ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202303.00659

Design of Metabolic Engineering Strategies for Biosynthesis (Postprint)

Authors: Ma Hongwu, Chen Xiulai, Yuan Qianqian, Liu Liming, Sun Jibin

Date: 2023-03-19T00:00:00+00:00

Abstract

The primary objective of metabolic engineering research is to efficiently synthesize target products by rewiring microbial metabolic networks. Owing to the complexity of cellular metabolic networks, identifying appropriate engineering targets from thousands of metabolic reactions and their regulatory circuits is exceedingly challenging, often necessitating extensive trial-and-error. Computational analysis of large-scale metabolic networks to design optimal biosynthetic pathways for specific products can facilitate the identification of suitable metabolic engineering strategies, reduce the empiricism inherent in the modification process, and accelerate the development of superior biosynthetic strains. This review addresses two fundamental questions: (1) how to design metabolic networks for the de novo synthesis of non-native products and enhancement of product yields, introducing metabolic engineering design methodologies based on computational network analysis; and (2) how to engineer strains for precise regulation of enzymatic reactions, presenting recent advances in dynamic metabolic engineering that employs designed genetic circuits to dynamically control metabolic pathway flux.

Full Text

Design of Metabolic Engineering Strategies for Biosynthesis of Valuable Products

MA Hongwu[1], CHEN Xiulai[2,3], YUAN Qianqian[1], LIU Liming[2,3,4]*, SUN Jibin[1]*

[1]Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
[2]State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi 214122, China
[3]Key Laboratory of Industrial Biotechnology (Ministry of Education), Jiangnan

University, Wuxi 214122, China
[4]National Engineering Laboratory for Cereal Fermentation Technology, Jiangnan University, Wuxi 214122, China

Abstract

One main objective of metabolic engineering is to rewire the metabolic network for efficient production of biochemicals. Due to the complexity of cellular metabolic networks, it is often not straightforward to identify the proper modification targets from thousands of metabolic genes. Therefore, a time-consuming trial & error process is often required for the successful development. Aided by computational modeling of large-scale metabolic networks, one can design optimal pathways for synthesis of objective products, reducing the uncertainty of development and thus accelerating the strain construction process. In this short text, we give brief introduction to metabolic engineering design methods from two aspects: how to modify an organism to produce new chemicals with higher yields, and how to improve the cellular adaptation to the changing process conditions by integrating gene circuits. The computer aided design approach together with automated genome edition technologies, will greatly enhance the efficiency of the construction of artificial cell factories.

Keywords: computational design, metabolic engineering, biosynthesis, dynamic regulation, metabolic network, synthetic biology

---

Introduction

The advent of genetic engineering has freed humanity from reliance on natural variation and screening, enabling purposeful genomic modifications that dramatically enhance strain performance. By introducing, knocking out, or finely tuning enzyme genes within metabolic networks, we can create novel cells that convert inexpensive raw materials into valuable target products—representing the core focus of current metabolic engineering and synthetic biology research. Biological conversion pathways from feedstock to product often involve dozens of enzymatic reactions. Conventional metabolic engineering strategies primarily involve overexpressing and optimizing key enzymes in product synthesis pathways, eliminating byproduct formation routes, and relieving product synthesis inhibition. Through these traditional approaches, numerous engineered strains have been developed for producing biochemicals (such as amino acids, organic acids, vitamins, and antibiotics), biofuels (such as ethanol), and biomaterials (such as polyhydroxyalkanoates and polylactic acid), driving industrial economic development into a new era based on renewable resources and green production methods.

The primary bottleneck in metabolic engineering research is the complexity of cellular metabolic networks. Naturally occurring cells are products of long-term evolution, typically containing thousands of enzyme genes with sophisticated

regulatory mechanisms. Modifying only a few genes along the feedstock-to-product pathway often fails to achieve desired outcomes. However, rapid advances in genome sequencing technology have enabled whole-genome sequencing of increasing numbers of organisms. Through genome annotation, we can quickly identify which enzymes an organism can encode and which reactions it can perform, thereby constructing genome-scale metabolic network models for the organism [1]. These genome-scale metabolic models, together with long-established cross-species metabolic reaction databases such as KEGG and MetaCyc, have laid the foundation for computational simulation and design of metabolic networks, providing novel strategies for synthetic biology and metabolic engineering that substantially accelerate targeted strain development and industrial application.

Synthesizing New Products: From Nothing to Something

Introducing foreign genes to enable an organism to synthesize a new product it could not previously produce is a common metabolic engineering approach. For example, Galanie et al. [2] achieved artificial synthesis of opioid drugs in yeast by introducing over ten genes from six different organisms (including plants, mammals, and bacteria). Using microbial cell factories to produce plant-derived natural products offers advantages: microorganisms grow faster, achieve higher yields per unit time and space, and can replace large-scale agricultural cultivation with smaller fermentation facilities, substantially reducing production costs.

Designing from scratch first requires identifying which foreign genes must be introduced into the chassis cell to establish a complete synthesis pathway for the new product. This necessitates integrating the chassis cell's metabolic network with metabolic reaction databases (such as KEGG and MetaCyc) and using computational methods to identify novel reactions needed in the chassis, thereby forming new pathways for product synthesis. For instance, Zhang et al. [3] incorporated KEGG reactions into the *E. coli* metabolic network, constructing a comprehensive model containing 7,316 reactions. Flux balance analysis using glucose as the substrate revealed that 1,777 heterologous products could be synthesized in *E. coli* by introducing appropriate foreign reactions. This study provides guidance for selecting suitable foreign genes to construct artificial *E. coli* cell factories. The method can also calculate results for other microbial chassis and determine the optimal host for a specific product (requiring fewer foreign genes, stronger precursor synthesis capability, etc.) through comparative analysis. Chatsurachai et al. [4] used *E. coli*, *Corynebacterium glutamicum*, and *Saccharomyces cerevisiae* as chassis cells, employing a stepwise expansion algorithm to add heterologous metabolites and reactions from the KEGG, BRENDA, and ENZYME databases into three strain metabolic models to synthesize new products. The expansion algorithm first identifies reactions whose reactants are present (but whose products are absent) in the strain's metabolic network, then adds these reactions to the model to enable

production of the new reaction products. This new reaction addition process is repeated iteratively on the expanded model until no further reactions can be added. Ultimately, for each chassis cell, a list of novel products it can synthesize and the new reactions required to generate each product are obtained. Using this method, the authors predicted that *E. coli* and *S. cerevisiae* are feasible hosts for 1,3-propanediol synthesis, requiring only the introduction of glycerol dehydrogenase and 1,3-propanediol oxidoreductase to achieve product synthesis.

These methods rely on expanding a specific chassis cell's metabolic network based on particular metabolic reaction databases; target metabolites and related reactions must exist in the database to design relevant pathways. However, many unknown biochemical reactions in nature are not included in these databases. For example, many enzymes such as aldolases exhibit broad catalytic activity and may catalyze transformations of other metabolites beyond known reactions. Through enzyme engineering, constructing enzymes with novel catalytic activities can enable new conversion processes to obtain new biological products, even synthesizing entirely new biomolecules that do not exist in nature. To address this limitation, researchers have proposed methods to design novel reactions based on biochemical reaction rules and compound structural features. For instance, Hatzimanikatis's group used cheminformatics to expand enzyme reactions based on KEGG database reactions, creating the ATLAS database [5], which contains over 130,000 novel biochemical reactions designed through structural similarity to known reactions. Using this database, they designed multiple new pathways for producing the biofuel butanone and compared the theoretical yields of different pathways to identify optimal routes [6].

Improving Product Yield: From Having to Optimizing

After endowing cells with new synthetic capabilities and achieving the "from nothing to something" transformation, metabolic engineering's next objective is to maximize product output while minimizing raw material consumption. The amount of product generated per unit of raw material (substrate) is commonly characterized as "yield," a crucial metric reflecting raw material costs, particularly for bulk chemicals where feedstock expenses dominate production costs. For some products synthesized through simple pathways, such as ethanol and lactic acid, yield-maximizing metabolic routes can be identified through intuitive observation. However, for most metabolites like nucleotides and aromatic amino acids, complex synthesis pathways involving byproducts and multiple precursors make theoretical yield calculation impossible through simple observation or manual computation. Furthermore, product synthesis involves ATP generation/consumption and redox balance, requiring additional substrate for energy and reducing power production or pathways to consume excess energy and reducing equivalents, further complicating yield calculations.

Based on metabolic network models, researchers have developed Flux Balance Analysis (FBA) to accurately calculate theoretical product yields and optimal

conversion pathways achieving these yields [7]. FBA employs stoichiometric relationships in metabolic network models (the molar ratios of reactants and products in a reaction) and reaction irreversibility constraints to solve constrained optimization problems, identifying optimal product synthesis pathways that maximize production while satisfying mass and energy balances. Since the resulting pathways are derived from holistic network analysis rather than conventional biochemical knowledge, they are more comprehensive and can account for energy balance, cofactor balance, multi-precursor balance, and byproduct recycling [8]. This method can identify multiple strategies achieving the same maximum yield, potentially surpassing our biochemical understanding to reveal novel pathways distinct from conventional routes, thereby providing new modification targets and strategies. For example, many biological products are synthesized from acetyl-CoA. When using glucose as substrate, acetyl-CoA is typically generated through pyruvate decarboxylation, the endpoint of glycolysis, causing loss of two carbon atoms from glucose as CO_2 that cannot enter the product. One glucose molecule (containing six carbon atoms) yields at most two acetyl-CoA molecules (four carbon atoms total). Therefore, even if all carbon from acetyl-CoA is converted to product, the maximum carbon molar yield can only reach 0.67 (Figure 1a [Figure 1: see original paper]). However, FBA calculations using the *E. coli* metabolic network model iJO1366 for the acetyl-CoA-derived product poly(3-hydroxybutyrate) (P3HB) achieved an optimal yield of 0.86, indicating the existence of novel pathways in the network that reduce carbon loss. The flux distribution in the optimal pathway is shown in Figure 1b, revealing that besides the conventional glucose-to-pyruvate-to-acetyl-CoA route via the EMP pathway, a threonine bypass pathway recycles formate and CO_2 from phosphoenolpyruvate back to pyruvate while generating an additional acetyl-CoA for product synthesis, thereby increasing product yield. Based on these computational results, Lin et al. [9] activated this cyclic pathway in *E. coli* through deregulation, nearly doubling P3HB yield.

Introducing foreign genes can not only enable cells to acquire new product synthesis capabilities from scratch but also create novel pathways to improve yields of existing products. Bogorad et al. [10] constructed a non-oxidative glycolysis (NOG) pathway in *E. coli* by introducing pentose/hexose phosphoketolase (Fxpk) from *B. adolescentis*. This pathway enables conversion of one glucose molecule into three acetyl-CoA molecules with complete carbon conservation (Figure 2 [Figure 2: see original paper]). After introducing this reaction into the *E. coli* iJO1366 model, we used FBA to calculate theoretical yields for hundreds of biological products, finding significant yield improvements for at least several dozen biochemicals. Guided by these computational results, we constructed *E. coli* strains containing the NOG pathway for P3HB, acetone, and isopropanol production, all showing substantially improved yields compared to parental strains [11,12]. Recently, Meadows et al. [13] introduced the NOG pathway into *S. cerevisiae* to construct a farnesene-producing cell factory, ultimately increasing the carbon molar yield of farnesene from glucose from 0.52 to 0.65.

Fine-Tuning Metabolic Networks: Dynamic Regulation

The reaction list and flux distribution obtained from FBA models represent ideal optimal reaction rates, serving as the ultimate goal for metabolic engineering—the actual enzyme expression levels and activities should precisely meet the requirements for achieving optimal metabolic flux, no more and no less. In practice, metabolic engineering can readily achieve "on/off" regulation or "static control" of enzyme activity, such as completely inactivating a biochemical reaction through gene knockout or increasing a reaction through foreign gene introduction. However, precisely regulating each enzyme's activity according to FBA-predicted flux requirements remains challenging. In reality, cells have evolved diverse dynamic regulatory mechanisms during long-term evolution, precisely controlling enzyme expression (induction and repression) and activity (activation and inhibition) in response to different environments to ensure metabolic networks adapt and achieve global optimization. For example, during fermentation, cells continuously adjust their metabolic states in response to environmental changes, forming distinct phases such as lag phase, exponential growth phase, and product formation phase, each with different requirements for gene/protein expression levels, metabolite concentrations, and flux distributions. Learning from naturally evolved dynamic regulatory mechanisms, researchers have recently proposed the concept and methodology of Dynamic Metabolic Engineering—dynamically regulating gene expression and activity according to intracellular states and environmental conditions to avoid accumulation or deficiency of intermediate metabolites and meet the differential demands of growth and product synthesis on flux distribution across various stages [14-16]. Through dynamic control of enzyme levels, metabolic pathway direction can be dynamically redirected (Figure 3a [Figure 3: see original paper]) or metabolic activity can be precisely balanced (Figure 3b).

Gene circuits, or transcriptional regulatory circuits, can turn gene expression on or off at specific times and cell densities, switching metabolic flux between endogenous pathways for cell growth and heterologous pathways for chemical production. Switching control primarily targets growth-essential genes, as they enable the most direct control between biomass formation and product biosynthesis. Besides controlling enzyme synthesis rates, gene circuits can also finely regulate enzyme degradation rates [17]. For instance, glucose-6-phosphate (G6P) is a critical node for metabolic flux control that can enter glycolysis via phosphofructokinase-1 (Pfk-1) for endogenous metabolism or enter heterologous pathways through synthetic routes. Controlling Pfk-1 activity through gene circuits could switch carbon flux from "growth mode" to "production mode." To achieve this, Brockman and Prather [18] designed a protein degradation circuit fusing Pfk-1 with an SsrA degradation tag that controls adaptor protein SspB binding, ClpXP protease recruitment, and proteolysis, thereby enabling dynamic control of Pfk-1 degradation rate. Applying this circuit to *E. coli* for inositol production increased inositol accumulation 2-fold compared to static control. Gene circuits enable dynamic regulation of metabolic pathways by con-

trolling essential gene expression levels, becoming an effective tool for industrial strain metabolic engineering design—rewiring endogenous metabolic networks to replace traditional static control, achieving time-dependent gene expression control and balanced interactions among multiple genes or pathways, with broad application prospects [17].

Excessive accumulation of metabolic intermediates may trigger side reactions, occupy important cellular resources (such as phosphates and cofactors), and cause direct or indirect cytotoxicity, adversely affecting cell growth or product synthesis. Gene circuits can utilize biosensors responsive to intermediate metabolite concentrations to activate downstream metabolic enzymes and repress upstream enzymes, matching synthesis and utilization rates of intermediate metabolites and improving overall pathway efficiency. For example, shikimate is an intermediate in aromatic amino acid metabolism, and its accumulation competes with aromatic amino acid metabolism. A shikimate-responsive gene circuit can control the key enzyme shikimate kinase (AroK) to balance pathway branches and achieve carbon flux equilibrium between shikimate accumulation and aromatic amino acid metabolism. Mahr et al. [19] constructed a metabolite-responsive circuit network by screening and assembling appropriate biosensor elements, including galactose-responsive and phenylalanine-responsive promoters. Application of this circuit increased phenylalanine production in engineered strain *E. coli* K-12 MG1655/pJC1-mtr sensor-type1 by 4.3-fold compared to the parental strain *E. coli* K-12 MG1655. Besides directly responding to intermediate metabolite concentrations, cell growth status and cofactor imbalance levels can also be used to construct biosensors.

Maximizing biosynthesis speed depends on both per-cell production capacity and total biomass. Typically, growth and production compete for the same cellular resources, requiring proper arrangement and balance to maximize overall biosynthesis speed. Quorum sensing (QS) systems—gene circuits responsive to cell density—provide new means to regulate the relationship between growth and production. For example, researchers assembled QS elements from *Pantoea stewartii* using characteristic promoters and ribosome binding sites of varying strengths to achieve on/off control of genes such as *pfk-1* at specific times and cell densities, increasing inositol production in engineered strain *E. coli* L19S by 5.5-fold compared to static control [20]. Engineered strain *E. coli* L19SA with QS circuit-controlled *aroK* expression increased shikimate production from 0 to 105 mg/L [20].

Metabolic network models and computational design methods based on them provide novel approaches and strategies for metabolic engineering strain modification. Through these computational design methods, researchers can select appropriate foreign genes to construct new pathways for synthesizing novel products from scratch, identify modification targets to optimize yields of existing products, or correlate environmental changes during cell fermentation with specific gene expression through gene circuits, enabling cells to maintain optimal states for different objectives and maximize target product formation. As un-

derstanding of cellular mechanisms and related regulatory systems deepens, it may become possible to establish whole-cell models that accurately describe and predict changes in intracellular metabolites, proteins, and gene expression levels, achieving fully model-guided metabolic engineering strategy design.

References

1. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 2009, 7(2): 129-143.
2. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. *Science*, 2015, 349(6252): 1095-1100.
3. Zhang X, Tervo CJ, Reed JL. Metabolic assessment of *E. coli* as a biofactory for commercial products. *Metab Eng*, 2016, 35: 64-75.
4. Chatsurachai S, Furusawa C, Shimizu H. An in silico platform for the design of heterologous pathways in nonnative metabolite production. *BMC Bioinformatics*, 2012, 13: 93.
5. Hadadi N, Hafner J, Shajkofci A, et al. ATLAS of biochemistry: A repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies. *ACS Synth Biol*, 2016, 5(10): 1155-1166.
6. Hatzimanikatis V, Li C, Ionita JA, et al. Exploring the diversity of complex metabolic networks. *Bioinformatics*, 2005, 21(8): 1603-1609.
7. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*, 2010, 28(3): 245-248.
8. McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol Syst Biol*, 2013, 9: 661.
9. Lin Z, Zhang Y, Yuan Q, et al. Metabolic engineering of *Escherichia coli* for poly(3-hydroxybutyrate) production via threonine bypass. *Microb Cell Fact*, 2015, 14: 185.
10. Bogorad IW, Lin TS, Liao JC. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature*, 2013, 502(7473): 692-695.
11. Yang X, Yuan Q, Zheng Y, et al. An engineered non-oxidative glycolysis pathway for acetone production in *Escherichia coli*. *Biotechnol Lett*, 2016, 38(8): 1359-1365.
12. Zheng Y, Yuan Q, Yang X, et al. Engineering *Escherichia coli* for poly-(3-hydroxybutyrate) production guided by genome-scale metabolic network analysis. *Enzyme Microb Technol*, 2017, 106: 1-11.

13. Meadows AL, Hawkins KM, Tsegaye Y, et al. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature*, 2016, 537(7622): 694-697.

14. Wall ME, Hlavacek WS, Savageau MA. Design of gene circuits: lessons from bacteria. *Nat Rev Genet*, 2004, 5(1): 34-42.

15. Zhang F, Keasling J. Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol*, 2011, 19(7): 323-329.

16. Keasling JD. Synthetic biology and the development of tools for metabolic engineering. *Metab Eng*, 2012, 14(3): 189-195.

17. Brockman IM, Prather KL. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol J*, 2015, 10(9): 1360-1369.

18. Brockman IM, Prather KL. Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. *Metab Eng*, 2015, 28: 104-113.

19. Mahr R, von Boeselager RF, Wiechert J, et al. Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. *Appl Microbiol Biotechnol*, 2016, 100(15): 6739-6753.

20. Gupta A, Reizman IM, Reisch CR, et al. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat Biotechnol*, 2017, 35(3): 273-279.

Author Biographies

MA Hongwu is a Professor at the Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences (CAS), and a recipient of the CAS Hundred Talents Program. He received his Ph.D. in biochemical engineering from Tianjin University in 2001, followed by ten years of research experience in Germany and the United Kingdom. His research focuses on computational analysis of large-scale biological networks and development of relevant computational tools. He has published over 30 papers in journals such as *Bioinformatics* and *Molecular Systems Biology*, with more than 1,000 SCI citations, and received the Thomson Reuters Highly Cited Researcher Award in 2014.
E-mail: ma_{hw}@tib.cas.cn

LIU Liming is a Professor at the State Key Laboratory of Food Science and Technology, Jiangnan University. He received his Ph.D. in fermentation engineering from Jiangnan University in 2006 and conducted postdoctoral research at Chalmers University of Technology from 2009-2010, focusing on systems biology. His research group specializes in systems biology, metabolic engineering, and enzyme engineering. He is supported by the Scientific and Technological

Innovation Leading Talents of the National "Ten Thousand Talents Program" and the National Natural Science Foundation of China.
E-mail: mingll@jiangnan.edu.cn

SUN Jibin is a Professor at the Tianjin Institute of Industrial Biotechnology, CAS, a recipient of the CAS Hundred Talents Program, and a Distinguished Professor of CAS. He received his Ph.D. from the German Research Centre for Biotechnology/Technical University of Braunschweig in 2004 and joined the Tianjin Institute of Industrial Biotechnology in 2008. His research focuses on systems and synthetic biotechnology of industrial microorganisms, including tool development and industrial applications.
E-mail: sunjibin@tib.cas.cn

Corresponding authors: LIU Liming, SUN Jibin

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.