetic constructs. The efficiency of seamless and sequence-independent strategies typically exceeds that of classical restriction/ligation protocols. For example, *uracil-specific excision reagent* (USER) cloning incorporates deoxyuridine into the 5′ prime ends of PCR products, followed by their excision to generate complementary overhangs that facilitate building-up long DNA sequences [59]. *Gibson assembly* was proposed for DNA synthesis in vitro, where overlapping DNA blocks are joined by the combined action of an exonuclease, a DNA polymerase and a ligase in a single isothermal step [60]. This technology was scaled for *de novo* assembly of a synthetic *Mycoplasma genitalium* genome [61,62]. Although *ligase chain reaction* (LCR) was developed in the 1990 s, a high-throughput assembly methodology, *multiplex LCR*, has been optimized to increase the number of DNA constructs that can be assembled [63]. Likewise, *Golden Gate assembly* harnesses type IIs restriction enzymes for joining DNA fragments [64]. Due to its high efficiency and owing to the adoption of modular cloning [65], Golden Gate assembly has become popular as it enables re-using and exchanging DNA parts between research groups. Modularity is particularly relevant for automating the construction of DNA large molecules—as epitomized by the automated assembly of 122 versions of 16 different gene clusters [66]. Fig. 2 (Build) summarizes the main SynBio technologies developed to this end.

5. Omics methodologies as the core of the *Test* stage

Multi-omics methodologies enable quantitative and qualitative analysis of each regulation layer in cellular systems (i.e. genes and genomes, transcripts, proteins, metabolites and metabolic fluxes distribution). A major driver of systems metabolic engineering is combining whole genome sequencing, measurement of cellular metabolite concentrations and identifying (potential) crosstalk between different strata of regulation [67]. Complex Omics approaches have evolved significantly over the last 20 years, with *next generation sequencing* (NGS) technologies playing an important role in genomics [68]. The first DNA sequencing technique (Sanger) based on chain termination [69] was later automated to open the door for commercial sequencing at large scale. High-throughput sequencing was established by the late 1990 s, allowing the sequencing of whole genomes in a very short period. Pyrosequencing empowered sequencing of the whole *M. genitalium* genome [70]. Both NGS throughput and coverage expanded enormously, and the cost per million bp has dropped accordingly [71]. *Ion Torrent* drastically increased NGS accuracy [72]. These sequencing platforms accelerated SynBio developments, especially when implementing novel pathways that rely on long DNA segments. NanoPore, developed by the end of the last century [73], has experienced a technological boost in the last decade that overcomes several shortcomings. Hence, NanoPore offers high-throughput, real-time, long-read and large-scale DNA sequencing [74].

Transcriptomics began concomitantly with the advent of DNA microarrays to investigate changes in gene expression levels [75], and allowed for the study of global changes in mRNA abundances, e.g. *E. coli* under different stresses [76]. Thereafter, RNA sequencing (RNA-Seq) emerged as an approach to deduce and quantify the transcriptome making use of deep-sequencing technologies [77–79]. These methods can be harnessed for quality control of DNA designs, engineered pathways and strains [80], yet continuous mRNA decay can distort quantifications and differential expression transcriptome analyses [81]. High-resolution transcriptomic profiling may include a combination of RNA-Seq and DNA microarrays [82]. Transcriptomes in individual bacteria were recently studied by implementing poly(A)-independent single-cell RNA-sequencing, which faithfully captured growth-dependent expression patterns in *Salmonella* and *Pseudomonas* cells across all RNA classes and genomic regions [83].

Moving from the transcript to the protein level, polypeptide detection and quantification provide a snapshot of the cell functionalities. Several techniques to detect and quantify proteins, starting with the foundational sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) technology [84], have been developed to this end. However, their throughput was not sufficient for the analysis of thousands of proteins until mass spectrometry (MS) was introduced in proteomics [85,86]. High sensitivity and accuracy was attained through *targeted proteomics*, aided by *in silico* prediction of fragments [87], using selected- and multiple-reaction monitoring (SRM/MRM) to detect individual fragments after liquid chromatography (LC) separation [88–91]. This methodology also displays a broad dynamic range, spanning several orders of magnitude, crucial for simultaneous detection of many intracellular proteins. In combination with chemically-produced or concatenated peptides generated from synthetic genes, targeted proteomics enables absolute protein quantification [92,93]. Paired with these efforts, deconvolution of mixture spectra was tailored to improve peptide identification, as a large amount of non-fragmented precursor ions are obtained upon acquiring MS/MS spectra. Data-dependent-analysis (DDA) was among the first approaches for effective spectra acquisition [94], selecting peaks with the highest intensities, followed by fragmentation and analysis of peptides within a specific mass range by tandem MS. Later, the introduction of data-independent-acquisition (DIA) enabled the isolation of a particular m/z window, conferring higher sensitivity and better reproducibility compared to DDA [95]. Recently, OpenSWATH leveraged acquisition power by implementing *Sequential Window Acquisition of All Theoretical Mass Spectra* (SWATH-MS) in an automated and high-throughput fashion [96,97]. The latest (and probably, the most robust) approach exploits deep neural networks combined with DIA [98], enabling deeper and highly-confident coverage when paired with rapid chromatographic methods.

As indicated for transcriptomics, single-cell proteomics emerged as an attractive development, yet it was challenged by limited sensitivity. Traditional proteomics requires the polypeptides from a given cell population to be pooled and analyzed together, hence variations among individual cells in the sample are masked by population-wide effects. To overcome these limitations, novel approaches both with high sensitivity and multiplexing capacity have been proposed for single-cell proteomic analysis [99]. A pioneering study [100] reported on the combined exploration of the single-cell transcriptome and proteome of *E. coli*. Furthermore, these approaches can be further applied in more complex systems, as exemplified by mapping of query datasets on top of a reference proteome atlas [101].

Metabolomics provides another level of essential information about overall physiology, not only as an overview of metabolites present in the cell, but also informing on metabolite accumulation and depletion as a response to genetic and environmental perturbations. Hence, metabolomics aids the identification of potential bottlenecks in metabolic pathways. Arguably, metabolomics flourished with the implementation of high-pressure liquid chromatography (HPLC), which replaced thin-layer chromatography (TLC), and with the switch from ultraviolet and flame-ionization detection to tandem MS during the late 1980 s and

early 1990 s [102]. Even today, key improvements in this field are driven by continuous technological advances in LC and MS technologies, towards faster separation and higher sensitivity, resolution and dynamic detection ranges [103]. A major (and only partially solved) challenge is the fast and efficient quenching of samples needed to detect as many metabolites per sample as possible [104]. Due to the diverse chemical nature and differences in metabolite concentration levels across organisms, measuring the complete metabolome through a single methodology is still difficult. Dedicated methodologies have been developed according to the properties of the metabolites of interest. Workhorse technologies are LC, hydrophilic interaction LC [105], reversed phase ion pairing chromatography [106] and gas chromatography (GC) separation coupled to MS, and numerous special applications complement a detailed picture of the cell metabolome—e.g. nuclear magnetic resonance (NMR) [107] or flow-injection (FIA)-MS [108–110]. Also, metabolomics can be *targeted* and *non-targeted*, each with inherent advantages and disadvantages. While non-targeted metabolomics detects all measurable metabolites in DIA, targeted metabolomics requires the prior selection of analytes of interest (i.e. "you only see what you are looking for"). Thus, non-targeted metabolomics detects more analytes and generates complex datasets, the analysis and interpretation of which are complex and time-consuming. Targeted metabolomics, on the other hand, makes use of SRM and is more sensitive and precise than non-targeted approaches [111]. Targeted methods generate easy-to-interpret datasets and relative and absolute quantification is feasible through the inclusion of internal standards. Recently, technical advances in high resolution tandem MS (HRMS) combined the strength of target and non-targeted metabolomics. Parallel, high-resolution acquisition of full-scan spectra facilitates metabolite discovery, identification and quantification [112,113]. The emergence of big data repositories for metabolite fragmentation spectra along with improved algorithms, e.g. spectral search and *in silico* tools, further facilitated identification in untargeted metabolomics [114–116]. While targeted metabolomics is (in general) a hypothesis-driven framework, untargeted metabolomics can reveal unexpected changes in the metabolism of engineered microorganisms [117]. Considering that much of microbial metabolism remains to be explored, the combination of metabolomic methodologies accelerates the DBTLc not only by highlighting the effect of modifications on the biochemical network, but also by providing fundamental information of the metabolome landscape of the host [118] that could fuel the next set of engineering efforts.

Metabolite abundance is the consequence of fluxes operating in the system. Metabolic fluxes cannot be measured directly, but they can be assessed through changes in metabolite concentrations or by detecting isotope distribution upon feeding isotopic labeled precursors (e.g. ^{13}C-labelled substrates). Fluxomics is based on the same detection methods as metabolomics, but also quantifying *isotopologues* (i.e. molecules sharing the same chemical structure but differing in their isotopic composition). The foundation for fluxomics was laid in the early 1990 s, when the first flux maps were determined based on flux balance analysis (FBA) [119,120], soon complemented by including isotope tracer experiments [121]. For a long time, fluxomics was only performed by a few laboratories on a limited number of biological systems, as it required highly-specific expertise in computational and experimental workflows [122]. Over time, however, publicly-available and easy-to-use software allowed wide access to fluxomics protocols [123–127]. During the last decade, the incorporation of automated, downscaled fluxomics has made high-throughput approaches possible [128–130]. This scenario strongly advocates for fluxomics to become a central, widespread analytical approach to explore cell factory performance within the DBTLc in a routinely fashion. Fig. 2 (Test) covers the pivotal technologies developed in the omics field.

6. Mechanistic modeling and machine learning to integrate Omics data in SynBio

SynBio requires both mechanistic and ML models to learn from omics data, informing the next round of strain engineering in the DBTLc. Mechanistic models represent biological components and their interactions, boosting interpretability, transparency and explainability. ML techniques identify features that differentiate strains and conditions, improving the accuracy of mechanistic models by pinpointing missing or inaccurate components or interactions, and ultimately guiding experimental designs based on data, model topology and simulations. ML models may not be interpretable, transparent or entirely explainable, but they facilitate making sense of omics data and improving biological knowledge. An overview of the major breakthroughs in both mechanistic modeling and ML in the context of SynBio is presented below and in a graphic visualization in Fig. 2 (Learn).

7. Mechanistic modeling and machine learning for SynBio: past, present and the way ahead

Analogous to models in physics and chemistry, the earliest models of biological systems included signaling [131] and gene expression networks [132], as well as metabolic reactions [133]. These interpretations took the form of ordinary and partial differential equations (ODEs and PDEs), and the most comprehensive models included network structure (i.e. topology, defined by how components interact), propensity for components to interact (i.e. thermodynamics) and rates at which they do so (i.e. kinetics). Kinetic models also combine mechanistic details (e.g. allosteric regulation) to provide accurate numerical simulations at both the biological component level and interaction rates; specifics of the kinetic formalism depend on the modeling framework [134]. Thus, kinetic models simulate the dynamics of a biological network and analyze the sensitivity of numerical simulations to model parameters [135]. However, a high level of detail requires significant amounts of (experimentally-curated) data for model parameterization, e.g. enzyme kinetics, enzyme, substrate and product concentrations, and thermodynamic information. Generating the data necessary to parameterize a kinetic model is technically challenging and cost-prohibitive [136,137]. Also, the computational and algorithmic complexity of parameterizing and simulating kinetic models grows with model size, relegating the scope of kinetic modelling to individual or just a few pathways. Several approaches have been explored to overcome these challenges. Expanding the model size using simplified reaction mechanisms allowed to identify computationally a minimal set of reactions capable to support bacterial growth and reactions to be modulated or knocked-out to overproduce a product of interest [138–140]. A different approach reduced the number of reactions while maintaining the scope of the model through model reduction [141–144]. Regardless of the method, the lack of available data and basic understanding of enzyme properties remain major bottlenecks towards full parametrization. In addition, most approaches implemented thus far result in context-specific models that may be difficult to extrapolate to other operational conditions.

Constraint-based analysis overcomes the challenges of scale and data availability through the assumption that the system is at steady state. The problem then morphs from an ODE or PDE system to a set of linear equations that can be solved using optimization techniques, e.g. linear programming. Genome-scale metabolic models (GSMM) are reconstructed based on the genome sequence of an organism [145,146] to calculate reaction and pathway fluxes based on mass balance (i.e. FBA) and enzyme capacity constraints [147,148]. Constraint-based reconstruction and analysis (COBRA) became an essential tool to compute optimal operativity of biological components given a set of inputs and network topology [149]. Additional constraints, based on thermodynamic and concentration data (i.e. thermodynamics-based FBA, *t*FBA), can be incorporated to refine flux allocation across the network [150]. A

plethora of constraint-based methods have been developed, including, just to name a few examples, flux variability analysis (FVA), parsimonious flux balance analysis (*p*FBA) and the previously-mentioned OptKnock [151,152]. The expanding palette of approaches accurately estimate ranges of steady-state reaction fluxes and compute optimal enzyme capacities to improve titers and yields.

GSMMs are limited to reactions within metabolism, and tend to ignore material and energy costs required for gene expression and protein synthesis, particularly relevant for engineering efforts. To fill this gap, metabolic and expression models have been formulated for well-studied organisms, e.g. *E. coli* [153,154] and *Corynebacterium glutamicum* [155], and approaches in this direction can be expected for alternative bioproduction *chassis*, e.g. *P. putida* [156–160]. A milestone in metabolic and expression model formalism is the direct inclusion of enzyme production costs in the overall mass balance. This feature mediates better predictions of pathway utilization, based on the costs of synthesizing the enzymes therein. Unfortunately, such metabolic and expression models are computationally challenging to solve, and require additional parameters (e.g. RNA synthesis and degradation rates), some of which are yet to be experimentally determined.

As indicated above, a shortcoming of constraint-based analysis is the inaccuracy of network-wide flux predictions. Metabolic flux analysis attempts to solve this problem by deriving fluxes from mass distribution vectors (MDVs), generated from isotope labeling experiments and assessed by LC-MS/MS or GC-MS [161,162]. Algorithm innovations, e.g. *elementary metabolite units* (EMU) [123], confidence interval estimation without exhaustive sampling [163], time-course data [164] and parallel-labeling experiments [165], rendered MFA computationally efficient for both ^{13}C- and ^{15}N-labeling tracer experiments. Thus, MFA can guide engineering strategies to divert fluxes into the desired products. Recent efforts expanded the size and scope of MFA to the genome-scale [148], e.g. by improving analytics to measure MDVs in reactions beyond central metabolism [166], model reduction strategies [144] and algorithms for genome-wide atom transfer prediction [167, 168]. When combined, these developments improved the resolution and scope of fluxes calculated using the current MFA toolbox [169–173].

In parallel to mechanistic modeling, ML assimilated solutions to challenges in computer vision, natural language processing (NLP) and board and video games solved by deep learning (DL). Unlike *linear classifiers*, *support vector machines* (SVMs) or *decision trees*, DL has been shown to serve as an universal function *approximator* [174]—stacking layers of neural networks on top of one another to gain abstraction and representational power as a function of network depth. DL fell out of favor with the ML community in the late 1990 s, surpassing traditional algorithms across all computer vision, NLP and game benchmarks only when a combination of general (e.g. back propagation [175]) and domain-specific algorithmic innovations (along with computer hardware improvements) was implemented to this end. Some of these innovations included convolution networks [176], image augmentation [177], attention [178] and NLP pre-training techniques [179]. In the biological domain, genomic and proteomic sequences are analogous to sequences of letters and therefore amenable to many of the NLP algorithms. Important examples of this sort include sequence labeling in genomics [180,181], sequence-to-feature [182] and sequence-to-structure predictions in proteomics [183]. A major milestone has been the direct prediction of protein structures from sequences, as AlphaFold2 [184] won the Critical Assessment of Structure Prediction Challenge (CASP14) by a large margin [185]. Similar bioinformatics tasks, involving inputs that can be modeled as those in computer vision and NLP, will surely see comparable improvements in the future.

The heterogeneity of omics data and the high degree of correlations between features render these datasets difficult for ML approaches. Unsupervised learning techniques that arrange and reduce dimensionality, e.g. *K*-means clustering [186], hierarchical clustering [187], principal component analysis (PCA) [188], independent component analysis (ICA) [189] and partial least squares (PLS) analysis [190], are workhorses for inferring class membership and omics dataset features. Classification algorithms, for instance, use transformed and reduced features as inputs to improve performance [191]. Alternatively, some omic data (e.g. metabolite concentrations) can be used to constrain mechanistic models and the simulation output (e.g. fluxes) can then be fed as inputs for classification algorithms [192]. Likewise, generative modeling [193] provides a framework for combining unsupervised dimensionality reduction with DL to infuse the benefits of abstraction and representation power (characteristic of DL) into unsupervised classification and feature importance identification [194]. Deep generative models, e.g. variational autoencoder (VAE, [195]) and generative adversarial network (GAN) [196], map data to probability distributions—and vice versa. The mapping process is analogous to a non-linear PCA, where high dimensional inputs are encoded to low dimensional latent spaces that can then be decoded to reconstruct the inputs. The latent space can be parameterized to capture both categorical and continuous data factors [197]. In addition, latent spaces from different datasets can be combined using latent arithmetic to generate data points not seen in the training datasets [198]. Deep generative modeling using VAEs impacted single-cell RNA sequencing, with various studies demonstrating how experimental noise and batch effects can be corrected within and across experiments [199,200]. Applications of clustering and feature identification have seen recent advances in the associated algorithms [201], indicating that the use of generative modeling in unsupervised learning has much to offer for omic analysis in the SynBio community.

Besides computer vision, NLP and games, DL enabled breakthroughs on ML tasks involving graphs. *Graph neural networks* (GNNs) operate directly on the graph structure through message passing, whereby node and relation attributes are propagated to their nearest neighbors for a selected number of iterations. Thus, it is possible to generate contextualized features for node and graph classification, as well as link prediction [113]. Two application domains with SynBio analogs include generating improved compound representations for drug discovery and product-recommendation systems. In the former, GNNs learn compound representations and similarity metrics between molecules from training data to yield more accurate property predictions [202] and to simulate chemical structures with desired properties [203]. In the latter, GNNs take advantage of the network formed between a given user's purchasing history and other customers purchasing history, to make relevant recommendations of potential products [204]. In both cases, previous knowledge (e.g. chemical structure in drug development and past purchasing histories in recommendation systems) is contextualized with current data. Features of omic datasets are related through rich networks. Gene and metabolite set enrichment analyses exploit the hierarchy and relations between genes, metabolites and biological processes to infer significance at the pathway level [205,206]. GNNs are suited to take advantage of biological network knowledge when analyzing omics data, and preliminary efforts indicate that this will be an actively explored research topic in biology in the future [207].

The examples above advocate a role for learning approaches in the DBTLc, an iterative process by nature. The choice of variables that should be changed in strain engineering is a non-trivial task, involving balancing *exploitation* (i.e. optimizing towards the best producing strain) and *exploration* (i.e. testing a diverse range of variables to gain better understanding of the system). These efforts involve a search space that cannot be exhaustively navigated. ML approaches can inform the design of experiments at each DBTLc iteration by learning the correct balance between exploitation and exploration and suggesting a recommended subset of variables to test experimentally. The importance of active learning has been demonstrated in synthetic chemistry [208], cell-free systems [209] and pathway engineering in yeast [210], where a combination of heuristics, mechanistic modeling and ML successfully fueled data-driven design of experiments.

8. Automation in the DBTLc for bacterial cell factory design

To deliver the SynBio promise of supporting a true bioeconomy, efficient microbial cell factories are required in high demand to supply market needs [211,212]. Replacing fossil fuels-related products currently in use by biologically produced counterparts and developing new-to-Nature chemicals is a driving force that pushes both companies and research centers to create novel technologies. High-quality, fast construction of cell factories calls for incorporating automation platforms to create phenotypes of interest in an effective and reliable way. Automated pipelines can be designed to increase throughput, reducing technical variability and improving data quality to this end. Such workflows find applications at each segment of the DBTLc—also connecting each cycle quadrant in the context of biofoundries [213], where pipelines of this sort will enable processing and analyzing hundreds to thousands of engineered strains. An integrated biofoundry requires operational flexibility, with easily reconfigurable systems to adapt to different biological systems while reducing human intervention [214]. In practical terms, this process involves deploying a robotic station equipped with liquid-handling mechanical arms to speed up workflows. State-of-the-art liquid handling devices are now able to pipette volumes within the micro- to milli-liter scale while providing the versatility required to build cell factories [215]. In the next section, the latest examples where automation was successfully incorporated into DBTLc workflows are discussed.

9. Latest development in automated pipelines for cell factory construction and testing

While some automated pipelines have only focused on an individual step within each DBTLc quadrant, others covered many steps simultaneously (Table 1). A recent study [216] illustrates the robustness of liquid-handling robots for high-throughput experiments. A flexible, open-source Python framework, PyHamilton, integrated complex liquid transfer patterns and systematized conventional laboratory procedures. The automated workflow was implemented to track up to 480 individual bacterial cultures — analyzing metabolic fitness landscapes across 100 different conditions — towards optimizing recombinant protein production. Another groundbreaking study showed how an integrated DBTLc pipeline can be adopted for optimizing bioproduction, as well as discovering novel metabolic pathway configurations. Here, (2 *S*)-pinocembrin (5,7-dihydroxyflavone) production by engineered *E. coli* strains was systematically optimized, reaching a flavonoid titer up to 88 mg L^{-1} after screening 65 variants out of 23,328 possible metabolic designs [217]. The overall approach entailed *in silico* selection of promising enzyme candidates, pathway assembly (aided by robotics and supported by a statistical method), rapid testing of product titers and cycle iteration through computational tools and laboratory automation (Fig. 3). Amyris Inc. recently described LILA, an automated scientist to handle all design and optimization steps of the DBTLc. LILA generates metabolic routes, identifies genetic elements for perturbation and informs (re-)design of microbial strains in a matter of seconds to minutes. Strains specified by LILA were built and phenotyped in a semi-automated in-house pipeline to yield the highest published titers for naringenin [218].

Rapid prototyping of engineered *chassis* in a semi-automated biomanufacturing process exploited a robotic platform to produce 17 building blocks over 85 days [219]. A timed "pressure test" was reported [220], whereby 3 months were allocated to engineer 215 microbial cell factories in a biofoundry in five species (i.e. *Saccharomyces cerevisiae*, *E. coli*, *Streptomyces albidoflavus*, *S. coelicolor* and *S. albovinaceus*) to produce 10 molecules. Likewise, semi-automated pipelines screened monoterpenes synthase libraries to identify best candidate variants, supported by robotic liquid handling paired with GC-MS analysis and automated data extraction [221]. Merging an integrated robotic system and ML algorithms enabled optimization of lycopene production by engineered *E. coli*. In this case, gene expression within the carotenoid biosynthesis pathway was tuned with optimization routines (paired predictive model and Bayesian algorithms) that resulted in high lycopene levels—while reducing the number of constructs to be evaluated to < 1 % of all 13,824 combinatorial possibilities. A pathway variant, leading to enhanced (1.7-fold) lycopene production increase was isolated with this method [222]. In an elegant approach to cell-free bioproduction, an active learning strategy was applied to explore a broad combinatorial space of ca. 4,000,000 buffer compositions to maximize protein production [223]. Here, the authors merged *exploitation* (i.e. buffer combinations with a low prediction accuracy) and *exploration* (i.e. buffer combinations predicted to maximize protein yields) to improve the output and decrease the model uncertainty. A big data collection was used to train an ML algorithm, achieving high quality prediction and improving protein production by 34-fold with a low-cost, home-made lysate.

10. Current DBTLc bottlenecks

A number of bottlenecks need to be solved in automated SynBio pipelines (Fig. 4), associated with limitations in robotic equipment that restrict task performance. Regardless of technical limitations, these technologies are still rather expensive, and costs involved in automating an entire laboratory are simply not affordable for many institutes or even private companies [224]. *In silico* pathway design, gene part selection, protein and enzymes engineering and *de novo* design of catalytic activities are individual DBTLc steps that suffer multiple limitations. Navigating the large catalog of genes and enzymes for designing a metabolic pathway can be a daunting task. Next, identifying suitable pathways is challenging as bacterial metabolism is inherently complex; crosstalk between routes (both native and engineered) is particularly difficult to predict [225]. Furthermore, the repertoire of hosts that can be used to create microbial cell factories is relatively limited, with < 10 conventional organisms widely adopted for such purpose [226]. Likewise, more efficient and standardized DNA assembly techniques are required to unveil context dependency, another typical SynBio problem [227,228]. Toxicity of final product(s) or intermediates generated during bioconversion is another major barrier in microbial engineering [229]. The engineered phenotypes must also be stable over time to permit scaling-up [230,231].

Once the cell factory is ready for testing, sample preparation and extraction methods come into place, and they are difficult to automate [232]. Moreover, some steps, e.g. culture inoculation, PCR amplification, plasmid transformation, replica-plating, plasmid curation, sample centrifugation, filtration and cell lysis are usually done off-line and still require human intervention to different extents. Furthermore, as different companies are developing proprietary technologies, robotics parts are not interchangeable or adaptable to other devices, which hampers their integration into other workflows [233]. Data extraction and interpretation can be equally challenging, due to the vast amount of data generated in a typical high-throughput experiment [234]. Analytical tools and experimental designs used for specific omic disciplines often lack versatility for integration across multiple omics layers [5].

Finally, a major challenge is our inability to accurately predict phenotypes from first principles (e.g. DNA modifications), together using small-scale experiments to forecast the behavior of cell factories at a larger scale [235]. Metabolic reconstructions and gene expression models, deployed to infer complex phenotypes, are both computationally demanding and call for the incorporation of additional parameters (e.g. RNA synthesis and degradation rates), which are difficult to obtain experimentally [236]. Linked to this effort, high-power computation is mandatory to model and predict the next-generation of microbial cell factories [237,238].

Table 1
Recent examples of automated *Design-Build-Test-Learn* workflows.

Host or system	Objective	Relevant features[a] at phase:				Reference
		Design	*Build*	*Test*	*Learn*	
E. coli	Systematic optimization of protein production[b] (GFP and RFP)	PyHamilton framework (Hamilton MicroLab STARlet 8-channel base model)	Transformation and inoculation	Media preparation Media-dispensing and dilution * Cultivation *	Feedback controller algorithm	[216]
E. coli	Mandelic and hydroxymandelic acid production	128 enzymes selected from 88 species encoding 50 different targets* 111 new gene parts, along with 25 parts and 18 plasmid backbones already in house * *In silico* tools: Retropath, RetroRules, Reaxys, Selenzyme, PartGenie, RBS calculator and PlasmidGenie	λ Red recombineering CRISPR technologies In-Fusion cloning Robot-assisted ligase cycling reaction * Transformation, replica plating and plasmid curing	Minion next-generation sequencing Cultivation * Enzymatic assays and pathway screening on a robot station * LC-TripleQuad–LC-MS/MS analysis LC-IMS QToF–LC-MS analysis GC-QToF–GC-MS analysis	DoE analysis Ordinary least square contrast regression analysis	[219]
E. coli	Monoterpenoid production	Not indicated	Megaprimer PCR In-Fusion cloning Transformation	Media-dispensing and colony picking ** Growth, induction and incubation * Sanger sequencing * GC-QTOF analysis *	Data analysis aided by machine learning (neural networks) *	[221]
E. coli	Lycopene production	Not indicated	Promoter mutagenesis Golden Gate assembly * Transformation *	Cell cultivation, colony picking and lycopene extraction * Colorimetric quantification of products	Machine learning (Bayesian optimization and Gaussian process) *	[222]
E. coli	Dodecanol production	Combination and modulation of three acyl-CoA/acyl-ACP reductases. *In silico* tools: J5, DeviceEditor, bioCAD, RBS calculator	DNA purification assisted by NIMBUS size selection robot ** Gibson assembly Golden Gate assembly reaction	MiSeq sequencing BioLector microbioreactor GC-MS analysis HPLC LC-MS/MS-QQQ analysis	ML regression approaches: random forest, polynomial, multilayer perceptron, and the TPOT meta-learner. ML to improve prediction: Ensemble Model Partial correlation analysis to evaluate RBS calculation	[250]
Cell-free system	Protein production[b] (GFP)	Not indicated	Golden Gate assembly Transformation	Cultivation, harvesting, lysate preparation, protein purification Cell-free reaction combinations *	Machine learning models *	[223]

[a] Automated steps are highlighted in bold, indicating whether they are either fully automated (*) or semi-automated (** , requiring human intervention). Abbreviations: GFP, green fluorescent protein; RFP, red fluorescent protein; DoE, design of experiments; LC, liquid chromatography; MS, mass spectrometry; IMS, ion mobility spectrometry; GC, gas chromatography; QToF, quadrupole time-of-flight.
[b] Green and red fluorescent proteins were used as model proteins for the optimization process.

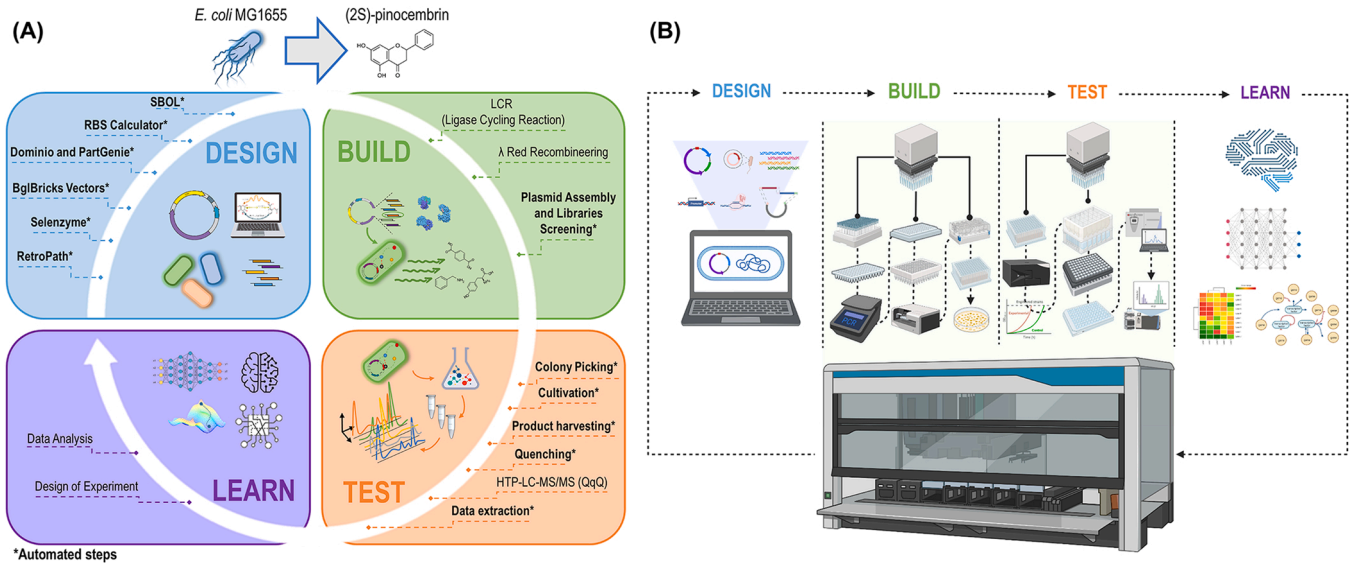

Fig. 3. (A) Design-build-test-learn cycle (DBTLc) workflow illustrated for the production of (2 *S*)-pinocembrin in engineered strains of *Escherichia coli* MG1655. Each quadrant of the cycle lists tools applied to build cell factories tailored for the production of flavanone. (B) "Linearization" of the DBTLc pipeline applied to the example of panel (A). The workflow started with *in silico* design of cell factories (based on *E. coli* as the *chassis*), including factorial pathway assembly. Next, the *in silico*-designed gene circuits were assembled by a robotic platform that also carried out the amplification, purification and transformation of DNA parts into the host. This operation was followed by cultivation of the engineered bacteria, sampling and processing of the samples for LC-MS/HPLC analysis. The last step integrated the massive amount of data generated, together with training routines to predict how a new model will behave making use of statistical analysis and machine learning (ML) algorithms. These activities concluded the first iteration of the DBTLc, paving the way for the next round.

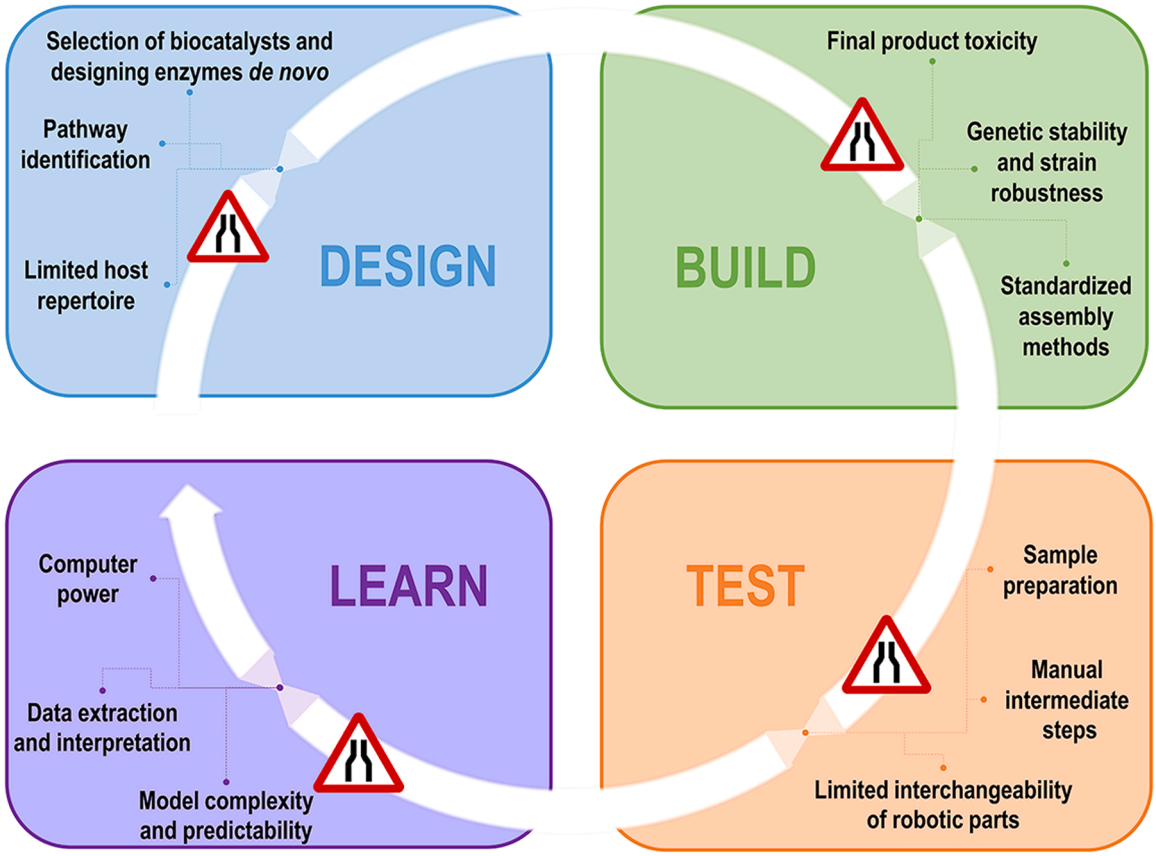

Fig. 4. Some of the current bottlenecks to be addressed in the classic design-build-test-learn cycle (DBTLc) towards the development of next-generation microbial cell factories.

11. Discussion and outlook

Automation of the DBTLc is bringing about a transition in the SynBio field. With novel technologies that allow generation of robust cell factories, incorporating automated steps into everyday laboratory work is no longer a dream. These developments have been triggered by the exponential increase of data that can be extracted from a single experiment. Yet, the limited throughput in data handling and interpretation calls for methodologies that can accelerate the process. ML provides the required prediction power to achieve this goal [239]. Automated, high-throughput approaches to generate reliable data and ML algorithms should be merged for rational design of cell factories endowed with a desired phenotype [240]. These developments can be expected over the next 10 years, as SynBio is blending mechanistic modeling and systems-level thinking to engineer biology. The most recent literature is brimming over with examples of mechanistic modeling and ML for omic analysis and experimental design, and a crossover of recent ML developments in computer vision, NLP, graphs and active learning into SynBio can be anticipated to support these efforts.

The compendium of biological components and databases describing their potential interactions is rather incomplete. Knowledge gaps in understanding of biology limit the ability to engineer living cells, and learning causality from data is essential to bridge these gaps. Identifying correlations from (noisy and context-specific) experimental data [241], or 'brute-force' search strategies (not scalable to large networks) [242], are partial solutions in this direction. The work of Pearl [243] and other pioneers in the field have shaped the modern DL framework [244]. Yet, the detail and scope of *in silico* design and simulation are also limited. Whole-cell, multi-level models are available for a few model organisms [245], laying the path towards modelling cell dynamics over multiple regulation layers. However, models for multicellular entities and microbial communities that include biochemical dynamics (i.e. kinetics beyond steady-state conditions) are virtually absent. Furthermore, almost all models used in the SynBio community assume constant cell volume, often ignoring the spatial organization of cells and tissues, and neglect other physical, chemical or electrical phenomena. Recent developments in graph networks (GNs) [246] and physics-informed neural networks (PINNs) [247] provide avenues of exploration, previously applied to physics [248] or very simple biological systems. Both GNs and PINNs blend mechanistic details for interpretability and explainability while integrating the scalability and scope of DL. Further work, integrating mechanistic modeling and DL, will enable more accurate designs and simulation in SynBio.

We envision a fully automated DBTLc, characterized by high-throughput and iterative workflows, paving the way to the long-standing ambition of SynBio to program phenotypes of interest from first principles. Self-driving labs, combining fully-automated experiments with AI to decide on the next set of experiments, may become a new paradigm of scientific research, as recently proposed [249]. The rational combination of individual approaches, as presented in this review, will facilitate these developments, providing, at the same time, valuable fundamental information on biological systems to fuel engineering efforts.

Declaration of interests

The authors declare that there are no competing interests associated with the contents of this article.

Acknowledgements

The authors would like to acknowledge the work by many researchers in the field of Synthetic Biology and Metabolic Engineering who have made authoritative contributions to the automation of the DBTLc, the work of whom could not always be cited here because of space reasons. The work at the Systems Environmental Microbiology group laboratory is supported by grants from the Novo Nordisk Foundation [NNF20CC0035580, *LiFe* (NNF18OC0034818) and *TARGET* (NNF21OC0067996)], the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 814418 (*SinFonia*) and the Cystic Fibrosis Trust, Strategic Research Centre Award–2019–SRC017 to P.I.N.

References

[1] Appleton E, Densmore D, Madsen C, Roehner N. Needs and opportunities in bio-design automation: four areas for focus. Curr Opin Chem Biol 2017;40:111–8. https://doi.org/10.1016/j.cbpa.2017.08.005.
[2] Martinelli L, Nikel PI. Breaking the state-of-the-art in the chemical industry with new-to-Nature products via synthetic microbiology. Microb Biotechnol 2019;12:187–90. https://doi.org/10.1111/1751-7915.13372.
[3] Densmore DM, Bhatia S. Bio-design automation: software + biology + robots. Trends Biotechnol 2014;32:111–3. https://doi.org/10.1016/j.tibtech.2013.10.005.
[4] Calero P, Nikel PI. Chasing bacterial *chassis* for metabolic engineering: a perspective review from classical to non-traditional microorganisms. Microb Biotechnol 2019;12:98–124. https://doi.org/10.1111/1751-7915.13292.
[5] Petzold CJ, Chan LJ, Nhan M, Adams PD. Analytics for metabolic engineering. Front Bioeng Biotechnol 2015;3:135. https://doi.org/10.3389/fbioe.2015.00135.
[6] Appleton E, Madsen C, Roehner N, Densmore D. Design automation in synthetic biology. Cold Spring Harb Persp Biol 2017;9:a023978. https://doi.org/10.1101/cshperspect.a023978.
[7] Carbonell P, Radivojevic T, García Martín H. Opportunities at the intersection of synthetic biology, machine learning, and automation. ACS Synth Biol 2019;8:1474–7. https://doi.org/10.1021/acssynbio.8b00540.
[8] Wang L, Dash S, Ng CY, Maranas CD. A review of computational tools for design and reconstruction of metabolic pathways. Synth Syst Biotechnol 2017;2:243–52. https://doi.org/10.1016/j.synbio.2017.11.002.
[9] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.
[10] Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. Nucleic Acids Res 2002;30:47–9. https://doi.org/10.1093/nar/30.1.47.
[11] Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The MetaCyc database. Nucleic Acids Res 2002;30:59–61. https://doi.org/10.1093/nar/30.1.59.
[12] Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 2003;84:647–57. https://doi.org/10.1002/bit.10803.
[13] Orsi E, Claassens NJ, Nikel PI, Lindner SN. Growth-coupled selection of synthetic modules to accelerate cell factory development. Nat Commun 2021;12:5295. https://doi.org/10.1038/s41467-021-25665-6.
[14] Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat Chem Biol 2011;7:445–52. https://doi.org/10.1038/nchembio.580.
[15] Maia P, Rocha M, Rocha I. *In silico* constraint-based strain optimization methods: the quest for optimal cell factories. Microbiol Mol Biol Rev 2016;80:45–67. https://doi.org/10.1128/MMBR.00014-15.
[16] Harder B-J, Bettenbrock K, Klamt S. Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. Metab Eng 2016;38:29–37. https://doi.org/10.1016/j.ymben.2016.05.008.
[17] Banerjee D, Eng T, Lau AK, Sasaki Y, Wang B, Chen Y, et al. Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. Nat Commun 2020;11:5385. https://doi.org/10.1038/s41467-020-19171-4.
[18] Kozaeva E, Volkova S, Matos MRA, Mezzina MP, Wulff T, Volke DC, et al. Model-guided dynamic control of essential metabolic nodes boosts acetyl-coenzyme A–dependent bioproduction in rewired *Pseudomonas putida*. Metab Eng 2021;67:373–86. https://doi.org/10.1016/j.ymben.2021.07.014.
[19] Carbonell P, Parutto P, Baudier C, Junot C, Faulon JL. RetroPath: automated pipeline for embedded metabolic circuits. ACS Synth Biol 2014;3:565–77. https://doi.org/10.1021/sb4001273.
[20] Carbonell P, Wong J, Swainston N, Takano E, Turner NJ, Scrutton NS, et al. *Selenzyme*: enzyme selection tool for pathway design. Bioinformatics 2018;34:2153–4. https://doi.org/10.1093/bioinformatics/bty065.
[21] Hon J, Borko S, Stourac J, Prokop Z, Zendulka J, Bednar D, et al. *EnzymeMiner*: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities. Nucleic Acids Res 2020;48:W104–9. https://doi.org/10.1093/nar/gkaa372.
[22] Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S. Gene designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinforma 2006;7:285. https://doi.org/10.1186/1471-2105-7-285.
[23] Czar MJ, Cai Y, Peccoud J, Writing DNA. with GenoCAD™. Nucleic Acids Res 2009;37:W40–7. https://doi.org/10.1093/nar/gkp361.
[24] Hillson N, Caddick M, Cai Y, Carrasco JA, Chang MW, Curach NC, et al. Building a global alliance of biofoundries. Nat Commun 2019;10:2040. https://doi.org/10.1038/s41467-019-10079-2.

[25] Knight T. Idempotent vector design for standard assembly of BioBricks. MIT Libraries. 2003. DSpace@MIT:hdl.handle.net/1721.1721/21168.
[26] Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, de las Heras A, et al. The standard european vector architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res 2013;41:D666–75. https://doi.org/10.1093/nar/gks1119.
[27] Martínez-García E, Fraile S, Algar E, Aparicio T, Velázquez E, Calles B, 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 2023;51:D1558–67. https://doi.org/10.1093/nar/gkac1059.
[28] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009;27:946–50. https://doi.org/10.1038/nbt.1568.
[29] Galdzicki M., Wilson M., Rodríguez C.A., Pocock M.R., Oberortner E., Adam L., et al. Synthetic biology open language (SBOL) version 1.1. 0. Technical Report 2012; BioBricks Foundation.
[30] Madsen C, Goñi-Moreno A, Umesh P, Palchick Z, Roehner N, Atallah C, et al. Synthetic biology open language (SBOL) version 2.3. J Integr Bioinform 2019;16:20190025. https://doi.org/10.1515/jib-2019-0025.
[31] Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. Science 2016;352:aac7341. https://doi.org/10.1126/science.aac7341.
[32] Ko YS, Kim JW, Lee JA, Han T, Kim GB, Park JE, et al. Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. Chem Soc Rev 2020;49:4615–36. https://doi.org/10.1039/D0CS00155D.
[33] Luo ZW, Lee SY. Metabolic engineering of *Escherichia coli* for the production of benzoic acid from glucose. Metab Eng 2020;62:298–311. https://doi.org/10.1016/j.ymben.2020.10.002.
[34] Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA 2000;97:6640–5. https://doi.org/10.1073/pnas.120163297.
[35] Zhang Y, Muyrers JP, Testa G, Stewart AF. DNA cloning by homologous recombination in *Escherichia coli*. Nat Biotechnol 2000;18:1314–7. https://doi.org/10.1038/82449.
[36] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the KEIO collection. Mol Syst Biol 2006;2. https://doi.org/10.1038/msb4100050.
[37] Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, et al. Update on the KEIO collection of *Escherichia coli* single-gene deletion mutants. Mol Syst Biol 2009;5:335. https://doi.org/10.1038/msb.2009.92.
[38] Court DL, Sawitzke JA, Thomason LC. Genetic engineering using homologous recombination. Annu Rev Genet 2002;36:361–88. https://doi.org/10.1146/annurev.genet.36.061102.093104.
[39] Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature 2009;460:894–8. https://doi.org/10.1038/nature08187.
[40] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012;337:816–21. https://doi.org/10.1126/science.1225829.
[41] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819–23. https://doi.org/10.1126/science.1231143.
[42] Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, Schweizer HP. A broad-host-range Flp-*FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. Gene 1998;212:77–86. https://doi.org/10.1016/s0378-1119(98)00130-9.
[43] Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat Biotechnol 2017;35:48–55. https://doi.org/10.1038/nbt.3718.
[44] Pyne ME, Moo-Young M, Chung DA, Chou CP. Coupling the CRISPR/Cas9 system with Lambda Red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. Appl Environ Microbiol 2015;81:5103–14. https://doi.org/10.1128/aem.01248-15.
[45] Ronda C, Pedersen LE, Sommer MOA, Nielsen AT. *CRMAGE*: CRISPR optimized MAGE recombineering. Sci Rep 2016;6:19452. https://doi.org/10.1038/srep19452.
[46] Blombach B, Grünberger A, Centler F, Wierckx N, Schmid J. Exploiting unconventional prokaryotic hosts for industrial biotechnology. Trends Biotechnol 2021;40:385–97. https://doi.org/10.1016/j.tibtech.2021.08.003.
[47] Wirth NT, Kozaeva E, Nikel PI. Accelerated genome engineering of *Pseudomonas putida* by I-*Sce*I–mediated recombination and CRISPR-Cas9 counterselection. Microb Biotechnol 2020;13:233–49. https://doi.org/10.1111/1751-7915.13396.
[48] Chen Y, Banerjee D, Mukhopadhyay A, Petzold CJ. Systems and synthetic biology tools for advanced bioproduction hosts. Curr Opin Biotechnol 2020;64:101–9. https://doi.org/10.1016/j.copbio.2019.12.007.
[49] Volke DC, Friis L, Wirth NT, Turlin J, Nikel PI. Synthetic control of plasmid replication enables target- and self-curing of vectors and expedites genome engineering of *Pseudomonas putida*. Metab Eng Commun 2020;10:e00126. https://doi.org/10.1016/j.mec.2020.e00126.
[50] Gaudelli NM, Lam DK, Rees HA, Sola-Esteves NM, Barrera LA, Born DA, et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nat Biotechnol 2020;38:892–900. https://doi.org/10.1038/s41587-020-0491-6.
[51] Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 2016;533:420–4. https://doi.org/10.1038/nature17946.
[52] Volke DC, Martino RA, Kozaeva E, Smania AM, Nikel PI. Modular (de) construction of complex bacterial phenotypes by CRISPR/nCas9-assisted, multiplex cytidine base-editing. Nat Commun 2022;13:3026. https://doi.org/10.1038/s41467-022-30780-z.
[53] Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol 2013;31:170–4. https://doi.org/10.1038/nbt.2461.
[54] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 2013;152:1173–83. https://doi.org/10.1016/j.cell.2013.02.022.
[55] Batianis C, Kozaeva E, Damalas SG, Martín-Pascual M, Volke DC, Nikel PI, et al. An expanded CRISPRi toolbox for tunable control of gene expression in *Pseudomonas putida*. Microb Biotechnol 2020;13:368–85. https://doi.org/10.1111/1751-7915.13533.
[56] Jakočiūnas T, Jensen MK, Keasling JD. System-level perturbations of cell metabolism using CRISPR/Cas9. Curr Opin Biotechnol 2017;46:134–40. https://doi.org/10.1016/j.copbio.2017.03.014.
[57] Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. Nat Commun 2018;9:2489. https://doi.org/10.1038/s41467-018-04901-6.
[58] Vo PLH, Ronda C, Klompe SE, Chen EE, Acree C, Wang HH, et al. CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. Nat Biotechnol 2021;39:480–9. https://doi.org/10.1038/s41587-020-00745-y.
[59] Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA. *USER FUSION*: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res 2007;35:e55. https://doi.org/10.1093/nar/gkm106.
[60] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5. https://doi.org/10.1038/nmeth.1318.
[61] Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science 2008;319:1215–20. https://doi.org/10.1126/science.1151721.
[62] Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci USA 2008;105:20404–9. https://doi.org/10.1073/pnas.0811011106.
[63] de Kok S, Stanton LH, Slaby T, Durot M, Holmes VF, Patel KG, et al. Rapid and reliable DNA assembly via ligase cycling reaction. ACS Synth Biol 2014;3:97–106. https://doi.org/10.1021/sb4001992.
[64] Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One 2008;3:e3647. https://doi.org/10.1371/journal.pone.0003647.
[65] Casini A, Storch M, Baldwin GS, Ellis T. Bricks and blueprints: methods and standards for DNA assembly. Nat Rev Mol Cell Biol 2015;16:568–76. https://doi.org/10.1038/nrm4014.
[66] Smanski MJ, Bhatia S, Zhao D, Park YJ, Woodruff LBA, Giannoukos G, et al. Functional optimization of gene clusters by combinatorial design and assembly. Nat Biotechnol 2014;32:1241–9. https://doi.org/10.1038/nbt.3063.
[67] Becker J, Wittmann C. From systems biology to metabolically engineered cells–an omics perspective on the development of industrial microbes. Curr Opin Microbiol 2018;45:180–8. https://doi.org/10.1016/j.mib.2018.06.001.
[68] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17:333–51. https://doi.org/10.1038/nrg.2016.49.
[69] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977;74:5463–7. https://doi.org/10.1073/pnas.74.12.5463.
[70] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437:376–80. https://doi.org/10.1038/nature03959.
[71] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53–9. https://doi.org/10.1038/nature07517.
[72] Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011;475:348–52. https://doi.org/10.1038/nature10242.
[73] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci USA 1996;93:13770–3. https://doi.org/10.1073/pnas.93.24.13770.
[74] Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. Nat Biotechnol 2012;30:344–8. https://doi.org/10.1038/nbt.2147.
[75] Floyd ET, DeLeo JM, Thompson EB. Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated RNAs. DNA 1983;2:309–27. https://doi.org/10.1089/dna.1983.2.309.
[76] Khodursky AB, Peter BJ, Cozzarelli NR, Botstein D, Brown PO, Yanofsky C. DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. Proc Natl Acad Sci USA 2000;97:12170–5. https://doi.org/10.1073/pnas.220414297.

[77] Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, et al. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genom 2006;7:246. https://doi.org/10.1186/1471-2164-7-246.

[78] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008;320:1344–9. https://doi.org/10.1126/science.1158441.

[79] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63. https://doi.org/10.1038/nrg2484.

[80] Robles JA, Qureshi SE, Stephen SJ, Wilson SR, Burden CJ, Taylor JM. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-sequencing. BMC Genom 2012;13:484. https://doi.org/10.1186/1471-2164-13-484.

[81] Herzel L, Stanley JA, Yao CC, Li GW. Ubiquitous mRNA decay fragments in *E. coli* redefine the functional transcriptome. Nucleic Acids Res 2022;50:5029–46. https://doi.org/10.1093/nar/gkac295.

[82] Kogenaru S, Yan Q, Guo Y, Wang N. RNA-seq and microarray complement each other in transcriptome profiling. BMC Genom 2012;13:629. https://doi.org/10.1186/1471-2164-13-629.

[83] Imdahl F, Vafadarnejad E, Homberger C, Saliba AE, Vogel J. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. Nat Microbiol 2020;5:1202–6. https://doi.org/10.1038/s41564-020-0774-1.

[84] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680–5. https://doi.org/10.1038/227680a0.

[85] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422:198–207. https://doi.org/10.1038/nature01511.

[86] Silva JC, Denny R, Dorschel C, Gorenstein MV, Li GZ, Richardson K, et al. Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a *SWEET* tale. Mol Cell Proteom 2006;5:589–607. https://doi.org/10.1074/mcp.M500321-MCP200.

[87] Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, et al. Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 2007;25:125–31. https://doi.org/10.1038/nbt1275.

[88] Picotti P, Aebersold R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 2012;9:555–66. https://doi.org/10.1038/nmeth.2015.

[89] Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. Cell 2009;138:795–806. https://doi.org/10.1016/j.cell.2009.05.051.

[90] Redding-Johanson AM, Batth TS, Chan R, Krupa R, Szmidt HL, Adams PD, et al. Targeted proteomics for metabolic pathway optimization: application to terpene production. Metab Eng 2011;13:194–203. https://doi.org/10.1016/j.ymben.2010.12.005.

[91] Lange V, Malmström JA, Didion J, King NL, Johansson BP, Schäfer J, et al. Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. Mol Cell Proteom 2008;7:1489–500. https://doi.org/10.1074/mcp.M800032-MCP200.

[92] Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci USA 2003;100:6940–5. https://doi.org/10.1073/pnas.0832254100.

[93] Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protoc 2006;1:1029–43. https://doi.org/10.1038/nprot.2006.129.

[94] Stahl DC, Swiderek KM, Davis MT, Lee TD. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. J Am Soc Mass Spectrom 1996;7:532–40. https://doi.org/10.1016/1044-0305(96)00057-8.

[95] Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat Methods 2004;1:39–45. https://doi.org/10.1038/nmeth705.

[96] Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinović SM, Schubert OT, et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol 2014;32:219–23. https://doi.org/10.1038/nbt.2841.

[97] Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteom 2012;11. https://doi.org/10.1074/mcp.O111.016717.

[98] Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods 2020;17:41–4. https://doi.org/10.1038/s41592-019-0638-x.

[99] Pham T, Tyagi A, Wang YS, Guo J. Single-cell proteomic analysis. Wiley Inter Rev Syst Biol Med 2021;13:e1503. https://doi.org/10.1002/wsbm.1503.

[100] Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 2010;329:533–8. https://doi.org/10.1126/science.1188308.

[101] Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Büttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol 2022;40:121–30. https://doi.org/10.1038/s41587-021-01001-7.

[102] Brotherton HO, Yost RA. Determination of drugs in blood serum by mass spectrometry/mass spectrometry. Anal Chem 1983;55:549–53. https://doi.org/10.1021/ac00254a030.

[103] Miggiels P, Wouters B, van Westen GJP, Dubbelman AC, Hankemeier T. Novel technologies for metabolomics: more for less. Trends Anal Chem 2019;120:115323. https://doi.org/10.1016/j.trac.2018.11.021.

[104] Nießer J, Müller MF, Kappelmann J, Wiechert W, Noack S. Hot isopropanol quenching procedure for automated microtiter plate scale ^{13}C-labeling experiments. Micro Cell Fact 2022;21:78. https://doi.org/10.1186/s12934-022-01806-4.

[105] Bajad SU, Lu W, Kimball EH, Yuan J, Peterson C, Rabinowitz JD. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J Chromatogr 2006;1125:76–88. https://doi.org/10.1016/j.chroma.2006.05.019.

[106] Coulier L, Bas R, Jespersen S, Verheij E, van der Werf MJ, Hankemeier T. Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography−electrospray ionization mass spectrometry. Anal Chem 2006;78:6573–82. https://doi.org/10.1021/ac0607616.

[107] Wishart DS. Quantitative metabolomics using NMR. Trends Anal Chem 2008;27:228–37. https://doi.org/10.1016/j.trac.2007.12.001.

[108] Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, Boughton BA, et al. Review of recent developments in GC-MS approaches to metabolomics-based research. Metabolomics 2018;14:152. https://doi.org/10.1007/s11306-018-1449-2.

[109] Fuhrer T, Heer D, Begemann B, Zamboni N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection–time-of-flight mass spectrometry. Anal Chem 2011;83:7074–80. https://doi.org/10.1021/ac201267k.

[110] Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics 2011;7:307–28. https://doi.org/10.1007/s11306-010-0254-3.

[111] Ribbenstedt A, Ziarrusta H, Benskin JP. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. PLoS One 2018;13:e0207082. https://doi.org/10.1371/journal.pone.0207082.

[112] Ramanathan R, Jemal M, Ramagiri S, Xia YQ, Humpreys WG, Olah T, et al. It is time for a paradigm shift in drug discovery bioanalysis: from SRM to HRMS. J Mass Spectrom 2011;46:595–601. https://doi.org/10.1002/jms.1921.

[113] Zhou J, Liu H, Liu Y, Liu J, Zhao X, Yin Y. Development and evaluation of a parallel reaction monitoring strategy for large-scale targeted metabolomics quantification. Anal Chem 2016;88:4478–86. https://doi.org/10.1021/acs.analchem.6b00355.

[114] Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites 2018;8:31. https://doi.org/10.3390/metabo8020031.

[115] Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, et al. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 2012;41:D781–6. https://doi.org/10.1093/nar/gks1004.

[116] Haug K, Salek RM, Steinbeck C. Global open data management in metabolomics. Curr Opin Chem Biol 2017;36:58–63. https://doi.org/10.1016/j.cbpa.2016.12.024.

[117] Teoh ST, Putri S, Mukai Y, Bamba T, Fukusaki E. A metabolomics-based strategy for identification of gene targets for phenotype improvement and its application to 1-butanol tolerance in *Saccharomyces cerevisiae*. Biotechnol Biofuels 2015;8:144. https://doi.org/10.1186/s13068-015-0330-z.

[118] Calero P, Gurdo N, Nikel PI. Role of the CrcB transporter of *Pseudomonas putida* in the multi-level stress response elicited by mineral fluoride. Environ Microbiol 2022;24:5082–104. https://doi.org/10.1111/1462-2920.16110.

[119] Vallino JJ, Stephanopoulos G. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol Bioeng 1993;41:633–46. https://doi.org/10.1002/bit.260410606.

[120] Varma A, Palsson BØ. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 1994;60:3724–31. https://doi.org/10.1128/aem.60.10.3724-3731.1994.

[121] Marx A, de Graaf AA, Wiechert W, Eggeling L, Sahm H. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnol Bioeng 1996;49:111–29. https://doi.org/10.1002/(sici)1097-0290(19960120)49:2<111::Aid-bit1>3.0.Co;2-t.

[122] Kohlstedt M, Becker J, Wittmann C. Metabolic fluxes and beyond—Systems biology understanding and engineering of microbial metabolism. Appl Microbiol Biotechnol 2010;88:1065–75. https://doi.org/10.1007/s00253-010-2854-2.

[123] Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metab Eng 2007;9:68–86. https://doi.org/10.1016/j.ymben.2006.09.001.

[124] Young JD. *INCA*: a computational platform for isotopically non-stationary metabolic flux analysis. Bioinformatics 2014;30:1333–5. https://doi.org/10.1093/bioinformatics/btu015.

[125] Zamboni N, Fendt SM, Rühl M, Sauer U. ^{13}C-based metabolic flux analysis. Nat Protoc 2009;4:878–92. https://doi.org/10.1038/nprot.2009.58.

[126] Young JD. ^{13}C metabolic flux analysis of recombinant expression hosts. Curr Opin Biotechnol 2014;30:238–45. https://doi.org/10.1016/j.copbio.2014.10.004.

[127] Rahim M, Ragavan M, Deja S, Merritt ME, Burgess SC, Young JD. *INCA 2.0*: a tool for integrated, dynamic modeling of NMR- and MS-based isotopomer measurements and rigorous metabolic flux analysis. Metab Eng 2022;69:275–85. https://doi.org/10.1016/j.ymben.2021.12.009.

[128] Fendt SM, Oliveira AP, Christen S, Picotti P, Dechant RC, Sauer U. Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. Mol Syst Biol 2010;6:432. https://doi.org/10.1038/msb.2010.91.
[129] Heux S, Poinot J, Massou S, Sokol S, Portais JC. A novel platform for automated high-throughput fluxome profiling of metabolic variants. Metab Eng 2014;25:8–19. https://doi.org/10.1016/j.ymben.2014.06.001.
[130] Klingner A, Bartsch A, Dogs M, Wagner-Döbler I, Jahn D, Simon M, et al. Large-scale ^{13}C flux profiling reveals conservation of the Entner-Doudoroff pathway as a glycolytic strategy among marine bacteria that use glucose. Appl Environ Microbiol 2015;81:2408–22. https://doi.org/10.1128/AEM.03157-14.
[131] Kollmann M, Løvdok L, Bartholomé K, Timmer J, Sourjik V. Design principles of a bacterial signalling network. Nature 2005;438:504–7. https://doi.org/10.1038/nature04228.
[132] de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 2002;9:67–103. https://doi.org/10.1089/10665270252833208.
[133] Covert MW, Knight EM, Reed JL, Herrgård MJ, Palsson BØ. Integrating high-throughput and computational data elucidates bacterial networks. Nature 2004;429:92–6. https://doi.org/10.1038/nature02456.
[134] Saa PA, Nielsen LK. Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. Biotechnol Adv 2017;35:981–1003. https://doi.org/10.1016/j.biotechadv.2017.09.005.
[135] Hartline CJ, Schmitz AC, Han Y, Zhang F. Dynamic control in metabolic engineering: theories, tools, and applications. Metab Eng 2021;63:126–40. https://doi.org/10.1016/j.ymben.2020.08.015.
[136] Lubitz T, Schulz M, Klipp E, Liebermeister W. Parameter balancing in kinetic models of cell metabolism. J Phys Chem 2010;114:16298–303. https://doi.org/10.1021/jp108764b.
[137] Murzin DY, Wärnå J, Haario H, Salmi T. Parameter estimation in kinetic models of complex heterogeneous catalytic reactions using Bayesian statistics. React Kin Mechan Catal 2021;133:1–15. https://doi.org/10.1007/s11144-021-01974-1.
[138] Burgard AP, Vaidyaraman S, Maranas CD. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol Prog 2001;17:791–7. https://doi.org/10.1021/bp0100880.
[139] Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. Genome Res 2004;14:2367–76. https://doi.org/10.1101/gr.2872004.
[140] Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 2006;8:1–13. https://doi.org/10.1016/j.ymben.2005.08.003.
[141] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 2007;3:121. https://doi.org/10.1038/msb4100155.
[142] Wang L, Birol I, Hatzimanikatis V. Metabolic control analysis under uncertainty: framework development and case studies. Biophys J 2004;87:3750–63. https://doi.org/10.1529/biophysj.104.048090.
[143] Wang L, Hatzimanikatis V. Metabolic engineering under uncertainty. I: framework development. Metab Eng 2006;8:133–41. https://doi.org/10.1016/j.ymben.2005.11.003.
[144] van Rosmalen RP, Smith RW, Martins dos Santos VAP, Fleck C, Suárez-Diez M. Model reduction of genome-scale metabolic models as a basis for targeted kinetic models. Metab Eng 2021;64:74–84. https://doi.org/10.1016/j.ymben.2021.01.008.
[145] Fang X, Lloyd CJ, Palsson BØ. Reconstructing organisms *in silico*: genome-scale models and their emerging applications. Nat Rev Microbiol 2020;18:731–43. https://doi.org/10.1038/s41579-020-00440-4.
[146] Hadadi N, Pandey V, Chiappino-Pepe A, Morales M, Gallart-Ayala H, Mehl F, et al. Mechanistic insights into bacterial metabolic reprogramming from omics-integrated genome-scale models. Syst Biol Appl 2020;6:1. https://doi.org/10.1038/s41540-019-0121-4.
[147] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28:245–8. https://doi.org/10.1038/nbt.1614.
[148] Hendry JI, Dinh HV, Foster C, Gopalakrishnan S, Wang L, Maranas CD. Metabolic flux analysis reaching genome wide coverage: lessons learned and future perspectives. Curr Opin Chem Eng 2020;30:17–25. https://doi.org/10.1016/j.coche.2020.05.008.
[149] Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgård MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat Protoc 2007;2:727–38. https://doi.org/10.1038/nprot.2007.99.
[150] Soh KC, Hatzimanikatis V. Constraining the flux space using thermodynamics and integration of metabolomics data. Methods Mol Biol 2014;1191:49–63. https://doi.org/10.1007/978-1-4939-1170-7_3.
[151] Bordbar A, Monk JM, King ZA, Palsson BØ. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet 2014;15:107–20. https://doi.org/10.1038/nrg3643.
[152] Rana P, Berry C, Ghosh P, Fong SS. Recent advances on constraint-based models by integrating machine learning. Curr Opin Biotechnol 2020;64:85–91. https://doi.org/10.1016/j.copbio.2019.11.007.
[153] Salvy P, Hatzimanikatis V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. Nat Commun 2020;11:30. https://doi.org/10.1038/s41467-019-13818-7.
[154] Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. *i*ML1515, a knowledgebase that computes *Escherichia coli* traits. Nat Biotechnol 2017;35:904–8. https://doi.org/10.1038/nbt.3956.
[155] Zhang Y, Cai J, Shang X, Wang B, Liu S, Chai X, et al. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. Biotechnol Biofuels 2017;10:169. https://doi.org/10.1186/s13068-017-0856-3.
[156] Belda E, van Heck RGA, López-Sánchez MJ, Cruveiller S, Barbe V, Fraser C, et al. The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic *chassis*. Environ Microbiol 2016;18:3403–24. https://doi.org/10.1111/1462-2920.13230.
[157] Nogales J, Mueller J, Gudmundsson S, Canalejo FJ, Duque E, Monk J, et al. High-quality genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. Environ Microbiol 2020;22:255–69. https://doi.org/10.1111/1462-2920.14843.
[158] Nikel PI, Pérez-Pantoja D, de Lorenzo V. Pyridine nucleotide transhydrogenases enable redox balance of *Pseudomonas putida* during biodegradation of aromatic compounds. Environ Microbiol 2016;18:3565–82. https://doi.org/10.1111/1462-2920.13434.
[159] Volke DC, Nikel PI. Getting bacteria in shape: synthetic morphology approaches for the design of efficient microbial cell factories. Adv Biosyst 2018;2:1800111. https://doi.org/10.1002/adbi.201800111.
[160] Volke DC, Turlin J, Mol V, Nikel PI. Physical decoupling of XylS/*Pm* regulatory elements and conditional proteolysis enable precise control of gene expression in *Pseudomonas putida*. Microb Biotechnol 2020;13:222–32. https://doi.org/10.1111/1751-7915.13383.
[161] Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting ^{13}C metabolite labeling patterns from cells. Curr Opin Biotechnol 2015;34:189–201. https://doi.org/10.1016/j.copbio.2015.02.003.
[162] Zamboni N, Fischer E, Sauer U. FiatFlux – a software for metabolic flux analysis from ^{13}C-glucose experiments. BMC Bioinforma 2005;6:209. https://doi.org/10.1186/1471-2105-6-209.
[163] Crown SB, Antoniewicz MR. Selection of tracers for ^{13}C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology. Metab Eng 2012;14:150–61. https://doi.org/10.1016/j.ymben.2011.12.005.
[164] Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. Biotechnol Bioeng 2008;99:686–99. https://doi.org/10.1002/bit.21632.
[165] Crown SB, Long CP, Antoniewicz MR. Integrated ^{13}C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. Metab Eng 2015;28:151–8. https://doi.org/10.1016/j.ymben.2015.01.001.
[166] Young JD, Shastri AA, Stephanopoulos G, Morgan JA. Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis. Metab Eng 2011;13:656–65. https://doi.org/10.1016/j.ymben.2011.08.002.
[167] Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 2019;20:1085–93. https://doi.org/10.1093/bib/bbx085.
[168] Ravikirthi P, Suthers PF, Maranas CD. Construction of an *E. coli* genome-scale atom mapping model for MFA calculations. Biotechnol Bioeng 2011;108:1372–82. https://doi.org/10.1002/bit.23070.
[169] McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, et al. Adaptive laboratory evolution resolves energy depletion to maintain high aromatic metabolite phenotypes in *Escherichia coli* strains lacking the phosphotransferase system. Metab Eng 2018;48:233–42. https://doi.org/10.1016/j.ymben.2018.06.005.
[170] McCloskey D, Young JD, Xu S, Palsson BØ, Feist AM. MID Max: LC–MS/MS method for measuring the precursor and product mass isotopomer distributions of metabolic intermediates and cofactors for metabolic flux analysis applications. Anal Chem 2016;88:1362–70. https://doi.org/10.1021/acs.analchem.5b03887.
[171] Barkai N, Leibler S. Robustness in simple biochemical networks. Nature 1997;387:913–7. https://doi.org/10.1038/43199.
[172] Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol Bioeng 2002;79:53–73. https://doi.org/10.1002/bit.10288.
[173] Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. Science 1999;283:381–7. https://doi.org/10.1126/science.283.5400.381.
[174] Kriegeskorte N, Golan T. Neural network models and deep learning. Curr Biol 2019;29:R231–6. https://doi.org/10.1016/j.cub.2019.02.034.
[175] Goh ATC. Back-propagation neural networks for modeling complex systems. Artif Int Eng 1995;9:143–51. https://doi.org/10.1016/0954-1810(94)00011-S.
[176] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. In: Arbib MA, editor. Handbook of Brain Theory and Neural Networks. MIT Press; 1995. p. 3361.
[177] Sworder DD, Singer PF, Doria D, Hutchins RG. Image-enhanced estimation methods. Proc Inst Elect Electron Eng 1993;81:797–814. https://doi.org/10.1109/5.257679.
[178] Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res 2000;40:1489–506. https://doi.org/10.1016/S0042-6989(99)00163-7.
[179] Lewis DD, Spärck Jones K. Natural language processing for information retrieval. Commun ACM 1996;39:92–101. https://doi.org/10.1145/234173.234210.
[180] Clauwaert J, Waegeman W. Novel transformer networks for improved sequence labeling in genomics. IEEE/ACM Trans Comput Biol Bioinform 2020;19:97–106. https://doi.org/10.1109/TCBB.2020.3035021.

[181] Iuchi H, Matsutani T, Yamada K, Iwano N, Sumi S, Hosoda S, et al. Representation learning applications in biological sequence analysis. Comput Struct Biotechnol J 2021;19:3198–208. https://doi.org/10.1016/j.csbj.2021.05.039.
[182] Nikam R, Gromiha MM. Seq2Feature: a comprehensive web-based feature extraction tool. Bioinformatics 2019;35:4797–9. https://doi.org/10.1093/bioinformatics/btz432.
[183] Magnan CN, Baldi P. SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. Bioinformatics 2014;30:2592–7. https://doi.org/10.1093/bioinformatics/btu352.
[184] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.
[185] Callaway E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. Nature 2020;588:203–4. https://doi.org/10.1038/d41586-020-03348-4.
[186] Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, et al. Visualization of omics data for systems biology. Nat Methods 2010;7:S56–68. https://doi.org/10.1038/nmeth.1436.
[187] Shen R, Wang S, Mo Q. Sparse integrative clustering of multilple omics data sets. Ann Appl Stat 2013;7:269–94. https://doi.org/10.1214/12-aoas578.
[188] Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp II DG, et al. DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics 2008;24:1556–8. https://doi.org/10.1093/bioinformatics/btn217.
[189] Yao F, Coquery J, Lê Cao KA. Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets. BMC Bioinforma 2012;13:24. https://doi.org/10.1186/1471-2105-13-24.
[190] Lê Cao KA, Rossouw D, Robert-Granié C, Besse P. A sparse PLS for variable selection when integrating omics data. Stat Appl Genet Mol Biol 2008;7:35. https://doi.org/10.2202/1544-6115.1390.
[191] Olson RS, Cava WL, Mustahsan Z, Varik A, Moore JH. Data-driven advice for applying machine learning to bioinformatics problems. Pac Symp Biocomput 2018;23:192–203.
[192] Nandi S, Subramanian A, Sarkar RR. An integrative machine learning strategy for improved prediction of essential genes in *Escherichia coli* metabolism using flux-coupled features. Mol BioSyst 2017;13:1584–96. https://doi.org/10.1039/C7MB00234C.
[193] Kingma DP, Rezende DJ, Mohamed S, Welling M. Semi-supervised learning with deep generative models. arXiv 2014. https://doi.org/10.48550/arXiv.41406.45298.
[194] Caron M, Bojanowski P, Joulin A, Douze M. Deep clustering for unsupervised learning of visual features. Springer International Publishing,; 2018. p. 139–56.
[195] Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv 2022. https://doi.org/10.48550/arXiv.41312.48611.
[196] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. arXiv 2014. https://doi.org/10.48550/arXiv.1406.2661.
[197] Smith AL, Asta DM, Calder CA. The geometry of continuous latent space models for network data. Stat Sci 2019;34:428–53. https://doi.org/10.1214/19-sts702.
[198] Wan C, Probst T, van Gool L, Yao A. Crossing nets: combining gans and vaes with a shared latent space for hand pose estimation. Proc IEEE Conf Comp Vis Pattern Recogn 2017;1:1196–205.
[199] López R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods 2018;15:1053–8. https://doi.org/10.1038/s41592-018-0229-2.
[200] Tan J, Doing G, Lewis KA, Price CE, Chen KM, Cady KC, et al. Unsupervised extraction of stable expression signatures from public compendia with an ensemble of neural networks. Cell Syst 2017;5(63–71):e66. https://doi.org/10.1016/j.cels.2017.06.003.
[201] Rohart F, Gautier B, Singh A, Lê Cao KA. *mixOmics*: an R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 2017;13:e1005752. https://doi.org/10.1371/journal.pcbi.1005752.
[202] Gao YK, Fokoue A, Luo H, Iyengar A, Dey S, Zhang P. Interpretable drug target prediction using deep neural representation. Intern J Conf Artif Intell Org 2018;1:3371–7.
[203] Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY. Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. J Chem Inf Model 2019;59:3981–8. https://doi.org/10.1021/acs.jcim.9b00387.
[204] Wang X., He X., Wang M., Feng F., Chua T.S., 2019. Neural graph collaborative filtering, In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, pp. 165–174.
[205] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50. https://doi.org/10.1073/pnas.0506580102.
[206] Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res 2010;38:W71–7. https://doi.org/10.1093/nar/gkq329.
[207] Jin S, Zeng X, Xia F, Huang W, Liu X. Application of deep learning methods in biological networks. Brief Bioinformat 2021;22:1902–17. https://doi.org/10.1093/bib/bbaa043.
[208] Coley CW, Thomas DA, Lummiss JAM, Jaworski JN, Breen CP, Schultz V, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science 2019;365:eaax1566. https://doi.org/10.1126/science.aax1566.
[209] Karim AS, Dudley QM, Juminaga A, Yuan Y, Crowe SA, Heggestad JT, et al. *In vitro* prototyping and rapid optimization of biosynthetic enzymes for cell design. Nat Chem Biol 2020;16:912–9. https://doi.org/10.1038/s41589-020-0559-0.
[210] Zhang J, Petersen SD, Radivojević T, Ramírez A, Pérez-Manríquez A, Abeliuk E, et al. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nat Commun 2020;11:4880. https://doi.org/10.1038/s41467-020-17910-1.
[211] Guo M, Song W. The growing U.S. bioeconomy: drivers, development and constraints. N Biotechnol 2019;49:48–57. https://doi.org/10.1016/j.nbt.2018.08.005.
[212] Patermann C, Aguilar A. The origins of the bioeconomy in the European Union. N Biotechnol 2018;40:20–4. https://doi.org/10.1016/j.nbt.2017.04.002.
[213] Tellechea-Luzardo J, Otero-Muras I, Goñi-Moreno A, Carbonell P. Fast biofoundries: coping with the challenges of biomanufacturing. Trends Biotechnol 2022;40:831–42. https://doi.org/10.1016/j.tibtech.2021.12.006.
[214] Chao R, Mishra S, Si T, Zhao H. Engineering biological systems using automated biofoundries. Metab Eng 2017;42:98–108. https://doi.org/10.1016/j.ymben.2017.06.003.
[215] Ortiz L, Paván M, McCarthy L, Timmons J, Densmore DM. Automated robotic liquid handling assembly of modular DNA devices. J Vis Exp 2017;130:e54703. https://doi.org/10.3791/54703.
[216] Chory EJ, Gretton DW, DeBenedictis EA, Esvelt KM. Enabling high-throughput biology with flexible open-source automation. Mol Syst Biol 2021;17:e9942. https://doi.org/10.15252/msb.20209942.
[217] Carbonell P, Jervis AJ, Robinson CJ, Yan C, Dunstan M, Swainston N, et al. An automated design-build-test-learn pipeline for enhanced microbial production of fine chemicals. Commun Biol 2018;1:66. https://doi.org/10.1038/s42003-018-0076-9.
[218] Singh AH, Kaufmann-Malaga BB, Lerman JA, Dougherty DP, Zhang Y, Kilbo AL, et al. An automated scientist to design and optimize microbial strains for the industrial production of small molecules. bioRxiv 2023. https://doi.org/10.1101/2023.01.03.521657.
[219] Robinson CJ, Carbonell P, Jervis AJ, Yan C, Hollywood KA, Dunstan MS, et al. Rapid prototyping of microbial production strains for the biomanufacture of potential materials monomers. Metab Eng 2020;60:168–82. https://doi.org/10.1016/j.ymben.2020.04.008.
[220] Casini A, Chang FY, Eluere R, King AM, Young EM, Dudley QM, et al. A pressure test to make 10 molecules in 90 days: external evaluation of methods to engineer biology. J Am Chem Soc 2018;140:4302–16. https://doi.org/10.1021/jacs.7b13292.
[221] Leferink NGH, Dunstan MS, Hollywood KA, Swainston N, Currin A, Jervis AJ, et al. An automated pipeline for the screening of diverse monoterpene synthase libraries. Sci Rep 2019;9:11936. https://doi.org/10.1038/s41598-019-48452-2.
[222] HamediRad M, Chao R, Weisberg S, Lian J, Sinha S, Zhao H. Towards a fully automated algorithm driven platform for biosystems design. Nat Commun 2019;10:5150. https://doi.org/10.1038/s41467-019-13189-z.
[223] Borkowski O, Koch M, Zettor A, Pandi A, Batista AC, Soudier P, et al. Large scale active-learning-guided exploration for in vitro protein production optimization. Nat Commun 2020;11:1872. https://doi.org/10.1038/s41467-020-15798-5.
[224] Genzen JR, Burnham CD, Felder RA, Hawker CD, Lippi G, Peck Palmer OM. Challenges and opportunities in implementing total laboratory automation. Clin Chem 2018;64:259–64. https://doi.org/10.1373/clinchem.2017.274068.
[225] Huang JF, Shen ZY, Mao QL, Zhang XM, Zhang B, Wu JS, et al. Systematic analysis of bottlenecks in a multibranched and multilevel regulated pathway: the molecular fundamentals of L-methionine biosynthesis in *Escherichia coli*. ACS Synth Biol 2018;7:2577–89. https://doi.org/10.1021/acssynbio.8b00249.
[226] Fatma Z, Schultz JC, Zhao H. Recent advances in domesticating non-model microorganisms. Biotechnol Prog 2020;36:e3008. https://doi.org/10.1002/btpr.3008.
[227] Ellis T, Adie T, Baldwin GS. DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol 2011;3:109–18. https://doi.org/10.1039/c0ib00070a.
[228] Hughes RA, Ellington AD. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. Cold Spring Harb Perspect Biol 2017;9:a023812. https://doi.org/10.1101/cshperspect.a023812.
[229] Salvachúa D, Johnson CW, Singer CA, Rohrer H, Peterson DJ, Black BA, et al. Bioprocess development for muconic acid production from aromatic compounds and lignin. Green Chem 2018;20:5007–19. https://doi.org/10.1039/C8GC02519C.
[230] Fernández-Cabezón L, Cros A, Nikel PI. Evolutionary approaches for engineering industrially-relevant phenotypes in bacterial cell factories. Biotechnol J 2019;14:1800439. https://doi.org/10.1002/biot.201800439.
[231] Rienzo M, Jackson SJ, Chao LK, Leaf T, Schmidt TJ, Navidi AH, et al. High-throughput screening for high-efficiency small-molecule biosynthesis. Metab Eng 2021;63:102–25. https://doi.org/10.1016/j.ymben.2020.09.004.
[232] Xia L, Yang J, Su R, Zhou W, Zhang Y, Zhong Y, et al. Recent progress in fast sample preparation techniques. Anal Chem 2020;92:34–48. https://doi.org/10.1021/acs.analchem.9b04735.
[233] Jessop-Fabre MM, Sonnenschein N. Improving reproducibility in synthetic biology. Front Bioeng Biotechnol 2019;7:18. https://doi.org/10.3389/fbioe.2019.00018.
[234] Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, et al. Systems biology and multi-omics integration: viewpoints from the metabolomics research community. Metabolites 2019;9:76. https://doi.org/10.3390/metabo9040076.

[235] Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother 2019;63. https://doi.org/10.1128/aac.00483-19.
[236] Greene JL, Wäechter A, Tyo KEJ, Broadbelt LJ. Acceleration strategies to enhance metabolic ensemble modeling performance. Biophys J 2017;113:1150–62. https://doi.org/10.1016/j.bpj.2017.07.018.
[237] Goñi-Moreno A, Nikel PI. High-performance biocomputing in synthetic biology–integrated transcriptional and metabolic circuits. Front Bioeng Biotechnol 2019;7:40. https://doi.org/10.3389/fbioe.2019.00040.
[238] Volke DC, Calero P, Nikel PI. *Pseudomonas putida*. Trends Microbiol 2020;28: 512–3. https://doi.org/10.1016/j.tim.2020.02.015.
[239] Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-generation machine learning for biological networks. Cell 2018;173:1581–92. https://doi.org/10.1016/j.cell.2018.05.015.
[240] Gurdo N, Volke DC, Nikel PI. Merging automation and fundamental discovery into the design–build–test–learn cycle of nontraditional microbes. Trends Biotechnol 2022;40:1148–59. https://doi.org/10.1016/j.tibtech.2022.03.004.
[241] Rohrer JM. Thinking clearly about correlations and causation: graphical causal models for observational data. Adv Methods Pr Psychol Sci 2018;1:27–42. https://doi.org/10.1177/2515245917745629.
[242] Porcelli M, Toint PL. BFO, a trainable derivative-free brute force optimizer for nonlinear bound-constrained optimization and equilibrium computations with continuous and discrete variables. ACM Trans Math Softw 2017;44:6. https://doi.org/10.1145/3085592.
[243] Pearl J. The seven tools of causal inference, with reflections on machine learning. Commun ACM 2019;62:54–60.
[244] Webb S. Deep learning for biology. Nature 2018;554:555–7. https://doi.org/10.1038/d41586-018-02174-z.
[245] Rees-Garbutt J, Chalkley O, Landon S, Purcell O, Marucci L, Grierson C. Designing minimal genomes using whole-cell models. Nat Commun 2020;11:836. https://doi.org/10.1038/s41467-020-14545-0.
[246] Battaglia PW, Hamrick JB, Bapst V, Sánchez-González A, Zambaldi V, Malinowski M, et al. Relational inductive biases, deep learning, and graph networks. arXiv 2018;1806:01261.
[247] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 2019;378:686–707. https://doi.org/10.1016/j.jcp.2018.10.045.
[248] Mao Z, Jagtap AD, Karniadakis GE. Physics-informed neural networks for high-speed flows. Comp Methods Appl Mech Eng 2020;360:112789. https://doi.org/10.1016/j.cma.2019.112789.
[249] Martin HG, Radivojevic T, Zucker J, Bouchard K, Sustarich J, Peisert S, et al. Perspectives for self-driving labs in synthetic biology. Curr Opin Biotechnol 2023; 79:102881. https://doi.org/10.1016/j.copbio.2022.102881.
[250] Opgenorth P, Costello Z, Okada T, Goyal G, Chen Y, Gin J, et al. Lessons from two design-build-test-learn cycles of dodecanol production in *Escherichia coli* aided by machine learning. ACS Synth Biol 2019;8:1337–51. https://doi.org/10.1021/acssynbio.9b00020.