



**An Assessment of Short-Term
Milestones in EBRC's 2019
Roadmap, *Engineering Biology***

**美国工程生物学研究联盟
对工程生物学研究路线图短期目标的评估**

中国科学院上海营养与健康研究所
上海生命科学信息中心
上海市生物工程学会
2023年4月

美国工程生物学研究联盟

对工程生物学研究路线图短期目标的评估

2019 年 6 月, 美国工程生物学研究联盟 (EBRC) 发布首份工程生物学技术路线图《工程生物学: 下一代生物经济的研究路线图》, 为未来 20 年工程生物学技术发展设定了预期目标, 并细分为 2 年、5 年、10 年和 20 年的阶段性里程碑。在该路线图发布后的第一阶段的里程碑 (2 年) 之际, EBRC 开展了该阶段目标进展的调查与评估, 将这 2 年来的进展与当初设定的目标进行了对比, 并于 2023 年 3 月发布该阶段的评估报告。

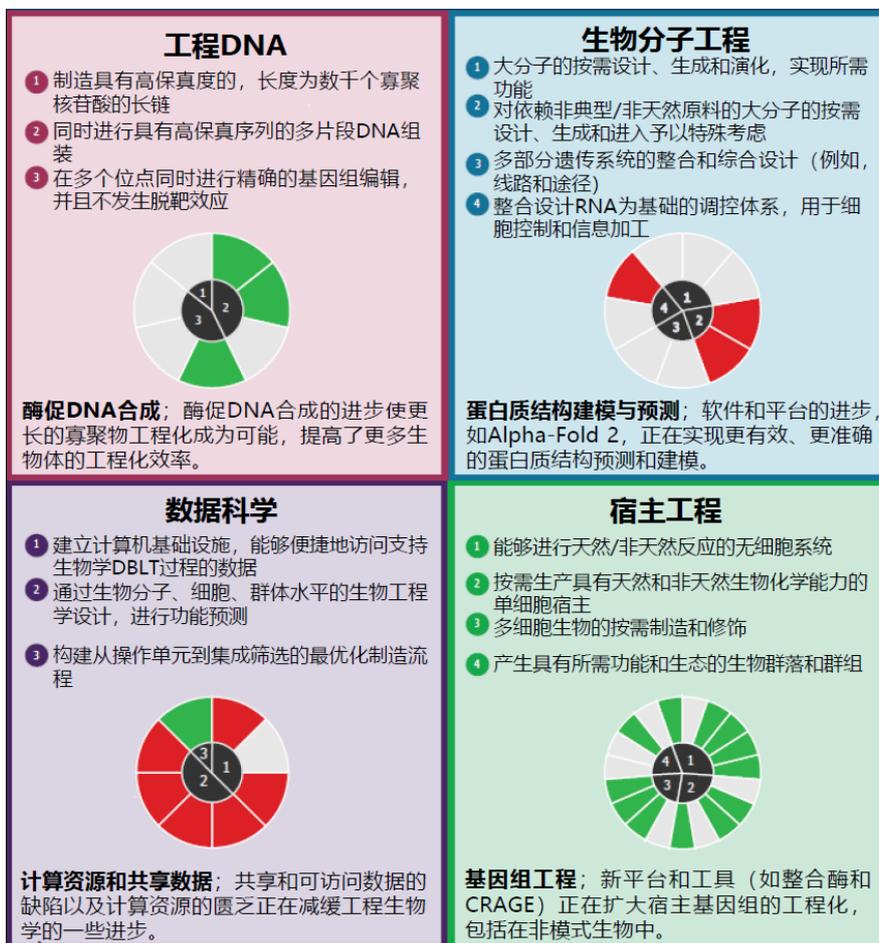


图 1 工程生物学研究路线图 2 年里程碑的评估

报告使用了 3 种颜色来评估发展达到计划预期的程度, 绿色: 指这一突破性能力的进展领先于路线图预测; 灰色: 指进展符合预期 (计划内); 红色: 则是指进展未及预期。总体上看, 技术主题“工程 DNA”和“宿主工程”的指标进展

较好, 而“生物分子工程”和“数据科学”则相反, 有较多方向都未及预期 (图 1)。报告强调, 在未来, “生物分子工程”和“数据科学”有可能会形成差异化的竞争壁垒, 是需要注意和加大投入的技术方向。

1. 工程 DNA: 基因编辑、合成和组装

工程 DNA 是一项基础技术, 也是生物分子工程和宿主工程进步的基础, 并将推动所有应用领域的工作。基因编辑、合成和组装 (工程 DNA) 的研究侧重于“工具的开发和改进, 实现染色体 DNA 合成和整个基因组的编辑”。之前路线图预测, 合成 DNA 的市场已经逐渐成熟。2019 年以来, 随着酶促 DNA 合成和底盘独立重组酶辅助基因组工程 (CRAGE) 等多项平台技术的进步, 这一预测仍然适用。

表 1 评估工程 DNA 2021 年里程碑进展

基因编辑、合成和组装的进展	
目标:	制造具有高保真度的, 长度为数千个寡聚核苷酸的长链
突破能力:	高效进行寡聚核苷酸合成, 从而提高寡聚核苷酸的数量、长度和保真度。
	<u>2021 年里程碑</u> : 合成一百万个长度为 200 个单体的寡聚核苷酸, 错误率应低于 1 个错误/500 核苷酸。
目标:	同时进行具有高保真序列的多片段 DNA 组装
突破能力:	预测性设计 DNA 序列, 实现更长、信息更丰富的 DNA 片段组装。
	<u>2021 年里程碑</u> : 将 DNA 序列设计与核苷酸组成优化相结合, 支持合成并保持遗传系统的功能。
突破能力:	用于一步法、持续进行长 DNA 片段组装和序列验证的方法。
	<u>2021 年里程碑</u> : 可靠组装 10000 个碱基对的非克隆 DNA 片段。
突破能力:	工程化遗传系统的合成、组装和功能测试管线。
	<u>2021 年里程碑</u> : 在低保真、易出错的遗传系统中实现所需功能。
目标:	在多个位点同时进行精确的基因组编辑, 并且不发生脱靶效应
突破能力:	在任何生物体中可靠地进行任何精确、特定的编辑 (单核苷酸多态性或基因替换), 同时不产生任何不必要的编辑, 编辑范围从单个碱基突变到整个路径的插入
	<u>2021 年里程碑</u> : 在模式生物中进行任何特定单碱基对改变。
突破能力:	在不同的时间段内、多种细胞和生物体中进行精确、可预测、可调控的基因表达。

 2021 年里程碑: 实现长期的基因抑制和活化。
突破能力: 能重复、高效地将编辑结构传送到特定的靶细胞或组织, 并控制编辑器的剂量和时间。
 2021 年里程碑: 提高编辑器功能, 使其在无序列要求下 (例如 PAM 序列) 获得与 2019 年最先进的功能相媲美。

注: 对 2021 年的每个里程碑都进行了评估以确定实现其进展的情况。4 个都填充绿色的圆圈表示 2021 年已经实现或接近完成目标, 3 个填充的圆圈表示在实现 2021 年里程碑方面已经取得了重大进展, 2 个填充的圆圈表示在实现 2021 年里程碑方面进展不大, 1 个填充的圆圈表示在实现 2021 年里程碑方面进展很小。在工程 DNA 中, 所有 2021 年的里程碑都已经实现/接近完成, 或者已经取得了重大进展。

1.1 工程 DNA 技术的亮点

通用的 DNA 组装工具包

DNA 组装是生物学研究必需的基础技术, 其进展将加速工程生物学的进步。在过去 20 年中, 组装小的 DNA 分子以形成更大构建体的技术库在迅速扩展, 包括 PCR 反应, 限制性核酸内切酶的消化、连接、II 型组装和 Gibson 组装。然而, 随着该领域的不断发展, 拥有适用于多个物种的通用的 DNA 组装工具包对研究人员来说就非常重要。随着非模式生物研究的增加, 研究人员可以利用通用 DNA 组装工具包有效地驯化这些生物, 而不是每次发现新生物体都需要创造新的 DNA 组装技术。在过去几年, DNA 组装技术的发展主要集中在创建平台和工具, 以加快物种质粒和试剂的使用。能够自动化重复实验有很大的好处, 可以努力将工程微生物的生产或功能扩大到工业化水平。尽管路线图中提到了 DNA 组装和自动化, 但是跨不同生物体的通用组装技术 (包括非模式生物的驯化) 取得了显著进展。

长读长测序 DNA 技术

能够读取 DNA 序列对于工程生物学技术的高效、准确和可重复至关重要。过去几年, DNA 测序技术发展迅速, 可以更便宜、更可靠, 并实现更高的通量。然而, 长读长测序技术 (longer-read sequencing technology) 正变得越来越重要。非模式生物, 特别是在早期研究工作中, 需要汇编和组装其基因组草稿。基因组测序仪共同获取新发现的生物体的不同 DNA 片段, 并创建一个冗长的 DNA 读取列表, 它们通过不同碱基对序列推断出属于该生物体的所有 DNA 测序片段。计算程序必须识别每个单独的 DNA 的读取以便准确地匹配到组成该生物体的相

应组织。这些不同的片段越短，由于其计算负担，这种匹配就越困难；相反，较长序列更容易处理，计算难度更小。纳米孔测序可以执行更长的读取时间，极大地实现非模式生物的基因组组装，从而形成基因组草稿。此外，许多纳米孔测试仪已经实现便携式，这样就可能用于现场研究或在实验室中难以驯化的非模式生物中使用。DNA 测序技术的进一步发展将继续显著加速合成基因构建体的验证和发现。人们越来越认识到，特定形式的 DNA 技术，如纳米孔测序，在工程生物学中发挥着巨大作用。

1.2 工程 DNA 进步的障碍

小型化、实验室便携式 DNA 合成硬件

合成 DNA 是工程生物学研究的关键，经常是从外部供应商订购。随着自动化和高通量技术，可以进行更多的工程生物学实验和分析，对合成 DNA 的需求也在增长。DNA 合成公司正在开发技术（例如，Twist Bioscience Technology, Evonetix），能够并行合成数以万计的 DNA 链，降低合成成本并缩短周转时间。实验室可以进一步增加其实验带宽，按需访问小型化的“台式”的高通量并行 DNA 合成设施，以更快地从实验设计转向实验实施。虽然台式合成仪已经问世一段时间，但新一代设备最近才进入市场或预计很快进入市场，这将使得能够更快、更准确地合成更多寡核苷酸（例如 DNA Script Syntax 系统）。未来，在不影响准确性的前提下，可以有快速打印更长的合成 DNA 链的分布式设备。这不仅可以使资源丰富的国家实验室更快地采取行动，而且可以在全球范围内使远离 DNA 合成供应商的研究人员显著减少其实验时间和成本。

建立基因编辑基准标准

生物体的基因编辑在科学和医学方面具有革命性潜力。然而，目前的指南缺乏对脱靶效应的精确验证标准。随着研究人员继续在实验室中使用基因编辑，需要建立构成“有效”基因编辑以及如何精准测量脱靶效应的标准。路线图强调基因编辑技术的发展是一项重要进步，但没有具体说明构成最小脱靶效应、效率化、基因组毒性以及与基因编辑实验相关的其他参数的基准标准。这种缺乏标准化的问题可能产生深远的影响，例如，生物系统生产生物分子的可重复性，减轻生物技术对人类健康的不利影响，以及促进健全、可重复的科学等多方面。由政府、行业、学术界和社会利益相关方组成联盟或网络，讨论适当的基准，以及确保安

全、可重复的标准，将为这一障碍提供潜在的补救措施。

用于组装和合成高 GC 含量的 DNA 的工具

具有高 GC 含量的 DNA 序列由于其复杂的二级结构、错误引物和错误退火，在 DNA 合成中可能存在很多问题。之前流行的短读测序技术在测序和组装这些高重复区域方面也遇到了困难，导致许多 GC 含量较高的生物（例如复杂的真核生物）的基因组序列不完整。对于高 GC 含量的 DNA，仍然需要强大的长读长测序技术和组装工具，以及识别错误和丢失序列的工具，特别是在非模式生物中。提高这些工具的普及程度也将有助于增加进一步研究基因组组装和遗传数据集的数量。

2. 生物分子工程：生物分子、途径和线路工程

路线图将分子生物学、途径和线路工程（生物分子工程）定义为专注于“单个生物分子本身的扩展性或新功能的工程目标和挑战”，并指出“成功的进展将包括通过从天然和非天然构建模块中按需生产功能性大分子，有针对性地设计复杂线路和途径，以及对系统动态的监控”。生物分子工程“一直在探索从自然界存在的东西中寻找不存在的东西”。围绕利用天然和非天然成分构建生物分子的一些新技术的出现，也一定程度上印证了这种观点。自 2019 年以来，生物分子工程取得重大进展，特别是在蛋白质结构预测和合成免疫学方面。除了路线图中设定的目标，机器学习技术等已经对许多生物分子工程应用产生了重要影响。

表 2 评估生物分子工程 2021 年里程碑进展

生物分子、途径和线路工程的进展	
目标：大分子的按需设计、生成和演化，实现所需功能	
突破能力：依据初始序列从头预测 RNA 结构、蛋白质结构、DNA/RNA 与蛋白质的复合物，并根据结构准确预测突变的可变性及其影响。	<p>●●●○ <u>2021 年里程碑</u>：根据初始序列（成功率>50%），在 5 埃范围内预测 300 个氨基酸的蛋白质和 200 个核苷酸的 RNA 结构域。</p> <p>●●○○ <u>2021 年里程碑</u>：提高力场和骨架取样算法，以及捕获转录后和翻译后修饰核苷酸和氨基酸力场的能力。</p>
突破能力：从头设计和/或预测大分子动力学以及大分子动态结构。	●●●○ <u>2021 年里程碑</u> ：改进能够整合实验数据的 RNA 动力学计算模型。
突破能力：高通量整合计算、实验和改进方案，用于细化所需的生物分子功能。	

●●●○	2021 年里程碑: 在模式生物中构建持续高突变率的体内连续 DNA 诱变和进化系统。
目标: 对依赖非典型/非天然原料的大分子的按需设计、生成和进入予以特殊考虑	
突破能力: 对长度达到 400 个碱基对、含非天然核苷酸的基因进行 PCR、逆转录、细胞复制和转录。	
●●○○	2021 年里程碑: 确认 A-T-G-C 碱基对功能的缺失。
突破能力: 拓展遗传密码系统,用于转录含有完全非天然氨基酸的 >100 个氨基酸的蛋白质,以及至少含有 4 种不同非天然氨基酸原料的蛋白质。	
●●●○	2021 年里程碑: 创造能够通过非天然氨基酸进行新型治疗活动的蛋白质。
目标: 多部分遗传系统的整合和综合设计(例如,线路和途径)	
突破能力: 在宿主生物体或细胞中,设计具有特定表达能力、高度稳定性的大型遗传系统(基因组),并整合全系统效益。	
●●●●	2021 年里程碑: 将基因表达相互作用整合到原核遗传系统设计中。
突破能力: 合理设计传感器组件、基因线路、代谢途径、信号级联和细胞分化通路。	
●●●○	2021 年里程碑: 对具有 10 多个复杂计算调节器的基因线路进行可靠地改造。
目标: 整合设计 RNA 为基础的调控体系,用于细胞控制和信息加工	
突破能力: 将核苷酸链置换技术应用于 RNA 实例化的细胞系统。	
●●●●	2021 年里程碑: 在细菌中实现 RNA 的级联链置换。
突破能力: 成功地将以细菌 RNA 为基础的遗传调控因子转移到真核和哺乳动物系统中。	
●●○○	2021 年里程碑: 第一代真核 RNA 为基础的调控因子:使用 RNA-RNA 相互作用和/或链置换并实现基因表达的 10 倍变化。
●●●●	2021 年里程碑: 创建 RNA 修饰机制,允许对 RNA 进行可编程的位点特异性修饰,专注于天然丰富的修饰(N6-甲基腺苷、2'-O-甲基化、假尿苷)。

注:在生物分子工程领域,2021 年的里程碑部分已经实现/接近完成,部分已经取得了重大进展,还有一部分进展并不显著。

2.1 生物分子工程技术发展亮点

蛋白质结构建模和预测的进步

过去几十年里,从氨基酸序列预测蛋白质结构的工具已经取得相当重要的发展,尽管预测通常收到准确性、速度和相对低效的同源性分析方法的限制。为了加快获得蛋白质结构三维模型所需的时间,研究人员开发了用于蛋白质结构预测的建模算法,并且在过去两年已经显示出这种建模的进步。其中最具代表性的发展之一是 AlphaFold 2,这是一个由 Alphabet 和谷歌神经网络 DeepMind 开发的人工智能程序。AlphaFold 2 利用人工智能深度学习技术来预测蛋白质结构,并

在其前身 AlphaFold 1 基础上显示出令人震惊的改进。在 2020 年的蛋白质结构预测关键技术评估 (CASP) 竞赛中, 基于衡量结构预测效率的基准, AlphaFold 2 正确预测了 60% 的蛋白质结构, 这些结构信息以前是未知的。AlphaFold 2 的开源软件和蛋白质组数据库已经发布, 可以通过 <https://github.com/deepmind/alphafold> 和 <https://alphafold.ebi.ac.uk/> 访问。除此之外, Baek 等人还开发了人工智能程序 RoseTTAfold, 可以生成高质量的蛋白质结构预测, 并能预测蛋白质复合物结构, 解决 X 射线晶体学和冷冻电子显微镜建模问题。这两项成果以及相关的研究工作, 极大地加速了基础研究, 为工程生物学的进步做出了贡献。AlphaFold 数据库包含有超过 2 亿个蛋白质结构预测, 现在可供所有研究人员免费试用 (<https://alphafold.ebi.ac.uk/>)。

机器学习用于改进生物分子功能

人们普遍预测机器学习可以促进工程生物学进步, 特别是通过结合蛋白质结构预测工具。除此之外, 机器学习算法还可以用于识别非天然氨基酸的潜在特性, 为代谢工程中的设计-测试-构建-学习循环推荐可用的菌株, 并确定 RNA toe-hold 序列响应所需的靶序列的有效性。例如, 研究人员结合机器学习来探索多个同时突变对蛋白质定向进化的影响。尽管工程生物学预测机器学习在预测生物分子结构-功能关系方面起着至关重要的作业, 但机器学习也在结构预测之外的生物分子工程中有更多应用。

2.2 生物分子工程进步的障碍

易受突变影响的脆弱基因线路

由于代谢负荷和毒性, 工程基因线路随着时间推移容易出现稳定性问题, 导致对合并环路选择性进化的压力。目前, 正在研究无数提高线路稳健性的方法, 包括如何使用测序技术来更好地监测基因线路中的突变, 已经通过隔离关键 DNA 序列的设计使基因线路更稳定。重要的是要认识到, 由于这些基因线路中许多具有工业目的的研究需要进行扩展, 因此它们的脆弱性就构成了生产的潜在风险。克服这一障碍的方法之一可能是建立可用作工程线路研究基准的弹性措施和标准。

3. 宿主工程: 宿主和群落工程

宿主细胞、有机体和系统、群落的工程化是工程生物学大多数应用的基础。

路线图中定义的宿主和群落工程(宿主工程)是“宿主细胞和生物体的标准和工程化所需的工具和技术的进步,以及这些系统和环境的整合与相互作用”。路线图预测,有充分的潜力实现利用传统和新模式生物的能力来设计有用的功能;过去两年中出现的几种平台技术(包括利用整合酶在非模式微生物中进行基因组修饰等)证明这一预测是相对可靠的。在路线图目标之外,已经开发出宿主和群落工程的基础技术,包括培养微生物群落的标准和非模式生物驯化的指南等。

表 3 评估宿主工程 2021 年里程碑进展

宿主和群落工程的进展	
目标: 能够进行天然/非天然反应的无细胞系统	
突破能力: 构建可复制、可比较的无细胞系统,用于在多种有机体,包括非模式宿主中进行生物工程化和生物制造实践应用。	
●●●○	<u>2021 年里程碑</u> : 完整表征细胞生长收获条件以及提取参数对无细胞细菌游离提取物的影响(例如蛋白质合成和天然遗传调控因子)。
突破能力: 构建细胞能力,包括支持 DNA 复制、转录、翻译、能量再生以及膜结构的分子子系统。	
●●●○	<u>2021 年里程碑</u> : 通过使用无细胞系统,制造(最小/合成)细胞所编码的所有组分。
突破能力: 用于蛋白质合成和生物制造的持久性、稳健性、低成本的无细胞体系。	
●●●○	<u>2021 年里程碑</u> : 在多个生物体及所有的生物领域,确定试剂在无细胞系统中的不确定性。
突破能力: 使用无细胞系统来设计基因元件和线路。	
●●●○	<u>2021 年里程碑</u> : 使用新一代测序读取方式定量表征基因设计产品在无细胞系统中的性能。
突破能力: 使用无细胞系统进行分散式、便携式、按需传感和制造。	
●●●●	<u>2021 年里程碑</u> : 使用内毒素含量较低的安全裂解液进行靶物质的传感和制造。
突破能力: 使用无细胞生物合成制造任何目标糖蛋白或代谢物。	
●●●●	<u>2021 年里程碑</u> : 构建用于糖基化组装的模块化、多功能的无细胞平台。
目标: 按需生产具有天然和非天然生物化学能力的单细胞宿主	
突破能力: 在可控制和调节的环境中,在任何时间制造任何宿主。	
●●●●	<u>2021 年里程碑</u> : 建立用于开发非模式生物细胞生存所需基质的方法。
●●●○	<u>2021 年里程碑</u> : 稳健地筛选除模式生物之外的有用骨架生物。
突破能力: 通过 DNA 传递和遗传修饰常规驯化非模式生物。	

	2021 年里程碑: 对现有方法和工具进行分类和测试, 用于在微生物/哺乳动物系统 (例如病毒载体、偶联、生化方法) 和植物系统 (例如基于农杆菌、基因枪、纳米材料的方法) 中进行 DNA 递送。
	2021 年里程碑: 开发可以在非模式生物中, 进行与 (标准方法) 有可比性的 DNA 交付的高通量方法。
	2021 年里程碑: 建立一套基因编辑工具, 用于在多种原始哺乳动物细胞中迅速插入和/或删除基因元件。
	2021 年里程碑: 在非模式生物中表征基础 DNA 组份从而改变表达强度。
突破能力: 通过设计或进化在细胞内构建和调控小分子生物合成。	
	2021 年里程碑: 识别用于实现特定类型化学性质的新型有机体, 或用于特定种类分子天然前体生物合成通路的有机体。
	2021 年里程碑: 用于精确调控已深入研究的系统的遗传表达。
突破能力: 细胞内代谢途径的空间控制或组织、非天然细胞器的构建。	
	2021 年里程碑: 靶向异种蛋白的工具以实现不同的亚细胞间隔。
突破能力: 产生和分泌任何具有糖基化以及其他翻译后修饰的蛋白。	
	2021 年里程碑: 能够生产实验室规模单糖形式所需蛋白的一种或多种微生物宿主。
目标: 多细胞生物的按需制造和修饰	
突破能力: 在群体中调控细胞分化和去分化的能力。	
	2021 年里程碑: 由两种或多种工程细胞类型组成的简单微组织或小群体所具有的按需、可再生功能。
突破能力: 表征并调控多细胞系统的三维结构。	
	2021 年里程碑: 对现有组分和标准测量程序进行表征从而评估功能。
突破能力: 在体细胞中实现稳定非遗传变化的能力。	
	2021 年里程碑: 将生物分子“效应器”(DNA、RNA、蛋白)转移至缓慢分裂或不分裂的细胞中。
突破能力: 通过胚系编辑, 实现可预测、精确、有针对性的可遗传编辑。	
	2021 年里程碑: 完成对所选宿主基因组的测序, 用于进行靶基因的编辑。
	2021 年里程碑: 在植物中定义并验证组织特异性的 DNA 组分。
目标: 产生具有所需功能和生态的生物群落和群组	
突破能力: 调控不同物种间、细胞间通讯的能力。	
	2021 年里程碑: 能够实现种内和种间通讯的, 严格控制的启动子-反应调节因子系统。
突破能力: 表征、操作和编程生物群落三维结构的能力 (包含多个物种的自然或可控的生	

物群落的“生态系统”。



2021 年里程碑：使用现有技术（包括宏基因组学、转录组学、蛋白质组学和质谱）更好地理解微生物群落的物种组成以及结构组成。

突破能力：控制或定义工程化微生物群落/群组的功能。



2021 年里程碑：结合同种特有功能，生产所需产品。

突破能力：通过群落成员的添加、移除或重组，针对性地修饰现有微生物群，实现新功能或解决微生态失调。



2021 年里程碑：使用现有技术（包括宏基因组学、转录组学、蛋白质组学和质谱），从广义上描述微生物群落的功能。

注：在宿主工程领域，2021 年的里程碑大部分已经实现/接近完成或已经取得了重大进展，少部分进展并不显著。

3.1 宿主工程技术发展亮点

底盘独立重组酶辅助基因组工程（CRAGE）

鉴定、修改和测试不同微生物生产有价值生物分子和代谢物的能力对于工程生物学的商业化和规模扩大至关重要。目前，无法同时快速筛选和研究几种微生物生产有用生物分子的能力，这是重大障碍。基于细菌水平传递遗传信息的能力，底盘独立重组酶辅助基因组工程（CRAGE）是将大型生物合成基因簇高效、准确地整合到细菌基因组中。这些生物合成基因簇通常产生次级代谢物，这些代谢物对于微生物生存是非必要的，但可以在面对环境或诱导压力时使微生物具有竞争优势。微生物利用这些生物合成基因簇合成产物的能力各不相同，因此需要确定哪些微生物最耐受并且最能够在工业环境中有效表达。此外，该工具可以使研究人员将生物合成基因簇整合到更多种微生物中，对生物分子合成进行筛选和比较研究。

使用整合酶编辑非模式生物的基因组

自 2019 年以来，对非模式生物驯化的研究日益增加，其中包括哪些原生表现出适合工业用途的生物。通常，这些非模式系统缺乏其研究适合的工具集。特别是缺乏染色体修饰机制，用以在非模式生物中产生有用的基因突变。尽管基因编辑技术的进展不断，例如 CRISPR-Cas9 显示出在多个物种中的应用潜力，但是许多生物在快速移植等方面面临障碍。过去几年，出现了位点特异性 DNA 整合酶。这些整合酶通过催化两个特定 DNA 序列之间的重组时间来运作，并且通常可以适应不同物种。某些整合酶家族，如大丝氨酸重组酶和丝氨酸整合酶，由

于它们不需要宿主的分子机制来执行重组, 因此可以在广泛的生物体中起作用。这种宽松的要求使整合酶能够更容易地修改非模式生物基因组, 例如恶臭假单胞菌。尽管需要一直开发通用工具包, 以便在任何所需物种中进行基因编辑, 但整合酶为研究人员探索非模式生物提供了帮助。

使用细胞融合技术创建有用的杂交宿主生物

尽管路线图中提到了蛋白质融合是推进工程生物学的重要工具, 但没有特别关注来自不同生物体的细胞融合。2019 年以来, 研究人员研究了细胞融合用以创建混合生物体, 其中混合了两种生物的特性。研究人员主要关注融合含有类似生物合成途径的细胞, 这些途径结合起来可以协同生产所需的分子。最近的一些进展也涉及到工程生物学领域的细胞融合。例如, 研究人员创建了一个动态基因组规模的代谢建模框架, 该框架评估了乙酰丁基梭菌和容达利梭状芽胞杆菌融合中性质的变化, 并通过模型预测了乙醇和异丙醇产量的改善; 研究人员提出了一种通过细菌细胞融合生成异核细胞的方案, 证明了通过合成膜相关脂肽控制细胞融合的特异性; 研究人员还描述了酵母球状体和哺乳动物 BHK-21 细胞直接的融合以及回收 Sindbis 病毒颗粒。这通常是一种复杂的程序, 需要非常昂贵的实验室试剂。细胞融合为生物体不同特征的结合提供了一种有趣的可能性, 为研究人员修改生物体的工具包增加了潜力。然而, 了解哪些生物适合细胞融合并预测杂交生物是否具有必要的生物分子维持正常运作是利用该技术的基本知识。

3.2 宿主工程进步的障碍

非模式生物驯化的工具和程序

宿主工程领域进展较少的一个方向是“在受控和调控环境中随时植入任何宿主”的能力。路线图预计, 在驯化非模式生物用于研究和应用的不同用途方面应该取得一些进展。在对非模式生物进行先进的宏基因组标准或基因组编辑之前, 需要开发更多表征技术、实验方案。由于存在这一瓶颈, 在基本分析方法和资源方面反而有了一些进步, 包括基础显微镜、细胞学和兼容工具开发 (例如质粒创建) 等。这些进步不仅可以应用于现有的模式生物, 而且可以适用于驯化非模式系统。随后, 许多实验室已经开始挑战驯化新物种, 一些研究小组也报道了相关的基本表征分析成果。然而, 任何生物体的工程化方面进展比预期要慢, 不管是模式生物还是新发现的物种。

4. 数据科学：数据整合、建模和自动化

数据整合、建模和自动化技术主题（数据科学）重点介绍了可以显著增强执行复杂分析和预测功能的能力，从而实现高级建模和自动化。工程生物学的数据科学面临着“需要新颖、更强大的计算工程和模型”的挑战。这些计算工具对于驾驭生物有机体固有的复杂性至关重要，例如模拟潜在的实验结果，设计合成生物分子的最佳途径以及创建简化的制造过程，但过去几年的研究还无法满足这一需求。除了路线图的目标，该领域还出现了用于单细胞分析的机器学习等基础技术，但路线图中设定的数据科学的能力目前在很大程度上仍然无法实现。

表 4 评估数据科学 2021 年里程碑进展

数据整合、建模和自动化的进展	
目标：建立计算机基础设施，能够便捷地访问支持生物学 DBLT 过程的数据	
突破能力：建立生物数据库以及分析方法，并建立标准化、可访问的储存库。	
● ○ ○ ○	<u>2021 年里程碑</u> ：在学术界和产业建立工程生物学数据访问和需求相关的沟通体系。
● ● ● ○	<u>2021 年里程碑</u> ：开发可查找、可访问、可互操作、可重复使用（FAIR）的数据标准并开放工程生物学储存库。
突破能力：建立通用的计算机基础设施，用于查找生物数据和通用 API，进行研究和分析。 该项突破能力没有任何与 2021 年里程碑相关的进展。	
突破能力：用于工程生物学系统的端到端、行业规范的设计软件平台。 该项突破能力没有任何与 2021 年里程碑相关的进展。	
目标：通过生物分子、细胞、群体水平的生物工程学设计，进行功能预测	
突破能力：根据大规模集成的数据框架，全自动化的进行分子设计。	
● ● ● ○	<u>2021 年里程碑</u> ：用于自动定向进化的，能够将大规模结果反馈至算法的、基于结构和比较分析的文库。
突破能力：使用多功能酶预测算法设计任何分子的生物合成线路（天然或非天然）。	
● ● ○ ○	<u>2021 年里程碑</u> ：能够确认任何有机分子中任何生物和化学通路的反式生物合成软件。
突破能力：可拓展的、数据驱动的宿主设计，用于实现天然生物分子高水平生产的复杂环境。	
● ● ● ○	<u>2021 年里程碑</u> ：实现上位性映射和合成交互的多宿主突变和筛选以及大规模的宿主优化。
● ● ○ ○	<u>2021 年里程碑</u> ：在适合的设计部署环境中获得更好的生理和健康数据。

突破能力: 设计功能化、自我支持的生态系统。



2021 年里程碑: 为合成元件选择有机体的数据驱动工具, 从而实现抗性、弹性活动。



2021 年里程碑: 直接进行重要群落的数据筛选, 为人体、农业、复杂生物反应器的设计提供信息。



2021 年里程碑: 用于确认跨组织网络和生态互动的建模工具。

目标: 构建从操作单元到集成筛选的最优化制造流程

突破能力: 实现信息工具、数据、自动平台的标准化, 用于高效、协作地使用和整合数据, 加快新产品的开发。



2021 年里程碑: 建立通讯和网络, 从而开发用于行业和学术界数据交流和自动化的民主化平台。

注: 在数据科学领域, 2021 年的里程碑目标没有一项实现或接近完成的。

4.1 数据科学技术发展亮点

生物分子和单细胞技术的机器学习分析

工程生物学需要计算工具来分析产生的大量数据, 特别是当前的自动化方法和并行实验可以全天候的创建数据。事实证明, 机器学习对于满足这日益增长的需求至关重要, 它能够推断知识和开发工具预测蛋白质结构, 并从单细胞转录组和表观遗传学数据中开发数学模型。机器学习能力的进步包括数据规范化、不同细胞类型的分类、破译基因调控网络以及多个数据源的互操作性。特别是, 单细胞 RNA 测序工具数据急剧增加, 开发了可用工具集(以及如何使用它们的描述)。鉴于这些分析机制, 创建高质量、标准化的数据集就至关重要, 这样可以确保整个领域的可重复性并解决单细胞方法的局限性(例如批量效应和脱落问题)。

4.2 数据科学进步的障碍

非模式生物驯化的高通量自动化

高通量和自动化程序, 例如将基因组编辑或途径引入生物体的程序, 可以大大提高生物体生产有价值分子或执行有效功能的效用。虽然不是完全高通量, 但是在有机体工程的自动化平台方面已经取得了进展, 例如 CRAGE 的开发等。尽管大肠杆菌和酿酒酵母等模式生物的自动化和高通量程序已经有了很大发展, 但这些工作流通常不会扩展到非模式生物, 这也加剧了非模式生物驯化的障碍。

模拟、预测和建模以指导实验规模

影响生物体功能的因素是多种多样的, 实验改动的次数可能对系统影响是非

常高的, 因此, 准确的调查既昂贵又耗时。为了解决这一障碍, 能够准确预测功能结果的模拟、预测和模型就成为工程生物学的关键技术。先进的建模可以帮助预测不同物种在共培养条件下的状况、遗传扰动对生化生产的影响, 或生物体对环境变化的响应等。最近的建模方法是利用机器学习和数据科学的功能来预测行为, 从而使研究人员通过设计最有可能实现目标的改动来节省时间、成本和精力。在相关评论中, 有研究人员描述了合成代谢途径的自动化工程, 特别强调了生物制造的最佳实验设计方法; 还有研究人员开发了优化生物铸造厂构建自动化管线的挑战和技术解决方案。这些都是目前面临的瓶颈问题, 许多团队也在研究相关内容, 希望创建可以为工程和系统生物学研究提供信息的人工智能系统。其中一个例子是诺贝尔图灵挑战赛, 旨在“开发一个高度自主的人工智能系统, 可以开展顶级科学研究, 与最优秀的人类科学家的操作没有质的区别”。持续开发模拟、预测和建模以指导所有技术主题的实验方案, 对于改进和加快整个领域是十分必要的。

用于工程生物学的可公开访问和共享数据

大型数据集在理解决定生物系统如何执行有用的细微特征方面发挥着关键作用。例如, 宏基因组学(研究来自混合生物群落的遗传物质集合)有助于产生数据, 以了解微生物群落和环境条件下如何相互作用。学术界、政府和行业研究人员之间共享大型分析数据集通常被认为是工程生物学进步的主要障碍。这些利益相关者之间的协调对于支持推进可获取的数据集是必要的, 特别是对私营企业分析非专利数据的激励措施。路线图强调了遵循 FAIR(可查找、可访问、可互操作和可重复使用)数据标准的数据系统的重要性。随着工程生物学寻求自动化和数据科学方法来加强分析, 研究人员需要将 FAIR 数据标准纳入其工作, 并创建开发的存储库来共享数据工具。

将具体情况和普遍政策的结合可能会发挥一定的作用。共享数据的激励措施可以作为参与共同利益计划(例如 BioMADE 和 Agile Biofoundary)的要求, 可能会促进不同研究机构建立共享标准。资助方和出版商对数据共享的明确要求也有助于加强数据共享实践。此外, 由政府或私人机构资助的可公开访问的数据集也可以使研究人员之间公平共享有用的数据。支持大型项目的数据集越来越多, 例如美国国立卫生研究院的共同基金数据生态系统(CFDE)。