

在这里填上你的标题

第一作者

单位 1 (到系所)

xxxx@xxxx.xxx

摘要: 随着合成基因线路结构的日益复杂, 基因线路行为变得难以预测。经研究发现, 基因线路和底盘细胞存在难以避免的耦合, 而这种相互作用往往导致了底盘细胞生理状态的改变并影响线路功能。依照传统的工程化思想设计基因线路的方法进入了一个瓶颈。本文阐述了原核底盘细胞与基因线路耦合的作用机制以及相关的定量工作。并进一步介绍了通过建立数理模型和正交化、模块化等手段预测或减轻两者之间耦合效应的相关研究进展。

关键词: 大肠杆菌; 合成生物学; 系统生物学; 底盘细胞; 基因线路; 相互作用; 生物物理模型; 正交化; 模块化

中图分类号: B81

文献标识码: A

1 引言

本世纪初两项创造性的工作, 拨动开关和压缩振荡子, 利用工程化的思想设计人工基因线路并使其实现特定的生理功能。秉承这一思想, 合成生物学在众多学科的交叉中诞生, 逐渐在生命科学领域的研究中受到广泛的关注。近二十年来该学科出现许多重要的突破: 一方面, 通过对基因组的挖掘, 人类得到大量定量刻画生物元件, 并设计出了越来越复杂的基因逻辑线路, 它们能响应特定的外界信号并输出预期的结果; 另一方面, 利用合成生物学工具, 人类对生命系统的工作机理有了更深层次的认识。在产业方面, 大量具有高价值的复杂产物通过合成基因线路实现了工业化生产, 有希望取代传统化工合成, 降低对环境的破坏和对化石能源的依赖性。

尽管如此, 我们对基因的解读能力仍处在较低的水平: DNA 序列提供的信息高度复杂, 通过人工组合、设计的复杂基因线路放入到底盘细胞后, 底盘细胞对特定信号的响应往往出乎我们的预期。细胞在执行命令的过程中存在许多细节是我们不清楚或难以定量描述的, 而这些细节也是基因线路功能实现的重要环节。基因线路同底盘之间存在种种联系, 线路功能依赖细胞的 DNA 复制系统、转录和翻译系统以及各种代谢产物前体。不同的菌株、不同的生长条件, 都会对基因的表达量产生影响, 而基因线路中不同基因的表达量决定了线路的输出结果; 同时, 外源的基因线路的表达对

于底盘细胞而言是一种“负担”，外源基因表达将占用有限的 DNA、RNA、蛋白质的合成机器以及各种底物，这种资源上的占用将改变底盘细胞内源的基因表达，进而改变底盘细胞的生理状态。因此，基因线路和底盘细胞两者存在天然的耦合。然而长期以来，我们对合成基因的线路设计思路主要集中在单独元件的定量和预测上，忽略了底盘细胞与基因线路元件之间的影响。所以，利用特定条件下的单个模体的定量数据来预测不同工作背景下在底盘细胞内的行为会产生偏差，导致合成基因线路设计不可避免的进入无限循环的“设计—构建—调试”中。不仅如此，许多工程菌，如环境治理工程菌、肠道菌群改造工程菌等，其工作环境往往不是标准的实验室条件，底盘细胞生理显然受环境的影响而发生了变化，导致很多基因线路无法实现预期功能^[1]。

基因线路设计进入到“瓶颈期”的同时，得益于系统生物学理论以及组学实验方法的发展，近年来，许多工作使得我们对于底盘细胞生理的理解有了更加全局、定量的认识。将底盘细胞的生理状态纳入到对基因线路行为的理解中，许多反直觉的问题迎刃而解^[2]；同时，这些理论也被应用到了合成基因线路设计的实践中，给合成基因线路设计带来了新的思路。除此之外通过开发正交性元件和通过更加精巧的模块化设计也为解决这一问题提供了新的新思路^[3]。本文将梳理近年来合成基因线路与底盘细胞相互耦合的理论研究，介绍底盘细胞与基因线路的相互作用，以及与之相关的定量技术，并进一步介绍近年来关于解开底盘-线路耦合的成果。基因线路设计进入到“瓶颈期”的同时，得益于系统生物学理论以及组学实验方法的发展，近年来，许多工作使得我们对于底盘细胞生理的理解有了更加全局、定量的认识。将底盘细胞的生理状态纳入到对基因线路行为的理解中，许多反直觉的问题迎刃而解^[3]；同时，这些理论也被应用到了合成基因线路设计的实践中，给合成基因线路设计带来了新的思路。除此之外通过开发正交性元件和通过更加精巧的模块化设计也为解决这一问题提供了新的新思路。本文将梳理近年来合成基因线路与底盘细胞相互耦合的理论研究，介绍底盘细胞与基因线路的相互作用，以及与之相关的定量技术，并进一步介绍近年来关于解开底盘-线路耦合的成果。

2 底盘细胞和基因线路的相互作用

细胞通过一系列的生理生化过程从外界环境中摄取维持生长所需的物质，通过同化作用转化为提供复制、转录、翻译和酶促反应等生物过程所需的资源。在不同底盘细胞中，或者同一底盘在不同的环境和生长条件中，细胞会将其资源以不同的策略分配给不同的环节，以最大限度地提高其适应性^[4-6]，分配策略的变化会导致线路的响应发生变化。同时基因线路作为外源基因，会竞争各种共享资源，如 RNA 聚合酶、核糖体和氨基酸，以及各种生化反应所需的酶和能量^[7, 8]对底盘细胞的生理状态产生影响。而且不同线路模块之间同样也存在竞争，这种竞争同样会导致线路的输出不符合预期 (图1 a)。本节我们将介绍关于线路-底盘相互作用的机制的研究成果，并讨论这些机制带来的效应。

2.1 底盘细胞对基因线路的影响

不同生物底盘之间的基因背景差异会导致标准化的基因元件的输出结果迥然不同^[9]，其差异甚至能达到一千倍以上^[10]，这为合成生物学的标准化设计提出了重大挑战。Moser 等研究指出转化到不同大肠杆菌菌株中的 AND 逻辑门会有出现不同的表现^[11]。而不同营养成分的培养基以及不同的培养条件也会对基因元件的响应产生重大改变。此外，细胞中蛋白的降解主要以细胞生长的方式对胞内蛋白进行稀释^[12]，因此细胞的生长速率对基因表达量的影响至关重要。而这种生长速率对基因线路的反馈作用，可能导致多种多样不可预知的结果。Tan 等人分析了由自激活突变体 T7 RNA 聚合酶组成的基因线路，虽然该自激活过程是非协同的，但该线路的基因表达水平却是双稳态的。然而，这一违反直觉的现象可以解释为，基因线路的表达引起了细胞生长速率的减慢，而同时细胞的生长对基因元件的表达存在着反馈调节，进一步增加蛋白的积累，从而实现了双稳态现象^[13]。

2.2 基因线路对底盘细胞的生长影响

人工基因线路的表达占用了底盘细胞生长代谢相关的资源，这就可能形成压力 (burden)^[14]，影响细胞的正常生长。当表达与生存不相关蛋白占细胞蛋白总量的约 30% 时，细胞会完全停止生长^[4, 15]。外源基因线路对 RNA 聚合酶和核糖体^[16] 的竞争是导致这种压力的主要因素。另外，外源基因的表达甚至会引起底盘细胞的整体资源分配策略的改变。细胞本身存在动态调节分配策略的机制^[17, 18]，当外源蛋白持续表达几代之后，大肠杆菌细胞可以利用 ppGpp 的调控机制，实现对全局性资源的重新分配^[18]。当 ppGpp 水平变得太高或太低时，细胞生长都会受到抑制。较高的 ppGpp 水平通过直接抑制核糖体合成来抑制细胞生长。相反，由于来自核糖体过量生产的资源竞争，较低的 ppGpp 水平可通过限制代谢蛋白的表达水平来抑制细胞生长。在引入外源基因线路的条件下，外源蛋白的持续表达使核糖体的生产达到顶峰，同时 ppGpp 的积累量也很低，这导致了核糖体的合成被调整以应对额外的蛋白质产生负荷。除此之外，元件与底盘细胞的特异互作可能导致毒性，比如同代谢过程中的酶发生互作等。

2.3 基因线路内部不同模块之间存在竞争

基因线路中不同的模块之间也会形成竞争^[19, 20]，当一个线路中所有的基因都在争夺这些有限的资源时，基因之间就会产生非预期的相互作用，这些相互作用可以极大地改变基因线路的预期行为，导致实验结果与模型预测完全不符。例如，Qian 等人的工作表明，当两个激活模块单独存在时，希尔函数可以很好的吻合实验结果。但是将两个模块串联形成更复杂的体系后，理论预测的结果甚至和实验结果完全相反^[21]。

2.4 其他效应

除了上述三种竞争效应外，底盘细胞和基因线路的互作还给耦合系统带来了一系列其他效应，例如，增加同基因群体的异质性，降低进化的稳定性以及毒性效应。

基因线路导致底盘细胞生理异质性增强

在细菌细胞中，资源的分配比例可能会随着个体的不同而发生变化，从而使一个细胞在执行的不同功能时存在随机化^[22]。例如，利用大肠杆菌内源降解系统 ClpXP 已经证明，底盘细胞内可能因为某些资源的紧缺时产生类似排队效应的随机过程，引起异质性以及意外的基因表达耦合^[19]。另外，在同基因型的群体中，细胞间微小的差异也会在群体水平上产生表型多样性^[23]。外源基因的表达使得这种随机扰动变得更加明显，从而导致底盘细胞的生理异质性增强。而这种异质性导致一小部分表达量相对较低的细胞获得相对较快的生长速度。这使得这类细胞在不断的传代培养中取得优势，最终导致合成基因线路的失效^[13](图1 b)。

系统的进化稳定性降低

细胞的生长压力带来的负向筛选作用会减弱人工基因线路的遗传稳定性。研究表明合成线路功能可以给细胞带来很大的适应性劣势 (fitness disadvantage)，并且发生适应性进化以减轻与表达外源蛋白质相关的代谢负荷或毒性。基因线路给底盘细胞带来压力同时导致底盘细胞生长速度往往低于野生型菌株，而突变导致的含有失效的基因线路的菌株往往具有更好的适应性，这样一来，在连续培养的过程中，带有期望功能的合成菌株会逐渐失去优势。Sleight 等人^[24] 评估了带有不同的调控元件的三种荧光蛋白在大肠杆菌中的进化稳定性，发现表达量低和带有更少的重复序列的基因线路具有更好的进化稳定性 (图1 c)。重复序列容易导致基因重组进而导致序列丢失，有必要建立更大的优质元件库，减少元件在同一基因线路中的重复使用，同时在设计的环节中加入同源序列检查的过程，避免元件组合的过程中产生意外的同源序列。

基因线路导致的细胞的毒性

在基因线路的测试过程，往往会发现有些基因原件或线路虽会严重影响底盘细胞的生长速率甚至导致细胞形态发生变化，如纤维化。一般认为这种现象是因为基因线路与底盘细胞关键生物过程发生了超出预料的相互作用。一种可能是，异源 DNA 中可能存在一些意外的功能序列，比如，dnaA box 会严重影响细菌的分裂^[25]，富含 AT 序列的 DNA 序列可能是意外的启动子，高频的转录导致细菌 RNA 聚合酶不足^[26]。另外，外源基因的表达产物也可能与底盘细胞的基因组意外的互作，例如 TetR 家族中一些阻遏的蛋白异常调节底盘细胞的基因表达^[27, 28]；高表达 dCas9 也会严重影响细胞生理，通过组学实验发现了高表达水平下 dCas9 非特异性结合到基因组上影响转录水

平^[29](图1 d), 另外也有实验表明 gRNA 只含有 9 碱基与目标位点匹配的情况下, 也能有效结合到基因组上引起转录下调^[30], 这对设计特异的 gRNA 带来了一定的困难。目前, 还没有系统性的定量毒性的方法, 对于这些意外的互作情况, 使用组学技术, 如 RNA-seq 等, 可以有效地进行评估。

综上所述, 基因表达与细胞生长之间存在着多方面的联系。对这些关系的认识可以提高对基因线路行为认识和预测预测, 为合成基因线路的理性设计提供指导。

3 预测或减轻合成基因线路与底盘细胞耦合的策略

由于底盘细胞与基因线路之间存在复杂的相互作用, 近年来不少工作利用数学模型解释了不同条件下细胞的行为, 将基因线路与底盘同时纳入到模型的框架中, 使得许多基因线路失效的案例得到了解释; 利用这些模型进行辅助设计基因线路, 也建立了许多对底盘和线路解耦合的方法。本节我们将介绍近年来出现的代表性的工作, 介绍它们如何用生物物理模型去解释复杂的现象以及预测或减轻耦合效应的策略。

3.1 建立相关模型

虽然, 用生物物理模型描述线路—底盘互作是近十几年随着合成生物学而发展出来的工具, 但模拟细胞内全部或主要生理过程的思想在很早就发展起来了。从进入二十世纪以来, 伴随着组学技术、生物定量技术的发展以及运算能力的提高, 出现了全盘考虑所有已知生化过程的数学模型, 其被称为全细胞模型 (whole cell model)^{[31][32]}。也有忽略大部分生化过程, 仅考虑关键的几个生化过程的粗粒化的模型 (coarse-grained model)^[4, 33, 34]。全细胞模型需要通过序列和实验确定大量的针对每个基因的参数, 然后再整合成较高尺度模块的参数, 继而形成整体模型; 而后者通过对模型简化, 抛弃大量的细节, 通过对少量实验参数的测定配合数学手段分析构建模型。前者由于参数涉及量庞大, 对计算需求量以及精度具有较高的要求, 若要形成准确的模型还需要大量的实验测定, 这限制了这个方法的发展; 粗粒化模型抓住生物过程中关键的限速、调控过程, 其参数确定相对较为简单, 且易于理解解释, 但由于其忽略了很多细节, 对其结果也需要谨慎。

基于基因组的全细胞模型

基于基因组的全细胞模型主要是随着组学技术的兴起发展起来的, 自 2000 年以来组学方面经历了构建全基因组网络结构、整合组学数据阶段, 到拥有一定的预测能力。2012 年 Karr 等人基于组学数据对 *Mycoplasma genitalium* 进行全细胞建模, 通过对细胞周期中的每个注释过的基因产物及反应都建立了描述性的方程, 该模型可以对基因组各基因与转录因子之间的互作等诸多细胞行为做出预测, 在这基础上通过对单基因扰动的模拟发现了已注释基因的新功能^[32]。

粗粒化模型

早期细胞生理学家关注的问题是如何预测特定环境下的细胞生长速度, 上世纪末开始一批人意识到单纯的以营养物质能量大小是无法解释对应实验结果的, 并注意到前体代谢物对生长的关键作用以及蛋白质合成和底物之间的调节效应, 相关内容总结于^[35]。Scott 等人基于实验现象将蛋白质组确定为核糖体相关部分、代谢部分和其它部分, 通过构建粗粒化模型得以解释实验中观察得到的现象^[4]。2015 年 Weiße 等人系统性地从能量角度整合并精简了相关因素, 同时对可能的线路与宿主之间的作用进行了模拟并做出了较为准确的预测^{10.1073/pnas}。2017 年, Liao 等人引入 ppGpp 作为全局调控因子, 进一步对营养物质切换时细胞生长的实验现象进行了解释, 并且引入了更多的竞争机制成功解释了相应的实验现象^[34]。

底盘与线路之间的互作模型

合成生物学发展起来以后, 所需要关注的问题不仅有细胞生长, 还引入了新的可观测现象—基因线路动力学, 并有了更精细的调控要求。底盘与线路之间的互作大体上同底盘内部的资源竞争有关, 早期关于外源表达基因对宿主生理的定量影响见于 1990 年 William 的工作^[36]。竞争方面的效应同经济领域和电子系统中较为类似, 相关规律也被应用于描述该行为^[14, 20]。具体的不同线路对竞争效应的响应也在各种文献中陆续被描述^[37, 38]。在上文中, 粗粒化模型相关文献中也能见到集中讨论底盘线路互作的内容^[33], 其中 Liao 等人的工作提到了, 当前还需要定量刻画基因线路对宿主的代谢等方面的特异性作用^[34]。

模型参数的确定方法

参数按获取来源可分实验得到参数和拟合得到参数。在全细胞模型中的参数与粗粒化模型中的参数略有不同。首先是数量上, 全细胞模型的参数要比粗粒化模型中的参数多得多, 生殖衣原体的全细胞模型整理了 900 篇相关文献中的 1900 个参数, 后续还涉及自己拟合的参数以及整合后的 28 个模组各自的参数^[32], 而粗粒化模型则要少得多, Lu 关于大肠杆菌的粗粒化模型中默认参数仅 20 多个^[34]。全细胞模型中的参数更为细节, 甚至包括超螺旋程度的影响, DNA-蛋白质亲和力和大小等, 但是大多数的参数都无法直接测得需要通过流平衡分析 (Flux Balance Analysis)^[39]、随机过程等数学手段推断得出^[40]。粗粒化模型中也存在很多无法通过实验测得的参数, 需要借助同样的数学工具进行推导, 粗粒化模型中涉及的实验参数的测定较为简便, 比如核糖体占比的测定, 以 RNA/蛋白质形式测得, 后续经过简单转换可得出核糖体占比^[4]。近些年, 无细胞系统凭借其方便的变量调节特点, 在细胞生理参数测定方面受到越来越多的关注^[3]。值得注意的是两种方式得到的参数都存在各自的争议, 实验测得参数由于实验方法和精度的限制可能无法准确反映真实数值; 流平衡分析是目前应用最多的参数拟合方法, 能较为准确的预测不少参数, 但是也有一些参数由于其基本假设同实际

情况而不符受到质疑, 比如细菌内部在营养较为缺乏的时候存在一定比例的冗余核糖体为了能在营养条件改善时更快的提升生长速度占据生长优势, 而并非以达到平衡为优化目标^[41]。

3.2 合成基因线路正交化

正交化是希望基因线路在实现其内部的逻辑调控的同时, 尽可能减少其内部元件与元件之间, 以及元件与底盘细胞自身系统的不必要的交互。基因线路会不可避免地使用底盘细胞的各种资源, 在特定情况下也需要建立与底盘细胞系统的接口, 比如在进行天然产物的生产中, 将对底盘细胞原有的代谢网络进行连接; 当前, 正交化的工作主要集中在对基因元件的正交设计中, 通过对基因元件的定量测试和定向进化, 使其发挥出更好的功能, 同时减少对底盘细胞的不必要干扰; 另一方面, 正交化基因线路, 减少了与底盘细胞的过多交互, 则在进行模拟预测的时可以减少参数, 提高系统的可预测性。通过在基因库中挖掘, 正交化的 DNA 复制、转录、翻译系统也不断涌现出来, 正交中心法则系统这一概念也被提出。正交中心法则系统希望中心法则中涉及的生物合成过程使用一套正交的系统完成, 可以类比为计算机科学中虚拟机的概念^[42]。

完整的 DNA 正交复制系统应当可以在 DNA—DNA 聚合酶层面实现正交, 即一种 DNA 聚合酶专门负责复制特定的 DNA 序列; 例如, 在酿酒酵母中, 源自于 *Kluyveromyces lactis* 的两个细胞质质粒能够利用相互独立的 DNA 聚合酶进行复制, 通过基因工程改造, 可以独立控制两个质粒的复制突变率^[43]。正交转录机器则早已广泛在合成生物领域使用, 噬菌体 RNA 聚合酶便是使用最多的元件, 这类 RNA 聚合酶识别特定的启动子序列进行转录, 大多为单体酶, 且具有很强的转录活性。同时他们也容易被改造成成为蛋白相互作用层次的逻辑门, 例如对其结构域进行分割, 融合其他类型的蛋白或结构域可以丰富其功能。例如, 将 T7 RNA 聚合酶分割为两个部分, 分别融合感光同源二聚蛋白结构域 (VVD), 可以使其在蓝光下形成二聚体从而激活转录^[44, 45]。类似的, 将 RNA 聚合酶与 DNA 识别结构域与锌指蛋白 (ZFNs), 转录样激活因子 (TALENs) 或 CRISPR Cas 蛋白等具有可设计能力的核酸酶进行组合, 可以扩展其使用的范畴, 例如拓展其转录动力学范围^[46]。另一类正交转录系统则基于 σ 因子, 利用不同来源的 ECF σ 因子, 通过共享 RNA 聚合酶的核心酶库, 不同家族的 ECF σ 因子对启动子的序列的不同亲和性, 可以设计出正交的 ECF σ 因子组合, 实现在转录层面的绝缘^[47]。蛋白的翻译涉及到大量复杂的生物过程, 通过理性化和定向进化等手段, 近年来也取得大量的进展。改造氨酰 tRNA 合成酶可以将非天然氨基酸引入到蛋白的序列中, 有希望给蛋白带来更多样的生理功能; 全合成基因组的手段, 成功将 64 种密码子压缩到了 61 种, 为编码非天然氨基酸打下了基础^[48]。通过改变核糖体 16S rRNA 的序列, 可以得到正交核糖体, 识别非典型的 RBS 序列, 从而绝缘合成基因线路和底盘细胞的基因表达^[49, 50]。

转录因子是合成基因线路设计中主要承担逻辑运算的元件, 这类元件识别特定的

序列激活或抑制下游基因的转录, 未经正交化的转录因子存在相互之间交叉识别位点的风险, 另外对底盘细胞管家基因相关序列的结合会影响细胞生长, 产生毒性。目前转录因子除了典型的阻遏蛋白和激活蛋白, 还有具有识别序列功能的蛋白, 如 dCas9、TALENs 等。但是这类蛋白往往具有细胞毒性; 通过突变 dCas9 的 PAM 识别区并融合阻遏蛋白 PhlF 的 dCas9_Phlf 可以增强其特异性减少毒性, 使细胞能耐受更高的表达量。另外, 通过定向进化等手段对阻遏蛋白进行优化, 也有助于降低细胞毒性^[51]。

对正交系统进行合理的设计, 可以部分解决蛋白表达过程中资源竞争的问题。在外源基因表达的过程中, 两个独立的基因线路因为共享底盘细胞的各种资源库, 因此两个基因线路的表达相互耦合, 可以用“等成本线”来描述。利用正交核糖体系统, 通过表达正交 16s rRNA 将底盘细胞的核糖体库分为两组, 其中一组倾向于翻译带有对应的正交 RBS 的基因线路, 利用这套系统减轻了由于基因表达过程种资源分配带来的耦合问题^[52]。

3.3 模块化基因线路设计

合成生物学的一大愿景是实现基因元件的即插即用^[53], 即通过对基因元件的排列组合实现复杂的调。除了在元件设计过程中的模块化, 更高层次的模块化进来得到了广泛的关注: 将功能线路, 例如逻辑门等, 进行模块化, 使得每一个实现特定功能的线路的输入输出在插入到基因线路中时不发生变化^[3]。进行功能模块化首要要解决回溯效力 (retroactivity)^[21], 基因线路中各个功能模块之间的相互作用主要由转录因子来实现, 而转录因子与下游模块的作用, 例如结合到基因位点将导致游离转录因子浓度, 进而扰动上游模块; 对于底盘细胞而言, 基因线路对转录翻译资源的消耗, 将改变细胞生长速度以及各种资源库的总量, 进而又影响了基因线中的各种动力学参数, 如转录翻译速度和稀释速度。总而言之, 功能模块化的目标是希望各功能板块间输入输出不受外部参数影响而变化。

负反馈调节可以使基因线路的稳态输出水平仅与自生元件的动力学性质决定, 而不受基因表达生成和稀释速率影响。当前大量基于反馈控制的解决方案已经被应用到基因线路的设计上, 可以归结为局部水平和全局水平两类 (图3)。

局部水平的反馈策略主要着眼于功能模块输出本身的稳定性, 希望利用负反馈系统来隔离模块与模块之间, 以及模块与底盘细胞之间的相互影响。例如, 使用转录后调控元件 sRNA 形成负反馈环, 使每一模块的稳态表达水平都得以固定, 使得它们对来自资源竞争产生的扰动具有抵抗的能力 (图3 b)^[54]。Aoki 等人则采用了对偶积分反馈的策略, 在保证了模块输出抗扰动能力同时, 还模的块稳态输出水平是动态可调的 (图3 c)^[55]。

基于全局的资源调控反馈策略能够根据基因线路表达带来的负担对所需的资源进行调控。在正交核糖体基因表达系统中加入反馈控制模块, 能够实现对正交核糖体与内源核糖体比例的动态分配, 使正交核糖体的比例按需分配给基因线路, 减轻基因线路内部的资源竞争带来的耦合 (图3 e)^[52]。

另一种反馈策略则由宿主生理状态决定, 根据代谢压力的大小来决定外源基因的表达。通过实验筛选出外源基因高表达时转录水平上调的启动子, 利用这类启动子激活下游基于 dCas9 的转录抑制系统, 从而确保外源基因线路对细胞的压力维持在一个较低的水平 (图3 d)^[56]。

在代谢工程领域, 为了增强基因线路的遗传稳定性, 会利用与生理状态偶联的正反馈回路 (图3 f)^[57, 58]。

除了反馈环这一策略外, 另一种基于转录调控的模块化手段则采用了非协调前馈环 (incoherent feedforward loop, iFFL), Segall-Shapiro 等人通过对 iFFL 模体的分析, 指出当 iFFL 模体中阻遏蛋白无协同效应时, 稳态下模体的输出将为一个常数而不依赖基因的拷贝数、细胞生长速度, 通过这一简单的策略, 可以使基因在不同拷贝数下实现稳定表达 (图3 g)^[59]。

4 总结与展望

过去 20 年来, 合成生物学领域涌现出了大量创造性的工作, 证明了利用合成生物学理论、方法改造生命为人类所用的可行性。未来, 合成基因线路需要能够适应更多的遗传背景迥异的底盘细胞, 以及更加复杂多变的使用环境。然而, 由于我们对底盘细胞缺乏更系统、全面的认识, 导致更为复杂的人工生命系统的开发工作效率低下。本文总结了当前关于细胞生理定量, 底盘细胞与底盘细胞相互耦合的机制, 以及如何避免因耦合引起的基因线路失效等方面的研究进展。我们认为在未来应当注重: 1. 更加系统、量化的底盘细胞生理研究, 理清底盘细胞在各种条件下, 如逆境、非稳态, 的资源调控机制, 扩大研究的底盘细胞的种类; 2. 挖掘、定量、优化更多的正交基因调控元件, 阐明不同遗传背景的底盘细胞基因元件的性能; 3. 开发更加精准、高通量的定量手段, 利用多层次的组学技术, 为定量刻画底盘细胞—基因线路相互作用机制提供更加可靠、全面的数据; 4. 建立更加精准的模型预测框架, 包括: 将底盘细胞与线路的模型描述整合在一起, 增强基因线路行为的可预测性; 以及从 DNA 序列生成环节出发, 检查序列组合中意外产生的功能序列, 如 dnaA 结合序列、启动子、转录因子结合序列等; 5. 规避“非预期”的底盘细胞与基因线路相互作用的解决方案, 如利用反馈控制机制设计更加模块化的基因线路。

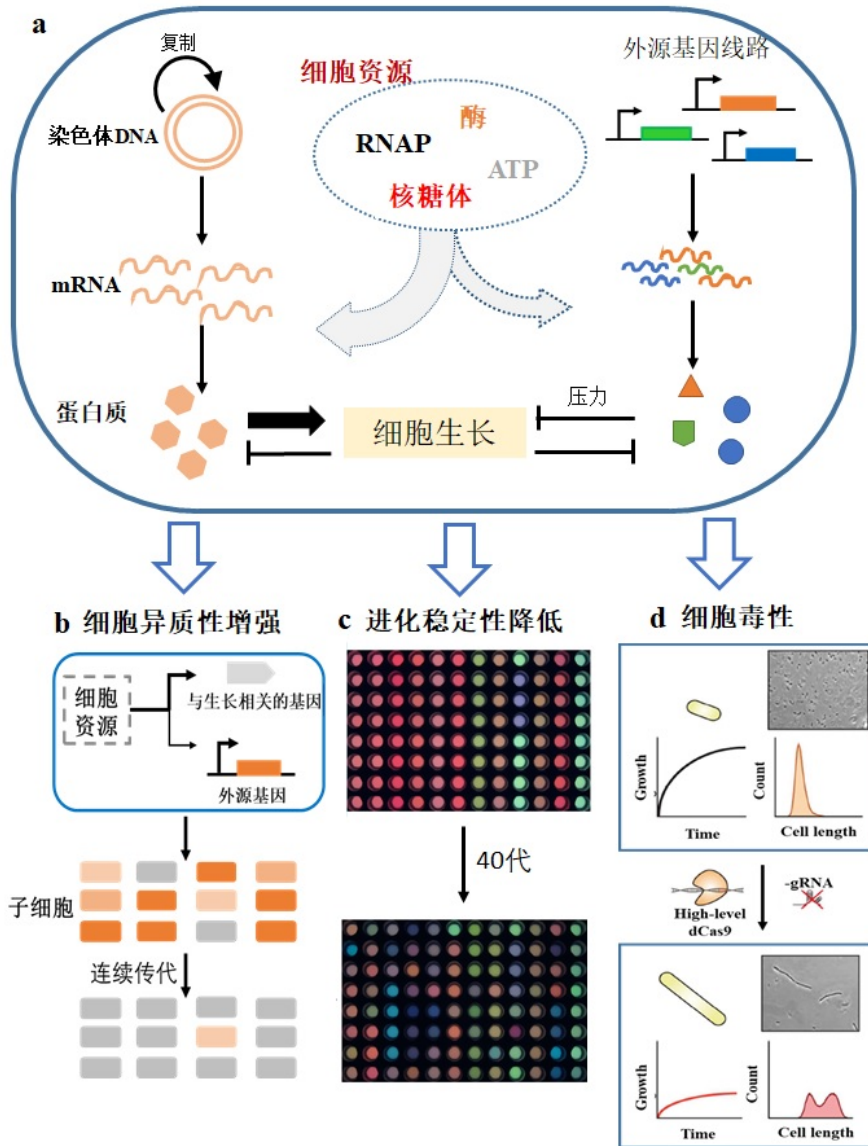


图 1: 底盘细胞和基因线路之间的相互作用。底盘细胞和外源基因线路通过竞争细胞内的资源实现耦合, 这种相互作用带来了一系列的效应: 细胞异质性增强, 进化稳定性降低 (图修改自^[24]) 以及细胞毒性 (图修改自^[28])

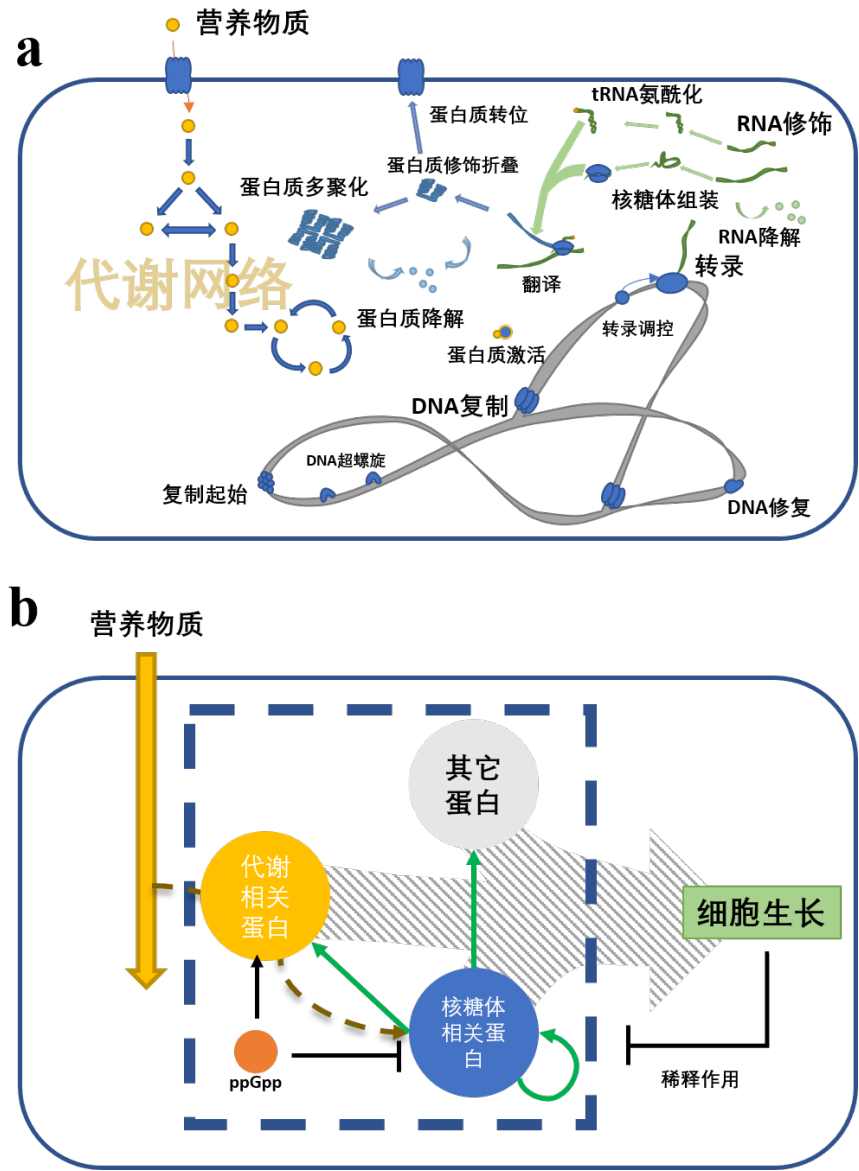


图 2: 全细胞模型与粗粒化模型涉及细胞生理过程示意图。a. 全细胞模型所涉及细胞生理过程, 其几乎涵盖了所有已知的生理反应, 通过数学手段对每个反应进行描述并整合成 28 个分区最终合并成完整的细胞模型 (图修改自^[32]) b. 粗粒化模型所涉及细胞生理过程, 目前普遍接受的概念是将蛋白质组分成三组: 核糖体相关蛋白, 代谢相关蛋白和其它蛋白, 细胞生长主要是由于蛋白质的增长导致, 主要耗能过程为转录及翻译, 以 ppGpp 为代表的调控机制平衡转录翻译所需的前体物质生产和转录翻译速率 (图修改自^[60])

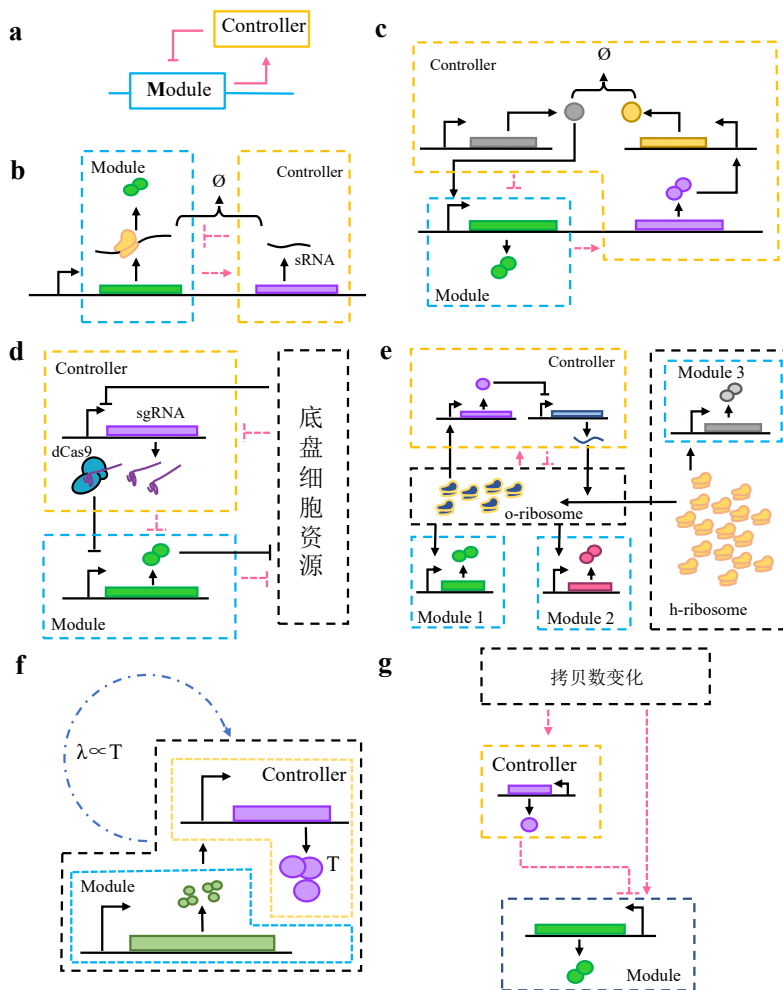


图 3: 模块化的基因线路设计。更高层次的模块化将基因线路中负责不同功能的局部线路称之为模块, (a) 中代表了一个负反馈环, 蓝色实线框内代表功能模块, 黄色实线框内代表控制器, 主要负责调节功能模块, 是局部水平负反馈 (b、c) 的基本模块; 基于转录后调控的局部负反馈设计 (b), 与输出模块一同转录的 sRNA 是报告基因 (绿色) 序列反义序列, 与 mRNA 结合, 引发降解, 使报告基因的表达水平与底盘细胞生理耦合; 基于对偶积分反馈的模块化设计 (c) 则靠控制器内的两个相互拮抗蛋白产物 (灰色和黄色) 来实现负反馈; 全局反馈控制系统 (d、e), 不同于局部反馈系统, 控制器直接感知底盘细胞的资源, (d) 中控制器内转录 sgRNA 的启动子能响应细胞的压力, 当模块中外源表达升高时, 转录水平上调, sgRNA 与胞内组成表达的 dCas9 蛋白结合, 靶向输出模块内的启动子形成负反馈调节; 正交翻译系统首先可以将模块 3 和其余输出模块解耦合, 控制器可以稳定正交核糖体 (o-ribosome) 的胞内浓度, 可以给模块 1 和模块 2 提供相对充足的翻译资源, 一定程度上减轻了模块 1 和模块 2 之间对资源的竞争; (f) 为生长偶联的正反馈环, 输出模块的产物可以激活控制器转录对于底盘而言为关键基因的 T, T 的表达水平与底盘细胞的生长速度正相关, 可以自然淘汰无法或表达过低的个体, 增强了进化稳定性; (g) 为利用 iFFL 设计的可以减少基因拷贝数引起的基因表达水平变化, 拷贝数的增加直接使控制器和输出模块的转录水平升高, 控制器的产物 TALEs 蛋白靶向模块内的启动子, 由于需要蛋白翻译、成熟等步骤, 造成拷贝数的激活作用与 TALEs 的抑制存在时间差。

参考文献

- [1] CARDINALE S, ARKIN A P. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems[J]. *Biotechnol. J.*, 2012, 7(7): 856-866.
- [2] BOADA Y, VIGNONI A, OYARZÚN D, et al. Host-circuit interactions explain unexpected behavior of a gene circuit.[J]. *IFAC-PapersOnLine*, 2018, 51(19): 86-89.
- [3] GRUNBERG T W, DEL VECCHIO D. Modular Analysis and Design of Biological Circuits[J]. *Curr. Opin. Biotechnol.*, 2020, 63: 41-47.
- [4] SCOTT M, GUNDERSON C W, MATEESCU E M, et al. Interdependence of Cell Growth and Gene Expression: Origins and Consequences[J]. *Science*, 2010, 330(6007): 1099-1102.
- [5] YOU C, OKANO H, HUI S, et al. Coordination of bacterial proteome with metabolism by cyclic AMP signalling[J]. *Nature*, 2013, 500(7462): 301-306.
- [6] HUI S, SILVERMAN J M, CHEN S S, et al. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria[J]. *Mol. Syst. Biol.*, 2015, 11(2): 784.
- [7] BORKOWSKI O, CERONI F, STAN G B B, et al. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology[J]. *Curr. Opin. Microbiol.*, 2016, 33: 123-130.
- [8] TOWBIN B D, KOREM Y, BREN A, et al. Optimality and sub-optimality in a bacterial growth law[J]. *Nat. Commun.*, 2017, 8: 1-8.
- [9] CARDINALE S, JOACHIMIAK M P, ARKIN A P. Effects of genetic variation on the *e. coli* host-circuit interface[J]. *Cell Rep.*, 2013, 4(2): 231-237.
- [10] VILANOVA C, TANNER K, DORADO-MORALES P, et al. Standards not that standard[J]. *J. Biol. Eng.*, 2015, 9(1): 17.
- [11] MOSER F, BROERS N J, HARTMANS S, et al. Genetic circuit performance under conditions relevant for industrial bioreactors[J]. *ACS Synth. Biol.*, 2012, 1(11): 555-564.
- [12] HINTSCHE M, KLUMPP S. Dilution and the theoretical description of growth-rate dependent gene expression[J]. *J. Biol. Eng.*, 2013, 7(1): 22.
- [13] TAN C, MARGUET P, YOU L. Emergent bistability by a growth-modulating positive feedback circuit[J]. *Nat. Chem. Biol.*, 2009, 5(11): 842-848.

- [14] CARBONELL-BALLESTERO M, GARCIA-RAMALLO E, MONTAÑEZ R, et al. Dealing with the genetic load in bacterial synthetic biology circuits: Convergences with the Ohm's law[J]. *Nucleic Acids Res.*, 2016, 44(1): 496-507.
- [15] VIND J, SØRENSEN M A, RASMUSSEN M D, et al. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: Expression from reporter genes does not always reflect functional mRNA levels[J]. *J. Mol. Biol.*, 1993, 231(3): 678-688.
- [16] LIU Q, SCHUMACHER J, WAN X, et al. Orthogonality and Burdens of Heterologous and Gate Gene Circuits in *E. coli*[J]. *ACS Synth. Biol.*, 2018, 7(2): 553-564.
- [17] SHACHRAI I, ZASLAVER A, ALON U, et al. Cost of Unneeded Proteins in *E. coli* Is Reduced after Several Generations in Exponential Growth[J]. *Mol. Cell*, 2010, 38(5): 758-767.
- [18] ZHU M, DAI X. Growth suppression by altered (p)ppGpp levels results from non-optimal resource allocation in *Escherichia coli*[J]. *Nucleic Acids Res.*, 2019, 47(9): 4684-4693.
- [19] COOKSON N A, MATHER W H, DANINO T, et al. Queueing up for enzymatic processing: Correlated signaling through coupled degradation[J]. *Mol. Syst. Biol.*, 2011, 7(1).
- [20] GYORGY A, JIMÉNEZ J I, YAZBEK J, et al. Isocost Lines Describe the Cellular Economy of Genetic Circuits[J]. *Biophys. J.*, 2015, 109(3): 639-646.
- [21] QIAN Y, HUANG H H, JIMÉNEZ J I, et al. Resource Competition Shapes the Response of Genetic Circuits[J]. *ACS Synth. Biol.*, 2017, 6(7): 1263-1272.
- [22] THOMAS P, TERRADOT G, DANOS V, et al. Sources, propagation and consequences of stochasticity in cellular growth[J]. *Nat. Commun.*, 2018, 9(1): 1-11.
- [23] KIM J, DARLINGTON A, SALVADOR M, et al. Trade-offs between gene expression, growth and phenotypic diversity in microbial populations[J]. *Curr. Opin. Biotechnol.*, 2020, 62: 29-37.
- [24] SLEIGHT S C, SAURO H M. Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits[J]. *ACS Synth. Biol.*, 2013, 2(9): 519-528.
- [25] KIMELMAN A, LEVY A, SBERRO H, et al. A vast collection of microbial genes that are toxic to bacteria[J]. *Genome Res.*, 2012, 22(4): 802-809.
- [26] LAMBERTE L E, BANIULYTE G, SINGH S S, et al. Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase[J]. *Nat. Microbiol.*, 2017, 2(January): 1-9.

-
- [27] STANTON B C, NIELSEN A A K, TAMSIR A, et al. Genomic mining of prokaryotic repressors for orthogonal logic gates[J]. *Nat. Chem. Biol.*, 2014, 10(2): 99-105.
- [28] HASNAIN A, BECKER D, SIBA A, et al. A data-driven method for quantifying the impact of a genetic circuit on its host[J]. *BioCAS 2019 - Biomed. Circuits Syst. Conf. Proc.*, 2019: 1-4.
- [29] CHO S, CHOE D, LEE E, et al. High-Level dCas9 Expression Induces Abnormal Cell Morphology in *Escherichia coli*[J]. *ACS Synth. Biol.*, 2018, 7(4): 1085-1094.
- [30] CUI L, VIGOUROUX A, ROUSSET F, et al. A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9[J]. *Nat. Commun.*, 2018, 9(1): 1-10.
- [31] CARRERA J, COVERT M W. Why Build Whole-Cell Models?[J]. *Trends Cell Biol.*, 2015, 25(12): 719-722.
- [32] KARR J R, SANGHVI J C, MACKLIN D N, et al. A Whole-Cell Computational Model Predicts Phenotype from Genotype[J]. *Cell*, 2012, 150(2): 389-401.
- [33] WEIßE A Y, OYARZÚN D A, DANOS V, et al. Mechanistic links between cellular trade-offs, gene expression, and growth[J]. *Proc. Natl. Acad. Sci. U. S. A.*, 2015, 112(9): E1038-E1047.
- [34] LIAO C, BLANCHARD A E, LU T. An integrative circuit – host modelling framework for predicting synthetic gene network behaviours[J]. *Nat. Microbiol.*, 2017, 2(12): 1658-1666.
- [35] MARR A G. Growth rate of *Escherichia coli*[J]. *Microbiol. Rev.*, 1991, 55(2): 316-333.
- [36] BENTLEY W E, MIRJALILI N, ANDERSEN D C, et al. Plasmid-encoded protein: The principal factor in the “metabolic burden” associated with recombinant bacteria[J]. *Biotechnol. Bioeng.*, 1990, 35(7): 668-681.
- [37] GOROCHOWSKI T E, AVCILAR-KUCUKGOZE I, BOVENBERG R A, et al. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes[J]. *ACS Synth. Biol.*, 2016, 5(7): 710-720.
- [38] NIKOLADOS E M, WEIßE A Y, CERONI F, et al. Growth Defects and Loss-of-Function in Synthetic Gene Circuits[J]. *ACS Synth. Biol.*, 2019, 8(6): 1231-1240.
- [39] KAUFFMAN K J, PRAKASH P, EDWARDS J S. Advances in flux balance analysis[J]. *Curr. Opin. Biotechnol.*, 2003, 14(5): 491-496.
- [40] BAPTIE A C, STUMPF M P. How to deal with parameters for whole-cell modelling[J]. *J. R. Soc. Interface*, 2017, 14(133): 20170237.

- [41] DAI X, ZHU M, WARREN M, et al. Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth[J]. *Nat. Microbiol.*, 2016, 2: nmicrobiol2016231.
- [42] LIU C C, JEWETT M C, CHIN J W, et al. Toward an orthogonal central dogma[J]. *Nat. Chem. Biol.*, 2018, 14(2): 103-106.
- [43] RAVIKUMAR A, ARRIETA A, LIU C C. An orthogonal DNA replication system in yeast[J]. *Nat. Chem. Biol.*, 2014, 10(3): 175-177.
- [44] HAN T, CHEN Q, LIU H. Engineered Photoactivatable Genetic Switches Based on the Bacterium Phage T7 RNA Polymerase[J]. *ACS Synth. Biol.*, 2017, 6(2): 357-366.
- [45] BAUMSCHLAGER A, AOKI S K, KHAMMASH M. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control[J]. *ACS Synth. Biol.*, 2017, 6(11): 2157-2167.
- [46] MCCUTCHEON S R, CHIU K L, LEWIS D D, et al. CRISPR-Cas Expands Dynamic Range of Gene Expression From T7RNAP Promoters[J]. *Biotechnol. J.*, 2018, 13(5).
- [47] RHODIUS V A, SEGALL-SHAPIRO T H, SHARON B D, et al. Design of orthogonal genetic switches based on a crosstalk map of σ s, anti- σ s, and promoters[J]. *Mol. Syst. Biol.*, 2013, 9(1): 702.
- [48] FREDENS J, WANG K, de la TORRE D, et al. Total synthesis of *Escherichia coli* with a recoded genome[J]. *Nature*, 2019.
- [49] RACKHAM O, CHIN J W. A network of orthogonal ribosome • mrna pairs[J]. *Nat. Chem. Biol.*, 2005, 1(3): 159-166.
- [50] AN W, CHIN J W. Synthesis of orthogonal transcriptiontranslation networks[J]. *Proc. Natl. Acad. Sci. U. S. A.*, 2009, 106(21): 8477-8482.
- [51] MEYER A J, SEGALL-SHAPIRO T H, GLASSEY E, et al. *Escherichia coli* “Marionette” strains with 12 highly optimized small-molecule sensors[J]. *Nat. Chem. Biol.*, 2019, 15(2): 196-204.
- [52] DARLINGTON A P S, KIM J, JIMÉNEZ J I, et al. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes[J]. *Nat. Commun.*, 2018, 9(1): 695.
- [53] ENDY D. Foundations for engineering biology[J]. *Nature*, 2005, 438(7067): 449-453.
- [54] HUANG H H, QIAN Y, DEL VECCHIO D. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand[J]. *Nat. Commun.*, 2018, 9(1): 5415.
- [55] AOKI S K, LILLACCI G, GUPTA A, et al. A universal biomolecular integral feedback controller for robust perfect adaptation[J]. *Nature*, 2019, 570(7762): 533-537.

-
- [56] CERONI F, BOO A, FURINI S, et al. Burden-driven feedback control of gene expression[J]. Nat. Methods, 2018, 15(5): 387-393.
- [57] RUGBJERG P, SARUP-LYTZEN K, NAGY M, et al. Synthetic addiction extends the productive life time of engineered Escherichia coli populations[J]. Proc. Natl. Acad. Sci. U. S. A., 2018, 115(10): 2347-2352.
- [58] XIAO Y, BOWEN C H, LIU D, et al. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis[J]. Nat. Chem. Biol., 2016, 12(5): 339-344.
- [59] SEGALL-SHAPIRO T H, SONTAG E D, VOIGT C A. Engineered promoters enable constant gene expression at any copy number in bacteria[J]. Nat. Biotechnol., 2018, 36(4): 352-358.
- [60] LIAO C, BLANCHARD A E, LU T. An integrative circuit-host modelling framework for predicting synthetic gene network behaviours /631/553/552 /631/553/2695 article[J]. Nat. Microbiol., 2017, 2(12): 1658-1666.

Title

First Author

affiliation 1

another affiliation

xxxxx@xxxx.xxx

Abstract

With the complexity of genetic circuits, the behavior of genetic circuits becomes unpredictable. Under research, an unavoidable coupling between the circuits and host cell was discovered, which changes the physiology state of host and influence the function of circuit. Traditionally applying engineering discipline to genetic circuit design has become a bottleneck. In this review, we introduce the coupling mechanism between the prokaryotic host cell and genetic circuits, together with some related quantitative works. The attempts to predict or reduce the coupling effects by modeling, orthogonalization and modular design are also summarized.