

特约评述

DOI: 10.12211/2096-8280.2026-046

糖合成生物学：从糖科学基础解析到糖基产品生物制造的范式变革

杜昱光¹, 李瑞莲¹, 王倬¹, 丁侃²

(¹ 中国科学院过程工程研究所生物药制备与递送全国重点实验室, 北京 100190; ² 中国科学院上海药物研究所, 上海 201203)

摘要：糖类作为自然界结构多样性最高的生物大分子类别，其糖链在细胞识别、信号转导和免疫调控中发挥不可替代的核心作用。然而，糖链的非模板驱动生物合成机制，糖苷键的连接位置、异头构型和分支模式均由定位在不同亚细胞区室的糖基转移酶协同决定，使其规模化制备长期受限于化学合成的高昂成本和天然来源的不均一性。合成生物学将糖基化系统重新定义为可设计、可重构的工程网络：通过糖基转移酶的基因组挖掘与蛋白质工程、糖核苷酸代谢网络的精细调控、以及微生物细胞工厂的系统优化，正在实现从单糖到复杂糖缀合物的精准生物制造。当前糖合成生物学面临三大核心瓶颈，包括糖基化元件库功能表征严重不完整、糖核苷酸胞内供给水平的精细调控精度不足、糖链的结构-活性关系远未系统解析。本文系统梳理了糖合成生物学四项核心使能技术，包括糖基转移酶元件工程、代谢途径重构、细胞工厂设计及无细胞合成体系，近二十年的关键突破，并讨论了在母乳寡糖、糖基化抗体、糖疫苗及功能多糖材料中的产业化进展。最后，本文分析 AI 驱动糖链结构预测与糖密码解析对糖合成生物学智能化设计的推动作用，展望该领域从经验驱动到智能设计的范式跃迁前景。

关键词：糖合成生物学；糖基转移酶；代谢工程；细胞工厂；糖核苷酸；母乳寡糖；无细胞合成

中图分类号：Q81 文献标志码：A

Glycosynthetic biology: paradigm shift from fundamental glycoscience interpretation to biomanufacturing of glycosyl products

DU Yuguang¹, LI Ruilian¹, Wang Zhuo¹, DING Kan²

(¹ State Key Laboratory of Biopharmaceutical Preparation and Delivery, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190; ² Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203)

Abstract: Carbohydrates constitute the most structurally diverse class of biological macromolecules in nature, with their glycan chains playing irreplaceable roles in cell recognition, signal transduction, and immune regulation. Unlike the template-driven biosynthesis of nucleic acids and proteins, glycan biosynthesis operates without a direct molecular

收稿日期: 2026-05-25 修回日期: 2026-07-26

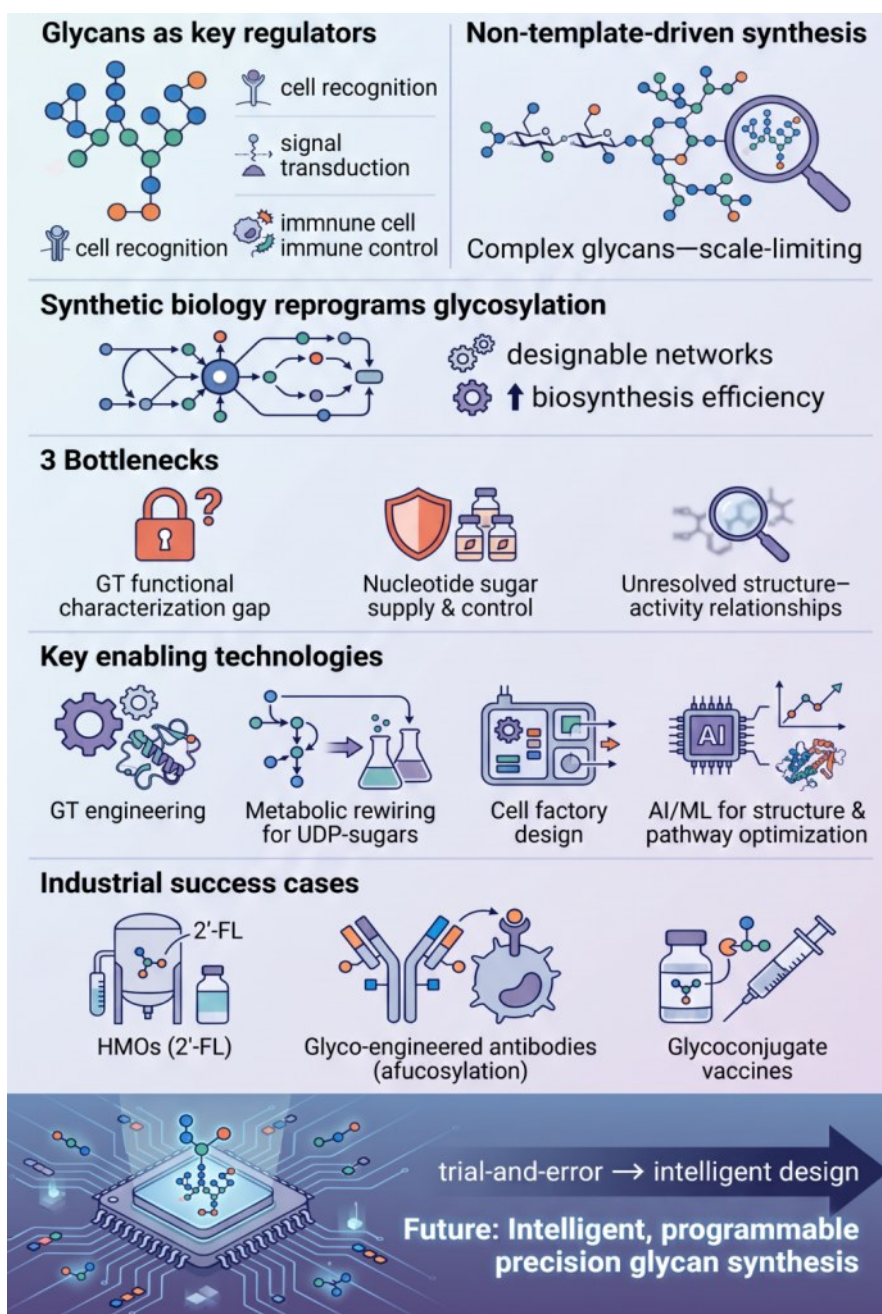
基金项目: 国家重点研发计划 (No. 2023YFC2308000)

引用本文: 杜昱光, 李瑞莲, 王倬, 丁侃. 糖合成生物学: 从糖科学基础解析到糖基产品生物制造的范式变革[J]. 合成生物学, 2026, 7. DOI: 10.12211/2096-8280.2026-046
Citation: DU Yuguang, LI Ruilian, Wang Zhuo, DING Kan. Glycosynthetic biology: paradigm shift from fundamental glycoscience interpretation to biomanufacturing of glycosyl products[J]. Synthetic Biology Journal, 2026, 7. DOI: 10.12211/2096-8280.2026-046

template. Glycosidic linkage positions (1→2, 1→3, 1→4, 1→6), anomeric configurations (α or β), branching patterns, and modification groups are collectively determined by glycosyltransferases (GTs) localized across distinct subcellular compartments. This multi-dimensional structural freedom endows carbohydrates with the high information density among biomolecules, yet it has historically constrained scalable production of complex glycans: chemical synthesis of glycans demands laborious protecting-group manipulations and stringent stereochemical control, while natural sources yield structurally heterogeneous mixtures that preclude systematic structure – activity relationship studies. Synthetic biology fundamentally reframes this challenge by redefining glycosylation systems as engineerable, reconfigurable networks amenable to rational design. Through genome mining and protein engineering of GTs, precision tuning of sugar nucleotide metabolic networks, and systematic optimization of microbial cell factories, the field is advancing toward the programmable biomanufacturing of structurally defined glycans, from simple monosaccharides to complex glycoconjugates, with unprecedented efficiency and precision. The defining contribution of glycosynthetic biology is methodological. It transforms glycoscience from a discipline that catalogs what nature provides into one that engineers what we need. For decades, glycans were the dark matter of molecular biology. They were structurally intractable, biosynthetically opaque, and functionally elusive, yet they sat at the center of development, immunity, infection, and tumorigenesis. Synthetic biology breaks this deadlock. By making homogeneous glycans accessible on demand, it enables the systematic elucidation of structure – function relationships that had been the missing foundation of the field. Three bottlenecks remain. The first is functional annotation of GTs. Over 500,000 sequences are catalogued in the CAZy database; fewer than 5% have been experimentally validated for substrate specificity and kinetics. Glycosylation pathway design therefore remains a screening exercise, not a rational assembly of standardized parts. The second is sugar nucleotide supply control. These metabolic networks are deeply embedded in central carbon metabolism. A perturbation in one pathway, be it GDP-fucose or CMP-Neu5Ac, ripples unpredictably through the entire system. Independent, dynamic, and precise control of multiple sugar nucleotide pools remains unsolved. The third is glycan structure – activity relationships. These defy the frameworks built for small molecules and proteins. The same epitope elicits different responses depending on backbone context, whether N-glycan, O-glycan, or glycolipid. Multivalency produces non-linear effects that cannot be extrapolated from monovalent affinities. This is not a data problem fixable by more measurements; it is an epistemological one requiring new conceptual tools. This review examines twenty years of progress through four enabling technologies. GT element engineering spans genomic mining against CAZy, structure-guided rational design, directed evolution, and ancestral sequence reconstruction. It has produced variants with orders-of-magnitude gains in catalytic efficiency, altered substrate specificity, and improved process stability. Metabolic pathway rewiring and cell factory design have delivered landmark achievements: the first transplantation of prokaryotic N-glycosylation into *E. coli*, the humanization of yeast for complex sialylated N-glycans, the CRISPR-enabled glycoform customization of CHO cells, and the de novo microbial biosynthesis of glycosaminoglycans including heparin and hyaluronic acid. Cell-free synthetic systems bypass the constraints of living chassis entirely. There are no membrane barriers, no metabolic cross-talk, no product toxicity. They uniquely enable the incorporation of non-natural sugar nucleotides for bioorthogonal applications. AI and machine learning are accelerating glycan structure prediction, GT function annotation, and pathway optimization. Yet their defining handicap is structural: glycan databases are orders of magnitude smaller than those for proteins or nucleic acids. This places a hard ceiling on what data-hungry models can currently deliver. Four product categories illustrate the translational arc from concept to commerce. Human milk oligosaccharides, led by 2'-fucosyllactose, represent the most mature success. Fermentation titers have surpassed 100 g/L; the global market exceeds one billion USD. Multiple species, including 3-FL, LNT, LNT, 3'-SL, and 6'-SL, have reached industrial production. Glycoengineered antibodies demonstrate glycan-level pharmacology. Removal of core fucose from IgG1 Fc enhances Fc γ RIIIa binding by approximately 50-fold, boosting ADCC from therapeutically marginal to clinically decisive. The FDA approval of afucosylated obinutuzumab for chronic lymphocytic leukemia validated this principle. Terminal galactosylation tunes complement activation; α 2,6-

sialylation extends serum half-life. These parameters are independently controllable, making the Fc glycan a multi-dimensional tuning interface. Glycoconjugate vaccines are being reshaped by recombinant bioconjugation. Engineered *E. coli* enzymatically transfers target oligosaccharides to carrier proteins, eliminating the chemical coupling and multi-step purification that dominate traditional conjugate vaccine manufacturing. Functional polysaccharides have fulfilled the promise of metabolic engineering at scale. Hyaluronic acid, produced at hundred-ton scale in engineered *Bacillus subtilis*, now serves ophthalmic surgery, joint injection, and aesthetic medicine. Its success validated the general design principle of push – pull UDP-sugar flux control. Underlying all four cases is a common trajectory: from empirical iteration toward predictive, computation-guided design, powered by the modular assembly of standardized biological parts.

Three grand challenges define the next decade. The first is deciphering the glyco-code, the rules that translate glycan



structure into biological function. Unlike the genetic code, this mapping is non-linear, multivalent, and context-dependent. Progress demands the convergence of glycan microarray technology, high-resolution mass spectrometry, and AI-driven pattern recognition. It also demands training data that currently does not exist in sufficient quantity. The second is transitioning from trial-and-error to intelligent design. This requires comprehensive, standardized GT kinetics databases. It requires embedding AI prediction models within automated biofoundry workflows. It requires low-resource machine learning methods that can learn from sparse, heterogeneous datasets, a fundamentally different challenge from the data-rich ecosystems of proteomics and genomics. The third is bridging the lab-to-plant gap. Cell factory robustness must withstand industrial fermentation conditions, including temperature swings, pH drift, and osmotic stress. Phage contamination remains an existential threat in bacterial chassis. Downstream purification still consumes 50 to 70 percent of total production costs.

Beyond these challenges, emerging frontiers are reshaping the boundaries of the field. C1 feedstock-driven glycan biosynthesis, using CO₂, methanol, or formate as sole carbon sources, promises to decouple carbohydrate production from arable land and plant photosynthesis. Non-natural glycan therapeutics, including fluorinated antivirals, thio-sugar glycosidase inhibitors, and amino-glycoside antibiotic potentiators, extend the accessible chemical space beyond nature's monosaccharide repertoire. Personalized glycan vaccines designed against patient-specific tumor glycoforms and CRISPR-like programmable glycan editing tools point toward a future where glycan structures can be edited as precisely as genes. Glycosynthetic biology is completing a fundamental transformation. It is moving from discovering what nature provides to designing and manufacturing what we desire. Carbohydrates are emerging from the neglected periphery of molecular biology to a frontier at the intersection of discovery and industry. The destination is clear: a standardized glycan operating system for the programmable, on-demand biomanufacturing of structurally defined glycans.

Keywords: Glycosynthetic Biology; Glycosyltransferase; Metabolic Engineering; Cell Factory; Sugar Nucleotide; Human Milk Oligosaccharides; Cell-Free Synthetic System

糖类(carbohydrates)是自然界中结构最为复杂的生物大分子之一。与核酸和蛋白质的线性模板驱动合成不同,糖链的生物合成在多个层次上展现出高度复杂性。糖苷键的连接位置(1→2、1→3、1→4、1→6)、异头构型(α或β)、分支模式和修饰基团的组合均由定位在不同亚细胞区室的GTs协同决定^[1, 2]。这种多维度的结构自由度使糖类成为生命体中信息密度最高的生物分子。Laine的经典计算表明,由6种常见单糖通过不同连接方式形成的己糖寡糖理论异构体可达 1.05×10^{12} ,远超相同单元数氨基酸或核苷酸所能形成的组合多样性^[3]。这种结构复杂性赋予了糖类在生命过程中无可替代的功能广度:细胞表面的糖萼(glycocalyx)编码细胞类型、发育阶段和生理状态的信息;蛋白质的糖基化修饰调控其折叠、稳定性、半衰期和生物活性;病原体-宿主识别、免疫

监视逃逸、肿瘤发生发展等关键过程中,糖链均扮演核心角色^[4]。

然而,糖科学长期面临"结构解析难、化学合成难、规模化制备难"的三重瓶颈^[5]。传统糖类研究遵循“解析-发现-合成”的经典范式:从天然材料中分离纯化糖类物质,通过核磁共振、质谱等技术解析其结构,探索其生物学功能,再尝试通过化学方法进行合成验证。这一路径面临三重根本性挑战:第一,天然糖链的微观不均一性(microheterogeneity)特性,使得获得结构均一的糖链样品极为困难;化学合成涉及繁琐的保护基操作和立体化学控制,合成效率低下;从天然来源提取则受限于原料和产率。这些根本性挑战使糖科学长期处于生命科学研究的“暗物质”领域。我们知晓糖类的重要性,却缺乏高效手段来制备和操控它们。

合成生物学 (synthetic biology) 的蓬勃发展正在根本性地改变这一格局。合成生物学的核心思想是将生命系统“工程化”，将生物系统重构为可预测、可编程的工程实体^[6, 7]。将这一范式引入糖科学领域，催生糖合成生物学这一新兴交叉学科。其核心目标是通过构建和优化糖基化生物系统，实现结构明确、功能特异的糖分子的高效、绿色、可编程合成。这一范式转换的本质，是从“发现自然已有的糖”转向“设计并制造想要的糖”。

过去二十年中，该领域的关键里程碑包括：细菌N-糖基化途径向大肠杆菌的成功移植^[8]、酿酒酵母糖基化系统的人源化改造^[9]、CHO细胞中均一糖型糖蛋白的CRISPR工程化生产^[10]、以及2'-岩藻糖基乳糖 (2'-FL) 等母乳寡糖 (HMOs) 达到吨级发酵规模^[11, 12]。在基础研究前沿，无细胞糖蛋白合成体系^[13]和从CO₂到淀粉的体外人工合成途径的出现^[14]，标志着糖合成生物学正在突破活细胞体系的固有限制。

1 糖合成生物学目前的核心瓶颈

尽管上述突破令人鼓舞，糖合成生物学仍面临若干核心瓶颈。其深层根源在于，糖链合成属于无分子模板驱动的生物大分子合成体系。这一内在特征使糖链具备高度的结构异质性与可塑性，也赋予了它们丰富多样的生物学功能。然而，当研究者试图将糖基化系统改造为工程化生产平台时，这些特性便转化为系统性的代价。

肽键和磷酸二酯键各仅有1种化学连接方式，这使得20种氨基酸和4种核苷酸的聚合可分别通过核糖体和DNA/RNA聚合酶的单一催化机制完成。糖苷键则涉及连接位置 (1→2/1→3/1→4/1→6等)、异头构型 (α/β)、供体糖和受体糖的多维组合，需要每种糖苷键类型匹配对应的糖基转移酶催化机制，而无法由“通用酶”统一完成。糖基转移酶与特定糖苷键反应的一对一关系是自然选择在给定化学约束下的最优解，但作为工程系统的代价，这意味着糖基转移酶元件必须被逐一表征。

蛋白质合成虽涉及20种氨酰-tRNA，但所有氨基酸通过同一类氨酰-tRNA合成酶 (aaRS) 家族加载到结构高度保守的tRNA支架上——底物多

样性被封装在统一接口之后，翻译机器只需识别一种分子类型。糖核苷酸体系则截然不同：每种糖核苷酸 (UDP-Glc、GDP-Fuc、CMP-Neu5Ac等) 拥有独立的合成途径，不存在统一的糖基载体。这一独立而复杂的途径结构，使糖核苷酸胞内调控天然分散，并与中心碳代谢 (糖酵解、核苷三磷酸代谢) 深度耦合，导致任何单一途径的工程扰动都可能引发全局代谢的不可预测响应。

核酸和蛋白质的序列与功能之间存在可解析的映射。三联体密码子决定氨基酸，氨基酸序列决定三维结构，三维结构决定功能。糖链的序列与功能则缺乏此类映射：同一糖表位在不同骨架 (N-糖链 vs O-糖链 vs 糖脂) 上可呈现完全不同的活性，多价效应的非线性叠加使单价亲和作用无法直接外推为生物学功能。这使糖链结构-活性关系的系统解析面临认识论层面的根本困境。

上述情况共同定义了糖合成生物学的工程化边界。不是单点技术突破可逐一解决的操作性困难，而是需要全新分析框架来驾驭的系统性挑战。这些挑战在不同技术层面上，集中表现为当前糖合成生物学的三个瓶颈问题。

瓶颈一：GT元件库功能表征严重不足。CAZy数据库已注释超过50万个GT序列 (分布于120余个家族)，但其中经实验验证功能的不足5%^[1]。绝大多数GT的底物特异性、动力学参数 (*k_{cat}*、*K_m*) 及表达兼容性仍属空白，使得糖合成途径的设计在很大程度上仍依赖试错筛选，而非基于标准化元件的理性组装。AlphaFold2等高精度结构预测工具正在加速远端同源GTs的功能推断，但其预测结果的实验验证通量仍远落后于序列发现速度^[15]。

瓶颈二：糖核苷酸供给调控精度不足。糖核苷酸作为糖基转移酶的必需活化供体，其胞内浓度和比例直接决定糖基化反应的热力学驱动力和产物分布^[16, 17]。糖核苷酸代谢网络与中心碳代谢密切交织，对单一途径的扰动可能引发全局代谢的不可预测响应。现有代谢工程策略尚难以实现对多种糖核苷酸胞内水平的独立、精准、动态调控。

瓶颈三：糖链结构-活性关系解析欠缺。糖链的结构-功能关系是非线性的、多价的、高度依赖

上下文的^[18]。糖链结构数据库的规模较蛋白质结构数据库小数个量级，且高质量标注数据极为匮乏，极大制约了基于数据驱动的构效关系挖掘^[19]。

为突破糖链非模板驱动合成的固有局限，亟需发展可计算的糖基转移酶元件模型与糖链结构预测工具，建立标准化的糖基转移酶元件库和糖核苷酸代谢模块，搭建高通量糖链分析平台及自动化实验设施，并构建与实验数据高度耦合的AI/

ML模型。由此，糖合成生物学有望从经验驱动的试错迭代范式，跃迁至数据驱动预测-设计新范式。

本文将围绕四项核心使能技术展开：糖基化元件工程、代谢途径工程与细胞工厂设计、无细胞合成与前沿体系，随后讨论产业化应用和AI/ML赋能，最后展望领域的核心挑战与未来方向（图1）。

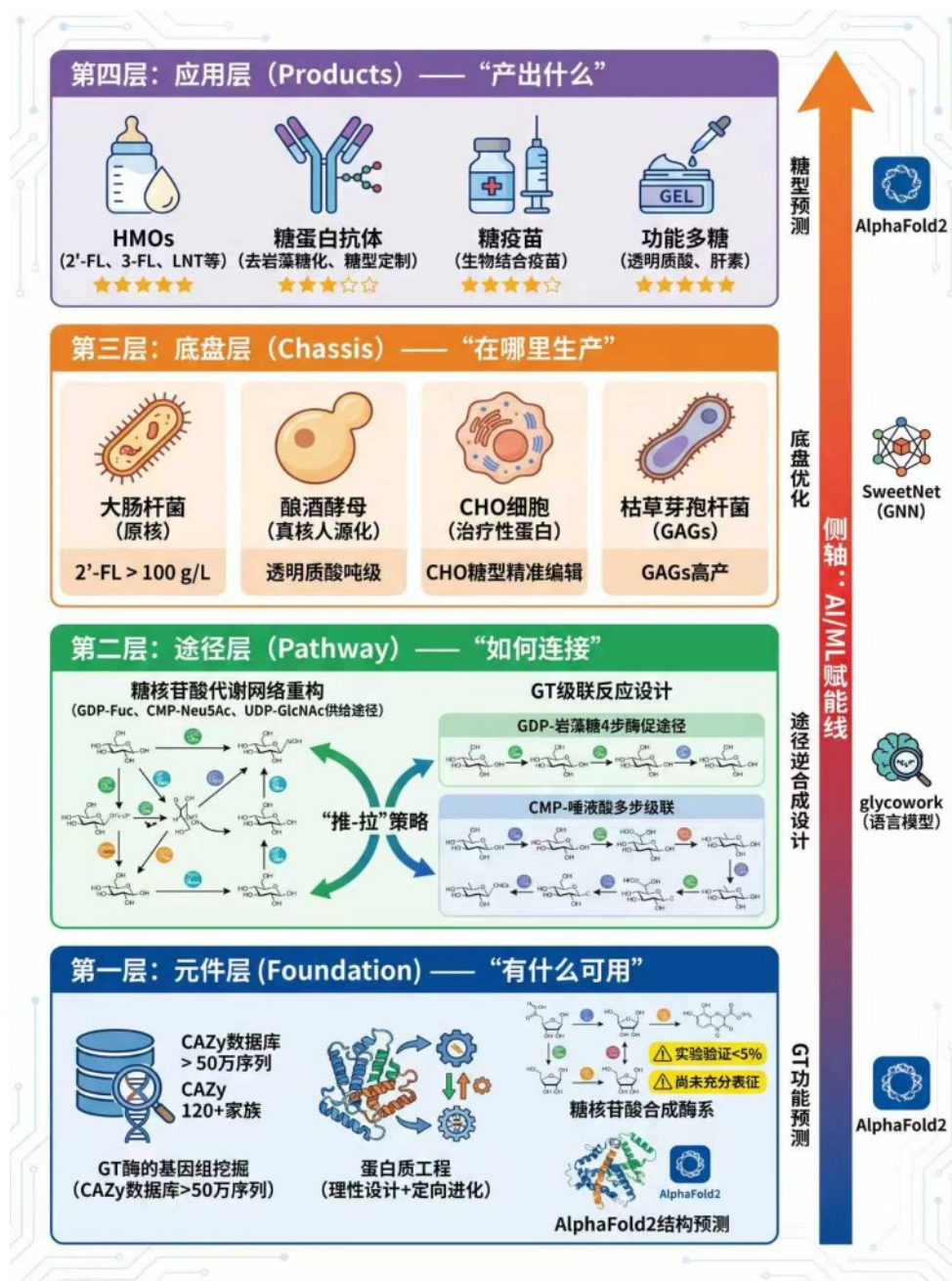


图1 糖合成生物学和新技术框架图

Fig.1 Framework Diagram of Glycosynthetic Biology and Emerging Enabling Technologies

2 糖基化元件工程

2.1 糖基转移酶的挖掘与CAZy资源

糖基转移酶 GTs (EC 2.4.x.y) 是催化糖苷键形成的核心酶超家族, 以糖核苷酸 (UDP-葡萄糖、GDP-岩藻糖、CMP-Neu5Ac 等) 为活化供体。GTs 采用两种催化机制: 反转型 (inverting) 通过 SN2-like 单步亲核取代直接翻转异头构型; 保持型 (retaining) 则通过双置换机制 (形成共价糖基-酶中间体) 保留异头构型。结构上, GTs 主要折叠为 GT-A 型 (含保守 DxD 基序, 依赖二价金属离子配位糖核苷酸的二磷酸基团) 和 GT-B 型 (两个 Rossmann 折叠域通过底物结合诱导的域间闭合催化糖基转移), 另有少量 GT-C 型膜蛋白折叠。这一结构-机制分类框架为 GT 工程改造提供了理性设计的分子基础。CAZy (Carbohydrate-Active enZymes) 数据库将 GTs 划分为超过 120 个序列家族 (截至 2025 年), 每个家族共享保守的结构折叠和催化机制, 其精细人工注释涵盖底物特异性、三维结构和催化残基, 为 GTs 的基因组挖掘提供了不可替代的基础设施^[1]。AlphaFold2 的高精度蛋白质结构预测能力进一步推动了基于结构的远端同源 GTs 发现和功能注释^[15]。

在工业应用中, 基因组挖掘已成功鉴定出多种关键 GTs。如甜叶菊 (*Stevia rebaudiana*) 来源的 UDP-糖基转移酶 UGT76G1, 在莱鲍迪苷 D (Reb D) 和莱鲍迪苷 M (Reb M) 的生物合成途径中催化多达 8 种不同的 1,3-葡萄糖基化反应, 其底物谱的"变色龙"特性为甜味剂的酶法定制合成提供了关键催化剂^[20]。

2.2 糖基转移酶的蛋白质工程

天然糖基转移酶的催化效率、稳定性、底物特异性和立体选择性往往难以满足工业需求。蛋白质工程通过结构指导的理性设计与定向进化的联合运用, 已成为突破天然酶性能天花板的核心手段^[21]。

以 GM1 神经节苷脂的化学酶法全合成为例, 研究者对关键糖基转移酶进行结构指导改造和过程工程优化, 实现了从简单前体出发的短路线全

合成^[22], 总产率显著优于传统化学合成路线。在更复杂的神经节苷脂合成中, GM3 (唾液酸乳糖神经酰胺) 的化学酶法合成通过 $\alpha 2, 3$ -唾液酸基转移酶对乳糖基受体的区域选择性唾液酸化实现; GD3 的合成则进一步依赖 $\alpha 2, 8$ -唾液酸基转移酶在 GM3 的唾液酸残基上引入第二个唾液酸; GT1b 等复杂神经节苷脂的合成需要多步糖基转移酶级联反应的有序组合^[23]。这些案例共同表明: 神经节苷脂化学酶法合成的核心挑战在于糖基转移酶的区域选择性和底物耐受性, 蛋白质工程通过重塑活性口袋的几何和化学环境来调节酶的底物谱和催化效率, 已成为突破这一瓶颈的系统性策略。

Zhang 等通过祖先序列重建策略, 获得了一种可高效催化 2-苯乙基- β -D-吡喃葡萄糖苷和红景天苷生物合成的祖先酶突变体 [18]^[24]。Lu 等通过对黄酮糖基转移酶 UGT75AJ2 的综合工程改造, 结合虚拟筛选和迭代定点突变, 获得催化活性提升 128 倍的突变体 Mut4-1, 同时通过结构指导的供体结合口袋改造实现了供体宽泛性和区域选择性糖基化^[25]。

糖基转移酶蛋白质工程的两种主流策略包括理性设计和定向进化, 各有其优势边界与固有局限, 理解其本质差异是实现高效酶改造的前提。

理性设计依赖于对 GT 三维结构和催化机制的深入理解, 通过定点突变、活性口袋重塑、loop 工程和结构域融合等手段精准调控酶的底物特异性、区域选择性和催化效率。其核心优势在于: 突变文库规模小 (通常 $10^2 - 10^3$ 个变体), 筛选成本低; 改造策略具有可解释性和可迁移性——对同家族其他 GT 的改造具有指导价值; 适合引入非天然底物偏好和改变区域选择性等需要精确结构操控的场景。然而, 理性设计的根本局限在于对 GT 结构-功能关系的认知深度——当活性位点残基与远端别构位点之间存在复杂的耦合网络时, 单一或少数位点的理性突变往往难以实现显著的催化性能跃迁。

定向进化方面, Arnold (2018 年诺贝尔化学奖) 系统阐述的蛋白质工程范式^[21]。通过引入随机突变并结合高通量筛选逐步富集有益突变, 在改善 GTs 热稳定性、有机溶剂耐受性和 pH 适应性方面成效显著。其核心优势在于不依赖先验结构

知识,能够挖掘远端残基协同突变和意外活性增益。这些不可能预先设计的突变组合往往产生远超理性设计的性能突破。然而,定向进化的根本瓶颈在于筛选通量:催化产物的常规检测依赖 HPLC 或 LC-MS,单日筛选通量通常仅为 $10^2 - 10^3$ 个变体,远低于定向进化所需的有效文库覆盖度(通常 $10^4 - 10^6$)。近年来发展的荧光标记底物法、生物传感器偶联检测策略以及微流控液滴筛选平台正在推动这一瓶颈的实质性突破。

实际应用中,两种策略的组合运用已被证明是最具成效的路线:通过理性设计构建智能化的突变文库,显著压缩筛选空间,再通过定向进化从精选文库中富集最优变体。Lu 等对 UGT75AJ2 的 128 倍活性提升正是这一组合策略的成功范例,先通过虚拟筛选锁定关键残基位点,再通过迭代定点突变积累有益突变^[25]。

2.3 糖核苷酸代谢工程

糖核苷酸是 GTs 的必需共底物,其胞内供给水平直接决定糖基化反应效率^[16, 17]。以 GDP-岩藻糖为例,其从头合成依次经 mannose-6-phosphate isomerase (ManA)、phosphomannomutase (ManB)、GDP-mannose pyrophosphorylase (ManC)、GDP-mannose 4,6-dehydratase (GMD) 和 GDP-4-keto-6-deoxymannose 3,5-epimerase/4-reductase (WcaG/FX) 的多步催化;CMP-唾液酸的合成则涉及 UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE/MNK)、N-acetylneuraminic acid synthase (NANS) 和 CMP-Neu5Ac synthetase (CSS 的多步级联转化^[26]。

GDP-岩藻糖途径强化,以大肠杆菌中 GDP-岩藻糖途径的经典工程化改造为例,通过共过表达 manB (phosphomannomutase)、manC (mannose-1-phosphate guanylyltransferase)、gmd (GDP-mannose 4,6-dehydratase) 和 wcaG (GDP-fucose synthase) 四个关键基因,可显著提升胞内 GDP-岩藻糖的积累水平,为 2'-FL 等岩藻糖基化产物的高效合成提供充足的活化供体池。该策略已成为 2'-FL 生产菌株构建的标准代谢工程方案^[11, 12, 27]。

CMP-Neu5Ac 途径优化,在大肠杆菌中重构 CMP-Neu5Ac 合成途径,通常涉及共过表达 neuB

(N-acetylneuraminic acid synthase)、neuC (UDP-GlcNAc 2-epimerase) 和 synX (CMP-Neu5Ac synthetase) 基因。值得注意的是,NeuB 催化的磷酸烯醇式丙酮酸 (PEP) 依赖性缩合反应是途径的主要限速步骤,PEP 的胞内可用性往往成为制约 CMP-Neu5Ac 合成通量的关键瓶颈。Zhang 等系统综述了唾液酸及唾液酸化 HMO 的微生物合成路径,包括 CMP-Neu5Ac 途径工程化改造与 PEP 供给优化策略^[28]。

UDP-GlcNAc 是 N-糖基化和 GAGs 生物合成中的核心供体。其从头合成从果糖-6-磷酸出发,经 GlmS (glucosamine-6-phosphate synthase)、GlmM (phosphoglucosamine mutase) 和 GlmU (UDP-GlcNAc pyrophosphorylase) 三步催化。代谢工程策略包括:(1) 强化 glmS-glmM-glmU 的共过表达提升途径通量;(2) 敲除 NagA (N-acetylglucosamine-6-phosphate deacetylase) 消除 UDP-GlcNAc 的分解回流;(3) 过表达 UDP-GlcNAc 转运蛋白以增强胞内可利用水平。这些策略在透明质酸 (HA) 的枯草芽孢杆菌发酵生产中得到了系统应用——UDP-GlcNAc 和 UDP-GlcA 的双途径协同强化是实现 HA 高产 (>50 g/L) 的关键代谢工程基础^[29]。

代谢工程优化策略包括:(1) 关键酶基因的过表达强化途径通量;(2) 竞争性途径的敲除减少碳流分流;(3) 反馈抑制抗性突变体的筛选;(4) 辅因子再生系统的整合^[30]。Zhu 在其关于哺乳动物细胞生物制药生产的综述中系统论述了表达系统选择对糖蛋白糖型的关键影响^[30]。模块化工程将复杂糖合成途径分解为独立的代谢模块分别优化后组合,是当前应对多目标协同优化的主流策略。此策略在稀有人参皂苷的酵母合成中得到了成功验证:研究者将三七 (*Panax notoginseng*) 的 GT 基因系统鉴定后,在工程酵母中重构了稀有人参皂苷的生物合成途径^[31]。

3 代谢途径重构与细胞工厂设计

3.1 原核底盘:细菌糖基化系统

大肠杆菌是糖合成生物学中应用最广泛的底盘之一。该领域的奠基性突破来自 Wacker 等^[8]:研究者发现空肠弯曲杆菌 (*Campylobacter jejuni*)

的N-糖基化基因簇 (*pgl* locus) 可功能性转移至大肠杆菌, 使原本不具备N-糖基化能力的宿主获得了完整的N-糖基化机器, 包括PglB寡糖基转移酶在周质空间催化预组装寡糖向受体蛋白天冬酰胺残基的整块转移。这一突破不仅首次在原核体系中实现了N-糖蛋白的从头合成, 更重要的是验证了糖基化途径可跨物种移植的概念, 为后续大肠杆菌作为糖蛋白生产底盘的系统开发奠定了理论基础。

Valderrama-Rincon 等进一步在大肠杆菌中构建了真核N-糖基化途径的原型^[32]: 通过引入酵母来源的甘露糖基转移酶和糖基水解酶, 实现了Man₃GlcNAc₂核心结构的合成, 为大肠杆菌作为糖蛋白生产底盘的可行性提供了概念验证。

3.2 HMO的从头合成

HMOs的大肠杆菌从头合成是糖合成生物学产业化最成功的案例。以2'-FL为例, 主要通过在大肠杆菌染色体中整合GDP-岩藻糖合成途径 (*manB/manC/gmd/wcaG*) 和 α 1,2-岩藻糖基转移酶基因序列, 构建可稳定遗传的生产菌株^[27], 通过转录组辅助的糖转运蛋白筛选, 过表达底物转运蛋白增强乳糖摄取效率^[33], 优化发酵工艺实现高产^[34]。LNT的合成则通过共表达 β 1,3-GlcNAcT (LgtA) 和 β 1,3-GalT (WbdO) 实现从乳糖到LNT II再到LNT的顺序合成^[35]。

当前HMO技术竞争的焦点已从转向合成成本优化。通过代谢途径精简减少副产物、开发廉价碳源替代乳糖、以及计算机辅助的途径通量优化。2'-FL, 3-FL、LNT、LNT、3'-SL和6'-SL等多种HMO也已实现产业化, 其中部分品种的发酵滴度达到30 - 80 g/L。Sprenger等^[12, 36]的综述系统总结了HMO微生物生产的工程策略与产业化进展。

3.3 真核底盘的糖基化人源化

酿酒酵母具备真核糖基化加工系统, 但野生型酵母产生高甘露糖型N-糖链, 在人体中具有强免疫原性。Hamilton等在酵母糖基化人源化方面取得了里程碑式突破^[9]: 通过敲除*OCH1*基因 (α 1,6-甘露

糖基转移酶) 消除高甘露糖延伸, 并引入人源GnT I、GnT II、UDP-GlcNAc转运蛋白、 β 1,4-半乳糖基转移酶 (β 1,4-GalT) 和 α 2,6-唾液酸基转移酶 (α 2,6-SiaT), 首次在酵母中实现了末端唾液酸化的复杂人源N-糖链合成。

CHO细胞是治疗性抗体和糖蛋白药物的主流生产底盘。Yang等利用CRISPR/Cas9介导的多重基因编辑, 系统性敲除CHO细胞中19个GT基因, 结合转录组学和糖组学分析, 实现了糖型精准工程化定制^[10]。Kotidis等整合了CHO细胞中全部糖基化相关酶的表达量、底物浓度和动力学参数, 构建了首个CHO细胞糖基化反应网络完整计算模型 (CHOglycoNET), 使得在计算机上预测特定基因编辑对糖型分布的影响成为可能^[37]。

3.4 糖胺聚糖 (GAGs) 的生物合成

糖胺聚糖 (glycosaminoglycans, GAGs), 包括肝素/硫酸乙酰肝素、透明质酸 (HA)、硫酸软骨素和硫酸皮肤素, 构成糖合成生物学中兼具基础研究深度和产业化广度的独特分支。肝素是临床上使用最广泛的抗凝药物之一, 传统上依赖猪小肠黏膜提取, 面临供应链生物安全风险 (如非洲猪瘟导致的肝素粗品短缺)。通过在大肠杆菌或酵母中重构肝素前体的合成途径, 从果糖-6-磷酸出发, 经KfiA (N-乙酰葡萄糖胺基转移酶)、KfiC (糖醛酸基转移酶) 等多步酶促反应, 结合体外化学/酶法磺酸化修饰 (N-乙酰葡萄糖胺/N-磺基转移酶、C5-差向异构酶、2-O/6-O/3-O磺基转移酶), 已实现生物工程肝素的克级制备^[38]。透明质酸 (HA, 由葡萄糖醛酸和N-乙酰葡萄糖胺交替连接而成) 的微生物发酵法生产, 通过在枯草芽孢杆菌或乳酸乳球菌中过表达透明质酸合酶 (HasA) 并强化UDP-GlcA和UDP-GlcNAc供给途径 (图2), 已达到百吨级商业规模, 广泛应用于眼科手术、关节腔注射和医美填充剂领域。

GAGs合成生物学的核心挑战在于: 磺酸化模式 (sulfation code) 的精确控制直接决定生物学活性, 例如肝素中特定3-O-磺酸化五糖序列 (GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S) 是抗凝血酶III高亲和力结合的必要条件^[39], 而特定

硫酸化模式的微生物酶法精准重现在技术上仍面临巨大挑战。

肝素生物合成涉及至少四种磺酸化转移酶(N-、2-O-、6-O-和3-O-磺基转移酶)，每种酶对底物序列和已存在的磺酸化修饰具有严格的依赖性。例如，3-O-磺基转移酶仅识别N-磺酸化且6-O-磺酸化的特定糖序列，而天然的特定亚型（如3-OST-1）仅作用于抗凝血酶III结合位点的五糖基序。这种多层级的底物识别规则使磺酸化转移酶的功能表征和工程改造面临严峻的挑战。

目前主流的磺酸化策略可归结为（1）全磺酸化策略，对heparosan前体进行化学或酶法定量磺酸化，获得高度硫酸化的产物。此路线工艺相对成熟，但产物磺酸化模式分布宽泛，难以精确控制特定活性基序的丰度。（2）位点选择性磺酸化策略，利用不同磺酸化转移酶的底物特异性差异，通过调节酶浓度、反应时间和底物浓度等参数，有序引入特定位置的磺酸基团。此路线理论上可精确控制磺酸化模式，但多步酶促反应的工序复杂性、中间产物纯化难度和总体产率限制了其工业化应用。两种策略的联合，先通过全磺酸化获得高硫酸化骨架，再通过选择性脱磺酸酶处理暴露特定活性基序，代表了当前最具前景的折中方案。

3.5 规模化生产中底盘菌株的稳定性

从实验室摇瓶到工业发酵罐，糖合成细胞工厂面临一系列工程生物学挑战，其中底盘菌株的长期遗传稳定性和代谢鲁棒性是最核心的制约因素。

高产菌株通常携带多个高表达外源基因盒和敲除突变，这些遗传操作在无选择压力条件下可能引发适应性回复突变、质粒丢失或重组导致的基因座失活。以2'-FL生产菌株为例，染色体整合策略虽显著提升了遗传稳定性，但长期连续培养中仍有部分克隆出现GDP-岩藻糖合成基因的表达衰减^[27]。

此外，高浓度目标产物对底盘细胞的渗透压胁迫和代谢毒性是导致产物滴度在放大过程中大幅低于实验室水平的常见原因。增强底盘细胞的胁迫耐受性，筛选耐高渗突变株、或理性强化相容性溶质的合成途径，是应对这一挑战的有效策略。

大肠杆菌作为工业化糖产品生产的首选底盘，面临噬菌体污染的持续威胁。工业发酵中噬菌体感染的爆发可导致整批发酵失败，造成重大经济损失。些工程考量虽非糖合成生物学独有的挑战，但在高密度、大规模、长周期发酵的生产场景中具有特殊重要性。

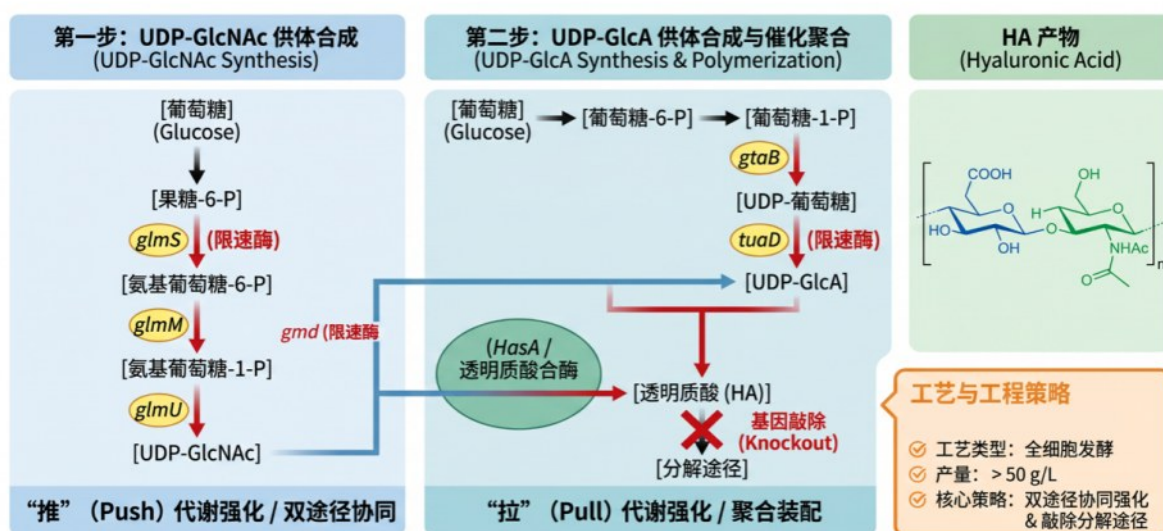


图2 透明质酸 (HA) 枯草芽孢杆菌合成代谢路线

Fig. 2 Metabolic Biosynthetic Pathway of Hyaluronic Acid (HA) in *Bacillus subtilis*

3.6 抗体糖工程的里程碑

Jefferis 在其经典综述中系统阐述了糖基化作为优化抗体治疗药物的多维度策略^[40]。Shields 等的里程碑式发现，去除 IgG1 Fc 段 N-糖链的核心岩藻糖可使抗体与 FcγRIIIa 的亲和力提升 50 倍，从而大幅增强 ADCC 活性，开启了抗体糖工程的新纪元^[41]。Li 等进一步阐明 Fc 糖工程调控 IgG 效应功能的精细分子机制^[42]。Huang 等发展了通用化化学酶法糖工程平台，利用内切糖苷酶和 GT 的级联反应，在完整 IgG 抗体上精确引入或去除特定糖基，实现了 ADCC、CDC 等多重效应功能的定量调控^[43]。Pereira 等系统综述了去岩藻糖化抗肿瘤抗体的开发进展^[44]。

4 前沿合成体系

4.1 无细胞糖蛋白合成

无细胞糖蛋白合成体系突破了活细胞代谢网络对途径设计的限制，研究者可在开放的体外环境中自由调配酶、辅因子和底物浓度，不受细胞膜屏障和毒性中间代谢产物积累的约束。Jaroentomeechai 等建立了首个高效的无细胞糖蛋白合成平台，在大肠杆菌无细胞转录-翻译体系中补充含有 PglB 寡糖基转移酶的细胞提取物，实现翻译、折叠和 N-糖基化的一步完成^[13]。Kightlinger 等将该平台发展为模块化系统，不同来源的 GT 和糖基化途径可作为独立插件组合，为定制化糖蛋白的快速原型设计提供了前所未有的灵活性^[45]。

无细胞体系的另一独特优势在于可直接利用非天然糖核苷酸前体，为制备带有非天然糖链的"生物正交"糖蛋白开辟了新路径。

4.2 糖核苷酸的体外酶法合成

体外多酶级联体系将纯化或粗提酶组分在反应器中组装，从廉价底物出发合成糖核苷酸或寡糖，是连接酶元件发现与工业化生产的关键技术桥梁。与微生物全细胞发酵相比，体外级联体系具有反应条件可控、副产物少、产物分离简单、无细胞代谢竞争等优势，但也面临酶稳定性、辅因子循环和多酶协同等工程化挑战。

Roberts 等系统综述了糖核苷酸合成中生物过程优化的最新进展和挑战，特别讨论了连续流动反应器中酶固定化和辅因子再生的策略^[46]。De Bruyn 等全面阐述了基于 UDP-糖的小分子糖基化生物催化策略^[16]。Ruffing 和 Chen 的早期综述提出过量表达糖基转移酶基因和强化核苷糖供应的"推-拉"策略，奠定了微生物代谢工程生产寡糖和多糖的理论框架^[17]。

4.3 从 CO₂到糖：C1原料的生物转化

Cai 等报道的从 CO₂到淀粉的无细胞化学酶法合成是糖合成生物学近年最具影响力的突破之一^[14]。该体外人工合成途径由 11 步核心酶促反应组成，能量效率远超自然光合作用的 Calvin 循环，为农业工厂化提供了技术原型。更早期的基础工作中，Antonovsky 等已在大肠杆菌中构建了从 CO₂到糖的合成代谢途径，通过引入 Rubisco 和 PRK 酶

表1 主流细胞工厂性能对比表
Table 1 Performance comparison of major cell factory chassis for glycan and glycoprotein production

对比维度	大肠杆菌	酿酒酵母	CHO 细胞	枯草芽孢杆菌
底盘优势	遗传操作成熟、生长快、成本低	真核分泌系统、蛋白折叠能力强	人源化糖基化、已获批生产	GRAS 安全菌株、蛋白分泌能力强
糖基化能力	原核简单糖基化（可移植 N-糖基化）	高甘露糖型→人源化改造	复杂 N-糖/ O-糖、可糖型定制	无天然糖基化（异源表达）
代表产品	2'-FL、3-FL、LNT、糖疫苗	人源化糖蛋白（早期）	治疗性抗体、促红细胞生成素	透明质酸、硫酸软骨素
发酵滴度	2'-FL: 121.4 g/L(64 h)；最高 141.27 g/L(45 h)	糖蛋白 mg 级	抗体 g/L 级	HA: 19.38 g/L
工业瓶颈	噬菌体污染、内毒素、无复杂糖基化	人源化程度有限、产量低	成本高、糖型异质性、周期长	产量偏低、遗传工具相对有限

验证了C1到C6碳骨架转化的生物学可行性^[47]。该方向的研究意义不止于替代传统农业的愿景。C1原料(CO₂、甲醇、甲酸)作为底物可将糖的生产从依赖植物光合作用转变为工业发酵模式。

4.4 非天然糖与生物正交化学

基于Bertozzi实验室奠基性工作的生物正交化学,使活细胞甚至活体动物中的糖链时空分辨标记和追踪成为可能^[48]。Laughlin等建立了叠氮糖代谢标记的标准方法学:将叠氮修饰的N-乙酰甘露糖胺(ManNAz)孵育细胞后,内源代谢途径将其活化为叠氮唾液酸并掺入新生糖链,随后通过Staudinger连接或点击化学反应实现可视化^[49]。Agard等发展的应变促进叠氮-炔基环加成(SPAAC)开创了无铜点击化学的先河^[50]。Baskin等将该技术推向活体动物成像^[51]。生物正交化学与糖合成生物学交叉融合,通过引入非天然糖前体并以生物正交反应激活,有望实现可编程的时空分辨糖链标记和靶向药物递送^[52]。

5 产业化应用

5.1 母乳寡糖(HMOs)

HMOs是糖合成生物学产业化最成功的案例之一,全球市场规模在2025年已突破10亿美元,预计在2030年前将以约12-15%的年复合增长率持续扩大。母乳中已鉴定出超过200种结构各异的寡糖,2'-FL含量最高(在人乳中约2-3 g/L)^[53]。当前,工程化大肠杆菌的2'-FL发酵滴度已超过100 g/L(时空产率约1.5-2.0 g/L/h),底物转化率可达0.5-0.6 g/g。

2'-FL的产业化成功是多重因素系统性协同的结果,其中尤其重要的是,2'-FL合成代谢途径相对简单,其合成仅需4个外源基因(GDP-甘露糖脱氢酶GMD、GDP-岩藻糖合酶WcaG/FX、乳糖通透酶LacY和 α 1,2-岩藻糖基转移酶FucT2),途径短则代谢负担低,支路产物少(图3)。除2'-FL外,3-FL、LNT的从头合成途径已实现^[35]、LNnT、3'-SL和6'-SL等多种HMO也已实现产

业化^[12]。

然而尽管HMO已实现吨级商业化生产,产业化挑战仍制约着品类扩展和成本下行。一是底物成本。主要产品依赖乳糖作为起始前体,乳糖占原料成本的40-60%,且价格受乳制品市场波动影响显著。开发以葡萄糖或蔗糖为底物的2'-FL从头合成途径是降低成本的关键方向;二是下游纯化成本。当前HMO的主流纯化工艺包括活性炭、离子交换色谱及模拟移动床等工艺,分离纯化本高。色谱纯化(离子交换、活性炭色谱和模拟移动床色谱的组合)占HMO纯化总成本的60%以上^[54]。低副产物菌株的开发(减少单糖、非靶向寡糖和培养基组分的残留)是从上游缓解纯化压力的核心策略。此外,结晶法和膜分离技术的整合正在逐步降低色谱在纯化流程中的成本占比。

以葡萄糖、蔗糖或C1原料(甲醇、甲酸)替代乳糖的HMO从头合成途径已有多篇概念验证报道,但碳转化效率和经济性仍远未达到与乳糖投料路线竞争的水平。当前技术竞争的焦点已从能否合成转向成本优化,通过代谢途径精简减少副产物,降低下游纯化成本,开发廉价碳源替代乳糖,以及计算机辅助的途径通量优化,以及连续发酵工艺开发。

5.2 糖疫苗

糖类抗原在疫苗开发中占据重要地位。多糖结合疫苗,比如将荚膜多糖共价偶联于载体蛋白(如CRM197、破伤风类毒素)以诱导T细胞依赖性免疫应答和高亲和力IgG抗体,是糖疫苗的成功典范^[55]。以肺炎球菌结合疫苗PCV13为例,其全球年销售额约60亿美元,是历史上最成功的疫苗产品之一。糖合成生物学提供了新路径:通过细胞工厂发酵获得结构均一的寡糖抗原,替代传统病原体培养提取法,提升疫苗批间一致性和安全性。Kay等综述了重组糖结合疫苗平台利用工程化大肠杆菌N-糖基化系统将目标寡糖直接转移至载体蛋白的最新进展,该生物结合技术省去了化学偶联步骤,已在志贺氏菌、链球菌等多种病原体疫苗开发中得到验证^[56]。Slovin等系统阐述了碳水化合物疫苗作为癌症免疫治疗策略的基础和

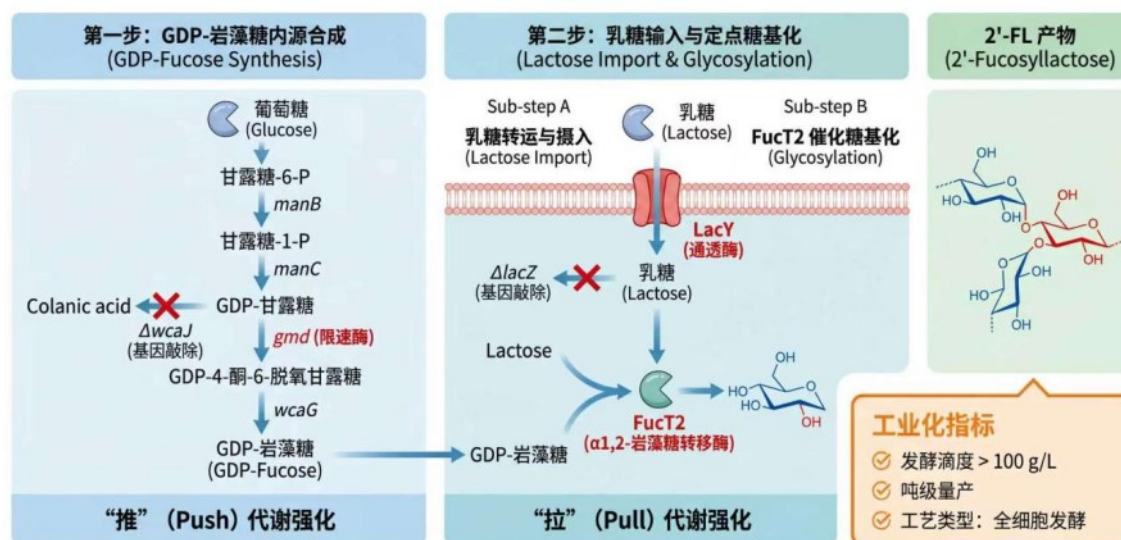


图3 2'-岩藻糖基乳糖微生物全细胞合成路线

Fig. 3 Whole-Cell Microbial Biosynthetic Pathway of 2'-Fucosyllactose

临床前研究^[57]。

糖疫苗领域仍面临不少产业化挑战。目前寡糖抗原的规模化合成的主要瓶颈仍是目标寡糖的体内滴度，远低于HMO发酵的百克升级别；此外，从实验室到GMP生产车间的放大过程中，糖基化效率和产物均一性的维持面临显著挑战；重组糖结合疫苗作为新兴产品类别，缺乏成熟的监管先例和标准化质量评价体系，各品种的审批路径存在较大的不确定性。

5.3 糖蛋白药物与抗体糖工程

绝大多数已获批的治疗性蛋白药物均为糖蛋白。以IgG1抗体为例，Fc段Asn297位保守N-糖链的核心岩藻糖缺失可使抗体与FcγRIIIa的亲合力提升50-100倍，从而将ADCC活性的EC50从约100 ng/mL降低至约1-10 ng/mL^[41]。去岩藻糖化抗CD20抗体Obinutuzumab (Gazyva) 已获FDA批准用于慢性淋巴细胞白血病，在III期临床试验中展现了优于利妥昔单抗的无进展生存期。末端半乳糖化通过与C1q结合增强CDC活性，而末端α2,6-唾液酸化通过遮蔽半乳糖残基被肝细胞ASGPR识别，可将抗体血清半衰期从约7-10天延长至约14-21天^[40]。糖基化工程因此成为下一代抗体药物的核心技术壁垒。通过精确定制Fc段N-糖链的岩藻

糖化、半乳糖化和唾液酸化水平，可独立调控ADCC、CDC和血清半衰期等多项药理学参数^[42, 43]。Pereira等系统综述了去岩藻糖化抗肿瘤抗体的开发进展，指出目前已有超过10种糖工程抗体处于临床开发阶段^[44]。

糖型均一性是糖蛋白药物质量控制的核心挑战。即使在同一发酵批次的同一重组蛋白中，N-糖链在末端半乳糖化、岩藻糖化和唾液酸化程度上呈现高度异质性。一个典型的治疗性IgG1分子可携带30种以上可分辨的糖型变异体。这种微观异质性虽然在其他生物制品中可被接受，但对于免疫效应精确依赖特定糖型的抗体药物，糖型的微观差异可能转化为疗效和免疫原性的不可控风险。实现目标糖型占比>90%的均一化生产仍依赖GT编辑、代谢重塑和培养条件优化的系统整合。

6 AI与机器学习赋能糖科学

6.1 糖链结构预测的深度学习模型

AI/ML方法正在深刻改变糖科学研究范式。近年来，多种专门针对糖链结构特点的深度学习模型相继问世。SweetNet采用图神经网络(GNN)将糖链建模为单糖节点和糖苷键连边的异构图，实现了糖链结构到免疫原性、受体结合亲和力等

功能性质的有监督预测^[58]。Bojar 团队开发的 glycowork 平台将语言模型 (Language Model, LM) 应用于糖链序列^[59]。基于 AlphaFold2 预测的蛋白质 3D 结构信息的糖链构象预测方法正在将糖链功能研究推向结构-动态-功能的三维层面,这是理解抗体 Fc 段糖链如何变构调控 FcγR 结合等关键生物学问题的重要方向。这些模型使糖链结构的计算预测部分可实现,但预测精度的提升仍严重受限于高质量标注训练数据的稀缺。

6.2 糖基转移酶功能预测

基于序列的预测方法通过训练 CAZy 数据库中已注释 GT 的序列与底物对应关系,实现在序列空间中推断未表征 GT 的可能底物谱。基于结构的预测方法受益于 AlphaFold2 的突破:针对目标 GT 预测其 3D 结构后,利用分子对接模拟候选供体核苷酸糖和受体底物在活性位点中的结合模式和能量,从而预测底物特异性。两种方法的互补构成了当前 GT 功能预测的方法框架。Gerloff 等建立了生物信息学预测与蛋白质工程实验验证相结合的 GT 活性鉴定流程,通过序列特征分析和结构建模预测候选 GT 的底物特异性,再经实验验证闭合“预测-验证”循环^[60]。近年来的预测准确率进展令人鼓舞:部分 GT 家族(如 GT1 家族的 UDP-糖基转移酶)的供体预测准确率已达 80-90%,但受体预测,特别是对复杂糖链多羟基环境的区域选择性预测,仍面临巨大挑战。

6.3 代谢路径优化与逆合成设计

AI 辅助的逆合成分析正在从有机合成化学向合成生物学渗透。具体策略包括:(1)从目标寡糖出发,通过反应规则逆向推导可能的 GT 级联反应序列;(2)基于通量平衡分析和基因组规模代谢模型,在约束条件下筛选最优途径组合,最大化理论产率、最小化 ATP 和 NAD(P)H 消耗、避免有毒中间代谢产物积累;(3)给定底物和产物,AI 推荐最优酶组合,包括基于 kcat/Km 数据和表达兼容性预测的预筛选。Rosano 和 Ceccarelli 的综述论述了微生物系统中重组蛋白表达的关键因素,为糖蛋白生产底盘的选择和优化提供了系统指

南^[61]。当前该方向的主要限制在于高质量 GT 动力学参数数据库覆盖的 GT 数量远小于 CAZy 数据库的总条目数,制约了 AI 模型的训练数据基础。

6.4 数据稀缺挑战与低资源方法

糖科学面临 AI 应用中最根本的瓶颈是高质量标注数据的严重不足。蛋白质结构数据库 (PDB) 已积累超过 20 万条结构,蛋白质序列数据库 (UniProt) 超过 2.5 亿条序列;而相应的糖链结构数据库(如 PDB 中的糖链结构子集和 GlyTouCan)的数据规模数量级更小。这一数据不足限定构成了所有 AI 方法的底层限制^[19]。

应对数据稀缺的低资源机器学习方法包括:(1)迁移学习:利用蛋白质序列的大规模预训练语言模型(如 ProtBERT、ESM-2)的潜在表征能力,通过微调迁移至糖链相关预测任务(如从蛋白质序列预测其 N-糖基化位点);(2)少样本学习:在仅有几十条标注糖链的条件下,通过与未标注糖链的对比学习(如 SimCLR 和 MoCo 框架)学习糖链的通用表征,再以极少标注样本微调;(3)主动学习:模型主动选择最有信息量的未标注糖链提交人工标注,以最少的标注代价最大化模型性能提升。数据标准化(GlyTouCan 为每条糖链分配唯一 ID 的 WURCS 格式)和开放共享(GlyGen 的数据整合平台和 MIRAGE 报告指南)的推进,是改善低资源困境的基础设施层面的迫切需求。

7 挑战与展望

尽管糖合成生物学取得了令人瞩目的进展,三个核心挑战构成了未来十年的主要研究前沿。

糖密码的深度解析。与遗传密码的三联体规则不同,糖链的“结构-功能”编码是非线性的、多价的、高度依赖上下文的。糖微阵列 (glycan microarray) 技术^[50]通过在芯片表面固定数百种结构明确的糖链并检测与荧光标记糖结合蛋白的差异结合谱,实现了糖-蛋白相互作用的高通量筛选,但所能覆盖的糖链结构仍远小于理论空间^[18]。质谱技术的进步正在缩小通量差距:

Ruhaak 等系统综述了 LC-MS/MS 在糖组学和糖蛋白质组学分析中的最新策略, 特别指出 PGC 色谱-串联质谱结合离子淌度分离已成为复杂糖链混合物高分辨率解析的黄金标准^[62]。Kailemia 等对寡糖质谱分析技术发展的全面综述为实验糖组学的方法选择提供了系统指南^[63]。

从试错到智能设计。当前糖合成途径的设计仍以经验驱动迭代为主流。实现向“预测-设计”范式的跃迁, 需要 (1) GT 元件的系统功能表征, 建立包含 kcat、Km、底物谱和表达兼容性的标准化数据库; (2) 将 AI 预测模型嵌入合成生物学自动化设施 (生物铸造厂), 实现全流程自动化; (3) 发展能够从有限标注数据中有效学习的低资源机器学习方法, 这是糖科学在数据稀缺条件下的特有挑战^[19]。

产业化瓶颈。从实验室到工业规模, 细胞工厂的工业鲁棒性 (温度、pH、渗透压波动的适应性) 和产物分离纯化成本是决定产品经济性的两大工程挑战。寡糖和糖蛋白产物的色谱纯化占总生产成本的 50-70%^[54], 高于大多数小分子药物和蛋白药物的纯化成本占比。其原因在于糖链产物的理化多样性迫使纯化流程采用多色谱联用, 每一步均有产率损失。

以 CO₂、甲醇、甲酸为单一碳源的微生物糖合成, 通过合成甲基营养途径 (RuMP 循环) 或还原性甘氨酸途径, 将成为未来十年最活跃的糖合成生物学方向之一。超越自然界已知单糖库的非天然糖类药物, 如氟代糖 (抗病毒)、硫代糖 (糖苷酶抑制剂) 和氨基糖衍生物 (抗生素增效剂), 正在成为糖合成生物学的新增长极, 但该方向对非天然糖核苷酸与 GT 的协同工程提出了高于现有能力的挑战。

个性化糖疫苗, 基于患者肿瘤特异性异常糖基化谱设计的患者定制化疫苗, 和糖基化靶向治疗 (如高甘露糖型抗体引导的巨噬细胞介导的肿瘤细胞清除) 代表了糖科学与精准医学融合的前沿。CRISPR-like 的可编程糖链修饰工具, 将核酸导向系统 (如 Cas9 的核酸识别模块) 与 GT 催化域融合, 实现特定基因座产物糖基化的定点编辑, 是糖生物学领域的技术目标。该方向的概念验证已在体外实验中初步实现, 但体内效率和脱靶控

制仍远未达到实用化水平。

8 结 论

糖类生物大分子以其惊人的结构多样性长期处于生命科学研究的边缘地带。这种边缘化并非源于其功能重要性的缺失, 恰恰相反, 细胞表面的糖萼、蛋白质的糖基化修饰、以及各类寡糖和多糖在发育、免疫、感染和肿瘤发生中扮演着不可替代的角色, 而是源于一个根本性的技术困境: 我们无法像操控核酸和蛋白质那样, 以可预测、可编程的方式合成结构均一的糖链。正是这一困境, 使得糖科学在基因组学和蛋白质组学蓬勃发展的数十年间, 始终未能建立起系统性的“结构-功能”关联知识体系。糖合成生物学的崛起, 正在从根本上打破这一僵局。本文系统综述的四项核心使能技术: GT 元件工程、代谢途径重构、细胞工厂设计及无细胞合成体系, 共同证明了一个关键论断: 糖链的复杂性与糖合成的可工程化之间不存在不可调和的矛盾。相反, 正是通过将糖基化系统拆解为可独立优化、可模块化组装的标准化生物部件, 合成生物学将糖类从传统认知中“难以复制的天然产物”成功转化为了生物制造工具箱中一个高度可控、可精确调参的工程对象。

这一范式变革的核心意义, 在于它完成了糖科学研究方法论的两次关键跃迁。第一次跃迁是从“发现自然已有的糖”到“设计并制造想要的糖”。在传统糖科学中, 研究者受限于天然来源的糖链样品, 它们通常是结构不均一的混合物, 微观不均一性使得构效关系研究寸步难行。糖合成生物学提供的逆向工作流彻底改变了这一局面: 研究者可以先在计算机上设计目标糖链的精确结构, 再通过模块化的酶元件库和代谢途径库, 在工程化底盘或体外反应体系中实现其精准生物合成。这种“设计-合成-功能验证”的闭环, 使得糖科学第一次获得了与基因组学、蛋白质组学对等的假说驱动研究能力。第二次跃迁是从“单途径优化”到“多维度系统集成”。早期糖工程研究往往聚焦于单一 GT 的过表达或单一糖核苷酸合成途径的强化, 而本文综述的大量成功案例, 从母乳寡糖的吨级发酵到抗体 Fc 糖型的精准定制表明,

高效的糖基化生物系统需要同时驾驭GT酶的底物特异性与催化效率、糖核苷酸前体的代谢通量分配、底盘细胞的分泌与折叠能力、以及发酵过程的参数耦合。这种多维度、多尺度的系统集成能力，正是合成生物学区别于传统代谢工程的核心标志。

在产业层面，糖合成生物学已完成了从概念验证到规模化应用的实质性跨越。以母乳寡糖为代表，2'-FL等核心产物的发酵滴度突破100 g/L，全球市场规模超过10亿美元^[11]，标志着糖类生物制造已具备与传统化学合成和天然提取法相竞争的经济可行性。在治疗性糖蛋白领域，去岩藻糖化抗体Obinutuzumab的临床成功充分验证了糖基化工程作为下一代抗体药物核心技术的战略价值^[44]。在糖疫苗领域，生物结合技术将寡糖抗原直接偶联于载体蛋白，省去了传统化学偶联的多步纯化与质量控制环节，为快速响应新型病原体提供了技术储备。在功能多糖材料领域，透明质酸的百吨级微生物发酵已广泛应用于眼科手术和医美填充剂，其背后的代谢网络优化策略，UDP-糖供体的“推-拉”式通量调控，已成为糖合成生物学的通用设计原则^[16]。这些产业化成就共同传递了一个清晰信号：糖合成生物学不再仅仅是学术探索的前沿阵地，而已成为驱动生物经济新增长极的核心引擎。

更深层次地看，糖合成生物学正在重塑合成生物学自身的学科边界和方法论体系。合成生物学的早期成功高度集中于核酸和蛋白质系统，从基因回路到代谢途径，其可预测性建立在中心法则的线性信息流基础之上。而糖基化系统呈现的是完全不同的复杂性范式：没有模板、没有直读的遗传编码、多个GTs在亚细胞区室中的时空协同、糖链结构的非线性分支特征，所有这些都挑战着传统合成生物学的线性思维。正是在应对这些挑战的过程中，糖合成生物学发展出了独特的方法论工具箱：基于CAZy数据库的GT元件系统化挖掘、祖先序列重建驱动的功能酶创制、糖核苷酸代谢网络的模块化重构、以及无细胞体系中多酶级联的空间组织策略。这些方法不仅服务于糖类的生物制造，更反过来为合成生物学处理其他“非模板”生物合成系统，如聚酮化合物、非

核糖体肽和萜类，提供了可迁移的技术范式和设计哲学。在这个意义上，糖合成生物学可以被视为合成生物学从“简单线性系统”走向“复杂非线性系统”的先行者和压力测试平台。

回顾本综述梳理的关键技术里程碑，一个贯穿始终的主题是：可编程性。从GT酶的底物特异性工程到代谢途径的逻辑门控，从细胞工厂的糖型定向设计到无细胞体系的模块化组装，糖合成生物学的每一步进展都在提升我们对糖基化过程的控制精度和预测能力。这种可编程性的逐步增强，最终将导向一个雄心勃勃的长期愿景，建立一个标准化的“糖质合成操作系统”，其中GT元件的生物学部件目录、糖核苷酸供体的合成模块、以及底盘细胞的糖基化线路图，将以类似于电子设计自动化(EDA)软件的方式供研究者调用和组合。届时，任何所需结构的目标糖链，无论是用于基础研究的探针分子、用于疫苗开发的抗原、还是用于治疗糖蛋白药物，都将能够通过合成生物学平台实现按需、快速、经济、绿色的精准制造。从CO₂到淀粉的体外人工合成途径^[14]更是证明了无生命系统实现超越自然光合作用效率的糖质合成是可行的，为构建完全可编程的体外糖合成体系提供了概念原型。

糖合成生物学的发展历程也为我们提供了一个宝贵的认识论启示：复杂性本身不应成为回避挑战的理由。恰恰是糖类那种令人生畏的结构多样性，催生了合成生物学领域最具创造力的技术突破。从“暗物质”到“富矿区”，糖的世界正在被重新发现、重新定义。这种重新发现的意义不仅在于它为药物开发和功能食品提供了新的物质基础，更在于它拓展了人类设计、操控和利用生命系统的能力边界。当糖链不再是不可控的复杂性，而成为可编程的信息流时，一个真正意义上的“糖世界”将在合成生物学的版图中徐徐展开。

参 考 文 献

- [1] DRULA E, GARRON M L, DOGAN S, et al. The carbohydrate-active enzyme database: functions and literature [J]. *Nucleic Acids Res*, 2022, 50(D1): D571-D7.
- [2] MOREMEN K W, TIEMEYER M, NAIRN A V. Vertebrate protein glycosylation: diversity, synthesis and function [J].

- Nature Reviews Molecular Cell Biology, 2012, 13(7): 448-62.
- [3] LAINE R A. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems [J]. *Glycobiology*, 1994, 4(6): 759-67.
 - [4] REILY C, STEWART T J, RENFROW M B, et al. Glycosylation in health and disease [J]. *Nat Rev Nephrol*, 2019, 15(6): 346-66.
 - [5] BOLTJE T J, BUSKAS T, BOONS G J. Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research [J]. *Nat Chem*, 2009, 1(8): 611-22.
 - [6] KHALIL A S, COLLINS J J. Synthetic biology: applications come of age [J]. *Nat Rev Genet*, 2010, 11(5): 367-79.
 - [7] NIELSEN J, KEASLING J D. Engineering Cellular Metabolism [J]. *Cell*, 2016, 164(6): 1185-97.
 - [8] WACKER M, LINTON D, HITCHEN P G, et al. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli* [J]. *Science*, 2002, 298(5599): 1790-3.
 - [9] HAMILTON S R, DAVIDSON R C, SETHURAMAN N, et al. Humanization of yeast to produce complex terminally sialylated glycoproteins [J]. *Science*, 2006, 313(5792): 1441-3.
 - [10] YANG Z, WANG S, HALIM A, et al. Engineered CHO cells for production of diverse, homogeneous glycoproteins [J]. *Nat Biotechnol*, 2015, 33(8): 842-4.
 - [11] BYCH K, MIKS M H, JOHANSON T, et al. Production of HMOs using microbial hosts - from cell engineering to large scale production [J]. *Curr Opin Biotechnol*, 2019, 56: 130-7.
 - [12] SPRENGER G A, BAUMGARTNER F, ALBERMANN C. Production of human milk oligosaccharides by enzymatic and whole-cell microbial biotransformations [J]. *J Biotechnol*, 2017, 258: 79-91.
 - [13] JAROENTOMECHAI T, STARK J C, NATARAJAN A, et al. Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery [J]. *Nat Commun*, 2018, 9(1): 2686.
 - [14] CAI T, SUN H, QIAO J, et al. Cell-free chemoenzymatic starch synthesis from carbon dioxide [J]. *Science*, 2021, 373(6562): 1523-7.
 - [15] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. *Nature*, 2021, 596(7873): 583-9.
 - [16] DE BRUYN F, MAERTENS J, BEAUPREZ J, et al. Biotechnological advances in UDP-sugar based glycosylation of small molecules [J]. *Biotechnol Adv*, 2015, 33(2): 288-302.
 - [17] RUFFING A, CHEN R R. Metabolic engineering of microbes for oligosaccharide and polysaccharide synthesis [J]. *Microb Cell Fact*, 2006, 5: 25.
 - [18] RILLAHAHAN C D, PAULSON J C. Glycan microarrays for decoding the glycome [J]. *Annu Rev Biochem*, 2011, 80: 797-823.
 - [19] BOJAR D, LISACEK F. Glycoinformatics in the Artificial Intelligence Era [J]. *Chem Rev*, 2022, 122(20): 15971-88.
 - [20] OLSSON K, CARLSEN S, SEMMLER A, et al. Microbial production of next-generation stevia sweeteners [J]. *Microb Cell Fact*, 2016, 15(1): 207.
 - [21] ARNOLD F H. Directed Evolution: Bringing New Chemistry to Life [J]. *Angew Chem Int Ed Engl*, 2018, 57(16): 4143-8.
 - [22] YU H, ZHANG L, YANG X, et al. Process Engineering and Glycosyltransferase Improvement for Short Route Chemoenzymatic Total Synthesis of GM1 Gangliosides [J]. *Chemistry*, 2023, 29(25): e202300005.
 - [23] ANDO H, KOMURA N. Recent progress in the synthesis of glycosphingolipids [J]. *Curr Opin Chem Biol*, 2024, 78: 102423.
 - [24] ZHANG T, TIAN S, GAO Z, et al. Engineering an Ancestral Glycosyltransferase for Biosynthesis of 2-Phenylethyl-beta-d-Glucopyranoside and Salidroside [J]. *J Agric Food Chem*, 2024, 72(36): 19966-76.
 - [25] LU Y, XIE L, XIE J, et al. Comprehensive engineering of novel glycosyltransferase for efficient, donor-promiscuous, and regioselective glycosylation of flavonoids [J]. *Sci Adv*, 2025, 11(40): eadu5064.
 - [26] CHEN X, VARKI A. Advances in the biology and chemistry of sialic acids [J]. *ACS Chem Biol*, 2010, 5(2): 163-76.
 - [27] BAUMGARTNER F, SEITZ L, SPRENGER G A, et al. Construction of *Escherichia coli* strains with chromosomally integrated expression cassettes for the synthesis of 2'-fucosyllactose [J]. *Microb Cell Fact*, 2013, 12: 40.
 - [28] ZHANG X, LIU Y, LIU L, et al. Microbial production of sialic acid and sialylated human milk oligosaccharides: Advances and perspectives [J]. *Biotechnol Adv*, 2019, 37(5): 787-800.
 - [29] JIA Y, ZHU J, CHEN X, et al. Metabolic engineering of *Bacillus subtilis* for the efficient biosynthesis of uniform hyaluronic acid with controlled molecular weights [J]. *Bioresour Technol*, 2013, 132: 427-31.
 - [30] ZHU J. Mammalian cell protein expression for biopharmaceutical production [J]. *Biotechnol Adv*, 2012, 30(5): 1158-70.
 - [31] JIANG Z, GAO H, LIU R, et al. Key Glycosyltransferase Genes of *Panax notoginseng*: Identification and Engineering Yeast Construction of Rare Ginsenosides [J]. *ACS Synth Biol*, 2022, 11(7): 2394-404.
 - [32] VALDERRAMA-RINCON J D, FISHER A C, MERRITT J H,

- et al. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli* [J]. *Nat Chem Biol*, 2012, 8(5): 434-6.
- [33] CHEN J, ZHAO L, YANG L, et al. Transcriptome-assisted sugar transporter screening for enhanced lacto-N-tetraose biosynthesis in *Escherichia coli* MG1655 [J]. *Int J Biol Macromol*, 2025, 321(Pt 2): 146190.
- [34] PRIEM B, GILBERT M, WAKARCHUK W W, et al. A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria [J]. *Glycobiology*, 2002, 12(4): 235-40.
- [35] BAUMGARTNER F, CONRAD J, SPRENGER G A, et al. Synthesis of the human milk oligosaccharide lacto-N-tetraose in metabolically engineered, plasmid-free *E. coli* [J]. *Chembiochem*, 2014, 15(13): 1896-900.
- [36] MENG J, ZHU Y, WANG H, et al. Biosynthesis of Human Milk Oligosaccharides: Enzyme Cascade and Metabolic Engineering Approaches [J]. *J Agric Food Chem*, 2023, 71(5): 2234-43.
- [37] KOTIDIS P, DONINI R, ARNSDORF J, et al. CHOglycoNET: Comprehensive glycosylation reaction network for CHO cells [J]. *Metab Eng*, 2023, 76: 87-96.
- [38] SHENG L L, CAI Y M, LI Y, et al. Advancements in heparosan production through metabolic engineering and improved fermentation [J]. *Carbohydr Polym*, 2024, 331: 121881.
- [39] GOTTSCHALK J, ELLING L. Current state on the enzymatic synthesis of glycosaminoglycans [J]. *Curr Opin Chem Biol*, 2021, 61: 71-80.
- [40] JEFFERIS R. Glycosylation as a strategy to improve antibody-based therapeutics [J]. *Nat Rev Drug Discov*, 2009, 8(3): 226-34.
- [41] SHIELDS R L, LAI J, KECK R, et al. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity [J]. *J Biol Chem*, 2002, 277(30): 26733-40.
- [42] LI T, DILILLO D J, BOURNAZOS S, et al. Modulating IgG effector function by Fc glycan engineering [J]. *Proc Natl Acad Sci U S A*, 2017, 114(13): 3485-90.
- [43] HUANG W, GIDDENS J, FAN S Q, et al. Chemoenzymatic glycoengineering of intact IgG antibodies for gain of functions [J]. *J Am Chem Soc*, 2012, 134(29): 12308-18.
- [44] PEREIRA N A, CHAN K F, LIN P C, et al. The "less-is-more" in therapeutic antibodies: Afucosylated anti-cancer antibodies with enhanced antibody-dependent cellular cytotoxicity [J]. *MAbs*, 2018, 10(5): 693-711.
- [45] KIGHTLINGER W, DUNCKER K E, RAMESH A, et al. A cell-free biosynthesis platform for modular construction of protein glycosylation pathways [J]. *Nat Commun*, 2019, 10(1): 5404.
- [46] ROBERTS T L, COSGROVE S C, MILLER G J. Advances and Challenges in Bioprocess Optimization for the Synthesis of Sugar Nucleotides [J]. *Chembiochem*, 2025, 26(20): e202500434.
- [47] ANTONOVSKY N, GLEIZER S, NOOR E, et al. Sugar Synthesis from CO₂ in *Escherichia coli* [J]. *Cell*, 2016, 166(1): 115-25.
- [48] PRESCHER J A, BERTOZZI C R. Chemistry in living systems [J]. *Nat Chem Biol*, 2005, 1(1): 13-21.
- [49] LAUGHLIN S T, BERTOZZI C R. Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation [J]. *Nat Protoc*, 2007, 2(11): 2930-44.
- [50] AGARD N J, PRESCHER J A, BERTOZZI C R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems [J]. *J Am Chem Soc*, 2004, 126(46): 15046-7.
- [51] BASKIN J M, PRESCHER J A, LAUGHLIN S T, et al. Copper-free click chemistry for dynamic in vivo imaging [J]. *Proc Natl Acad Sci U S A*, 2007, 104(43): 16793-7.
- [52] KUFLEITNER M, HAIBER L M, WITTMANN V. Metabolic glycoengineering - exploring glycosylation with bioorthogonal chemistry [J]. *Chem Soc Rev*, 2023, 52(2): 510-35.
- [53] BODE L. Human milk oligosaccharides: Every baby needs a sugar mama [J]. *Glycobiology*, 2012, 22(9): 1147-62.
- [54] SHARMA V, MOTTAFEGH A, JOO J U, et al. Toward microfluidic continuous-flow and intelligent downstream processing of biopharmaceuticals [J]. *Lab Chip*, 2024, 24(11): 2861-82.
- [55] RAPPUOLI R, DE GREGORIO E, COSTANTINO P. On the mechanisms of conjugate vaccines [J]. *Proc Natl Acad Sci U S A*, 2019, 116(1): 14-6.
- [56] KAY E, CUCCUI J, WREN B W. Recent advances in the production of recombinant glycoconjugate vaccines [J]. *NPJ Vaccines*, 2019, 4: 16.
- [57] SLOVIN S F, KEDING S J, RAGUPATHI G. Carbohydrate vaccines as immunotherapy for cancer [J]. *Immunol Cell Biol*, 2005, 83(4): 418-28.
- [58] BURKHOLZ R, QUACKENBUSH J, BOJAR D. Using graph convolutional neural networks to learn a representation for glycans [J]. *Cell Rep*, 2021, 35(11): 109251.
- [59] BOJAR D, POWERS R K, CAMACHO D M, et al. Deep-Learning Resources for Studying Glycan-Mediated Host-Microbe Interactions [J]. *Cell Host Microbe*, 2021, 29(1): 132-44 e3.

- [60] GERLOFF D L, ILINA E I, CIALINI C, et al. Prediction and verification of glycosyltransferase activity by bioinformatics analysis and protein engineering [J]. STAR Protoc, 2023, 4(1): 101905.
- [61] ROSANO G L, CECCARELLI E A. Recombinant protein expression in microbial systems [J]. Front Microbiol, 2014, 5: 341.
- [62] RUHAAK L R, XU G, LI Q, et al. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses [J]. Chem Rev, 2018, 118(17): 7886-930.
- [63] KAILEMIA M J, RUHAAK L R, LEBRILLA C B, et al.

Oligosaccharide analysis by mass spectrometry: a review of recent developments [J]. Anal Chem, 2014, 86(1): 196-212.



第一作者及通讯作者：杜昱光（1963—），男，研究员，博士生导师。研究方向为糖链合成生物学与营养糖工程研究，推动功能糖产品在农业、食品及医药领域产业化应用。

E-mail: ygdu@ipe.ac.cn