



合成生物学十大未来挑战

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

2022年3月

合成生物学十大未来挑战

编者按：合成生物学走过了 20 年的历程，已经进入蓬勃发展阶段，不仅推动了生物经济的重大创新，还促进了生物医学和生物技术的进步：多家合成生物学企业以 10+亿美元的估值上市；基因组工程已经渗透到生命科学研究的各个领域；DNA 的编写、编辑和重新编码正在帮助解决长期存在的遗传疾病，更是在新冠疫苗研发中发挥了重要作用。英国帝国理工的合成生物学团队 2021 年 8 月在 *Engineering Biology* 发表综述论文，讨论了合成生物学相关的 10 项技术，包括从制造合成生命、细胞模拟器、定制基因组，到自动化、深度学习和进化控制等工程生物学新方法；这些技术的预期和进步与合成生物学未来的研究和投资密切相关。研究团队希望通过整理这份清单，分析未来可能面临的挑战以及展望技术的发展。

1. 自动化和工业化

合成生物学推动生物元件标准化发展的主要包括 2 个原因：可重用性（reusability）和易于 DNA 组装（ease of DNA assembly）。标准化的建立能够实现 DNA 构建的自动化，并扩大规模，以制造可并行测试的大量工程细胞。这种构建和测试过程的“工业化”是合成生物学领域长期以来追求的目标，但目前仍然不常见。那么未来将会如何改变呢？

一些生命铸造厂和大公司已部署了自动化的“设计-构建-测试-学习”（DBTL）循环工具。除了各种计算机辅助设计（CAD）工具外，表征过的元件库也有助于元件的选择和基因结构的设计。例如，液体处理机器人能以高精度和快速高通量传输微升、纳升甚至皮升的试剂，简化了复杂的实验流程并提高了实验的重复性。数据处理工具对实验设置进行编码并整理实验元数据，确保错误的可追溯性。统计分析软件可解析大量数据、形成深层见解并优化实验设计（DoE）策略，这将有助于未来实验的开展。

未来十年，会有越来越多的工作将这些工具组合集成，形成高效、多功能的管线，快速生成合成生物学解决方案。同时，伴随而来的是合成生物学新时代日益增长的数字化，以及对于更好的数据基础设施、管理和策略的需求。数据交换形式、本体论、实验操作、实验数据和元数据报告的标准化，不仅能促进 DBTL 循环的信息流动和不同平台间的数据

流动，还能实现不同研究团体和企业间更好地进行开源软件开发和数据共享。除此之外，自动化还将推动基因元件的标准化及其性能评估的需求，该领域的长期目标是实现可预测性，这也可能促进不同生命铸造厂中主要宿主细胞和菌株更好的标准化。

应对这些挑战最快速的是几十家生命铸造厂和公司，他们拥有自动化生产所需的资金和技术。作为“云端实验室”运营的生命铸造厂和公司承诺向合作伙伴和研究人员提供相关软件 and 平台，帮助他们加快 DBTL 迭代周期，缩短整个领域的开发时间。云端实验室通过改良工程菌株或进行大规模筛查实验，可以让那些不具备合成生物学专业知识或实验设施的人员快速启动科研项目。Ginkgo Bioworks 和微软公司第二工作站 (Microsoft's Station B) 就是在合成生物学领域研发集成的生命铸造厂的典型案例；Synthace 和 Riffyn 等公司在实验室自动化和数据管理方面也已经占据一席之地。目前，所有这些案例的重点都是在“学习”和“设计”，而非“构建”和“测试”，这也符合该领域长期发展愿景。然而，合成生物学现阶段的问题在于是否能够民主化地推进此项进展。对企业来说，以专业工程承包商的身份进行竞争时并非是为所有人开放；学术型生命铸造厂同样也面临挑战，他们的博士生和博士后通过云端实验室进行实验和工作，其观点与传统科研机构的也不尽相同。因此，更需要积极帮助学术界和工业界关注和资助合成生物学，以推动该领域未来向自动化这种更有效的工作方式转变。

2. DNA 设计的深度学习

机器学习、人工智能 (AI) 和深度学习是现今期刊论文、项目资助和初创公司中常见术语。生命科学早期的显著成果是利用深度学习对显微镜图像进行分类，以帮助预测蛋白质结构及其可用作抗生素的药物分子结构。深度学习对合成生物学影响最大的领域便是 DNA 设计，因为在 DNA 中进行遗传编码其实是语言问题。

设计 DNA 是合成生物学的关键。然而，即便是编辑简单的基因，也会让科学家犯难，但深度学习可以在不知道确切细节的情况下将一类数据转换为另一类数据，甚至可以在缺少一些系统基础知识的情况下准确转换。自然语言处理 (NLP) 模型 GPT-3 可以通过“小样本学习” (few-shot learning) 生成令人信服的“人类”文本，已经引发伦理争议。这展示了深度学习网络在解释和生成 DNA 序列等更复杂语言任务中的巨大潜力。

相比其他学科，合成生物学在 DNA 设计领域更有优势。由于基因组编辑和 DNA 合成技术的进步，读取和写入数据不仅局限于所收集到的数据，还可以生成更具意义的训练数据并嵌入深层次的理解。作为机器学习的范式，主动学习有助于补充学习模型，并且能够集成到自动化工作流程中。已有几家生物科技公司部署了这样的方法，针对性地搜集生物序列空间。此外，内置数据的管理还减少了训练模型所需的数据量。

未来深度学习模型可以帮助人们改变现有方法，将 DNA 序列标准化并用作生物铸造组件的模块。另一方面，深度学习本身能受益于人们对 DNA 序列更为深入的理解。随着人们对 DNA 进化更好的理解，期待能从中获益并带来更好的设计。

3. 全细胞模拟设计

随着组学的发展，生物学相关数据每年呈指数级增长。十年前，研究最小细胞的研究人员率先利用这些数据模拟了生殖支原体中所有基因和蛋白质的行为，这种细菌的基因组有大约 500 个基因。研究人员开发了细胞中所有关键过程的数学模型，利用组学数据将这些模型参数化并集成到细胞周期的动态模拟中。系统生物学这一进展让人们对于细胞资源利用有了新的认识。对合成生物学来说，最令人激动的或将是预测当基因从基因组中删除或引入时，细胞会发生什么。

想象一下，如果可以利用在线提供的千兆字节组学数据对常见工程模型生物的细胞及其功能进行类似或更好的模拟，将会发生什么呢？从理论上讲，可能会改变合成生物学设计的全过程，可以先设计、模拟和优化基因回路和代谢途径，再对其预期性能进行模拟，对宿主细胞环境影响进行测试。“宿主细胞维度” (host cell dimension) 将成为基于模型设计的参数，帮助合成生物学家更好地考虑细胞内工程的连带效应 (knock-on effects)，例如，资源利用、代谢产物通量、基因调控的可追溯性等。在某种程度上，系统生物学和合成生物学的进步已接近这个目标。基因组规模的代谢模型现在常用于推进许多微生物系统中的生物合成细胞工程，其中最先进的模型包括表达酶的资源成本。全细胞模拟是这些模型的逻辑扩展，可用于代谢工程之外的更复杂的合成生物学研究，例如，转录因子水平变化对逻辑回路的影响。

如果支原体中使用的方法是扩展的，全细胞模拟或将作为生物设计工具。未来，大肠杆菌的全细胞模拟是最令人期待的，它将为合成生物学家提供第一个真正的测试平台，以了解如何设计工程菌株和基因线路设计，并在理想情况下减少工程细胞的实验试错。

然而，人们也需要认识到，模拟并不总能给出正确的预测。模拟可能只是作为缩小设计空间的方法之一，但希望通过更多人的使用和测试，其预测能力将不断提高。模拟对于从头设计和构建定制基因组也至关重要。信息存取和访问也将是关键问题。运行生殖支原体的约 500 个基因的细胞周期对计算已经提出了很高的要求，随着基因数量上升（大肠杆菌有 1000+ 个基因），网络复杂性也呈指数增长，会使模拟计算时间激增。随着这一进展的推进，学术界需要思考如何将复杂的模拟变成研究中可访问的设计工具。

4. 生物传感：随时随地检测

自然界充满了生物自然感知和响应的信息，将这些机制纳入工程细胞制造生物传感器，已经成为 20 年来合成生物学研究的主要内容。此外，研究人员还在开发活细胞传感技术。当人们能够随时随地检测任何事物，将在各个方面彻底改变世界，不仅是对研究领域，而且还包括大流行预防、个人健康评估、甚至地球健康方面。

如何才能获得通用、模块化的生物传感架构呢？研究人员们已经利用不同形状和尺寸的蛋白质制造生物传感器，最近还通过 Toehold 开关和 CRISPR 酶制造了检测 DNA 和 RNA 序列的生物传感器。改进天然蛋白质、寡核苷酸配体和其它大分子的表征是一个很好的开始，将极大扩展用于检测的目标分析物。最近几项具有里程碑意义的研究表明，基于蛋白的通用传感器是可行的，并开始建立核糖开关设计和定向进化。此外，AlphaFold、Rosetta、RaptorX 等在蛋白结构预测方面也取得重大进展，未来能够调整蛋白和 RNA 的灵敏度和特异性，以实现针对不同配体的传感。

生物体响应不同刺激的差异基因表达也可以用作间接捕获有关细胞响应环境复杂信息的手段，而无需为目标分子制造单个特定传感器。通过分析细胞的转录组状态，可以了解其不同“遭遇”的特征表达谱，为研究那些不存在特定生物感应器的物体提供途径。在很多方面，这类似于人们闻到食物时的反应——鼻子里有许多不同分子的传感器，使大脑能探测到更复杂的刺激的特征模式。基因逻辑回路和基于重组酶的记忆系统可以设计到细胞中，响应转录组的模式，并将这些信号转化成可以读取的信号。

更具挑战性的问题是如何广泛部署和读取生物传感器的数据。首先,要考虑生物传感器的寿命。封装或冻干全细胞或无细胞生物传感器有助于延长寿命,并且可在没有冷链的情况下运输。将传感基序设计成自然抗压的生物结构(例如孢子和生物膜),也可以在恶劣环境下实现传感功能。第二个考虑的因素是生物传感器的易用性。理想的生物传感器需要最少的预处理和后处理,自主且易于部署。第三个考虑的是从生物传感器获取数据。比色法报告可能对即时诊断最有用。有图像捕捉和处理能力的智能手机的普及,将推动比色法输出的生物传感阵列的现场分析。将生物传感器的信息传递到外部电子单元可能是长期目标,这将促进对生物的实时监测。实现这一目标不一定需要物理的生物-电子接口,例如植物和土壤发出信号,由多光谱成像的无人机和卫星可进行远程记录。

5. 实时精确控制进化

合成生物学的类比和语言构成了工程生物学的研究框架,就像设计汽车和计算机一样。但大自然不是工程师,它更像一个修补匠,利用进化修正和改进生物系统,虽然概念上简单,但进化实现了极其复杂的系统,并根据环境实现了优化。这也是生物学的重要特征,研究人员应该致力于将其纳入合成生物学的研究中。

目前,大多数合成生物学研究项目是设计和构建细胞或菌株,使其实现生物合成或生物传感等功能,但生命本质上是动态的,可能会出现突变等问题。即使是在一个控制良好的环境中进行小规模进化,突变问题仍无法完全避免。因此,在生物设计中应利用进化,研究不同环境和应用中如何控制和引导工程化的基因、细胞和有机体发挥作用。

该领域所需的生物技术工具正在快速涌现。基于 CRISPR 系统和黑色素瘤抗原(MAGE)等相关进展,使人们能够设计、指导和控制生命系统中某些位点和基因的突变;深度学习模型能预测 DNA 和氨基酸变化将如何影响基因调控、蛋白质结构和功能;定向进化可以产生数百种酶,使宿主适应更好的代谢生物合成,甚至是优化基因回路。此外,基因定向进化的分子工具也已出现。设想构建的工程细胞,其目标基因和染色体具有外部可控进化的功能,可以随时对其优化使之适应新环境,或是在原设计基础上不断进行改进。这样的系统可以快速合成并进行构建,在应用中不断进化,实现自我优化。

具有进化的合成生物学显然是非常有吸引力的,但它也需要一种全新的思考和设计方式。合成生物学家热衷于对 DNA 设计的可测量结果进行表征、预测和标准化,例如,靶

蛋白质水平、代谢物产量和指数增长率等，但是如何设计、预测、评估“进化潜能” (evolutionary potential) 呢？可能首先要转变思维范式，如果在这方面能够有所进展，合成生物学未来 20 年的工作方式将发生革命性变化。

6. 细胞群和多细胞系统

合成生物学的挑战之一是设计细胞的分化和特异性，从而在合成多细胞系统中实现有效的“分工”。迄今为止，合成生物学主要集中在通过单一任务设计群体中的每个细胞。研究人员其实在以往实验中已经报道了生物膜的程序边缘检测和基于群体的振荡器等技术，以及针对更复杂的多细胞功能的研究，例如，基于群体的运算 (consortia-based computation)、分子图灵结构、合成形态发生等。

群落感应系统是细胞间通信的方式，可进行时间、动态和正交水平上的控制。例如，AHL (N-酰基高丝氨酸内酯) 介导的细菌群落感应系统已有所应用。真核生物的细胞间的信息是通过生长素或者基于 GPCR (G 蛋白偶联受体) 的信号进行传导。近期的研究也表明，哺乳动物细胞中通过 Notch 信号通路系统的组合式传感器是非常有效的。研究已经明确了如何构建一个通信回路，其中最有效的是基因组件的模块化工具包，为许多细胞提供可靠、动态和正交的细胞间通讯。然而，仅凭这些并不能构建可预测的群落和多细胞系统，考虑新的因素 (例如成长竞争、共生关系、细胞合作、休眠和避免突变) 或将变得越来越重要。

类似于神经网络在复合学习 (composite learning) 中的互相连接和支持，工程细胞的共培养也提供了相似的优势。共培养可以让不同菌株作为功能模块提高天然化合物的生产效率，而不需要每个细胞都承担复杂代谢途径的全部负担。“细胞作为模块”的方法可以提供“即插即用”的有机体和细胞群体，也可投入新系统中发挥作用，就像计算机代码中可重用的函数。实现这样的系统需要确定细胞群落组分间动态范围的兼容性，以及其通用性和可组合性。

具有多种相互作用的工程细胞系统为推进合成生物学提供了巨大的动力，但这些并不仅限于共培养或相互依赖的细胞群落。植物和动物等多细胞生物的复杂性、稳健性和多功能性已经展现了生物自身系统的强大功能。人们通过编程 DNA 指导细胞连接并以不同方

式成长的能力正在不断成熟。随着在可预测生物群落的共培养和工程化方面的发展,应该同时推进组织和器官的理性设计,使不同层面的工程化进展实现交叉融合。

7. 定制和动态合成基因组

合成生物学大规模解决的下一个领域是合成基因组。10 多年前, J Craig Venter 研究所首次合成了具有合成基因组的细胞。此后, 研究人员陆续实现了只含必需基因的最小化支原体基因组, 人工合成的工程化酵母染色体, 使用 61 种密码子合成的大肠杆菌基因组。合成基因组的标准方法已经开始出现, 尽管目前 DNA 合成成本和工作量仍然较大, 但已经可以下降到可以由单个团队实现百万碱基染色体的制造。下一个 10 年, 合成基因组学也将开始瞄准多细胞生物(例如哺乳动物和植物)的基因组。

目前, 大多数合成基因组项目旨在提供新知识, 使人们对基因组编码、内容和组织有新的见解。未来, 随着 DNA 合成越来越便宜以及自动化和可扩展的 DNA 组装出现, 面向应用的合成基因组设计成为新挑战。密码子删减的基因组可以提供遗传密码扩展到应用的能力, 尤其是在细胞制造的蛋白质中添加非经典氨基酸。为其它应用定制的基因组或将紧随其后, 例如, 为特定合成生物项目定制的基因组。随着该领域出现更大的多细胞生物基因组, 可能只需要定制那些特定作用的基因。例如, 细胞疗法越来越受欢迎, 但出于安全考虑, 未来更希望从最小合成基因组的细胞中实现这些疗法, 除去与治疗功能无关的基因, 且避免意外风险。

然而, 设计定制基因组并非易事。全细胞模拟和组学实验中的深度学习有望更快接近这一时代, 有助于更好地从头“组合”所需基因和调控序列。也许在此之前, 将会出现更为激进的想法, 即动态基因组, 可以有意将重组酶及其结合位点、其他变异热点纳入现有的基因组, 然后触发这些基因组随着时间推移发生物理变化, 以新的方式发挥作用。随着细胞分化, 可以通过删除多余基因来进行精简。不久的将来, 定向进化和动态合成基因组将是可并行发展的两种强大工具。

8. 人造细胞

虽然研究人员为现有细胞合成新基因组付出了相当大的努力, 但一个更为基本的合成细胞方法已在化学和生物化学领域中出现, 该方法纯粹通过组合生化成分来重建活细胞的

基本功能。目标是“自下而上”地生产人工细胞，以实现生物学中基础细胞的可编程合成模拟。这似乎是一项艰巨的任务，但在过去 10 年中，人工细胞研究的基础已经解决了许多实际问题，为不断增多的复杂研究项目提供了发展空间。

通过“构建”来帮助理解是贯穿合成生物学的重要理念，也是人工细胞研究的主要驱动力。当人们从惰性分子中创造出一个自我复制的实体时，合成细胞将帮助区分非生物与生物之间的屏障。在更实用的层面上，还将帮助人们更好地了解天然细胞的工作原理。利用人工细胞（例如脂质囊泡原细胞）已经可以研究分子生物学的各个方面。随着微流体、化学区室化、DNA 合成等领域的快速发展，这一领域有望成熟，帮助人们更好地研究简化生物系统的工程化。

虽然在原始细胞中实现大规模生物行为（例如细胞周期）仍旧很棘手，但研究人员正对通过构建微结构实现吞噬作用、细胞间通讯，甚至自我复制等功能。巨型脂质囊泡为构建早期细胞样囊泡的实验打下了基础。作为人工细胞器的囊泡也激发了细胞中合成和自然成分的混合，例如充当半人工生物反应器的脂质体。这些模块可进一步开发新的实验方法，例如与新型化学信使进行细胞间通讯。总的来说，区块化在细胞内构建了稳定且不受混杂因素干扰的环境，并通过实验创造能够计算、自我组装，且有朝一日能够实现自我复制的系统。

随着人工细胞研究成果的不断涌现，研究团队开始合作构建细胞，包括美国的 Build-a-Cell 和欧洲的 SynCellEU。现在，是时候开始思考人工细胞实现后有哪些应用。对于合成生物学，可定制的原始细胞最终可能会取代天然微生物系统，成为合成生物学工程系统的首选。合成细胞自身的可编码系统也可以作为一种工具，或许可以将生物学中进化、自主性、自我再生等结合在一起。

人工细胞研究与“异源生物学”（xenobiology）密切相关。过去 10 年中，DNA、RNA 和肽中使用非经典生物化学分子修饰细胞已经在小规模实验中取得成功。人工细胞或将成为研究非标准核酸、转化糖（flipped sugar）、手性肽以及不同密码子排列的理想载体。

9. 具有 DNA 编码特性的材料

由细胞群落构成的工程活性材料 (ELMs) 是新兴领域。细胞作为工厂推动了合成生物学行业的一些重大成果, 并促进了与食品、时尚等其他领域的交叉发展。细菌纤维素、蜘蛛丝和蘑菇等材料是用作纺织品、家具、先进建筑物中建筑耗材可持续性的基础。此外, 得益于 iGEM 项目, 利用微生物大规模生产材料的范围正在不断发展。除了大肠杆菌, 合成生物学主要相关生物和相关细菌 (例如葡糖醋杆菌和芽孢杆菌) 都是分泌纤维素和高度耐用淀粉样纤维的优质生产者。不久的将来, 遗传工具包的扩展将加速可用于工程材料生产的微生物的研发。

虽然由生物生产的材料替代品构成的绿色领域已成为可能, 但真正面临变革的是生物学作为制造学科, 其优势在于细胞不仅可以当作微型材料工厂, 还可以在分子水平上修改材料, 在制造技术方面几乎没有额外成本。通过光遗传学或细胞间信号传递进行微图案化, 对机器来说或许是一项复杂任务, 但对微生物而言却很容易。这种图案化对于医疗保健应用或需要微米级特征的先进电子产品十分有用。此外, 还能在材料中引入诸如颜色和疏水性等高级特性, 就像在细胞中插入额外 DNA 触发功能蛋白和酶的表达一样容易。当然, 能够大规模生产新材料的制造技术也面临挑战, 一旦克服限制条件, 这种新型材料将更具竞争优势。几家合成生物材料初创企业都在与减碳行业合作。

这些进展和工作为工程活性材料未来的复杂性奠定了基础。例如, 软体机器人是用一种光响应的凝胶工程材料制成, 其中含有能够通过光脉冲触发跳动的细胞。类似这样的项目也影响着人们对新型材料的看法。理想情况下, 如果合成生物学可以实现标准化的重大进步, 就可以通过细胞内的模块化 DNA 编码程序来定义和设计材料的特性和功能。

10. 为可持续发展目标设计有机体

生物技术一直在改善着人们的生活。过去几十年, 科研人员将大部分精力花在研究人类繁荣和长寿上。利用合成生物学改善健康是毋庸置疑的, 毕竟当人们听到“生物学”时, 首先想到的是自己的身体。生物学不仅仅是人类健康, 在下个世纪, 生物学或将是预防气候灾害和保护地球环境的关键, 而合成生物学将成为解决这些问题的核心。

值得庆幸的是，可持续性已经成为合成生物学的关键目标和趋势。全球推动的循环经济将依赖工业的生物转型，人们也正积极开发和应用工程生物帮助投资人、公司和国家实现可持续的发展目标。工程生物有潜力创造可再生能源、延长产品生命周期、保护环境和生物多样性、提高粮食可持续性以及改善健康等。合成生物学的进步正加快生物技术创新，在保护环境的基础上为现有经济体创造新价值。合成生物学企业正在成为重要领域并发挥积极作用，例如固氮、二氧化碳和沼气收集、塑料回收、植物性产品和污染修复等。植物基肉类和人造肉也是减少人类对地球影响的重要方式。这些新型、更可持续、更合乎道德的食品生产方式需要合成生物学的不断创新。

然而，前进的道路充满挑战。决策者需要有先见之明，不断争取社会对以生物为基础的策略的支持。许多技术挑战也仍然存在，例如，在能源和制造业，扩大生物技术规模仍不简单，需要一套专业的体系将生物创新从实验室规模转化到工业规模；需要对生物进行工程改造降低生物技术成本。这些挑战需要各国决策者和投资者的长期支持，以及研究人员持续不断的探索。可以相信，该领域的回报也将是巨大的，不仅是对环境，也对国家经济。因此，所有的合成生物学家都应该准备好迎接这些挑战。

结语

基于合成生物学领域的十大进展，希望在不久的将来，可以完全改变人们的工作和生活方式。然而，如果不及时解决新技术带来的社会和政治问题，将很难最大限度发挥这些科学进步的作用。因此，合成生物学领域的研究人员需要经常与公众接触、倾听不同观点，并及时调整研究重点。当涉及像地球健康这样的重大问题时，地球上的每个人都是利益相关者，所有观点都应加以考量。当为获得安全使用适当的监管时，世界才将接受合成生物学。接下来的几代人需要负责任地开发这些技术，理性、公平和安全地使用这些技术来保护自然，而非消耗自然。

刘晓 朱成姝 编译自 *Engineering Biology*