迈向补料分批代谢网络控制遗传学的建模、优化与预测控制框架

Sebastián Espinel-Ríos[1] | Bruno Morabito[2] | JohannesPohlodek[3] | KatjaBettenbrock[1] | Steffen Klamt[1] | Rolf Findeisen[3]

[1]生物网络分析与重设计，马克斯·普朗克复杂技术系统动力学研究所，德国马格德堡[2]横河Insilico生物技术有限公司，德国斯图加特[3]控制与信息物理系统实验室，达姆施塔特工业大学，德国达姆施塔特

通信作者
电子邮件：rolf.findeisen@iat.tu-darmstadt.de

arXiv:2302.02177v1 [eess.SY] 4 Feb 2023

摘要

生物技术为有价值产品的可持续制造提供了诸多机遇。优化生物过程的工具箱包括胞外过程要素，例如生物反应器设计和运行方式、培养基配方、培养条件、进料速率等。然而，这些要素往往不足以实现最佳过程性能或精确控制产品组成。可以采用代谢工程和基因工程方法，在细胞内层面进行优化。但这些方法往往具有静态特性，在应用于动态过程或发生扰动时便会失效。此外，许多生物过程通过经验进行优化，实施时几乎没有反馈控制来抵消扰动。网络控制遗传学这一概念带来了优化生物过程的新可能：它可以通过外部输入（例如光遗传学中的光强）在线调节代谢相关蛋白的基因表达。本文将网络控制遗传学与基于模型的优化和预测控制相结合，以优化动态生物过程。为此，我们提出采用动态约束模型，将代谢反应、资源分配和诱导型基因表达的动态特性纳入其中。我们构建基于模型的最优控制问题，以寻找最优过程输入。此外，我们提出采用模型预测控制，通过在线反馈应对不确定性。我们聚焦于补料分批过程，其中底物进料速率是一个额外的优化变量。作为仿真示例，我们展示了对 ATP 酶复合体进行光遗传控制，以动态调节强制 ATP 浪费，从而调整产物得率和生产率。

关键词：
代谢网络控制遗传学、动态代谢控制、基于约束的建模、最优控制、模型预测控制、状态估计、光遗传学。

1 引言

近年来，对可持续生物技术产品的需求显著增长（Hughes 和 Jones，2020年；Wohlgemuth 等，2021年）。尽管一些生物过程已取得商业成功（Jullesson 等，2015年；Sanford 等，2016年），但许多生物过程仍在早期阶段遭到淘汰，因为它们的竞争力不及传统技术。由此自然引出一个问题：如何优化生物过程的效率?

在细胞外或宏观层面，生物过程优化的工具箱包括选择生物反应器操作模式、进行生物反应器设计、优化培养条件（pH、温度等）、配制培养基、确定最优补料曲线和初始浓度等（例如，参见 Azimi 等人（2019年）；Behera 等人（2019年）；Vandermies 和 Fickers（2019年））。这些优化策略能够影响整体细胞代谢。但仅凭这些策略本身往往难以针对特定代谢要素（例如关键代谢通量）进行调控，同时又不影响细胞的其他功能。

基于动态模型的优化和预测控制策略可用于挖掘生物过程的动态潜力。动态优化可用于寻找最优动态运行条件，例如 del Rio-Chanona 等人（ 2019 年）、 Jabarivelisdeh 等人（ 2018 年）、 Jabarivelisdeh 和 Waldherr （ 2016 年）、Nimmegeers 等人（2018年）以及 Ryu 等人（2019年）的研究。反馈控制方案，尤其是 Jabarivelisdeh 等人（2020年）、Jabarivelisdeh 和 Waldherr（2018年）以及 Morabito 等人（2019年、2022年、2021年）所采用的预测控制方法，可用于抵消未知扰动，例如补料条件变化或未建模的动态，同时最大限度地提高生产效率并确保过程性能稳定。

在细胞内 或微观层面，生物过程优化工具箱包括用于重构代谢通路的代谢工程和基因工程方法。经典的静态 代谢工程旨在提高细胞的产物得率，但往往以降低生物量产率为代价，因为底物通量会从生物量生成反应转向目标产物途径。这不可避免地会降低批次式生物反应器中的体积生产率（Lalwani 等，2018年；Venayak 等，2015年）。此外，基于静态代谢控制原则（这些原则通常是在稳态假设下推导出来的）设计动态过程，可能导致代谢失衡（Cui 等，2021年）。

通过外部输入诱导表达代谢相关蛋白，已成为微观层面生物过程优化的一种颇具前景的动态 自由度（Hartline 等，2021年；Lalwani 等，2018年；Shen 等，2019年）。光遗传学的应用日益普及，即利用光调控基因表达（Hoffman 等，2022年）。借助光遗传学，可通过调节酶表达来开启或关闭代谢通路中的通量（Lalwani、Ip 等，2021年；Tandar 等，2019年；Zhao 等，2021年）。还可以通过调节（抗）毒素蛋白的表达，直接影响细胞生长（Lalwani、Kawabe 等，2021年）。在后一类光遗传学案例中，最优 光输入值是通过析因实验及类似的启发式方法确定的，进而形成了例如两阶段或三阶段发酵。此外，输入以开环方式施加，即没有在线反馈或纠正措施。考虑到基因表达通常具有随机性（De Vrieze 等，2020年），且可能存在过程扰动和批次间变异，以这种方式运行的生物过程可能重现性较差、产品质量中等或较差，且失败风险较高。

受这些挑战的启发，一些作者提出了网络控制遗传学方案，利用计算机辅助反馈控制来补偿不确定性。在这种情况下，纠正措施在细胞外部计算，例如由计算机辅助控制器计算（Hsiao 等，2018 年；Khammash，2022 年）。据我们所知，网络控制遗传学在生物技术中的应用迄今仅限于例如通过光遗传学控制荧光蛋白和生长调节蛋白（如参与必需氨基酸合成的酶或赋予抗生素耐药性的蛋白质）的表达（Gutiérrez Mena 等，2022 年；Milias-Argeitis 等，2016 年）。

人们认为，这一方向的下一步具有巨大潜力，即实现代谢网络生物系统，也就是强调动态代谢工程应用（Carrasco-López 等，2020 年）。因此，我们希望将网络控制遗传学的应用范围扩展到需要动态操纵代谢通量的情境（例如，最大化体积生产率、达到目标产物得率、使产物比例达到给定值等），同时能够补偿扰动和工艺变化。我们旨在采用基于模型的优化和预测控制方法，充分发挥代谢网络生物系统的潜力，并同时考虑网络遗传学输入和补料速率等传统过程输入。

因此，建立一个足够好的模型，将诱导型基因表达与代谢通量分布的变化及潜在的资源负担联系起来，对于推进我们的探索至关重要。遗憾的是，迄今为止，在网络控制遗传学领域使用的模型都非常简单，且通常基于现象关系（例如，参见 GutiérrezMena 等（2022）；

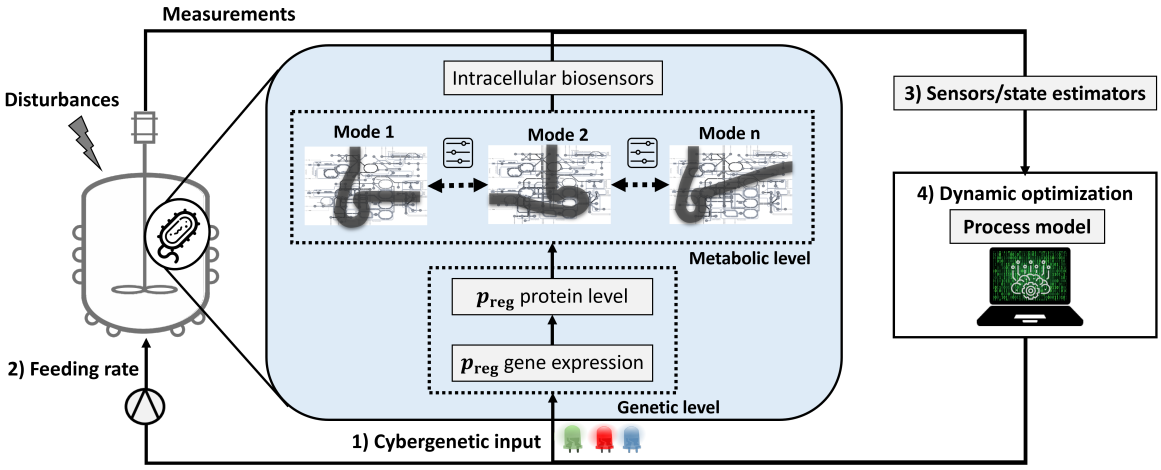


图 1 补料分批模式下代谢网络控制遗传学概览。酶等关键代谢相关蛋白$p_{\mathrm{reg}} \in \mathbb{R}^{n_{p_{\mathrm{reg}}}}$受可诱导基因表达系统调控，从而利用外部输入（如光强）随时间实现不同的代谢模式。基于模型的优化可为被控对象确定最优输入。过程结果可通过（生物）传感器和状态估计器进行监测。重复求解优化问题即可实现反馈控制。

Lovelett 等（2021）；Milias-Argeitis 等（2016））。我们认为，这些模型无法捕捉基于模型的控制代谢通量 控制遗传系统所需的所有重要现象。

因此，作为本研究的核心贡献，我们提出了一种代谢网络控制遗传学建模框架，并将其与基于模型的优化和预测控制相结合，以动态调节细胞内代谢通量，实现生物过程优化。为不失一般性，我们重点关注补料分批过程，因为与纯批次操作相比，这类过程具有优势。补料分批过程通过浓缩进料向生物反应器供应新鲜培养基，从而延长生产阶段。这为系统提供了额外的动态输入（进料速率），并可提高生产率，使产品流的浓度更高。此外，这也是处理底物抑制过程的一种高效方式（Doran 2013；S. Liu 2020）。所提出的补料分批代谢网络控制遗传学平台由四个主要部分组成（见图 1）：1）能够动态诱导基因表达的信息遗传输入；2）可调节的底物进料流；3）用于监测和估计过程状态的在线（生物）传感器和状态估计器；4）以闭环、全自动方式运行的基于模型的优化。

本文其余部分安排如下。第2节概述一种动态约束型控制遗传建模方法，该方法整合了代谢、资源分配和诱导型基因表达。所建立的模型用于支持代谢网络控制遗传学的基于模型的优化、反馈控制和状态估计（第3节和第4节）。第5节评估我们的框架，研究在大肠杆菌的厌氧乳酸发酵过程中对 ATP 酶 Fı-亚基[1] 进行光遗传调控，以提高产率和生产率。我们考虑光穿透不均匀的补料分批培养模式，并采用开环和反馈控制。请注意，在此前的一项研究中，我们仅探讨了均匀光穿透条件下经典批次过程中的开环优化（Espinel-Ríos, Morabito, Pohlodek, 等，2022）。

2 补料分批代谢网络控制遗传学的动态优化与控制建模

我们采用一种扩展的基于约束的建模方法，以捕捉代谢、资源分配和诱导型基因表达的综合动态。基于约束的模型通常是欠定的（Gottstein 等，2016；Klamt 和 von Kamp，2022）。因此，这类模型通常被表述为带有符合生物学规律的目标函数、并

[1]从现在起，我们将把 ATP 酶/基因的 Fı-亚基称为“ATP 酶/基因”。

受约束条件限制的优化问题。与动力学模型相比，基于约束的模型通常所需的参数数量更少（例如通量边界），结构也可能更简单（Saa 和 Nielsen（2017）；Yasemi 和 Jolicoeur（2021））。

不失一般性，我们假设细胞由代谢酶、核糖体和配额元素[2]组成。这些生物量组分包含在摩尔向量 $\boldsymbol{p} \in \mathbb{R}^{n_p}$ 中，因此生物量 $\boldsymbol{B} \in \mathbb{R}$ 的量可计算为

$$\boldsymbol{B}(t) = \upsilon_L(t)(\boldsymbol{b}^T \boldsymbol{p}(t)), \tag{1}$$

其中，t 为时间，$\upsilon_L \in \mathbb{R}$ 为生物反应器体积，$\boldsymbol{b} \in \mathbb{R}^{n_p}$ 为 $\boldsymbol{p}$相应分子量组成的向量。因此，$\boldsymbol{b}^T \boldsymbol{p}(t)$ 是一个标量，表示 g/L中的生物量浓度。请注意，向量和矩阵使用粗体表示，标量变量和参数则使用非粗体表示。

我们在 $\boldsymbol{p}$ 中汇集受调控蛋白 $\boldsymbol{p}_{\mathbf{reg}} \in \mathbb{R}^{n_{p_{\mathrm{reg}}}}$ 的浓度以及其余未受调控生物量组分 $\boldsymbol{p}_{\mathbf{unr}} \in \mathbb{R}^{n_{p_{\mathrm{unr}}}}$的浓度，使得 $\boldsymbol{p}(t) = \left[\boldsymbol{p}_{\mathbf{reg}}(t)^{\mathsf{T}}, \boldsymbol{p}_{\mathbf{unr}}(t)^{\mathsf{T}}\right]^{\mathsf{T}}$。在本文中，“受调控”是指蛋白质表达受到控制遗传学调控，并通过适当的遗传系统（例如光诱导基因表达系统）进行外部调节（Lindner 和 Diepold（2021）；Z. Liu 等（2018））。请注意，受调控蛋白可以包括直接参与代谢通路的酶，这是本文的重点；原则上也可以包括用于调节生长的（抗）毒素蛋白或赋予抗生素耐药性的蛋白质。“未受调控”的生物量组分通常包括例如不受控制遗传学调控的代谢酶、核糖体和配额元素。下文中，若上下文已明确，我们将省略变量对时间的依赖关系。

关于可调控基因表达系统，我们区分控制器操控的输入 $\boldsymbol{u}_s \in \mathbb{R}^{n_u}$（例如光强）与生物反应器内细胞感知到的值 $\boldsymbol{u}_c \in \mathbb{R}^{n_u}$。这种区分是必要的，因为根据输入和生物反应器的特性，$\boldsymbol{u}_s = \boldsymbol{u}_c$ 可能并不成立。这在大型装置中尤为重要，因为其中的条件往往不够均一；此外，细胞接收到的输入值也可能取决于细胞密度，例如受浊度影响。

在第5节中，我们将推导 $\boldsymbol{u}_c$，以便在光遗传学背景下考虑光穿透。为保持一般性，我们认为细胞感知到的输入由函数 $\boldsymbol{f}_u : \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_{\theta_u}} \to \mathbb{R}^{n_u}$ 给出，该函数将输入源处的输入 $\boldsymbol{u}_s$ 映射为平均输入 $\bar{\boldsymbol{u}}_c$。因此，

$$\bar{\boldsymbol{u}}_c = \boldsymbol{f}_u(\boldsymbol{u}_s, \boldsymbol{x}, \boldsymbol{\theta}_u), \tag{2}$$

其中，$\boldsymbol{x} \in \mathbb{R}^{n_x}$ 原则上可包括所有模型状态，$\boldsymbol{\theta}_u \in \mathbb{R}^{n_{\theta_u}}$ 包含 $\boldsymbol{f}_u(\cdot)$ 可能涉及的参数，而 $\bar{\boldsymbol{u}}_c$ 是在完全混合条件下生物反应器中 $\boldsymbol{u}_c$ 的平均值。引入 $\bar{\boldsymbol{u}}_c$ 可简化模型，使其仅描述随时间而非随空间的变化，同时仍能考虑平均输入梯度。

我们将受调控蛋白量的变化描述为

$$\frac{\mathrm{d}\left(\upsilon_L \boldsymbol{p}_{\mathbf{reg}}\right)}{\mathrm{d}t} = \boldsymbol{F}_{\mathbf{reg}}(\boldsymbol{B}, \bar{\boldsymbol{u}}_c) - \boldsymbol{D}_{\mathbf{reg}}(\boldsymbol{p}_{\mathbf{reg}}), \tag{3}$$

其中，$\boldsymbol{F}_{\mathbf{reg}} : \mathbb{R} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_{p_{\mathrm{reg}}}}$ 和 $\boldsymbol{D}_{\mathbf{reg}} : \mathbb{R}^{n_{p_{\mathrm{reg}}}} \to \mathbb{R}^{n_{p_{\mathrm{reg}}}}$ 分别表示蛋白质生成和降解的函数。

细胞具有转录因子，这些因子能够以特定信号所决定的速率在活性状态和非活性状态之间切换。当活性转录因子与受调控基因的启动子区域结合时，它们可以激活或抑制转录过程。核糖体将转录过程产生的信使核糖核酸翻译成蛋白质。在大肠杆菌等细菌中，转录和翻译高度耦合，这意味着翻译与活跃转录同时发生（Scull 等，2021；Yang 等，2019）。因此，我们提出将这两个过程合并为集总剂量-反应函数 $\boldsymbol{\eta} : \mathbb{R}^{n_u} \to \mathbb{R}^{n_{p_{\mathrm{reg}}}}$，于是

$$\boldsymbol{F}_{\mathbf{reg}}(\boldsymbol{B}, \bar{\boldsymbol{u}}_c) = \boldsymbol{B}\boldsymbol{\eta}(\bar{\boldsymbol{u}}_c). \tag{4}$$

对于受调控蛋白的降解，我们同时考虑生长导致的细胞稀释效应和蛋白质周转的内在作用。后者由 $\boldsymbol{D}_{\mathbf{reg}}$ 描述。另一方面，受调控蛋白的稀释已在我们的建模框架中隐式考虑。也就是说，细胞内受调控蛋白所占的比例会因其余生物量组分的生成而被稀释 。

我们通过动态酶成本通量平衡分析（deFBA），将受调控蛋白的动态与整体代谢和细胞资源分配联系起来。deFBA 是一种考虑资源分配约束的基于约束的代谢框架（Jabarivelisdeh 等，2020；Jabarivelisdeh 与 Waldherr，2018；Waldherr 等，2015）。

[2]配额元素包括 DNA、脂质、碳水化合物、非催化蛋白质以及其他小分子。

胞外代谢物的量建模为

$$\frac{\mathrm{d}\left(v_L \boldsymbol{z}\right)}{\mathrm{d}t} = F_{\mathrm{in}}\boldsymbol{z_{\mathrm{in}}} + v_L(\boldsymbol{S_z V}) - \boldsymbol{D_z}(\boldsymbol{z}), \tag{5}$$

其中，$\boldsymbol{z} \in \mathbb{R}^{n_z}$ 是胞外代谢物的摩尔向量，$\boldsymbol{S_z} \in \mathbb{R}^{n_z \times n_V}$ 是 $\boldsymbol{z}$ 的化学计量矩阵，$\boldsymbol{V} \in \mathbb{R}^{n_V}$包括转运反应、代谢反应和生物量生成反应的通量，单位为摩尔/时间。$\boldsymbol{D_z} : \mathbb{R}^{n_z} \to \mathbb{R}^{n_z}$表示 $\boldsymbol{z}$的降解，$F_{\mathrm{in}} \in \mathbb{R}$ 是补料速率，$\boldsymbol{z_{\mathrm{in}}} \in \mathbb{R}^{n_z}$ 包含 $\boldsymbol{z}$的进料浓度。在模型中，我们假设进料仅包含底物。

细胞需要投入资源来制造自身组分。因此，改变某种受调控蛋白的表达预计会影响其他生物量组分的生成速率以及由此产生的代谢通量分布，因为细胞内的资源有限且由各组分共享。请注意，将生物量生成反应纳入网络，是捕捉资源成本的一种方式，因为我们明确考虑了合成所有生物量组分所需的化学计量前体和能量当量。有鉴于此，未受调控的生物量组分的量满足

$$\frac{\mathrm{d}\left(v_L \boldsymbol{p_{\mathrm{unr}}}\right)}{\mathrm{d}t} = v_L(\boldsymbol{S}_{\boldsymbol{p_{\mathrm{unr}}}} \boldsymbol{V}) - \boldsymbol{D_{\mathrm{unr}}}(\boldsymbol{p_{\mathrm{unr}}}), \tag{6}$$

其中，$\boldsymbol{S}_{\boldsymbol{p_{\mathrm{unr}}}} \in \mathbb{R}^{n_{p_{\mathrm{unr}}} \times n_V}$ 是 $\boldsymbol{p_{\mathrm{unr}}}$ 的化学计量矩阵，$\boldsymbol{D_{\mathrm{unr}}} : \mathbb{R}^{n_{p_{\mathrm{unr}}}} \to \mathbb{R}^{n_{p_{\mathrm{unr}}}}$ 描述了 $\boldsymbol{p_{\mathrm{unr}}}$的降解。对于受调控蛋白，我们添加以下约束

$$\boldsymbol{V}_{\boldsymbol{p_{\mathrm{reg}}}} - \frac{\mathrm{d}\boldsymbol{p_{\mathrm{reg}}}}{\mathrm{d}t} = \boldsymbol{0}, \tag{7}$$

其中，$\boldsymbol{V}_{\boldsymbol{p_{\mathrm{reg}}}} \in \mathbb{R}^{n_{p_{\mathrm{reg}}}}$ 包含相应的受调控蛋白生成反应通量。

我们考虑胞内代谢物含量的准稳态动力学

$$\frac{\mathrm{d}\left(v_L \boldsymbol{m}\right)}{\mathrm{d}t} = v_L(\boldsymbol{S_m V}) = \boldsymbol{0}, \tag{8}$$

其中，$\boldsymbol{m} \in \mathbb{R}^{n_m}$ 是　胞内代谢物的摩尔向量，$\boldsymbol{S_m} \in \mathbb{R}^{n_m \times n_V}$ 是 $\boldsymbol{m}$的化学计量矩阵　　.

$\boldsymbol{p_{\mathrm{unr}}}$ 中由酶催化的反应的代谢通量受相应的催化酶浓度和催化常数（$\boldsymbol{k_{\mathrm{cat}}} \in \mathbb{R}^{n_{\mathrm{cat}}}$）约束

$$\sum_{j \in \mathrm{cat}_{\mathrm{unr}}} \left| \frac{V_j}{k_{\mathrm{unr},j}} \right| \leq p_{\mathrm{unr}_i}, \quad \forall i \in [1, n_{p_{\mathrm{unr}}}], \tag{9}$$

其中，$\mathrm{cat}_{\mathrm{unr}}(i)$ 表示由酶催化的所有反应，p_{unr_i}，而 $|\cdot|$ 指绝对值算子。由酶催化的反应的代谢通量在$\boldsymbol{p_{\mathrm{reg}}}$ 中受到以下约束:

$$\sum_{j \in \mathrm{cat}_{\mathrm{reg}}(i)} \left| \frac{V_j}{k_{\mathrm{reg},j}} \right| = p_{\mathrm{reg}_i}, \quad \forall i \in [1, n_{p_{\mathrm{reg}}}], \tag{10}$$

其中，$\mathrm{cat}_{\mathrm{reg}}(i)$是由$p_{\mathrm{reg}_i}$催化的反应集合。

我们考虑酶饱和条件。因此，式（9）将酶浓度与催化常数的乘积作为代谢通量的上界。尽管式（9）是不等式约束，但在假设我们能够控制由p_{reg_i}催化的通量的情况下，我们在式（10）中使用等式约束。换言之，我们将这一自由度从细胞转移给外部控制器。

生物量干重的 $\varphi_Q \in [0, 1]$ 部分对应于一种集总配额化合物 $p_Q \in \mathbb{R}$

$$\varphi_Q(\boldsymbol{b}^T \boldsymbol{p}) \leq p_Q,\ p_Q \in \boldsymbol{p_{\mathrm{unr}}}. \tag{11}$$

由于底物进料速率 F_{in}，生物反应器液体体积会随时间发生变化

$$\frac{\mathrm{d}v_L}{\mathrm{d}t} = F_{\mathrm{in}}. \tag{12}$$

代谢通量受到生物学可行的上下界约束

$$\boldsymbol{V_{\mathrm{min}}} \leq \boldsymbol{V} \leq \boldsymbol{V_{\mathrm{max}}}. \tag{13}$$

同样，我们考虑动态状态的可行边界

$$\boldsymbol{p_{\mathrm{min}}} \leq \boldsymbol{p} \leq \boldsymbol{p_{\mathrm{max}}},\ \boldsymbol{z_{\mathrm{min}}} \leq \boldsymbol{z} \leq \boldsymbol{z_{\mathrm{max}}},\ v_{L_{\mathrm{min}}} \leq v_L \leq v_{L_{\mathrm{max}}}. \tag{14}$$

系统在初始过程时间 t_0 的条件为

$$\boldsymbol{p}(t_0) = \boldsymbol{p_0},\ \boldsymbol{z}(t_0) = \boldsymbol{z_0},\ v_L(t_0) = v_{L0}. \tag{15}$$

总而言之，我们将所得的补料分批代谢网络控制系统动态约束模型表示为以下一般动态优化问题[3]

$$\max_{\boldsymbol{V}(\cdot)} \quad \int_{t_0}^{t_0+\Delta t_{\text{bio}}} F_{\text{V}}(\cdot)\,\mathrm{d}t, \tag{16a}$$

$$\text{s.t.} \quad \text{Eqs. (1)} - \text{(15)}, \tag{16b}$$

其中， $F_{\text{V}}(\cdot)$ 是细胞优化的目标函数，通常假设其目标是使细胞生长最大化；$V(\cdot)$ 是所得代谢通量分布的函数。求解这一动态优化问题，即可模拟并预测细胞的行为，如第5节所示。我们将以此动态约束模型为基础，对过程进行优化和控制。

3 代谢网络控制遗传学的最优控制

基于推导出的动态约束模型，我们旨在寻找最优输入轨迹，以驱动细胞代谢，使由代价函数 $J(\cdot)$ 描述的期望性能指标最大化。将所有过程输入（操纵变量）汇总为 $\boldsymbol{u}_p$，即 $\boldsymbol{u_p} := [\boldsymbol{u_s}(\cdot)^\mathsf{T}, F_{\text{in}}(\cdot)]^\mathsf{T}$，并将所有模型参数汇总为向量 $\boldsymbol{\theta}$。请注意， $\boldsymbol{x}$ 包含所有动态状态，即 $\boldsymbol{x} := [\boldsymbol{p}_{\mathbf{reg}}^\mathsf{T}, \boldsymbol{z}^\mathsf{T}, \boldsymbol{p}_{\mathbf{unr}}^\mathsf{T}, v_L]^\mathsf{T}$。为找到被控对象的最优输入，我们构建一个最优控制问题

$$\max_{\boldsymbol{u_p}} \quad \boldsymbol{J}(\cdot), \tag{17a}$$

$$\text{s.t.} \quad \text{(16)}, \tag{17b}$$

$$\mathbf{0} \le \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u_p}, \boldsymbol{\theta}). \tag{17c}$$

求解（17）属于双层最优控制问题，因为（16）中的动态约束模型本身就涉及优化。式（17c）描述了额外的状态约束和输入约束。

$\boldsymbol{J}(\cdot)$ 可根据具体目标以多种方式定义。例如，可以最大化产量、维持期望的设定值，或跟踪参考轨迹等。式（17c）可包含物理、安全或经济相关的过程约束。如果过程以批处理模式运行，则可将 $\boldsymbol{u_s}(\cdot)$ 设为唯一的优化自由度。式（17）中的最优控制问题属于开环优化，因为计算最优输入轨迹时仅使用过程状态的初始条件，随后将该轨迹施加于被控对象，且不进行反馈。这样便无法对未知扰动或模型-对象失配作出响应。

4 网络控制遗传学的模型预测控制

在考虑补料分批过程时，我们采用收缩时域模型预测控制（MPC）来减轻过程不确定性（例如模型-对象失配和扰动）的影响（Findeisen &Allgöwer 2002；Rawlingsetal. 2020），即缓解开环控制面临的挑战。

4.1 收缩时域模型预测控制

在模型预测控制中，最优控制问题会在给定的采样时刻反复求解。在这些采样时刻，系统状态会被测量，或由观测器估计。由此引入了反馈，因为当前系统状态的信息会传递给控制器，从而可以采取校正控制动作。设 t_k 为进行测量的采样时刻。不失一般性，我们假设状态测量值在等间隔采样时刻可用，即 $t_k :=$

[3]该模型可通过设置 $\dot{v}_L = 0$ 来描述批次系统。对于连续过程，可以加入一项离开生物反应器的额外流量。

kh_s 其中 $k \in \mathbb{N}_0$，且 h_s 为固定的采样间隔。此外，我们假设控制器预测至终止时间 $t_f := Nh_s$，其中 $N \in \mathbb{N}$ 为预测时域内的步数。因此，预测时域会在每个采样时刻缩短。时刻 t_k 的时域递减型模型预测控制写作

$$\min_{\boldsymbol{u_p}} \quad J(\cdot), \tag{18a}$$

$$\text{s.t.} \quad \max_{\boldsymbol{V}(\cdot)} \int_{t_k}^{t_k+\Delta t_{\text{bio}}} F_{\text{V}}(\cdot)\,\mathrm{d}t, \tag{18b}$$

$$\text{s.t.} \quad \text{Eqs. (1)- (14)}, \tag{18c}$$

$$\boldsymbol{x}(t_k) = \tilde{\boldsymbol{x}}_{\boldsymbol{k}}, \tag{18d}$$

$$\text{Eq. (17c)}, \tag{18e}$$

其中，$t \in [t_k, t_f]$ 和 $\tilde{\boldsymbol{x}}_{\boldsymbol{k}}$ 表示 $\boldsymbol{x}$的测量值。

我们假设，可以根据所施加的补料速率轻松监测培养体积，并且有多种在线传感器可用于测量细胞外代谢物浓度（Fung Shek 和 Betenbaugh，2021年；Reardon，2021年；Reyes 等，2022年）。因此，利用现有技术在技术上可以监测 v_L 和 $\boldsymbol{z}$。然而，通常没有可用于测量完整细胞内生物量组成的商用传感器。为应对这一挑战，已有研究提出了一些状态估计器，用于重构生物量组成（Espinel-Ríos、Morabito、Bettenbrock 等，2022年；Jabarivelisdeh 等，2020年）。下一节将简要介绍全信息估计器的应用——这是一种基于优化的估计器，会考虑过程动态和过程约束，以及过去和当前的测量值。有关更多详情，请参阅 Espinel-Ríos、Morabito、Bettenbrock 等（2022年）。

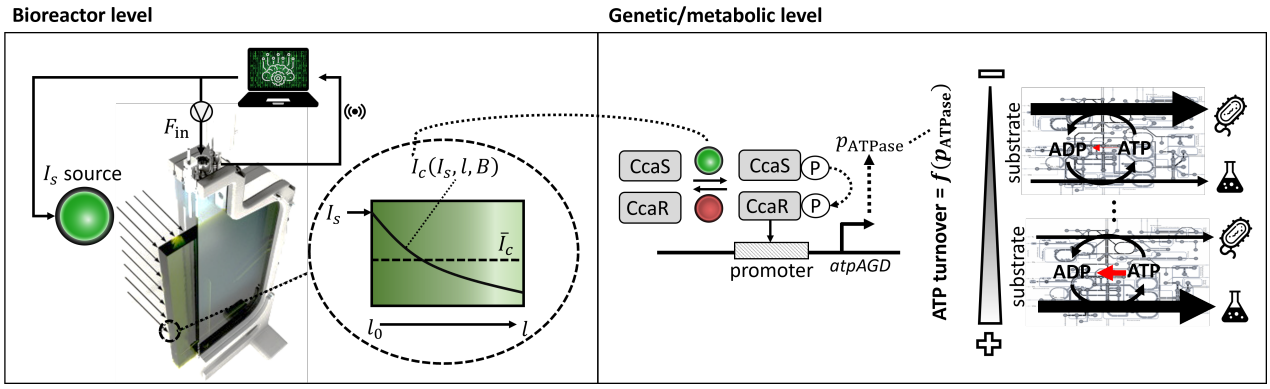


图 2 使用 Ccas/CcaR 光遗传系统调节 ATP 酶表达的示例概览。左侧展示了所考虑的平板式（光）生物反应器和代谢物联网遗传控制方案，其中包括培养物内部的平均光强。右侧展示了 ATP 酶不同表达水平对细胞代谢的影响。生物反应器图片改编自 Pfaffinger 等人（2016年）。

4.2 重建未测量的细胞组分

令 $(\cdot)_i$ 为在时间 t_i 计算得到的一般优化变量。我们将动态方程 (3)、(5)、(6) 和 (12) 汇总到向量函数 $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u_p}, \boldsymbol{\theta})$ 中。与模型预测控制一样，为了简化表述，我们假设估计器采用等间隔采样时刻，但也可以采用非等间隔采样时刻。在时间 t_k，可通过求解以下优化问题来构建全信息估计器

$$\min_{\boldsymbol{x_0}, \boldsymbol{\theta}, \boldsymbol{w}} \quad \left\| \begin{bmatrix} \boldsymbol{x_0} \\ \boldsymbol{\theta} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{x}}_{\boldsymbol{0}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix} \right\|_{\mathbf{P}}^2 + \sum_{i=0}^{k} \|\boldsymbol{y}(t_i) - \tilde{\boldsymbol{y}}_i\|_{\mathbf{R}}^2 + \tag{19a}$$

$$+ \|\boldsymbol{w}_i\|^2_{\mathbf{Q}}$$

$$\text{s.t.} \quad \max_{\boldsymbol{V}(\cdot)} \int_{t_i}^{t_i+\Delta t_{\text{bio}}} F_{\text{V}}(\cdot)\,\mathrm{d}t \tag{19b}$$

$$\text{s.t.} \quad \boldsymbol{x}(t_i + h_s) = \boldsymbol{x}(t_i)^{\mathsf{T}} + \int_{t_i}^{t_i+h_s} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}_p, \boldsymbol{\theta})\,\mathrm{d}t + \boldsymbol{w}_i, \tag{19c}$$

$$\mathbf{0} \le \boldsymbol{c}(\boldsymbol{x}(t_i), \boldsymbol{u}_p(t_i), \boldsymbol{\theta}), \tag{19d}$$

$$\boldsymbol{y}(t_i) = \boldsymbol{h}(\boldsymbol{x}(t_i), \boldsymbol{u}_p(t_i), \boldsymbol{\theta}), \tag{19e}$$

$$\mathbf{0} \le \boldsymbol{g}(\boldsymbol{x}(t_i), \boldsymbol{u}_p(t_i), \boldsymbol{\theta}), \tag{19f}$$

其中 $\|\boldsymbol{a}\|^2_{\mathbf{A}} := \boldsymbol{a}^{\mathsf{T}}\mathbf{A}\boldsymbol{a}$ 和 $i \in [0, ..., k]$。 $\mathbf{P}$、 $\mathbf{R}$ 和 $\mathbf{Q}$ 是维度适当的权重矩阵， $\boldsymbol{c} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u+1} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_c}$ 是模型约束， $\boldsymbol{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u+1} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_y}$ 是测量方程。优化变量包括初始条件 $\boldsymbol{x}_0$、参数 $\boldsymbol{\theta}$ 和状态噪声 $\boldsymbol{w} := [\boldsymbol{w}_0^{\mathsf{T}}, ..., \boldsymbol{w}_k^{\mathsf{T}}]^{\mathsf{T}}$。我们用 $(\cdot)^*$ 表示全信息估计问题的解，用 $(\hat{\cdot})$ 表示变量的 先验信息。借助 $\boldsymbol{x}_0^*$、 $\boldsymbol{w}^*$ 和 $\boldsymbol{\theta}^*$，我们重建 t_k 时刻的状态，这些状态可用于模型预测控制。全信息估计器会考虑所有测量值；而如果仅使用给定时间窗口内的测量值，则称为移动时域估计器（Elsheikh 等，2021年；Rawlings 等，2020年）。

5 示例：通过光遗传控制大肠杆菌的ATP酶进行厌氧乳酸发酵

我们研究以葡萄糖为底物的大肠杆菌*E. coli* 厌氧乳酸发酵，并对 ATP 酶复合物进行光遗传控制，参见图 2。我们只有一种受调控蛋白，因此，$\boldsymbol{p}_{\text{reg}} := p_{\text{ATPase}}$。后者负责催化 ATP 水解为 ADP 的反应。我们聚焦于大肠杆菌*E.coli* KBM10111 菌株，该菌株经基因工程改造，敲除了*adhE* （醛醇脱氢酶）、*ackA* （乙酸激酶）和*pta* （磷酸乙酰转移酶）（Hädicke 等，2015年）。在这些条件下，丙酮酸合成乳酸是平衡糖酵解过程中产生的氧化还原辅因子的必要途径。由于糖酵解产生三磷酸腺苷净增益，乳酸生产与三磷酸腺苷合成相关（见图 3）。在产物途径与三磷酸腺苷净生成相关的情况下，研究表明，强制 三磷酸腺苷周转或浪费 三磷酸腺苷，可通过抵消三磷酸腺苷损失来增加底物摄取量以及三磷酸腺苷生成途径中的代谢通量（例如，参见 Boecker 等（2021年）；Hädicke 等（2015年）；Zahoor 等（2020年））。因此，动态调节 ATP酶表达，进而调节 ATP酶通量，可用于调控生物过程中的产物得率和体积生产率（Espinel-Ríos, Bettenbrock, 等，2022年；Espinel-Ríos, Morabito, Pohlodek, 等，2022年）。

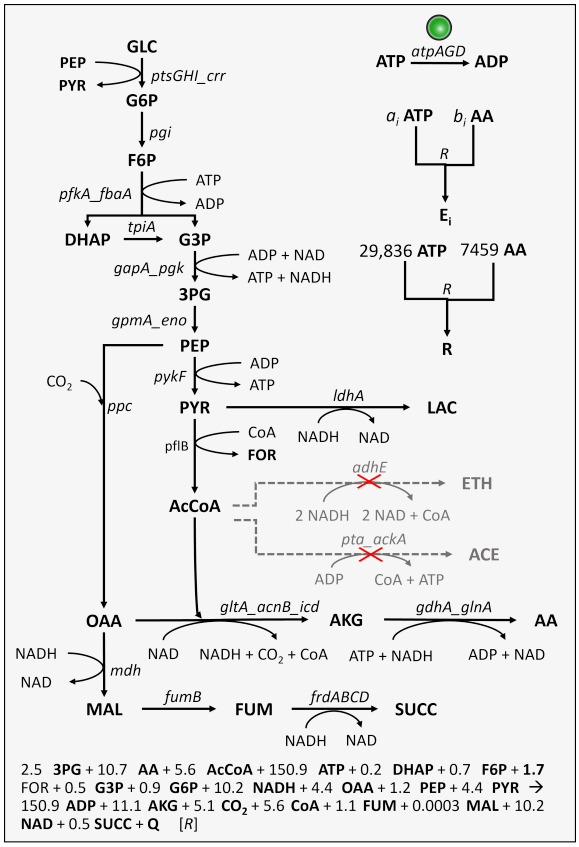
我们考虑使用 Ccas/CcaR 光遗传系统（Olson 等人，2014 年）来调节 ATP酶表达。CcaS 是一种感受器组氨酸激酶，可被绿光激活（λ_{535nm}）。活化的 CcaS 会使其对应的响应调节蛋白 CcaR 磷酸化。磷酸化后的（活化）CcaR 可使目标基因发生转录。相反，CcaS 会被红光灭活（λ_{650nm}），¯从而阻断转录。下面令 $\boldsymbol{u}_s := I_s$ 和 $\bar{\boldsymbol{u}}_c := I_c$，其中 I_s 是由¯控制器调节的绿光光强， I_c 是 I_c的平均值，即生物反应器内细胞感知到的绿光光强。因此，过程输入包括一个信息遗传输入和底物进料速率。

此外，我们考虑一种平板式光生物反应器，它由两块平面和其间的狭窄间隙构成，从而形成一个矩形通道（Chanquia 等人，2021 年）。该生物反应器的一侧由绿光光源照射。已知这种几何结构能使单位培养体积的受光面积最大化，因此非常适用于光遗传学。

5.1 模型与过程考量

我们假设 ATP酶表达的剂量-反应函数遵循希尔函数（Olson 等人，2014 年 ）

$$\eta_{\text{ATPase}}\left(\bar{I}_c\right) = \alpha + \beta \frac{\bar{I}_c^{\delta}}{K^{\delta} + \bar{I}_c^{\delta}}, \tag{20}$$



3PG：3-磷酸-D-甘油酸；AA：氨基酸；AcCoA：乙酰辅酶A；ACE：乙酸；ADP：二磷酸腺苷；AKG：α-酮戊二酸；ATP：三磷酸腺苷；CO2：二氧化碳；CoA：辅酶A；DHAP：磷酸二羟丙酮；ETH：乙醇；Ei：酶i；F6P：果糖-6-磷酸；FOR：甲酸；FUM：延胡索酸；G3P：甘油醛-3-磷酸；G6P：葡萄糖-6-磷酸；GLC：葡萄糖；LAC：乳酸；MAL：苹果酸；NAD：烟酰胺腺嘌呤二核苷酸；NADH：还原型NAD；OAA：草酰乙酸；PEP：磷酸烯醇式丙酮酸；PYR：丙酮酸；Q：配额；R：核糖体；SUCC：琥珀酸。

图 3 大肠杆菌 *KBM10111*厌氧乳酸发酵的资源分配模型示意图。催化物种以斜体表示（例如，*atpAGD* 指 ATP 酶的相关基因）。部分酶通过下划线符号合并表示。灰色部分表示受阻通路。改编自 Espinel-Ríos、Bettenbrock 等人（2022年）。

其中， α 是与输入无关的基础产量速率（例如，由于启动子泄漏或组成型表达）， β 是依赖于输入的最大产量速率， K 是饱和常数， δ 是希尔系数。

我们假设，由于细胞会干扰光束，生物反应器内部的光穿透并不均匀。令 l 表示生物反应器两块板之间的距离。假设光线垂直照射受光平面，且培养物充分混合，我们在无穷小距离 $\mathrm{d}l$上进行衡算。从 l_0 到 l 积分后，得到

$$I_c(l, B) = I_c(l_0, B)e^{-a_\lambda B(l-l_0)}, \tag{21}$$

其中， $I_c(l_0, B) = I_s$， a_λ 是一个集总的、以生物量为基准的常数，用于计入光散射和吸收效应。请注意，¯后一个方程的推导与朗伯-比尔定律的推导类似（Hofmann 等人，2014年）。我们通过对 $I_c(l, B)$ 在 0 到 l 之间求积分平均值，得到 I_c

$$\bar{I}_c(B) = \frac{I_s}{a_\lambda Bl}\left(1 - e^{-a_\lambda Bl}\right). \tag{22}$$

乳酸发酵的动态基于一个现有的 deFBA 模型（Espinel-Ríos, Bettenbrock, 等，2022）。资源分配模型的概要见图 3。该模型总体包含 34 个通量：16 个代谢反应和 18 个生物量生成反应。后者中，16 个反应用于合成催化酶，1 个用于核糖体，另 1 个用于集总配额化合物。模型考虑了 $\boldsymbol{z}$ 中的 5 种物质（葡萄糖、乳酸、甲酸、琥珀酸和二氧化碳）、 $\boldsymbol{m}$ 中的 18 种物质以及 $\boldsymbol{p}$ 中的 18 种物质。细胞组成（g/g 生物量）为：0.06 的催化酶、0.38 的非催化酶，

0.27 的核糖体以及 0.29 的其他组分（DNA、脂质、碳水化合物等），因此 $\varphi_Q = 0.38 + 0.29$。该模型已包含一个生成 ATP 酶的反应。生物量生成反应的成本以氨基酸和 ATP 当量表示。模型参数包括 34 个催化常数和 18 个分子量。此外，模型在式（16）中采用了 $F_{\mathrm{V}} = B(t)$。

此外，在方程（5）至（6）中，我们假设降解可忽略；而对于 ATP 酶

$$D_{\mathrm{ATPase}} = d_{\mathrm{ATPase}} p_{\mathrm{ATPase}} v_L, \tag{23}$$

其中，d_{ATPase} 是一个恒定的 ATPase降解速率。

表1总结了 CcaS/CcaR模块和平板式生物反应器的模型参数，以及过程初始条件。代价函数旨在使过程结束时的乳酸浓度最大化，即 $J = z_{\mathrm{LAC}}(t_f)$，其中 $t_f = 30$ h，共有 12 个控制动作（$N = 12$）。此外，我们还对输入施加盒约束，即$I_s = [0,1]\,\mathrm{W/m^2}$和 $F_{\mathrm{in}} = [0,1]\,\mathrm{L/h}$。我们在优化中进一步加入约束 $z_{\mathrm{GLC}}(t_f) = 0$，以确保所有葡萄糖（即供料底物）都被完全消耗。最后，生物反应器体积不得超过最大工作体积容量$v_{L_{\max}}$，因此我们加入约束 $v_L \leq v_{L_{\max}}$。

关于优化问题数值解的说明：问题 (16)、(17) 和 (18) 的解是最优输入函数。因此，这些问题是无限维的，通常难以求解。一种获得解的方法是采用有限维方法（Findeisen & Allgöwer 2002; Rawlingset al. 2020）。在我们的案例中，这通过假设输入为分段常数，并采用基于拉格朗日插值多项式的正交配置对常微分方程进行离散化来实现，Waldherr et al. (2015) 对此提供了依据。通过对内层优化问题应用 Karush-Kuhn-Tucker条件，问题 (17)、(18) 和 (19) 中的双层优化被转化为带互补约束的数学规划（单层优化）（Dempe & Franke 2019）。最终的优化问题使用 Python、CasADi (Andersson et al. 2019) 和 IPOPT (Wächter & Biegler 2006) 求解。

表1 名义模型的相关参数和初始条件。

Item	数值	Unit	参考文献/注释
δ	2.490	1	Olson 等（2014年）
K	0.138	$\mathrm{W/m^2}$	Olson 等（2014年）
α	$2 \cdot 10^{-6}$	mmol/g/h	注释1
β, S1	$1 \cdot 10^{-4}$	mmol/g/h	注释 1
β, S2	$2.5 \cdot 10^{-5}$	mmol/g/h	注释 1
β, S3	$1 \cdot 10^{-5}$	mmol/g/h	注释 1
d_{ATPase}	$6.3 \cdot 10^{-2}$	1/h	注2
l	0.022	m	注3
$v_{L_{\max}}$	45	L	注3
a_λ	$1 \cdot 10^{-2}$	$\mathrm{m^2/g}$	注4
$z_{\mathrm{GLC,in}}$	2220	mM	——
$x_{\mathrm{GLC}}(0)$	139	mM	——
$x_{\mathrm{LAC}}(0)$	0	mM	——
$x_{\mathrm{CO_2}}(0)$	0	mM	——
$x_{\mathrm{FOR}}(0)$	0	mM	——
$x_{\mathrm{SUCC}}(0)$	0	mM	——
$B(0)$	0.59	g/L	——
$\boldsymbol{p}(0)$	注5	mM	——
$v_L(0)$	30	L	——

注1。根据不同的诱导强度情景（Si）下可行的动态通量平衡分析模拟推断出生物学上合理的数值（Espinel-Ríos, Bettenbrock, 等，2022）。**注2**。估算为 $d_{\mathrm{ATPase}} = \frac{\ln(2)}{t_{0.5}}$，其中 $t_{0.5}$ 为ATP酶蛋白半衰期（Benito et al. 1991）。**注3**。基于中试规模平板式光生物反应器设计（Koller et al. 2018）。**注4**。假设为生物学上合理的数量级。估算为微藻典型参数值的约 1/30（Blanken等，2016年）。**注5**。基于资源平衡分析，使用 $B(0)$估算（Jabarivelisdeh等，2020年）。

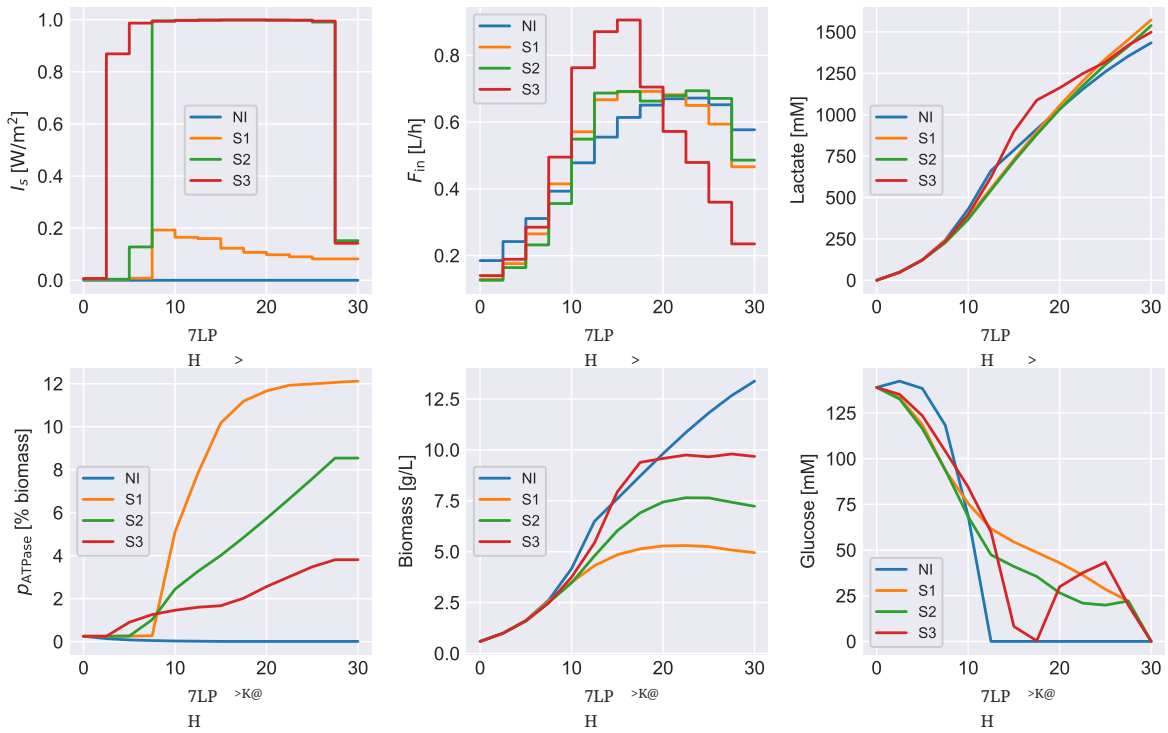


图4 无 模型-对象失配的开环补料分批模拟。场景：NI——无诱导，S1——高强度、S2——中等强度、S3——低强度可诱导CcaS/CcaR系统。

5.2 开环最优光遗传控制

图 4 展示了在不存在模型-被控对象失配的情况下，补料分批发酵的开环优化结果。我们展示了四种情景：

1. S1：高强度诱导型 CcaS/CcaR 系统——高 β 值。2. S2：中等强度诱导型 CcaS/CcaR 系统——中 β 值

3. S3：低强度诱导型 CcaS/CcaR 系统——低 β 值。4. NI：无可诱导ATP酶——不存在CcaS/CcaR系统。

NI 情况下最终乳酸浓度为 1434.3 mM，而 S1 达到 1572.3 mM （↑ 10 %）、S2 1538.5 mM（↑ 7 %）和 S3 1498.4 mM （↑ 4 %）。请注意，到所有发酵结束时，生物反应器体积均达到允许的最大值，且葡萄糖已完全耗尽。这意味着总体而言，补加并消耗的葡萄糖净量相同。因此，在先前的场景中，产物滴度的相对增益也按比例对应于产物得率和体积生产率的提升。

此外，在 S1 中，ATP酶的最高浓度为细胞干重的 12.1 %，而在 S2 和 S3 中则分别为 8.5 % 和 3.8 %。不难预见，CcaS/CcaR系统的诱导能力越强，ATP酶表达的净增量就越高，因此产物得率的净增幅也越大。以往关于单阶段分批发酵中 ATP 动态周转的研究表明，产物得率的提高与体积生产率的下降相关（Espinel-Ríos, Bettenbrock, 等，2022；Espinel-Ríos, Morabito, Pohlodek, 等，2022）。本文表明，在补料分批系统中，可以通过 ATP 周转机制提高产物得率，并且 通过引入补料速率提高体积生产率。

与 NI 相比，S1、S2 和 S3 情景下的最终生物量浓度分别降低了 63、46 和 28 %。这可以归因于 ATP 周转导致的生物量产率降低，以及生产 ATP 酶所需成本可能带来的资源负担，两者共同作用所致。需要注意的是，底物补料也会产生稀释效应。总体而言，ATP 周转速率的提高尽管降低了生物量生长速率，却提升了最终乳酸滴度。

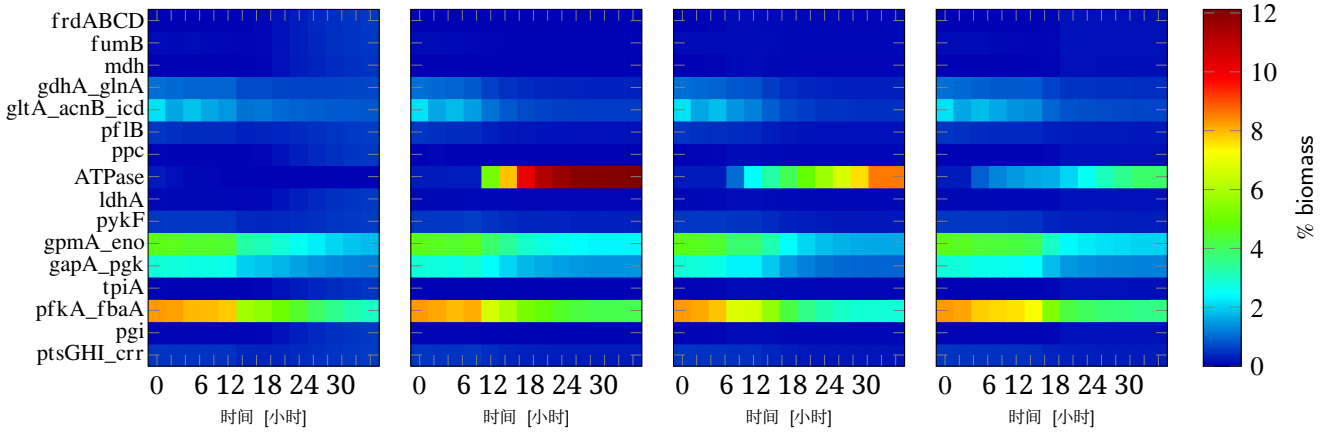


图 5 图 4 中补料分批发酵的酶表达热图，以生物量干重为基准。从左到右依次为：NI–无诱导、S1–高强度、S2–中等强度和 S3–低强度的可诱导CcaS/CcaR系统。

在所有诱导情景中，补料速率起初逐渐增加，随后在发酵中期左右开始持续下降。这样既能积累足够的生物量，又能维持较低的 ATP 酶诱导水平。随后，为避免生物量过度稀释，随着 ATP 酶表达水平升高，补料速率逐渐降低。

我们基于模型的优化的一个优点是考虑了资源分配约束。图 5 展示了 ATP 酶表达相关的资源分配现象，其中给出了开环发酵过程中动态变化的酶组成谱。请注意，ATP 酶的诱导导致非调控酶重新分配。例如，我们比较一下酶 frdABCD、 fumB、 mdh、ppc、 ldhA、 tpiA 和 pgi 的谱图。在 NI 发酵中，这些酶似乎略有积累，而在 ATP 酶诱导的情况下，它们的浓度则维持在较低水平。在 S1 情景中，这种效应更为明显，因为 ATP 酶的表达水平较高。

5.3 | 对抗 模型不确定性和扰动——光遗传闭环控制

开环控制无法应对模型不确定性、意外扰动和工艺变化。因此，我们现在评估缩短预测时域的模型预测控制在应对系统不确定性方面的性能。分析仅针对表现最佳的高强度诱导型 CcaS/CcaR 系统。我们通过将酶 pfkA_fbaA、gpmA_eno、gapA_pgk、gltA_acnB_icd 和 gdhA_glnA 的催化常数乘以 0.98，引入模型-对象失配，这会略微降低发酵速率。我们还分别将 δ 和 d_{ATPase} 降低至原来的 0.97 和 0.98；后者会降低希尔函数的陡峭程度，前者则使 ATP 酶降解速度变慢。植物仿真使用修改后的 模型，而控制器使用名义模型。我们考虑两种模型预测控制方案：

1. 模型预测控制 1：所有状态都可以在线测量，且不存在测量噪声。
2. 模型预测控制 2：ATP酶浓度、生物量干重和胞外代谢物可以在线测量。测量值中加入了高斯白噪声（1 % 标准差）。通过全信息估计[5]估计细胞组成。将重建的细胞组成与在线测量值一起传递给模型预测控制器。

模型预测控制（MPC）仿真结果如图 6 所示。我们还绘制了开环情景（存在模型-对象失配）作为参考。开环控制器最终得到的乳酸浓度为 1449.5 mM，且有 64.3 mM 净未消耗葡萄糖。施加的光强使 ATPase 酶浓度升至细胞干重的 12.6 %，但随后略微下降至 11.9 %。

[4]ATP酶浓度可以通过荧光生物传感器等方式测量（Kim 等，2021年）。[5]为简化起见，在全信息估计中，我们仅考虑目标函数 (19a) 中的第二项。我们忽略状态噪声，并假设模型参数恒定，因此不对其进行估计。矩阵 **R** 选为单位矩阵。

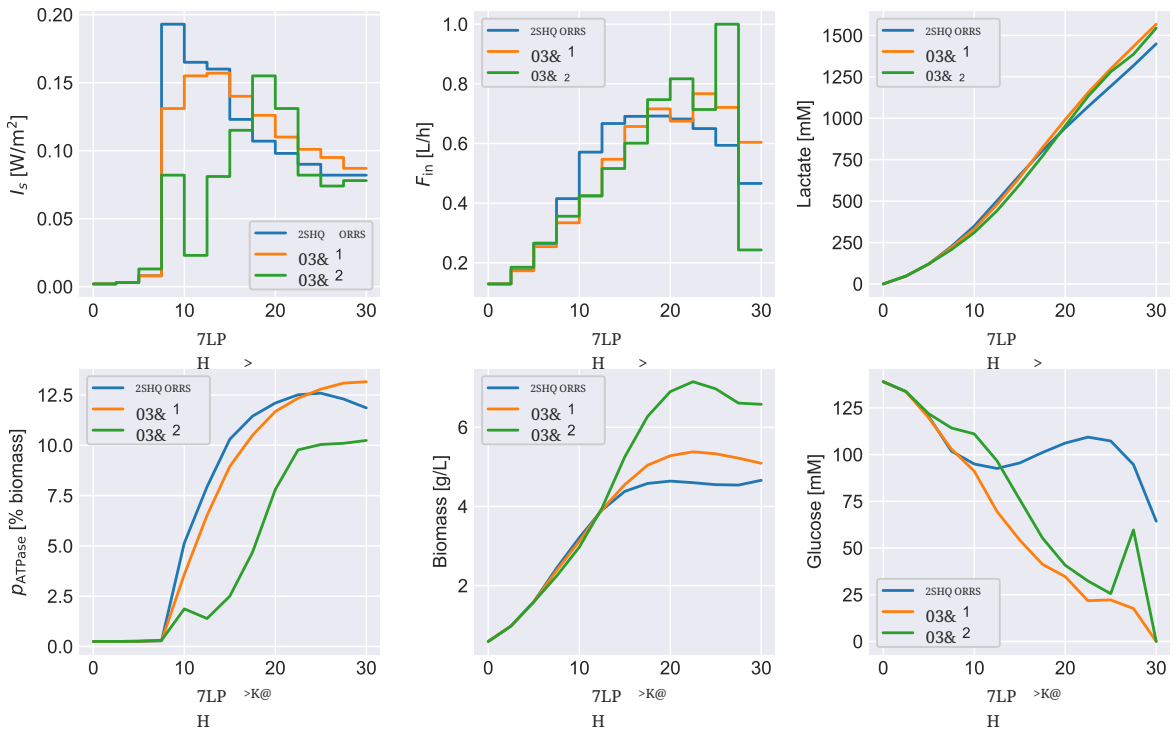


图6 高强度诱导型 CcaS/CcaR 系统在模型不确定性下的闭环补料分批模拟。模型预测控制 1：采用全状态测量且无测量噪声。模型预测控制 2：在存在测量噪声的情况下，使用 p_{ATPase}、$\boldsymbol{B}$ 和 $\boldsymbol{z}$ 的测量值；使用全信息估计器估计 $\boldsymbol{p}$。图中还展示了开环情形（无在线校正动作）。

MPC 1 中施加的光强使细胞内 ATPase酶组分最终超过了开环发酵中达到的数值。修正后的光强和进料速率共同作用，使最终乳酸滴度达到 1567.8 mM。与开环优化相比，这一结果提升了 8 %。此外，MPC 1 在过程结束时没有未消耗葡萄糖。请注意，MPC 1 情景的结果非常乐观，因为假设能够进行全状态测量，且不存在测量噪声。

模型预测控制 2 的结果更符合实际应用情况，即采用状态估计并考虑测量噪声。图7展示了不同采样时刻所选物种的细胞组成估计值。我们还计算了估计值的标准误差（SE）[6]。总体而言，全信息估计器很好地跟踪了生物量组分的浓度趋势。不过需要指出的是，通常随着过程推进，估计效果会逐渐改善。也就是说，在过程开始的约三分之一阶段，估计精度较低（例如，酶 pfkA_fbaA、gltA_acnB_icd 和 gdhA_glnA的估计曲线）。估计效果逐步改善，是因为估计时域不断增长，可用测量值也随之增加。这也解释了为什么在过程开始时，相较于模型预测控制 1 和开环优化，控制器 ′的预测偏差较大。模型预测控制 2 达到约 10.2 %的细胞内ATPase酶浓度，从而促进了更多的生物量积累。控制器调整了补料速率，以避免过程结束时仍有未消耗的葡萄糖。模型预测控制 2 得到的最终乳酸浓度为 1544.4 mM，与模型预测控制 1 情形下达到的数值非常接近。

模型预测控制（MPC）仿真表明，采用基于模型的反馈控制，并可选择结合状态估计方法，能够在存在系统不确定性的情况下提升代谢网络生物系统的过程性能。

[6] $\mathrm{SE}_{\mathrm{FIE}} = \sqrt{\frac{\sum_{j=0}^{N-1}(p_{i,j} - p^*_{i,j})^2}{E_i}}$，$\forall i \in [1, n_{p_i}]$，$E_i$：$p_i$ 估计值的总数。

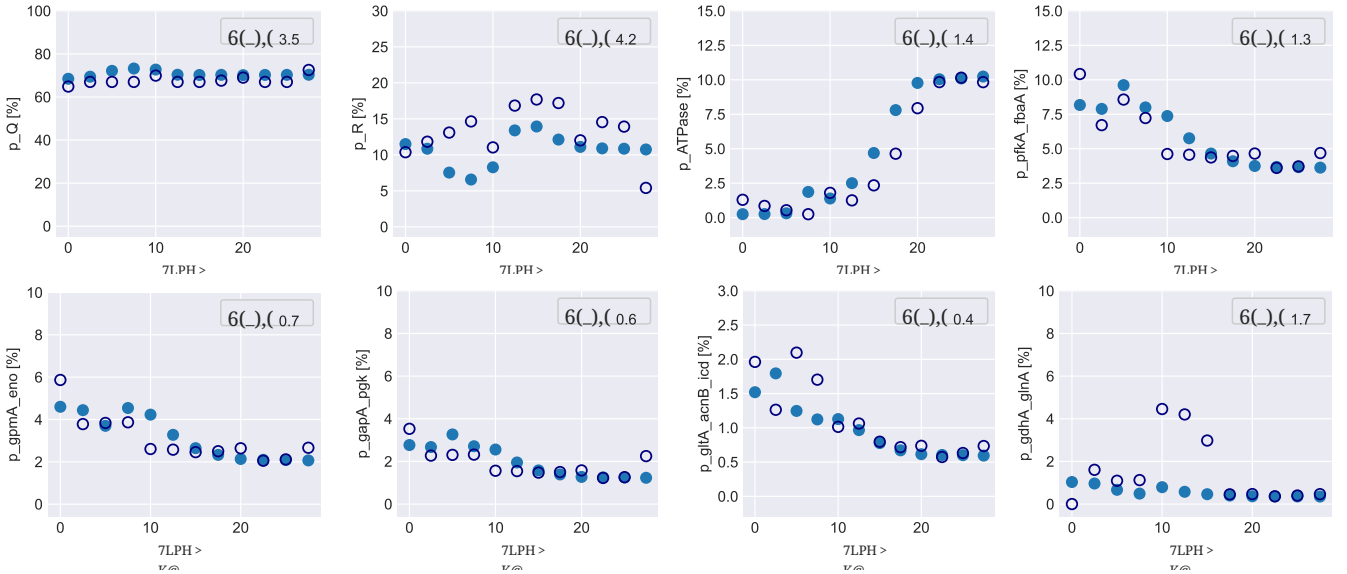


图 7 细胞组分占细胞干重的百分比在线估计。仅显示含量最高的八种组分。实心圆：精确值。空心圆：状态估计值。估计的标准误差（SE_FIE）。

6 结论与展望

我们提出将网络控制遗传学与基于模型的优化和预测控制相结合，用于动态代谢工程应用。所提出的代谢赛博遗传学框架利用在线代谢调控的理念，通过动态调节与代谢相关的细胞内蛋白质的基因表达来实现这一目标。为此，我们开发了一个动态约束建模框架，将代谢反应、资源分配以及外源基因表达调控的动态过程整合起来。该模型结合了基于模型的优化、预测控制和估计方法，以促进代谢网络生物系统的实施。

通过光遗传调控 ATP酶基因表达，实现对细胞 ATP周转的动态控制，这凸显了该技术的潜力。我们表明，对光强和底物进料速率进行最优控制，可以提升产物滴度和体积生产率等方面的过程性能。此外，我们还证明，通过模型预测控制引入反馈，有助于抵消系统不确定性。

我们认为，所概述的代谢赛博遗传学框架为新的、更先进的生物技术应用开辟了道路，在这些应用中，需要在整个过程中操纵代谢通量。这实际上与动态代谢工程方法相一致，并超越了传统的网络遗传学方案；在传统方案中，受调控蛋白（例如荧光报告蛋白）并未直接参与代谢通路。此外，所提出框架的基于模型这一特点，有助于缩短工艺开发时间、降低其成本，并实现更稳健、一致且灵活的运行。

我们目前正在开发结合机器学习的代谢赛博遗传学方法，并将其应用范围扩展至合成微生物群落。最后，我们计划通过实验验证所提出的框架，验证内容包括文中所述案例研究及其他相关生物过程。

7 致谢

本研究得到了国际马克斯·普朗克过程与系统工程高级方法研究生院（IMPRS ProEng）以及由欧盟-欧洲区域发展基金资助的 DIGIPOL 项目的支持。

参考文献

Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., & Diehl, M.（2019）。CasADi：用于非线性优化和最优控制的软件框架。Math Program Comput，11(1)，1–36。Azimi, H., Tezel, H., & Thibault, J.（2019）。通过膜渗透汽化原位回收 ABE发酵液中的丁醇。Chem Eng Res Des，150，49–64。Behera, B., Acharya, A., Gargey, I. A., Aly, N., & P, B.（2019）。可持续生物燃料生产中微藻培养的生物过程工程原理。Bioresour Technol Rep，5，297–316。Benito, B., Moreno, E., & Lagunas, R.（1991）。静息酵母细胞中质膜ATP酶及其激活系统的半衰期。Biochem Biophys Acta - Biomembr，1063(2)，265–268。Blanken, W., Postma, P. R., de Winter, L., Wijffels, R. H., & Janssen, M.（2016）。预测微藻生长。《藻类研究》，14，28–38。Boecker, S., Harder, B.-J., Kutscha, R., Pflügl, S., & Klamt, S.（2021）。提高三磷酸腺苷周转可提升大肠杆菌中2,3-丁二醇的合成产率。Microb Cell Factories，20(1)，63。Carrasco-López, C., García-Echauri, S. A., Kichuk, T., & Avalos, J. L.（2020）。光遗传学和生物传感器为代谢网络控制遗传学奠定基础。Curr Opin Biotechnol，65，296–309。Chanquia, S. N., Vernet, G., & Kara, S.（2021）。用于培养和合成的光生物反应器：规格、挑战与展望。Eng Life Sci，elsc.202100070。Cui, S., Lv, X., Xu, X., Chen, T., Zhang, H., Liu, Y., . . . Liu, L.（2021）。用于代谢通路动态调控的多层遗传回路。ACS Synth Biol，10(7)，1587–1597。del Rio-Chanona, E. A., Ahmed, N. R., Wagner, J., Lu, Y., Zhang, D., & Jing, K.（2019）。用于补料分批生物过程动态优化的基于物理原理与数据驱动建模技术比较。Biotechnol Bioeng，116(11)，2971–2982。Dempe, S., & Franke, S.（2019）。采用KKT方法求解双层优化问题。《优化》，68(8)，1471–1489。De Vrieze, J., De Mulder, T., Matassa, S., Zhou, J., Angenent, L. T., Boon, N., & Verstraete, W.（2020）。微生物学中的随机性：管理不可预测性以实现可持续发展目标。Microb Biotechnol，13(4)，829–843。Doran, P. M.（2013）。生物过程工程原理（第2版）。阿姆斯特丹；波士顿：Elsevier/学术出版社。Elsheikh, M., Hille, R., Tatulea-Codrean, A., & Krämer, S.（2021）。生物过程应用中多速率移动时域估计方案的比较综述。Comput Chem Eng，146，107219。Espinel-Ríos, S., Bettenbrock, K., Klamt, S., & Findeisen, R.（2022）。通过约束模型优化和三磷酸腺苷周转预测控制，最大限度提高分批发酵效率。AIChE J.，68(4)，e17555。Espinel-Ríos, S., Morabito, B., Bettenbrock, K., Klamt, S., & Findeisen, R.（2022）。用于监测细胞组成动态变化的软测量。IFAC-PapersOnLine（即将出版）。Espinel-Ríos, S., Morabito, B., Pohlodek, J., Bettenbrock, K., Klamt, S., & Findeisen, R.（2022）。分批发酵中ATP浪费的强化：ATP酶基因表达的最优控制与动态调节。IFAC-PapersOnLine，55(7)，174-180。Findeisen, R., & Allgöwer, F.（2002）。非线性模型预测控制简介。载于《第21届比荷卢系统与控制会议》（第11卷，第119–141页）。Fung Shek, C., & Betenbaugh, M.（2021）。把握生物过程脉搏：哺乳动物细胞培养物的近线和在线监测。Curr Opin Biotechnol，71，191–197。Gottstein, W., Olivier, B. G., Bruggeman, F. J., & Teusink, B.（2016）。从单一生物到微生物群落的基于约束的化学计量建模。J R Soc Interface，13(124)，20160627。Gutiérrez Mena, J., Kumar, S., & Khammash, M.（2022）。通过光遗传反馈动态控制细菌共培养组成。Nat Commun，13(1)，4808。Hartline, C. J., Schmitz, A. C., Han, Y., & Zhang, F.（2021）。代谢工程中的动态控制：理论、工具与应用。《代谢工程》，63，126–140。Hoffman, S. M., Tang, A. Y., & Avalos, J. L.（2022）。光遗传学照亮微生物工程应用。Annu Rev Chem Biomol Eng，13(1)，373–403。Hofmann, A., Simon, A., Grkovic, T., & Jones, M.（2014）。光谱方法。载于（第10–94页）。剑桥大学出版社。

Hsiao, V., Swaminathan, A., & Murray, R. M.（2018年）。合成生物学的控制理论：合成生物学系统表征、控制设计和控制器实现的最新进展。IEEE Control Syst, 38(3), 32–62。Hughes, S. R., & Jones, M. A.（2020年）。生物燃料和生物技术衍生大宗化学品的全球需求：技术、市场与挑战。载于 A. A. Vertès, N. Qureshi, H. P. Blaschek, & H. Yukawa（编），Green Energy to Sustainability（第1版，第137–156页）。Wiley。Hädicke, O., Bettenbrock, K., & Klamt, S.（2015年）。强制ATP无效循环提高了大肠杆菌厌氧乳酸生产的比生产率和产率：通过ATP浪费提高产率和生产率。Biotechnol Bioeng, 112(10), 2195–2199。Jabarivelisdeh, B., Carius, L., Findeisen, R., & Waldherr, S.（2020年）。基于约束的建模与状态估计的生物过程自适应预测控制。Comput Chem Eng, 135, 106744。Jabarivelisdeh, B., Findeisen, R., & Waldherr, S.（2018年）。基于动态代谢-遗传网络模型的补料分批生物反应器模型预测控制。IFAC-PapersOnLine, 51(19), 34–37。Jabarivelisdeh, B., & Waldherr, S.（2016年）。在动态优化方案中使用基于约束的模型提高生物过程生产率。IFAC-PapersOnLine, 49(26), 245–251。Jabarivelisdeh, B., & Waldherr, S.（2018年）。基于代谢-遗传网络模型和双层动态规划的生物过程生产率优化。Biotechnol Bioeng, 115(7), 1829–1841。Jullesson, D., David, F., Pfleger, B., & Nielsen, J.（2015年）。合成生物学和代谢工程对精细化学品工业生产的影响。Biotechnol Adv, 33(7), 1395–1402。Khammash, M. H.（2022年）。网络控制遗传学：遗传控制系统的理论与应用。Proc IEEE, 110(5), 631–658。Kim, H., Ju, J., Lee, H. N., Chun, H., & Seong, J.（2021年）。基于荧光蛋白的基因编码生物传感器。Sensors, 21(3), 795。Klamt, S., & von Kamp, A.（2022年）。分析并解决代谢网络通量平衡分析中的不可行性。Metabolites, 12(7), 585。Koller, A. P., Wolf, L., Brück, T., & Weuster-Botz, D.（2018年）。在平板气升式光生物反应器中利用栅藻属进行生物量生产放大的研究。Bioprocess Biosyst Eng, 41(2), 213–220。Lalwani, M. A., Ip, S. S., Carrasco-López, C., Day, C., Zhao, E. M., Kawabe, H., & Avalos, J. L.（2021年）。利用光遗传控制lac操纵子进行细菌化学品和蛋白质生产。Nat Chem Biol, 17(1), 71–79。Lalwani, M. A., Kawabe, H., Mays, R. L., Hoffman, S. M., & Avalos, J. L.（2021年）。光遗传控制微生物群落种群以生产化学品。ACS Synth Biol, 10(8), 2015–2029。Lalwani, M. A., Zhao, E. M., & Avalos, J. L.（2018年）。代谢工程中动态控制的当前与未来模式。Curr Opin Biotechnol, 52, 56–65。Lindner, F., & Diepold, A.（2021年）。细菌中的光遗传学——应用与机遇。FEMS Microbiology Reviews, fuab055。Liu, S.（2020年）。生物过程工程：动力学、可持续性与反应器设计（第3版）。阿姆斯特丹，荷兰；马萨诸塞州剑桥：Elsevier。Liu, Z., Zhang, J., Jin, J., Geng, Z., Qi, Q., & Liang, Q.（2018年）。用光编程细菌——合成生物学中的传感器与应用。Front Microbiol, 9, 2692。Lovelett, R. J., Zhao, E. M., Lalwani, M. A., Toettcher, J. E., Kevrekidis, I. G., & L. Avalos, J.（2021年）。用于代谢工程应用的酵母光遗传回路动力学建模。ACS Synth Biol, 10(2), 219–227。Milias-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P., & Khammash, M.（2016年）。用于精确稳健调节基因表达和细胞生长的自动化光遗传反馈控制。Nat Commun, 7(1), 12546。Morabito, B., Kienle, A., Findeisen, R., & Carius, L.（2019年）。不确定生物技术过程的多模式模型预测控制与状态估计。IFAC-PapersOnLine, 52(1), 709–714。Morabito, B., Pohlodek, J., Kranert, L., Espinel-Ríos, S., & Findeisen, R.（2022年）。利用HILO-MPC对生物反应器进行高效、简便的高斯过程支持随机模型预测控制。IFAC-PapersOnLine, 55(7), 922–927。Morabito, B., Pohlodek, J., Matschek, J., Savchenko, A., Carius, L., & Findeisen, R.（2021年）。面向重复过程的风险感知机器学习支持模型预测控制与开环优化。IFAC-PapersOnLine, 54(6), 321–328。Nimmegeers, P., Vallerio, M., Telen, D., Impe, J., & Logist, F.（2018年）。参数不确定性下生物反应器的交互式多目标动态优化。Chem Ing Tech, cite.201800082。Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R., & Tabor, J. J.（2014年）。利用光学编程的基因表达信号表征细菌基因回路动力学

。Nat Methods, 11(4), 449–455。Pfaffinger, C. E., Schöne, D., Trunz, S., Löwe, H., & Weuster-Botz, D.（2016年）。平板气升式光生物反应器中微藻面产率的基于模型优化。Algal Res, 20, 153–163。Rawlings, J., Mayne, D., & Diehl, M.（2020年）。模型预测控制：理论、计算与设计（第2版）。圣巴巴拉：Nob Hill Publishing, LLC。Reardon, K. F.（2021年）。生物制造中细胞和底物的实用监测技术。Curr Opin Biotechnol, 71, 225–230。Reyes, S. J., Durocher, Y., Pham, P. L., & Henry, O.（2022年）。用于监测、控制和改进细胞培养过程的现代传感器工具与技术。Processes, 10(2), 189。Ryu, K. H., Kim, B., & Lee, J. H.（2019年）。光自养条件下脂质生产的微藻培养策略基于模型的优化。Comput Chem Eng, 121, 57–66。Saa, P. A., & Nielsen, L. K.（2017年）。代谢动力学模型的构建、建立与分析：建模框架综述。Biotechnol Adv, 35(8), 981–1003。Sanford, K., Chotani, G., Danielson, N., & Zahn, J. A.（2016年）。可再生化学品的规模放大。Curr Opin Biotechnol, 38, 112–122。Scull, C. E., Dandpat, S. S., Romero, R. A., & Walter, N. G.（2021年）。转录核糖开关整合不同时间尺度，以调控细菌基因表达。Front Mol Biosci, 7, 607158。Shen, X., Wang, J., Li, C., Yuan, Q., & Yan, Y.（2019年）。动态基因表达工程作为通路工程工具。Curr Opin Biotechnol, 59, 122–129。Tandar, S. T., Senoo, S., Toya, Y., & Shimizu, H.（2019年）。用于控制大肠杆菌中心代谢通量的光遗传开关。Metab Eng, 55, 68–75。Vandermies, M., & Fickers, P.（2019年）。解脂耶氏酵母重组蛋白生产的生物反应器规模策略。Microorganisms, 7(2), 40。Venayak, N., Anesiadis, N., Cluett, W. R., & Mahadevan, R.（2015年）。通过动态控制进行代谢工程。Curr Opin Biotechnol, 34, 142–152。Waldherr, S., Oyarzún, D. A., & Bockmayr, A.（2015年）。结合基因表达的代谢网络动态优化。《理论生物学杂志》, 365, 469–485。Wohlgemuth, R., Twardowski, T., & Aguilar, A.（2021年）。生物经济逐步向前发展——全球之旅。《新生物技术》, 61, 22–28。Wächter, A., & Biegler, L. T.（2006年）。大型非线性规划内点滤波线搜索算法的实现。Math Program, 106(1), 25–57。Yang, S., Kim, S., Kim, D.-K., Jeon An, H., Bae Son, J., Hedén Gynnå, A., & Ki Lee, N.（2019年）。转录与翻译促进基因位点重定位至大肠杆菌拟核周边。Nat Commun, 10(1), 5131。Yasemi, M., & Jolicoeur, M.（2021年）。细胞代谢建模：基于约束的稳态方法与动力学方法综述。Processes, 9(2), 322。Zahoor, A., Messerschmidt, K., Boecker, S., & Klamt, S.（2020年）。在酿酒酵母中基于ATP酶实现强制ATP浪费，以提高乙醇生产。Biotechnol Biofuels, 13(1), 185。Zhao, E. M., Lalwani, M. A., Chen, J.-M., Orillac, P., Toettcher, J. E., & Avalos, J. L.（2021年）。用于光诱导代谢调控的光遗传放大电路。ACS Synth Biol, 10(5), 1143–1154。

