Strategies in Engineering Sustainable Biochemical Synthesis through Microbial Systems

Yoseb Song and Kristala LJ Prather

Addresses:

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

Corresponding author:

Prather, Kristala LJ (kljp@mit.edu)

Abstract

Growing environmental concerns and the urgency to address climate change have increased demand for the development of sustainable alternatives to fossil-derived fuels and chemicals. Microbial systems, possessing inherent biosynthetic capabilities, present a promising approach for achieving this goal. This review discusses the coupling of systems and synthetic biology, to enable the elucidation and manipulation of microbial phenotypes for the production of chemicals that can substitute for petroleum-derived counterparts and contribute to advancing green biotechnology. The integration of artificial intelligence with metabolic engineering to facilitate precise and data-driven design of biosynthetic pathways is also discussed, along with the identification of current limitations and proposition of strategies for optimizing of biosystems, thereby propelling the field of chemical biology towards sustainable chemical production.

Keywords:

Metabolic engineering, Systems Biology, Synthetic Biology, Microbe. Machine learning, Genome engineering

Introduction

Since the advent of the Industrial Revolution, human reliance on fossil feedstocks for energy and chemical production has markedly increased. These resources, essential for transportation, manufacturing, and daily life, have driven technological advancements and significantly improved living standards. However, this dependency has also led to significant environmental issues, such as climate change and pollution, primarily due to the combustion of fossil fuels and the persistence of non-biodegradable waste. In light of these challenges, the development of sustainable and eco-friendly methods for energy and chemical synthesis has become crucial. A promising approach involves leveraging microorganisms as biocatalysts to produce target chemicals, mirroring nature's synthesis of diverse compounds [1]. Additionally, the use of renewable and non-edible feedstocks, including carbon dioxide, plastics, and waste, offers a sustainable alternative, while eliminating the need for harmful solvents and non-reusable catalysts often found in traditional chemical production methods [2].

Advances in biotechnology and metabolic engineering have facilitated the tailoring of microorganisms to produce a wide range of chemicals previously obtained from petrochemical routes. Recent progress in systems and synthetic biology has provided deep insights into biological systems and the tools for precise genetic engineering, propelling the field of metabolic engineering forward [1]. Systems biology has benefitted from advancements in sequencing technology, enriching the available data for analysis and enhancing our understanding of the genotype-to-phenotype in target organisms. Synthetic biology, utilizing this knowledge, has introduced methods to alter gene expression and regulation through advanced genetic engineering techniques, enabling effective modifications [3]. Moreover, the integration of artificial intelligence with complex molecular and cellular networks has addressed numerous

challenges, laying a solid foundation for designing, engineering, and optimizing the production of desired chemicals [4]. The synergy of these disciplines has facilitated sustainable processes for producing a wide range of products, from bulk chemicals and natural products to biofuels, through microbial systems [5-7].

This review delineates essential strategies for the microbial synthesis of target chemicals, focusing on host selection, system analysis, rational design, and process optimization (**Figure 1**). Illustrated through recent case studies, these strategies highlight the current trends and potential future developments in the field. Lastly, we discuss the challenges faced and propose viable solutions to enhance the efficiency of bio-based production systems.

Selecting the Host System

Selecting the appropriate host organism is a pivotal step in the successful biological production of chemicals. The host essentially serves as the foundational 'hardware,' endowed with intrinsic metabolic pathways, gene expression mechanisms, and the necessary resources to support biochemical reactions [8]. The selection process involves a comprehensive evaluation of various factors, notably the host's growth capabilities and the feasibility of genetic modifications. The ability of the host to utilize different substrates, maintain growth under specific pH and temperature conditions, and its resilience to stress are crucial considerations. For example, modifying host systems to accommodate higher temperatures and acidic pH has led to enhanced biochemical production of compounds such as 2,3-butanediol and citramalate, surpassing the yields obtained with previous hosts [9,10]. Furthermore, hosts that exhibit increased tolerance to toxic substrates have demonstrated a six-fold increase in 2,3-butanediol production compared to their wild-type counterparts [11].

Model organisms have been preferred for chemical production due to the vast array of genetic information and tools (**Figure 1b**). *Escherichia coli* and *Saccharomyces cerevisiae* stand out as the most extensively employed hosts over the past three decades [2]. The wealth of tools available for these strains, such as rigorously validated *in silico* models, including whole-cell models, and detailed gene annotations, facilitates thorough and rational engineering efforts [12-14]. The availability of optimized genetic engineering tools streamlines the development process, eliminating the time-consuming task of identifying appropriate toolboxes for each host and thereby conserving valuable research resources [15]. This suite of advantages has enabled the synthesis of complex compounds, ranging from bioplastics and biofuels to natural products [16-18].

Non-model organisms, which may be less characterized or recently isolated, present unique opportunities for biochemical production due to their distinctive physiological traits and inherent production capabilities. The ability to process a wide array of substrates, including those that are toxic, can significantly reduce the need for extensive genetic engineering. Notably, acetogens and methanotrophs exploit C1 substrates, such as carbon monoxide (CO), carbon dioxide (CO_2), and methanol, for biomass accumulation, thus playing a role in reducing greenhouse gas emissions and fostering a sustainable bio-economy by demonstrating the production of 3 g/L/h of acetone and isopropanol with 90% selectivity [19-21]. The natural propensity of non-model organisms to produce substantial quantities of target chemicals is frequently attributed to their extensive metabolic capabilities and finely tuned endogenous pathways, a result of evolutionary adaptations to challenging environments [22-24]. An example of utilizing metabolic capabilities is *Yarrowia lipolytica*, which naturally accumulates lipids up to 30% of dry cell weight, while engineered strains can accumulate lipids to more than 50% of

the cell weight [22]. Recent advancements in genome sequencing have enabled the detailed characterization of these organisms, providing insights into their physiology and potential for novel chemical synthesis [23,24].

The advent of chassis systems in synthetic biology introduces a tailored platform through strategic gene integration and deletion, aimed at simplifying genomic complexity and enhancing system predictability for more manageable control. This strategy focuses on redesigning host systems for specific functions while minimizing unnecessary genetic load, thus optimizing energy use for the desired outcomes. Although this approach necessitates a deep understanding of the host system, which can sometimes be a limiting factor, chassis systems developed from well-characterized organisms have shown promising results [25]. Techniques such as sequential deletion have led to significant genome reductions in species like *E. coli*, enhancing cell density and threonine production (**Figure 1b**) [26,27]. Moreover, a genome-reduced *Pseudomonas putida* strain has exhibited beneficial outcomes in applications like plastic upcycling and lignin valorization [28,29]. Concurrently, the reduction in DNA synthesis costs and improvements in DNA assembly techniques have led to the development of innovative strategies for chassis reconstruction [30]. Investigations into the minimal cell JCVI-syn3.0, complemented by strategic gene insertions, have clarified the minimal gene set essential for cellular replication, marking a significant breakthrough in delineating the core requirements for minimal cell viability [31].

Understanding the Host System

Following the selection of a suitable host system, it is important to understand and predict the organism's phenotype for informed and rational design. The essence of systems biology lies in

understanding the organism holistically and predicting biological responses under specific conditions (**Figure 1c**).

Advanced sequencing technologies have enabled the acquisition of vast amounts of information, providing molecular-level insights and unraveling complex biological processes within a system (**Figure 2**). This is particularly crucial for non-model organisms, where existing research may be limited. At the same time, developments in mass spectrometry have facilitated detailed metabolite analyses, offering valuable insights for the optimization of biochemical production. Collectively, these advancements allow for the integration and interpretation of data, paving the way for more feasible rational engineering with an expanded array of genetic components for fine-tuning molecular regulations (**Figure 2a - d**) [1].

Recent studies demonstrated that the multi-omics approach offers a comprehensive view of the host system (**Figure 2a**) [21,32]. Genome sequencing, the first step, provides the foundational blueprint essential for understanding the system by predicting phenotypic responses and facilitating genome mining. Resources such as antiSMASH and MIBiG can reveal previously unknown pathways or metabolites, including novel natural products and vaccine adjuvants (**Figure 2e**) [33-36]. Integrating genomic data with transcriptomics reveals patterns of gene expression, determining which genes are actively transcribed under given conditions (**Figure 2a**). This combination not only identifies genetic compositions but also quantifies gene expression and expands the molecular regulatory toolbox (**Figure 2b**). By leveraging native promoters, the genetic toolkit for CO_2-fixing archaea and bacteria has been broadened, enabling the utilization of CO_2 as the sole substrate for biochemical production [32,37]. Moreover, exploring the translational step reveals the actual genes translated into proteins, bridging the gap between DNA and protein caused by post-transcriptional regulation and facilitating rational

engineering. Such investigations provide insights into the translational efficiency of genetic parts, allowing for tailored protein synthesis and resource allocation for bioproduction [21,32].

The development of *in silico* models has revolutionized our ability to calculate phenotypic responses through simulation, encompassing both metabolic networks and whole-cell levels (**Figure 2e**). These models serve as instrumental tools in identifying candidates for achieving desired phenotypes (**Figure 1c**). Genome-scale models, constructed using genomic data as a backbone and incorporating other omics data, mathematically represent the organism under targeted conditions by calculating fluxes (**Figure 1c**). These models suggest ways to rewire pathways to minimize carbon loss and increase the biosynthesis rate of desired products, while reducing the accumulation of undesired byproducts [38,39]. A well-constructed model can rationalize metabolic engineering strategies for various biochemical productions, such acetone, heme, and isopropanol production [21,40]. Metabolomics data further enrich these models by providing kinetic insights into reactions within the host system, enabling more accurate flux predictions and phenotypic response forecasting [41]. Moreover, the integration of machine learning (ML) in pathway design and deep learning (DL) in enzyme design has significantly enhanced our understanding of cellular responses, thus improving engineering efficiency through more accurate predictions [1].

The development of artificial intelligence in systems biology has introduced novel applications for designing host systems to alter phenotypes and boost biochemical production (**Figure 2e**) [42]. Integrating these technologies with genome-scale model reconstruction and analysis streamlines traditionally labor-intensive processes such as gap filling, making simulations more feasible and accurate [43]. Furthermore, ML-based simulation methods surpass conventional flux balance analysis, offering improved phenotype predictions [44]. The design of

metabolic pathways, a process traditionally marred by the need for extensive manual labor, has been significantly improved. Contemporary tools, employing both template-based and freeform approaches, facilitate the efficient selection of biochemical production routes [4]. These tools leverage ML and DL algorithms to recommend suitable enzymes, either by identifying similarities or by predicting the enzymatic functions, as indicated by EC numbers, pertinent to the enzymatic reactions in question [45]. Predictive models based on ML can ascertain the properties of these selected enzymes, including kinetic parameters, providing valuable insights that guide further engineering efforts [4]. In cases where an appropriate enzyme is lacking, the field of *de novo* protein engineering offers a possible solution. Recent advancements in *de novo* design techniques enable the creation of enzymes tailored to specific functional requirements, thereby expanding the possibilities for biochemical synthesis and host system modification [46,47].

Optimizing Biochemical Production

Following the investigative and design phases, optimizing biochemical production necessitates comprehensive engineering at both the molecular and cellular levels. Despite the rapid advances in genetic engineering, plasmid-based systems remain potent for manipulating host organisms, including *E. coli* and non-model organisms (**Figure 2d**). These systems regulate gene expression at both the transcriptional level, where DNA is transcribed into RNA, and the translational level, where RNA is translated into proteins, both of which are essential for synthesizing complex biochemicals (**Figure 2c**) [16,32]. Employing plasmid-based systems, alongside native or synthetic biological parts derived from multi-omics analyses, has proven effective in host engineering [32,48,49]. In addition, a notable application is activating silent

biosynthetic gene clusters through heterologous expression, facilitating the production of complex natural products and biofuels [7].

Genome modification to introduce genetic cassettes has gained favor for its ability to ensure the stability of heterologous devices, reduce the metabolic burden associated with maintaining plasmids, avoid antibiotic resistance issues, and control gene copy numbers (**Figure 1d**). Traditional recombination systems, including Lambda Red and RecET, have been widely used for DNA integration into genomes. The advent of CRISPR-Cas systems has revolutionized genetic engineering by streamlining processes and enhancing efficiency and with demonstrated utility across different kingdoms of life. For example, CRISPR systems have been developed not only for bacteria but also for filamentous fungi, demonstrating the versatility and broad applicability of these tools [15,50]. CRISPR-Cas9, in particular, facilitates precise genomic alterations through double-stranded DNA breaks, followed by homology-directed repair for knock-in and knockout applications, even in non-model organisms [15,50]. Integration efficiency can be further augmented by combining CRISPR with conventional systems like Lambda Red [15].

In addition to creating double-strand DNA breaks, the Cas9 protein has been modified (Asp10Ala) to function as nCas9, targeting single DNA strands for nicking rather than cleaving (**Figure 1d**). This adaptation facilitated the development of base editors (BEs), such as cytosine base editors (CBE) and adenine base editors (ABE), which convert cytosine to uracil or adenine to inosine, respectively, altering nucleotide pairs from C:G to T:A or A:T to G:C. A further iteration, glycosylase base editors (GBE), emerged by fusing nCas9 with cytidine deaminase and uracil-DNA glycosylase, enabling the conversion of cytosine to adenine or guanine. Additionally, the prime editor system, incorporating nCas9 (His840Ala), reverse transcriptase,

and prime editing sgRNA (pegRNA), offers the capability to insert, delete, or mutate specific genomic sequences as dictated by pegRNA design [15]. These BEs have shown potential for multiplex genome-wide targeting, enhancing organismal fitness under stress [51]. Recent advancements have introduced tools capable of accurately predicting the efficiency of ABE, CBE, and GBE, facilitating precise single nucleotide engineering [52].

Gene expression modulation has been effectively achieved using CRISPR interference (CRISPRi) and activation (CRISPRa) systems (**Figure 1d**). By deactivating the Cas protein (Asp10Ala and His840Ala mutations), specific genomic regions can be targeted to suppress or enhance gene expression. CRISPRi, particularly useful for downregulating essential genes, serves as an alternative to knockout strategies and has become a popular tool for functional genomics screens. It leverages genome-wide targeting of sgRNAs to pinpoint genes that, when modulated, improve fitness and biochemical output [53]. CRISPRa, on the other hand, employs cellular machinery to activate silent biosynthetic gene clusters (BGCs), boosting the expression of key genes for complex antibiotic production [54].

The integration of CRISPR systems with transposons has enabled efficient incorporation of foreign DNA into both model and non-model organism genomes. This methodology extends to diverse microbial communities, showcasing the potential for broad-scale genomic editing [55,56]. Additionally, combining error-prone DNA polymerase with nCas9 has diversified target gene sequences, creating a vast library of variants that have led to increased production of compounds like L-proline [57].

Evolutionary approaches, particularly adaptive laboratory evolution (ALE), offer significant potential for phenotype enhancement in host systems. The production of toxic biochemicals and inefficient substrate utilization often hampers efficiency, burdening the host.

ALE enables organisms to adapt to and thrive under such stressors, improving overall fitness. This technique has been instrumental in enabling non-model organisms to tolerate otherwise inhibitory substrates. One example is gaining tolerance to the toxic substrate CO via ALE, then utilizing native promoters obtained from omics analysis to integrate heterologous gene expression and optimize gene expression in the non-model organism, leading to enhanced 2,3-butanediol production from C1 substrates [11]. The integration of ALE with automated culturing systems further increases the likelihood of achieving desirable phenotypes in bio-systems [58].

Conclusion and Perspectives

The integration of systems and synthetic biology for the engineering of biological systems has witnessed rapid development, propelled by advancements in sequencing and genome engineering technologies. The field of artificial intelligence has further revolutionized our capacity to analyze vast datasets, uncovering previously overlooked insights and establishing relationships between various variables. This has significantly improved our ability to predict the responses of biosystems to specific conditions.

Despite these strides, several challenges remain that must be addressed to design and construct biosystems that function optimally. A primary concern is the need to broaden the spectrum of host systems. Through environmental isolation and genome analysis, a diverse array of potential host systems has been identified. The advent of high-throughput culturomics, enhanced by machine learning and automation, offers a robust methodology for correlating physiological traits with genomic data [59]. Automation could potentially play a critical role in streamlining the culturing and analysis processes, possibly reducing the time and labor required for large-scale experiments. This efficiency may be beneficial for studying non-model

organisms, which often have unique growth requirements and are less well-characterized. By automating repetitive tasks and integrating advanced data analysis tools, researchers can rapidly screen and optimize conditions for these organisms, facilitating their establishment as viable hosts for biochemical production. Furthermore, automated systems can handle high-throughput screening of genetic variants and environmental conditions, accelerating the potential discovery of optimal strains and pathways for target compound synthesis. However, the exploration of additional non-model organisms is imperative to discover more efficient methods for producing novel or complex biochemicals. Co-culture systems present a potential strategy for achieving desired phenotypes, though they also introduce a level of complexity that may be difficult to control.

Optimization of pathway design is another critical area for improvement. While ML and DL have advanced our ability to predict cellular responses and devise production pathways, challenges such as enzyme kinetics, gene regulation optimization, resource allocation, toxicity of intermediates and final products, and economic viability persist. Tools for enzyme identification exist, yet integrating these enzymes into biosystems to optimize production, ensure enzyme activity and solubility, and provide necessary substrates and cofactors requires meticulous flux rewiring and can only be refined through experimental validation. As experimental data accumulate, the precision in pathway design is expected to improve.

The processes involved in host selection, biosystem investigation, and engineering are labor-intensive and time-consuming, rendering them challenging for a single research lab to undertake independently. The demand for standardized biological parts, accessible analytical tools, and rapid build-test infrastructures is growing. Efforts are underway to develop a universal

platform that merges systems and synthetic biology, yet the accessibility and affordability of such infrastructure are limited [60].

Richard Feynman, a renowned theoretical physicist, once stated, "What I cannot create, I do not understand." This philosophy resonates profoundly with the endeavor to engineer biosystems with specific phenotypes, whether for complex biochemical production or bioremediation. The advancements discussed in this review bring us closer to the goal of accurately and efficiently engineering desired biosystems. By addressing the challenges outlined, we can deepen our understanding and develop bio-platforms capable of supplanting traditional chemical production methods reliant on fossil resources.

Declaration of Competing Interest

The authors declare that there is no competing financial interest.

Acknowledgments

This work was supported by the US National Science Foundation through the Division of Molecular and Cellular Biosciences (Grant No. MCB-2218259).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Han T, Nazarbekov A, Zou X, Lee SY: **Recent advances in systems metabolic engineering**. *Curr Opin Biotechnol* 2023, **84**:103004.
2. Nurwono G, O'Keeffe S, Liu N, Park JO: **Sustainable metabolic engineering requires a perfect trifecta**. *Curr Opin Biotechnol* 2023, **83**:102983.
3. Han YH, Kim G, Seo SW: **Programmable synthetic biology tools for developing microbial cell factories**. *Curr Opin Biotechnol* 2023, **79**:102874.
4. Yu T, Boob AG, Volk MJ, Liu X, Cui H, Zhao H: **Machine learning-enabled retrobiosynthesis of molecules**. *Nature Catalysis* 2023, **6**:137-151.
5. Bannister KR, Prather KL: **Engineering polyester monomer diversity through novel pathway design**. *Curr Opin Biotechnol* 2023, **79**:102852.
6. Lee SY, Kim HU, Chae TU, Cho JS, Kim JW, Shin JH, Kim DI, Ko Y-S, Jang WD, Jang Y-S: **A comprehensive metabolic map for production of bio-based chemicals**. *Nature Catalysis* 2019, **2**:18-33.
7. Li X, Gadar-Lopez AE, Chen L, Jayachandran S, Cruz-Morales P, Keasling JD: **Mining natural products for advanced biofuels and sustainable bioproducts**. *Curr Opin Biotechnol* 2023, **84**:103003.
8. Martinez-Garcia E, de Lorenzo V: ***Pseudomonas putida* as a synthetic biology chassis and a metabolic engineering platform**. *Curr Opin Biotechnol* 2024, **85**:103025.
9. Sheng L, Madika A, Lau MSH, Zhang Y, Minton NP: **Metabolic engineering for the production of acetoin and 2,3-butanediol at elevated temperature in *Parageobacillus thermoglucosidasius* NCIMB 11955**. *Front Bioeng Biotechnol* 2023, **11**:1191079.
10. Wu ZY, Sun W, Shen Y, Pratas J, Suthers PF, Hsieh PH, Dwaraknath S, Rabinowitz JD, Maranas CD, Shao Z, et al.: **Metabolic engineering of low-pH-tolerant non-model yeast, *Issatchenkia orientalis*, for production of citramalate**. *Metab Eng Commun* 2023, **16**:e00220.
11. Jin S, Kang S, Bae J, Lee H, Cho BK: **Development of CO gas conversion system using high CO tolerance biocatalyst**. *Chemical Engineering Journal* 2022, **449**:137678.
12. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, et al.: ***i*ML1515, a knowledgebase that computes *Escherichia coli* traits**. *Nat Biotechnol* 2017, **35**:904-908.
13. Lu H, Li F, Sanchez BJ, Zhu Z, Li G, Domenzain I, Marcisauskas S, Anton PM, Lappa D, Lieven C, et al.: **A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism**. *Nat Commun* 2019, **10**:3586.
14. Elsemman IE, Rodriguez Prado A, Grigaitis P, Garcia Albornoz M, Harman V, Holman SW, van Heerden J, Bruggeman FJ, Bisschops MMM, Sonnenschein N, et al.: **Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies**. *Nat Commun* 2022, **13**:801.

15. Teng Y, Jiang T, Yan Y: **The expanded CRISPR toolbox for constructing microbial cell factories**. *Trends Biotechnol* 2024, **42**:104-118.
16. Park SY, Eun H, Lee MH, Lee SY: **Metabolic engineering of *Escherichia coli* with electron channelling for the production of natural products**. *Nature Catalysis* 2022, **5**:726-765.
* This paper demonstrates engineered bacteria designed to produce lutein, (+)-nootkatone, apigenin, and L-DOPA by developing an electron channeling system that facilitates efficient electron transfer between P450 enzymes and redox carriers. Additionally, the authors enhanced the production by increasing the flux through the heme pathway, achieving a lutein titer of 218.0 mg/L and a productivity rate of 5.01 mg/L/h.
17. Ma Y, Zu Y, Huang S, Stephanopoulos G: **Engineering a universal and efficient platform for terpenoid synthesis in yeast**. *Proc Natl Acad Sci U S A* 2023, **120**:e2207680120.
18. Ni C, Prather KLJ: **Consistent biosynthesis of D-glycerate from variable mixed substrates**. *Metab Eng* 2024, **82**:41-48.
19. Ljungdahl LG: **A life with acetogens, thermophiles, and cellulolytic anaerobes**. *Annu Rev Microbiol* 2009, **63**:1-25.
20. Sahoo KK, Katari JK, Das D: **Recent advances in methanol production from methanotrophs**. *World J Microbiol Biotechnol* 2023, **39**:360.
21. Liew FE, Nogle R, Abdalla T, Rasor BJ, Canter C, Jensen RO, Wang L, Strutz J, Chirania P, De Tissera S, et al.: **Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale**. *Nat Biotechnol* 2022, **40**:335-344.
** This study presents the scale-up of C1-fixing *Clostridium autoethanogenum* for the production of acetone and isopropanol using industrial waste gas feedstocks. Achieving high production rates (3 g/L/h) and selectivity (90%), the authors employed pathway optimization from an industrial strain collection, omics-based strain enhancement, cell-free experimentation, computational modeling, and process refinement. The work demonstrates the microorganism's potential for sustainable and efficient commodity chemical production from waste feedstocks.
22. Park YK, Ledesma-Amaro R: **What makes *Yarrowia lipolytica* well suited for industry?** *Trends Biotechnol* 2023, **41**:242-254.
23. Viacava K, Qiao J, Janowczyk A, Poudel S, Jacquemin N, Meibom KL, Shrestha HK, Reid MC, Hettich RL, Bernier-Latmani R: **Meta-omics-aided isolation of an elusive anaerobic arsenic-methylating soil bacterium**. *ISME J* 2022, **16**:1740-1749.
24. Henriksen N, Schostag MD, Balder SR, Bech PK, Strube ML, Sonnenschein EC, Gram L: **The ability of Phaeobacter inhibens to produce tropodithietic acid influences the community dynamics of a microalgal microbiome**. *ISME Commun* 2022, **2**:109.
25. Ma S, Su T, Lu X, Qi Q: **Bacterial genome reduction for optimal chassis of synthetic biology: a review**. *Crit Rev Biotechnol* 2023:1-14.
26. Iwadate Y, Honda H, Sato H, Hashimoto M, Kato J: **Oxidative stress sensitivity of engineered *Escherichia coli* cells with a reduced genome**. *FEMS Microbiol Lett* 2011, **322**:25-33.
27. Hirokawa Y, Kawano H, Tanaka-Masuda K, Nakamura N, Nakagawa A, Ito M, Mori H, Oshima T, Ogasawara N: **Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli***. *J Biosci Bioeng* 2013, **116**:52-58.28. Bao T, Qian Y, Xin Y, Collins JJ, Lu T: **Engineering microbial division of labor for plastic upcycling**. *Nat Commun* 2023, **14**:5712.

* This study highlights the effectiveness of engineered microbial consortia as a platform for polymer upcycling. The authors developed synthetic consortia consisting of two engineered *Pseudomonas putida* strains that utilize terephthalic acid and ethylene glycol. These consortia efficiently degrade polyethylene terephthalate hydrolysate and produce polyhydroxyalkanoates and cis-cis muconate. Compared to monoculture systems, these consortia demonstrate enhanced deconstruction of plastics.

29. Lee S, Kang M, Jung CD, Bae JH, Lee JY, Park YK, Joo JC, Kim H, Sohn JH, Sung BH: **Development of novel recombinant peroxidase secretion system from *Pseudomonas putida* for lignin valorisation**. *Bioresour Technol* 2023, **388**:129779.
30. Venter JC, Glass JI, Hutchison CA, 3rd, Vashee S: **Synthetic chromosomes, genomes, viruses, and cells**. *Cell* 2022, **185**:2708-2724.
31. Pelletier JF, Sun L, Wise KS, Assad-Garcia N, Karas BJ, Deerinck TJ, Ellisman MH, Mershin A, Gershenfeld N, Chuang RY, et al.: **Genetic requirements for cell division in a genomically minimal cell**. *Cell* 2021, **184**:2430-2440 e2416.
32. Song Y, Bae J, Jin S, Lee H, Kang S, Lee J, Shin J, Cho S, Cho BK: **Development of highly characterized genetic bioparts for efficient gene expression in CO_2-fixing *Eubacterium limosum***. *Metab Eng* 2022, **72**:215-226.
33. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, et al.: **antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation**. *Nucleic Acids Res* 2023, **51**:W46-W50.
34. Terlouw BR, Blin K, Navarro-Munoz JC, Avalon NE, Chevrette MG, Egbert S, Lee S, Meijer D, Recchia MJJ, Reitz ZL, et al.: **MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters**. *Nucleic Acids Res* 2023, **51**:D603-D610.
35. Yee DA, Niwa K, Perlatti B, Chen M, Li Y, Tang Y: **Genome mining for unknown-unknown natural products**. *Nat Chem Biol* 2023, **19**:633-640.
36. Martin LBB, Kikuchi S, Rejzek M, Owen C, Reed J, Orme A, Misra RC, El-Demerdash A, Hill L, Hodgson H, et al.: **Complete biosynthesis of the potent vaccine adjuvant QS-21**. *Nat Chem Biol* 2024.

* This study demonstrates how an integrated approach combining bioinformatics, biochemistry, and synthetic biology can produce the valuable yet structurally complex natural product, vaccine adjuvant QS-21, in an engineered system. By mining the *Quillaja saponaria* genome, the authors identified enzymes responsible for the biosynthesis of a unique arabinofuranosylated acyl chain, which is critical for its immunostimulatory activity.

37. Xu Q, Du Q, Gao J, Chen L, Dong X, Li J: **A robust genetic toolbox for fine-tuning gene expression in the CO(2)-Fixing methanogenic archaeon *Methanococcus maripaludis***. *Metab Eng* 2023, **79**:130-145.
38. Zhang Y, Wang X, Odesanmi C, Hu Q, Li D, Tang Y, Liu Z, Mi J, Liu S, Wen T: **Model-guided metabolic rewiring to bypass pyruvate oxidation for pyruvate derivative synthesis by minimizing carbon loss**. *mSystems* 2024:e0083923.
39. Roell GW, Schenk C, Anthony WE, Carr RR, Ponukumati A, Kim J, Akhmatskaya E, Foston M, Dantas G, Moon TS, et al.: **A High-Quality Genome-Scale Model for *Rhodococcus opacus* Metabolism**. *ACS Synth Biol* 2023, **12**:1632-1644.
40. Ishchuk OP, Domenzain I, Sanchez BJ, Muniz-Paredes F, Martinez JL, Nielsen J, Petranovic D: **Genome-scale modeling drives 70-fold improvement of intracellular heme**

production in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 2022, **119**:e2108245119.
41. Chen Y, Gustafsson J, Tafur Rangel A, Anton M, Domenzain I, Kittikunapong C, Li F, Yuan L, Nielsen J, Kerkhoven EJ: **Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0**. *Nat Protoc* 2024.
* This study demonstrates an integrated computational approach that combines genome-scale modeling, enzyme kinetics, and omics data analysis to reconstruct enzyme-constrained metabolic models, thereby enhancing phenotype prediction. By leveraging existing databases, deep learning predictions, and detailed enzyme information, the GECKO Toolbox enables the construction of models with constraints that limit metabolic flux.
42. Greener JG, Kandathil SM, Moffat L, Jones DT: **A guide to machine learning for biologists**. *Nat Rev Mol Cell Biol* 2022, **23**:40-55.
43. Chen C, Liao C, Liu YY: **Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning**. *Nat Commun* 2023, **14**:2375.
44. Faure L, Mollet B, Liebermeister W, Faulon JL: **A neural-mechanistic hybrid approach improving the predictive power of genome-scale metabolic models**. *Nat Commun* 2023, **14**:4669.
* This study introduces a hybrid approach named Artificial Metabolic Networks (AMNs) that combines neural network and mechanistic modeling to enhance phenotype predictions from genome-scale metabolic models. The authors demonstrate that AMNs can accurately predict growth rates based on media compositions and metabolic gene knockouts, surpassing the performance of traditional constraint-based modeling and standard machine learning methods.
45. Kim GB, Kim JY, Lee JA, Norsigian CJ, Palsson BO, Lee SY: **Functional annotation of enzyme-encoding genes using deep learning with transformer layers**. *Nat Commun* 2023, **14**:7370.
46. Yeh AH, Norn C, Kipnis Y, Tischer D, Pellock SJ, Evans D, Ma P, Lee GR, Zhang JZ, Anishchenko I, et al.: ***De novo* design of luciferases using deep learning**. *Nature* 2023, **614**:774-780.
** This papers introduces the *de novo* computational design of highly active and specific luciferase enzymes for chemically synthesized luciferin substrates. Utilizing a novel "family-wide hallucination" approach, the study generates a multitude of idealized protein scaffolds featuring diverse binding pockets. By integrating the designed active sites within these scaffolds, the authors developed LuxSit, a notably compact (13.9 kDa) and thermostable luciferase, exhibiting high substrate specificity and catalytic efficiency comparable to native luciferases for the synthetic luciferin diphenylterazine.
47. Hossack EJ, Hardy FJ, Green AP: **Building Enzymes through Design and Evolution**. *ACS Catalysis* 2023, **13**:12436-12444.
48. LaFleur TL, Hossain A, Salis HM: **Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria**. *Nat Commun* 2022, **13**:5159.
* This study demonstrates a computational model that precisely predicts bacterial transcription initiation rates and start sites for any σ^{70} promoter sequence. The model synthesizes data from massively parallel assays, which measured transcription from over 14,000 designed promoters, with biophysical modeling of RNA polymerase/σ^{70} interactions with promoter DNA motifs. Additionally, it incorporates machine learning to quantify 346 interaction energy parameters.
49. Lin G-M, Voigt CA: **Design of a redox-proficient *Escherichia coli* for screening terpenoids and modifying cytochrome P450s**. *Nature Catalysis* 2023, **6**:1016-1029.

* This study demonstrates the engineering of *Escherichia coli* strains designed to enhance the production and screening of a broad array of terpenoids. The authors implemented modifications to boost precursor supply, supplied diverse redox partners for cytochrome P450 enzymes, and introduced regulatory systems to optimize metabolic pathways.The study successfully identified bacterial cytochrome P450s capable of replicating plant-like oxidation patterns on the ent-kaurene scaffold.

50. Yuan Y, Cheng S, Bian G, Yan P, Ma Z, Dai W, Chen R, Fu S, Huang H, Chi H, et al.: **Efficient exploration of terpenoid biosynthetic gene clusters in filamentous fungi**. *Nature Catalysis* 2022, **5**:277-287.
51. Liu Y, Wang R, Liu J, Lu H, Li H, Wang Y, Ni X, Li J, Guo Y, Ma H, et al.: **Base editor enables rational genome-scale functional screening for enhanced industrial phenotypes in *Corynebacterium glutamicum***. *Sci Adv* 2022, **8**:eabq2157.

** This study demonstrates a genome-scale functional screening strategy in *Corynebacterium glutamicum*, utilizing a cytosine base editor alongside a library of approximately 12,000 sgRNAs, which target 98.1% of the organism's genes. The study identifies both known and novel genes that confer resistance to 5-fluorouracil, 5-fluoroorotate, oxidative stress, and furfural.

52. Kim N, Choi S, Kim S, Song M, Seo JH, Min S, Park J, Cho SR, Kim HH: **Deep learning models to predict the editing efficiencies and outcomes of diverse base editors**. *Nat Biotechnol* 2023.
53. Shin J, Bae J, Lee H, Kang S, Jin S, Song Y, Cho S, Cho BK: **Genome-wide CRISPRi screen identifies enhanced autolithotrophic phenotypes in acetogenic bacterium *Eubacterium limosum***. *Proc Natl Acad Sci U S A* 2023, **120**:e2216244120.
54. Ameruoso A, Villegas Kcam MC, Cohen KP, Chappell J: **Activating natural product synthesis using CRISPR interference and activation systems in *Streptomyces***. *Nucleic Acids Res* 2022, **50**:7751-7760.
55. Alalmaie A, Diaf S, Khashan R: **Insight into the molecular mechanism of the transposon-encoded type I-F CRISPR-Cas system**. *J Genet Eng Biotechnol* 2023, **21**:60.
56. Rubin BE, Diamond S, Cress BF, Crits-Christoph A, Lou YC, Borges AL, Shivram H, He C, Xu M, Zhou Z, et al.: **Species- and site-specific genome editing in complex bacterial communities**. *Nat Microbiol* 2022, **7**:34-47.
57. Long M, Xu M, Qiao Z, Ma Z, Osire T, Yang T, Zhang X, Shao M, Rao Z: **Directed Evolution of Ornithine Cyclodeaminase Using an EvolvR-Based Growth-Coupling Strategy for Efficient Biosynthesis of l-Proline**. *ACS Synth Biol* 2020, **9**:1855-1863.
58. Hirasawa T, Maeda T: **Adaptive Laboratory Evolution of Microorganisms: Methodology and Application for Bioproduction**. *Microorganisms* 2022, **11**.
59. Huang Y, Sheth RU, Zhao S, Cohen LA, Dabaghi K, Moody T, Sun Y, Ricaurte D, Richardson M, Velez-Cortes F, et al.: **High-throughput microbial culturomics using automation and machine learning**. *Nat Biotechnol* 2023, **41**:1424-1433.
60. Tellechea-Luzardo J, Otero-Muras I, Goni-Moreno A, Carbonell P: **Fast biofoundries: coping with the challenges of biomanufacturing**. *Trends Biotechnol* 2022, **40**:831-842.

Figures

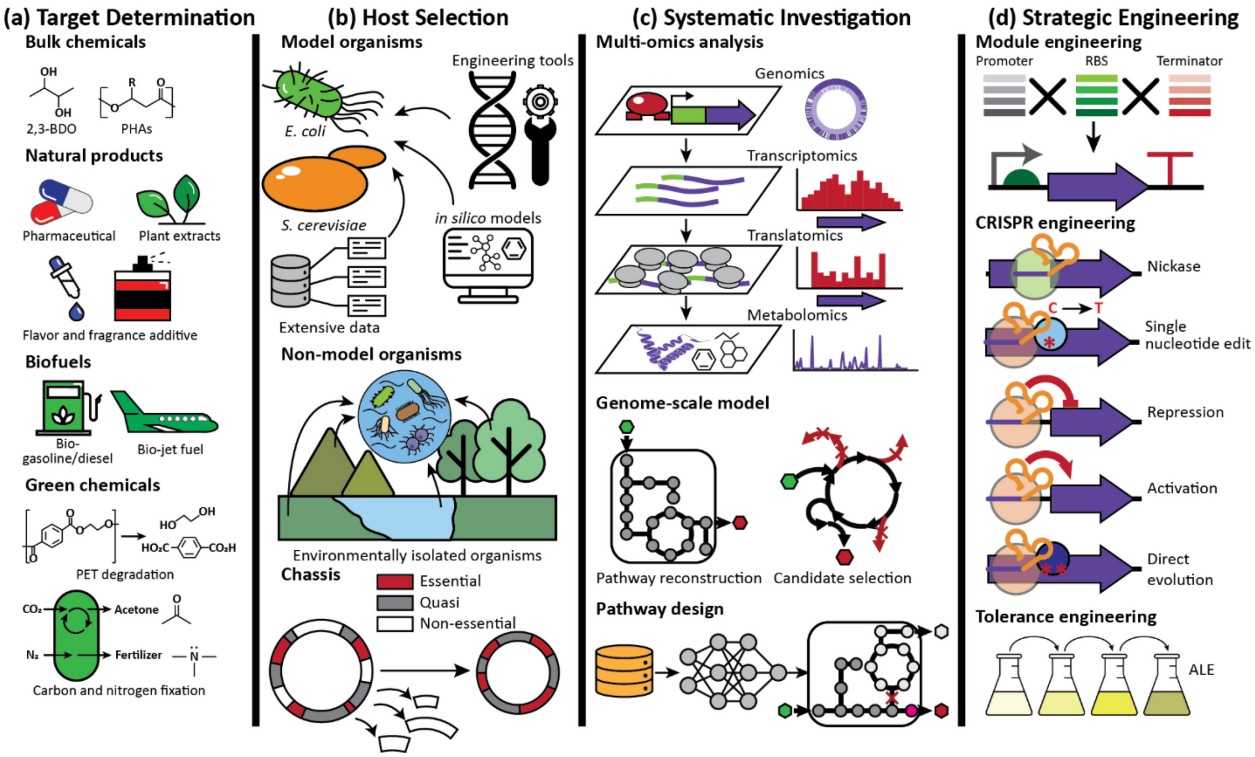

Figure 1. Schematic representation of reconstructing biosystems for targeted phenotypes. (**a**) Target Determination: Establish the objective for the biosystem, whether to synthesize bulk chemicals, natural products, biofuels, or to undertake eco-friendly activities such as plastic degradation or the utilization of non-canonical substrates. (**b**) Host Selection: Commonly utilized model organisms include *Escherichia coli* and *Saccharomyces cerevisiae*, while non-model organisms encompass those recently isolated from the environment or not extensively characterized. An additional option is a chassis organism, streamlined by the removal of non-essential genomic regions to optimize the host system for the desired function. (**c**) Systematic Investigation: Through multi-omics analysis, an intricate understanding of the regulatory systems governing the biosystems is achieved. Genome-scale modeling offers a comprehensive perspective of the host, enabling the identification of candidate modifications to augment desired biochemical production. Database utilization alongside artificial intelligence provides an

efficient methodology for determining the most effective production pathways. (**d**) Strategic Engineering: Engineering efforts typically include modular manipulation using biological parts for gene expression regulation, CRISPR-mediated genome editing for precise modifications, and tolerance engineering to improve organismal robustness and fitness for the production of desired compounds.

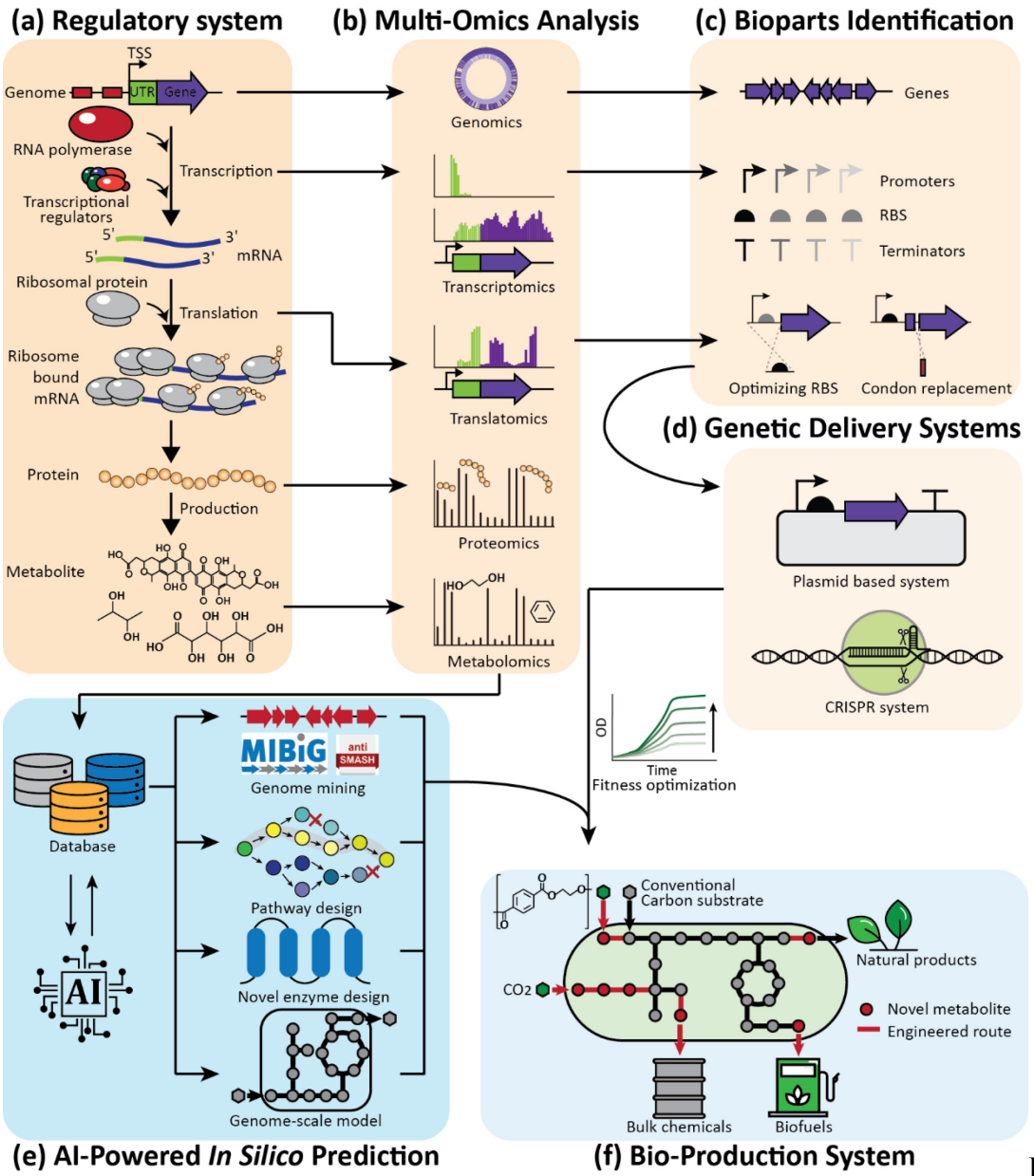

Figure 2. **From omics analysis to host system engineering.** (**a**) Regulatory Systems: The flow from genomic DNA through RNA transcription, protein translation, and resultant metabolite production within the biosystem. (**b**) Multi-Omics Analysis: Comprehensive multi-omics approaches unveil the host system's intricate regulatory networks, spanning genomics, transcriptomics, translatomics, proteomics, and metabolomics. (**c**) Biological parts Identification: Leveraging multi-omics data facilitates the acquisition of biological parts, elucidating gene

functions, the properties of individual components, and translational bottlenecks. (**d**) Genetic Delivery Systems: Plasmid-based systems and CRISPR technologies are employed to introduce and integrate genetic components into the biosystem. (**e**) AI-Powered *In Silico* Prediction: Multi-omics data are collated and processed through artificial intelligence algorithms, predicting optimal strategies for host system design. (**f**) These strategies, integrated with engineering techniques and fitness optimization processes, enable the construction of host systems capable of processing diverse substrates and producing a range of compounds, from bulk chemicals to biofuels and natural products.