



上海科技大学
ShanghaiTech University

合成形态发生学：使用合成生物学方法构建特定大小的细胞群体



上海科技大学
ShanghaiTech University

本科毕业论文（设计）

题 目：_____合成形态发生学：_____

使用合成生物学方法构建特定大小的细胞群体

学生姓名：_____陈铭哲_____

学 号：_____60971060_____

入学年份：_____2015_____

所在学院：_____生命科学与技术学院_____

攻读专业：_____生物科学_____

指导教师：_____钟桂生_____

上海科技大学

2019 年 5 月



上海科技大学
ShanghaiTech University

THESIS

Subject: Synththetic Morphogenesis:

Engineering Cell Population with Specific Size

Student Name: Mingzhe Chen

Student ID: 60971060

Year of Entrance: 2015

School: School of Life Science and Technology

Major: Biology

Advisor: Guisheng Zhong

ShanghaiTech University

Date: 2019 / 5



合成形态发生学：使用合成生物学方法构建特定大小的细胞群体

摘要

合成生物学作为一门在过去二十年间活跃发展的新兴学科，旨在通过重新布线细胞内部的基因回路以服务于人类特定的工程目的。在过去二十年间，合成生物学的发展主要集中在以单细胞尺度，研究人员以单个细胞作为研究对象，通过对细胞内部的转录-翻译等生物过程的建模，来协助设计服务于特定目的的基因回路。随着近年来对细胞间相互作用以及对细胞间通讯的研究越发深入，如何以细胞为基本单位，构建具有特定组织结构功能的细胞群体，成为最近几年来合成生物学的一个新的研究方向。然而，如何以细胞为基本单位，对多细胞群体的生物过程进行改造，建模和研究，到目前为止，尚未有成系统的研究和讨论。什么是编码多细胞群体发育的基本工程模块，以及如何实现这些模块以重建形态发生过程的学科，我们称之为合成形态发生学。本文使用从一个单一细胞发展为一个具有特定大小的细胞群体的过程为例，提出工程化实现所需要的基因回路设计，对细胞生长与基因转录-翻译过程的参数进行验证，应用时间尺度分离，利用 Morpheus 进行建模，最后设计对应的实验对细胞通讯和质粒稀释两部分进行验证。

关键词：合成生物学 形态发生 群体数量控制

Synthetic Morphogenesis: Engineering Cell Population with Specific Size

ABSTRACT

Synthetic biology is a discipline which aims to serve specific purposes by rewiring the genetic circuit inside the cell. The recent development of synthetic biology exhibits the ability to engineering the more complex cell to cell interaction, such as cell-to-cell communication and the cell-to-cell contact by mechanical force. However, until now, there have not been comprehensive discussion on the basic principle needed to reconstruct organism from the multicellular level and how to use such a principle to reconstruct morphogenesis from the synthetic biology perspective, which we call it synthetic morphogenesis. Here, I try to define some basic modules which could be used as the cornerstone for synthetic morphogenesis and design and validate the corresponding genetic circuits to execute the individual module's function which can be coordinated to generate cell population with specific size.

Key words: Synthetic biology, morphogenesis, cell population control

目 录

- 第一章：从合成生物学到合成形态发生-----1
 - 1.1 背景介绍-----1
 - 1.1.1 传统合成生物学：以单细胞生物过程为研究对象-----1
 - 1.1.2 合成形态发生学：以细胞群体过程为研究对象-----1
 - 1.2 构建可控的具有特定大小的细胞球体-----1
 - 1.2.1 目标-----1
 - 1.2.2 原理-----1
 - 1.2.3 合成生物学实现-----1
 - 1.2.4 重要性与创新性-----1
- 第二章：模拟，数值验证与实验-----1
 - 2.1 利用 Morpheus 进行建模与数值模拟： -----1
 - 2.1.1 建模与 Morpheus 介绍-----1
 - 2.1.2 动力学参数确定与时间尺度分离-----1
 - 2.1.3 Morpheus 过程-----1
 - 2.1.4 讨论与反思-----1
 - 2.2 实验设计与验证-----1
 - 2.2.1 细胞间长距离通讯系统测试-----1
 - 2.2.2 非稀释质粒和稀释质粒系统构建-----1
- 第三章：反思与讨论-----1
- 参考文献-----1
- 致谢-----1
- 附录-----1

第一章 从合成生物学合成形态发生

1.1 背景介绍

1.1.1 传统合成生物学：以单细胞生物过程为研究对象

合成生物学是一门将传统电子工程的研究范式与分子生物学和遗传学的研究内容相结合的新兴工程学科。尽管它致力于使用定量方式分析生物过程与生物系统系统，但是与定量和计算生物学的研究范式不同的是，合成生物学倾向于“重新构建生物过程”，也就是说，对改造后的生物系统进行定量的工程学分析，而不是对细胞内生的生物系统进行测量和分析。尽管合成生物学的产品是服务于特定的目的，但是对已有生物系统进行工程化改造的过程，往往有意无意中能够提高我们对复杂生物系统的理解。

合成生物学借鉴计算机科学中的抽象化的思维逻辑，将标准化的“输入/输出”（I/O）系统引入生物学研究，通过解耦联已有的生化过程，将生物过程与其周围的细胞环境抽离开来，作为单独的研究对象。这种将电子工程的 I/O 概念引入合成生物学，将细胞内部的生物过程解耦联成一个个单独的“黑盒子”的研究方法，使研究人员能够专注于特定生物过程的输入和输出，并且对该过程进行定量化的研究和考量。在过去的二十年中，大多数合成生物学家专注于设计细胞内的遗传回路，如振荡器，前馈环路和触发开关等，尽管很多时候测量的是 bulk property，但其实质仍是以单细胞作为研究对象，细胞与细胞之间的关系和结构并不重要。[1] 但同时也有一些尝试工程化细胞群体行为的研究，例如同步震荡过程，群体感应，图灵图案的形成等等。这些尝试都取得了卓有成效的研究。但是这些研究或研究细胞群体行为的整体性质，不注重细胞群体内部的精细结构，或依赖于随机过程形成的结构图案，而没有对如何诱导细胞形成特定样式进行系统性的讨论。

过去关于合成生物学的绝大部分研究是在诸如大肠杆菌和酵母等的简单的模式生物中进行的，因为这些底盘生物作为科学界几十多年来的研究对象，其遗传学背景相对而言简单明晰，而遗传操作手段也相对成熟。然而，哺乳动物细胞例如中国仓鼠卵巢细胞（CHO）以及人胚胎肾细胞系（HEK293），作为更加高级的细胞系，具有与人类细胞更为相似的遗传背景。以这些细胞作为底盘生物的研究，为未来合成生物学在医药行业中的应用奠定了基础，而其本身所具有的形成多细胞结构的潜能性，使其成为可靠的编码多细胞群体复杂行为的底盘生物。

1.1.2 合成形态发生学：重新编码细胞群体结构

纵观生物进化的历程，从最简单的单细胞生物，如作为原核生物的大肠杆菌以及作为真核生物的酿酒酵母，到相对而言具有均一性质的多孔动物，再到具有三胚层分化，复杂组织架构的高等动物例如哺乳动物和鸟类，生物有机体的发展规律是从简单的各自营生的单细胞生物，向具有复杂结构的，多细胞分工合作明确的有机体发展。

先前的合成生物学则专注在单细胞层面的研究，而近年来随着编码细胞间相互作用（包括细

胞间信号传递与细胞间机械力介导的黏连)过程的发展,对于合成生物学家而言,接下来的问题是,我们是否能够通过编码细胞内部基因回路,调整细胞-细胞界面的相互作用,工程化具有某种图案样式特性的复杂多细胞系统?我们如何在时间与空间上对复杂多细胞系统的生物发育过程进行建模和考察?我们在何种程度上,能够复现生物界已有的复杂多细胞系统,又在何种程度上能够设计自然界尚不存在的多细胞结构?

1.2 构建可调控的具有特定大小的细胞球体

1.2.1 目标:

本文的目标是构建一个由单一细胞分裂而成的系统,该系统能通过细胞分裂,形成具有特定数量的群体。一旦群体数量达到一定值,该多细胞群体将停止分裂或诱导细胞程序性死亡(Programmed cell death, PCD)以维持特定的数量。一旦该细胞球体的一部分受损,该系统可以通过分裂恢复到与先前相同的大小。该多细胞系统的数量理论上可以通过外部激活因子调控基因回路的表达来加以控制。

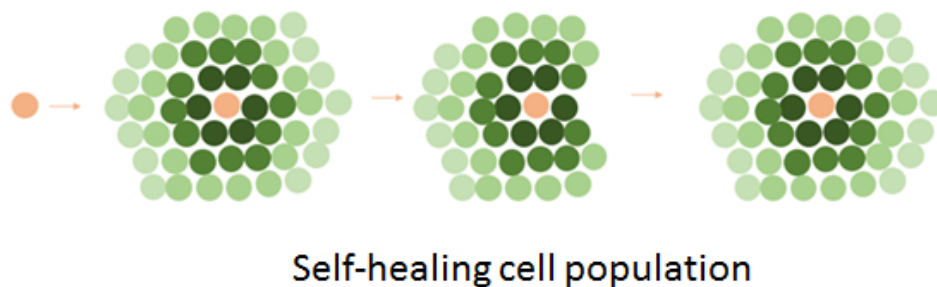


图 1-1: 具有自愈功能的,一定细胞数量的细胞群体

1.2.2 原理:

本文提出了三种不同的功能设计,来用以达成产生具有特定群体数量的细胞球体的目的,包括:(1)非对称细胞分裂(2)细胞与细胞间通信系统(3)细胞程序性死亡或细胞周期控制。

- (1) 非对称细胞分裂指,对于母细胞而言,其每次分裂将成为两种不同的细胞类型,我们定义他们为:“母细胞”和“子细胞”。母细胞指具有与分裂前相同性质的细胞,而子细胞则具有与分裂前细胞不同的性质。在这里,我们定义子细胞每次分裂都会形成两个相同的子细胞,因此可以想象,对于从单一母细胞分裂而成的整个系统将包含一个母细胞和数个子细胞

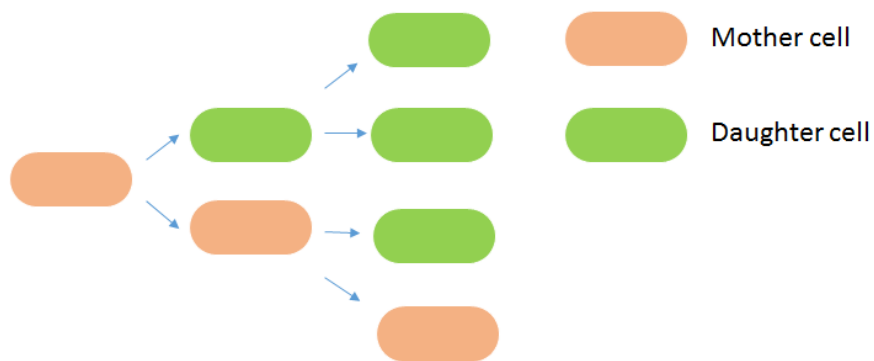
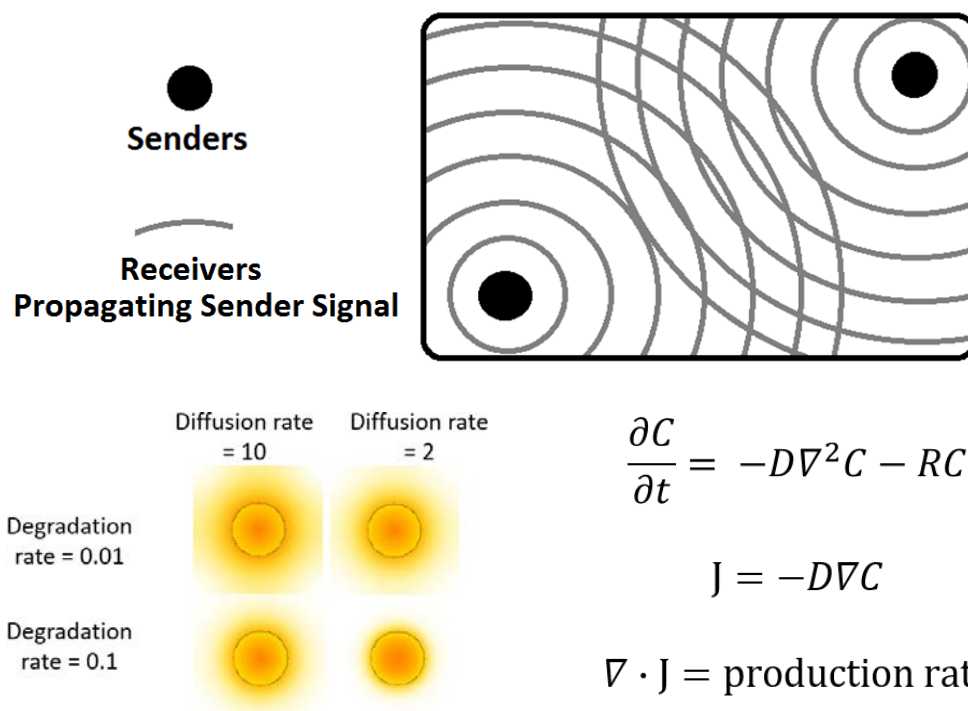


图 1-2: 母细胞进行不对称细胞分裂, 分裂为两种细胞类型, 而子细胞分裂为与原细胞一致的细胞类型

(2) 信号传递系统。母细胞需要产生特定的维持细胞生存状态的信号, 该信号可以以细胞间接接触的形式, (例如通过改造 Notch-delta 通路, 工程化的 synNotch 系统)^[2,3] 被传递给周围的子细胞, 抑或以类似于群体感应的, 长距离细胞通讯的形式, 生成特定的“存活信号梯度”, 来维持“子细胞”的生存状态^[4]。

对于扩散性的信号系统而言, 信号分子的浓度梯度与信号强度由母细胞生成信号分子的速率, 信号分子的局部扩散常数 D , 信号分子在扩散过程中的局部衰变速率 R 共同决定。



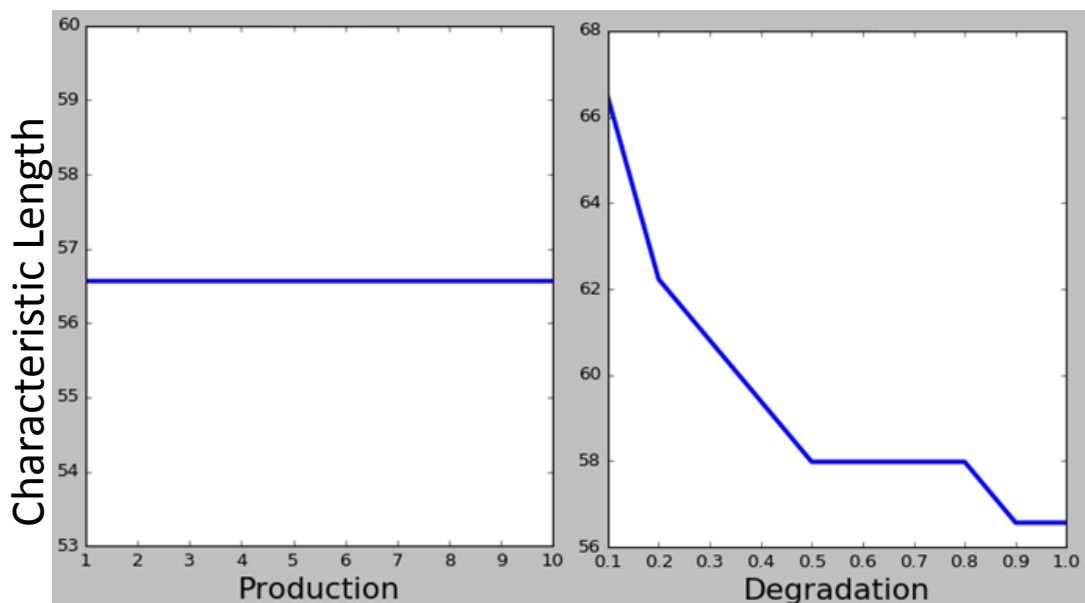


图 1-3: 扩散示意图（上），扩散模拟（左）和扩散方程（右）与特征长度随生产速率和降解速率的变化（下）（retrieved from Katherine, modeled by Morpheus）

上图是利用 Morpheus 模拟在不同的扩散速率以及衰减速率情况下，在达到稳定状态后信号分子浓度的梯度以及强度。通过分析模拟可以看出，当信号分子的生成速率不变的情况下，扩散速度越快，降解速度越慢，信号分子的扩散范围越广。反之当信号分子扩散速度很慢，而降解速度很快的话，则只能扩散有限的距离。当信号的生成速率提高时，整体的信号强度都将提高，但是特征长度（当浓度达到平均浓度时，距离信号源的距离）与生产速率无关，反而与降解速率呈负相关。这表明，当信号降解速率提高时，信号浓度将在更短距离内达到平均值。

- (2) 细胞周期抑制或细胞程序性死亡：在原核生物中已有利用群体感应系统接应自杀开关来调控群体数量震荡变化的实例，而相关的死亡开关在真核生物中的研究尚缺。在真核细胞中，有工程化核糖体开关，构建诱导性基因表达回路将细胞限制在特定细胞时期的研究。也有通过细胞周期蛋白将细胞捕获于某一特定细胞时期的研究。在这里，我们需要构建基于真核细胞的死亡开关回路诱导细胞程序性死亡，抑或抑制细胞的正常分裂。

1.2.3 合成生物学实现

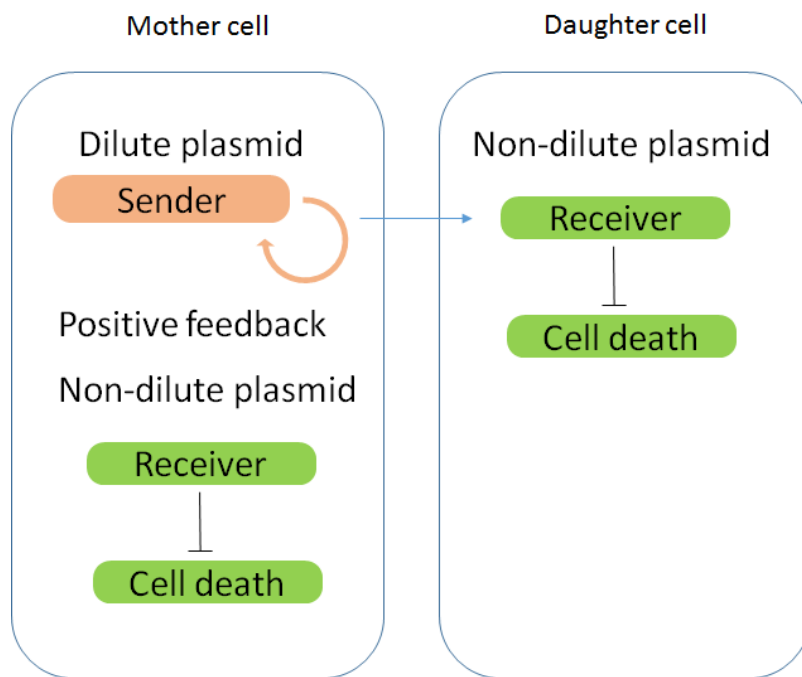


图 1-4：合成生物学实现设计概述

该设计的核心是，设计两种不同的质粒来分别搭载两套基因回路，利用质粒是否能够在细胞中复制的特点，制造母细胞与子细胞的基因表达差异。

母细胞同时携带不具有复制特性的稀释性的质粒以及能够随细胞周期一同复制的非稀释性质粒。当细胞仅含有一个单一的稀释性质粒时，在下次细胞分裂的过程中，该单个质粒，会随机进入分裂出的其中一个细胞中；而非稀释性质粒则会伴随细胞周期一同复制，平均分配至两个细胞中。因此同时包含稀释性质粒与非稀释性质粒的细胞，将具有母细胞的特性，而仅包含非稀释性质粒的细胞，将具有子细胞的特性。

稀释性质粒将携带细胞通讯系统中的“发射侧”装置，负责转录翻译催化合成信号小分子的蛋白质，或者合成对应的 synnotch 信号（分别对应细胞长距离通讯和细胞接触两种不同类型的信号）。而非稀释性质粒将携带对应的“接收”装置，一旦接收到的信号强度低于一定阈值，将会启动下游的细胞程序性死亡系统。

考虑到从单个质粒产生的蛋白质将难以检测，在设计实验阶段可以在稀释性质粒上引入正反馈回路，来扩大单个质粒造成的基因表达差异。抑或可以利用蛋白质-DNA 相互作用和蛋白质的多聚化作用，将多个稀释性质粒交联在一起。在原核生物中，利用相关原理的非对称分裂已经实现^[5]，而在真核生物中还未有过相关的尝试。

1.2.4 重要性与创新性：

先前已有相关利用群体感应系统控制细胞群体数量的实例^[4]，但是群体感应系统倾向于形成均一的，各向同性的信号浓度分布，而该系统则能被用以构建形成信号分子的浓度梯度，来控制群体数量。在使用群体感应系统构建各向同性的信号浓度的模型中，信号的绝对强度作为单一变量，调控整体细胞的行为，所有细胞的行为受统一的信号强度调控，呈现“0 或 1”式行为，而在此模型中所形成的“同心圆”式浓度梯度模型，则将信号的浓度与空间信息相整

合，为后续基因回路的接入提供广阔的参数空间，并且为之后编码复杂的生物结构，例如在不同的信号强度下诱导不同的基因表达等行为，提供了更多的自由度。

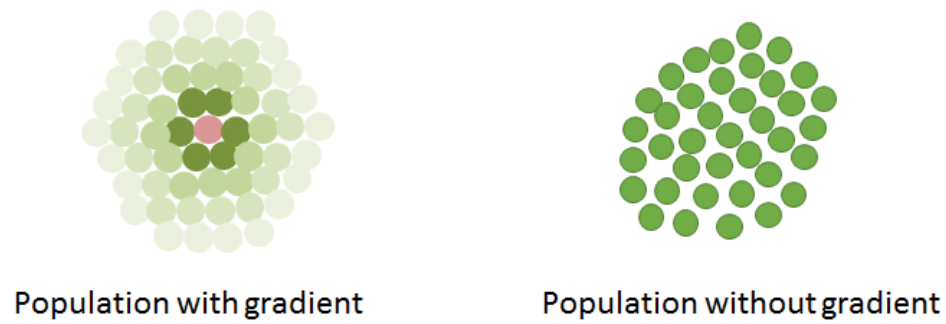


图 1-5 基于非对称分裂形成的信号浓度梯度与基于群体感应形成的均一性信号强度示意图

第二章 模拟，数值验证与实验

2.1 利用 Morpheus 进行建模与数值模拟

2.1.1 建模与 Morpheus 介绍

对单细胞生物过程的研究，以细胞内的调控-转录-翻译过程为主，通常是利用微分方程建立数学模型，将生物系统的动力学和热力学参数代入模型进行数值模拟。在单细胞尺度上，这些生物过程的时间尺度往往从数秒（转录，翻译速度）至数小时（真核生物蛋白质半衰期）不等，空间尺度则从蛋白质的纳米级别至真核生物的数十数百微米不等。

然而，对多细胞体系进行建模则相当不同，从时间与空间尺度来讲，多细胞过程的空间尺度往往远高于数微米，根据细胞数量不同从微米级别到毫米即便，而其形态发生的时间尺度则往往要以小时计数。在模型的构建上，更多的细胞间相互作用例如细胞间通讯系统，细胞间力学相互作用等因素需要被纳入考虑。

研究尺度的变化，同时也导致了我们的研究与工程对象的变化。在单细胞尺度下，研究者感兴趣的是基因作为基础元件对整体网络的影响，例如如何调谐基因回路中的启动子强度使得输出蛋白质表达量能够满足下游基因回路的调控需求。而在多细胞体系中，我们不再关注细胞内部的基因调控过程，而更着重于这种过程如何在细胞界面和细胞的时空尺度上发挥作用，例如细胞间的相互黏连，抑或细胞间直接接触通讯，细胞间长距离通讯等过程。这些细胞界面的性质，而非细胞内部具体过程的性质，直接决定了细胞群体所形成的不同特征的多细胞结构。

Morpheus 作为一种可模拟这种多细胞生物过程的软件，同时拥有模拟细胞内部生化反应过程，模拟细胞界面细胞间相互作用过程以及模拟分子扩散过程的能力。使用该软件，我们可以通过对单细胞内部的生物过程进行设计，来模拟单细胞演化形成具有固定大小的细胞群体的生物过程。

2.1.2 动力学参数确定和时间尺度分离 (time scale separation):

Morpheus 使用数值法模拟微分方程随时间的演化，因此当同时处理多个细胞体系中数量众多的微分方程时，会对每个细胞内部的生物过程分别模拟，大大提高了计算量和模拟所需的时间，难以得出模拟结果，因此应用一定的假设来简化模拟的过程是十分必要的。

在这里，我将论证细胞内部的基因调控-转录-翻译过程的时间尺度，将其与细胞周期的时间尺度相比较，尝试提出以下假设：对于多细胞过程的时间尺度而言，假设相关基因的表达水平一直处于稳态抑或迅速达到激活所需的阈值是合理的。

通过查阅相关资料^[6]，并且计算相关的生物学尺度，利用时间尺度分离的方法来说明该假设的合理性：

哺乳动物细胞的大小根据细胞类型不同从数十微米到数千微米长度不等。细胞内蛋白质等生物大分子的扩散常数约为 $10 \text{ } \mu\text{m}^2 / \text{s}$ ，因此根据扩散方程，我们可以得到扩散的时间尺度约

从数十秒（对于数十微米的细胞）至数小时（对于数千微米的细胞）不等。实验所使用的细胞（HEK293）的直径约为 14 微米，因此假设其扩散尺度在数十秒内是相当合理的。

对于小分子而言，其扩散常数约为 $1000 \text{ } \mu\text{m}^2 / \text{s}$ ，因此对于 HEK293 细胞系，其可以在一秒内完成扩散。

$$time = \frac{length\ scale^2}{Diffusion\ constant} \quad (2-1)$$

真核生物内部的转录和翻译速率尺度约为 $10 \text{ nt} / \text{s}$ 和 $10 \text{ aa} / \text{s}$ ，平均蛋白质大小约为 300aa，因此我们可以合理地假设蛋白质的生产速率时间尺度约为数分钟，HEK293 细胞的体积约为 $1000\mu\text{m}^3$ ，因此一个蛋白质所对应的分子内浓度约为：

$$\begin{aligned} & \text{Concentration for 1 molecule} \\ &= \frac{1 \text{ molecule}}{10^{23} \text{ molecule} \times 1000\mu\text{m}^3 \times \frac{10^{-18}\mu\text{m}^3}{\text{m}^3} \times \frac{10^3\text{L}}{\text{m}^3}} \\ &= 10^{-11}\text{M} = 10^{-2}\text{nM} \end{aligned} \quad (2-2)$$

而细胞内蛋白质的有效浓度约为 1nM 至 1 μ M，其对应的蛋白质的拷贝数为 100 至 10000。倘若 DNA 拷贝数为 1，而一个 DNA 同一时间只能转录一个 mRNA，mRNA 也只能同时产生一个蛋白质，则其时间尺度约为 5min 至 8h，而考虑到 DNA 与 mRNA 可能同时存在多个拷贝，而 mRNA 也能够合成多个蛋白质，因此其实际时间尺度可能会更小，这表明我们目标蛋白质的浓度可以在很短的时间内达到其效应浓度。

哺乳动物细胞中，RNA 与蛋白质的降解速率相较于其合成速率相对较慢。以 HeLa 细胞系为例，其 mRNA 的半衰期约在十小时左右内，而蛋白质的半衰期则更长，约为三十小时左右。从之前的计算我们可以发现，蛋白质合成的速度很快，因此相对较长的半衰期表面其达到稳态后的表达量会相当高，即便缓慢的降解速率使得蛋白质需要更长的时间到达稳态。

尽管蛋白质降解速率的限制导致其到达稳态所需要的时间与细胞分裂的时间尺度相近，但是倘若我们考察“蛋白质达到效应浓度所需要的时间”，该过程仅受限于转录-翻译速率，而该时间尺度远小于细胞分裂所需的时间，因此假设其在细胞分裂前就能够迅速达到效应浓度也是合理的。

因此，我们可以在模拟时，对基因调控表达的生物过程与细胞的分裂过程进行时间尺度分离的处理：当我们考虑细胞的分裂与死亡时，假设蛋白质和小分子的扩散和蛋白质的合成已达到效应浓度，因此直接调用稳态时蛋白质分子和小分子的效应浓度，而不再单独对其反应过程进行述职模拟。

对于细胞的程序性凋亡而言，细胞自噬的激活将在 30 分钟内发生，而程序性死亡的时间尺度，根据情况不同，从数小时至 48 小时不等。细胞自噬的时间尺度与分裂的时间尺度相似，但是目前的模型尚且无法模拟细胞死亡后对周围环境的影响。因此这可能导致最后的模拟结果与实际实验产生明显的差别。

2. 1. 3 Morpheus 过程:

该程序所模拟的生物学过程，是从单一或多个“母细胞”开始，该母细胞的每次分裂，将会形成两个具有不同性质的细胞，其中一个将保持原有“母细胞”的性质，而另一个细胞拥有与母细胞不同的“子细胞”的属性。对于之后的每次分裂，母细胞会通过不对称分裂产生一个母细胞和一个子细胞，而子细胞则只能均等分裂为两个相同的子细胞。

母细胞携带有产生长程细胞通讯元件的质粒，其转录翻译而成的蛋白质，将合成特定的信号小分子，信号分子通过扩散作用向周围传递，在 2D 模拟过程中，扩散作用与小分子的降解速率共同决定了信号分子的浓度梯度。子细胞不能产生该信号小分子，只能检测信号小分子的浓度，倘若小分子在子细胞处的局部浓度小于一定阈值，细胞程序性凋亡就会被激活。

在不考虑对改造细胞间相互作用力的情况下，我们设置母细胞与子细胞之间的机械力相互作用强度和子细胞之间的相互作用强度相同。

模拟结果:

小分子信号浓度和细胞间接触两种不同的信号传导系统被分别模拟，来预测细胞的图案形成。模拟过程经历了足够的时间，来保证细胞群体数量达到稳定水平。在小分子信号模拟过程中，通过调节细胞对小分子信号的敏感性，可以控制细胞死亡开关的响应阈值，从而调控形成的细胞群体数量。

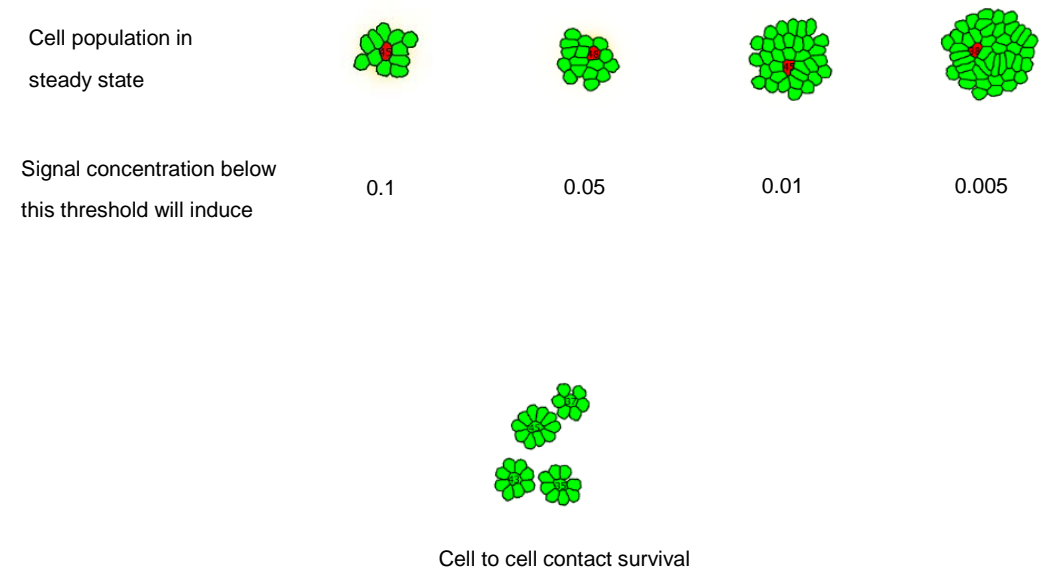


图 2-1：不同细胞死亡阈值产生不同大小的细胞群体（上）以及通过细胞间接触信号形成的四个小细胞群体（数字示母细胞分裂次数）

结果表明，当细胞的死亡阈值降低时，较低的信号浓度便足以维持子细胞的生存，抑制死亡开关，而细胞群体的数量相较于较高的死亡阈值也会有所增加。如果通过同时提高信号分子的合成速率和表观扩散速率来增加局部信号浓度，在稳定状态下也会产生更大规模的细胞群体。在模拟过程中，每个细胞分裂的方向是随机的，母细胞会是分裂而来的两个细胞中的仍以一个。由于这种分裂的随机性，在模拟的初始过程中，母细胞与子细胞的相对位置并不是固定的，但是当群落到达稳定状态时，由于细胞凋亡的作用，最终将形成以母细胞为中心，子细胞环绕四周的图案。

对于细胞间接触信号的模拟过程，只有当子细胞与任何一个母细胞接触时才能存活。最终细胞群落将形成相对固定的双层“同心圆”式结构，其到达稳态的速率以及模式的相对稳定性皆优于信号浓度梯度诱导图案形成的过程。然而目前这种设计只适用于形成双层同心圆结构群体，并不能任意调节细胞群落的大小。

Morpheus 模拟过程的视频请参见附录。

2.1.4 讨论与反思：

在模拟过程中我们引入了数个假设，根据实际情况不同这些假设有可能并不适用于所有情况，因而造成模拟结果与实际的实验结果有所出入，因此我们需要在此对这些假设加以讨论：

（1）在模拟过程中，细胞一旦死亡就会被立刻从当前位置清除，并且不会占据原先空间，或者对周围的细胞或环境造成影响，但实际上细胞凋亡需要数小时至数十小时，并且其死亡后产物也有可能对周围细胞的继续培养产生不可预知的影响。

这种影响可能可以通过精巧的结构设计来避免，例如，相较于在封闭容器中培养细胞，使用连续液流培养基培养细胞可以尽快移除细胞凋亡产物，或者通过同时上调多个凋亡通路的表达来加快细胞凋亡的速度。

（3）信号浓度梯度的形成是由信号分子的扩散和降解共同决定的，只有当信号的浓度梯度足够大时（ $\mu\text{M}/\mu\text{m}$ ）才能将细胞群落的大小控制在可观测的范围内。因此，当信号小分子扩散速率较快，而降解速率较慢的情况下，信号浓度的下降梯度较小，从而导致需要很远的距离才能衰减到效应浓度，在这种情况下，信号分子诱导的下游基因表达在短距离内可能表现为各项同性，导致结果不明显。

分子的扩散常数 D 由 Stocks-Einstein 方程描述：

$$D = \frac{kT}{6\pi R\mu} \quad (2-3)$$

其中 kT 为热力学扰动， R 为分子半径， μ 为液体的粘滞系数

一种解决办法是选择扩散速率较慢的信号分子，例如蛋白质大分子等，通过增加半径 R 减小扩散常数；在不改变小分子类别的情况下，也可以通过改变介质的粘滞系数来改变扩散速率。

另一方面，根据先前的讨论，提高小分子的降解速率，例如表达降解该种小分子的酶，也能够有效减小特征长度。

2.2 实验设计与验证：

由于时间和条件的限制，本文基于先前的研究，对基于哺乳动物细胞的群体感应系统进行测试，同时尝试构建稀释性与非稀释性质粒系统，并且对实验结果进行定量分析。

2.2.1 细胞间长距离通讯系统测试：

原核生物的长距离通讯系统例如群体感应系统，然而到目前为止，可以被运用于哺乳动物细胞的正交性长距离通讯系统尚未发表。根据先前的初步研究，研究人员发现，植物能够合成一种名为 phloretin 的小分子，该分子在自然界起到广泛的抑菌作用。而部分微生物如 *Pseudomonas putida* 则已经协同演化出了一套响应 phloretin 的外排系统，当 phloretin 存在时，受体 TtgR 能够响应其信号，上调通道蛋白 TtgABC 的表达，将该小分子排出体外^[8]。

研究人员通过广泛的筛选，发现利用 *Aradopsis thaliana* 中的 4-钴酸盐：辅酶 A (CoA) 连接酶 (4CL1) 和 *Hypericum androsaemum* 中的查尔酮合成酶 CHS，可以将 phloretic acid 催化为 phloretin。

而 *Pseudomonas putida* 中的 TtgR 受体则被克隆出来，用以构建依赖于 phloretin 的激活/抑制系统。当 Phloretin 不存在时，二聚化的 TtgR-dimer 会结合于特定的 DNA 序列，而当 Phloretin 与 TtgR 结合时，TtgR-dimer 复合物将被打开，TtgR 从 DNA 位点脱离。若将该 DNA 元件置于核心启动子的保守序列之间，则 TtgR 与 DNA 元件的结合将抑制启动子启动下游基因表达，而 Phloretin 的加入会解离 TtgR-dimer-DNA 复合物，从而激活下游基因表达，而若将 DNA 元件置于核心启动子之前，并且将 TtgR 和转录激活结构域结合，则能够 TtgR 与 DNA 的结合能够激活下游基因表达，而 phloretin 的加入则会抑制下游基因表达^[9]。

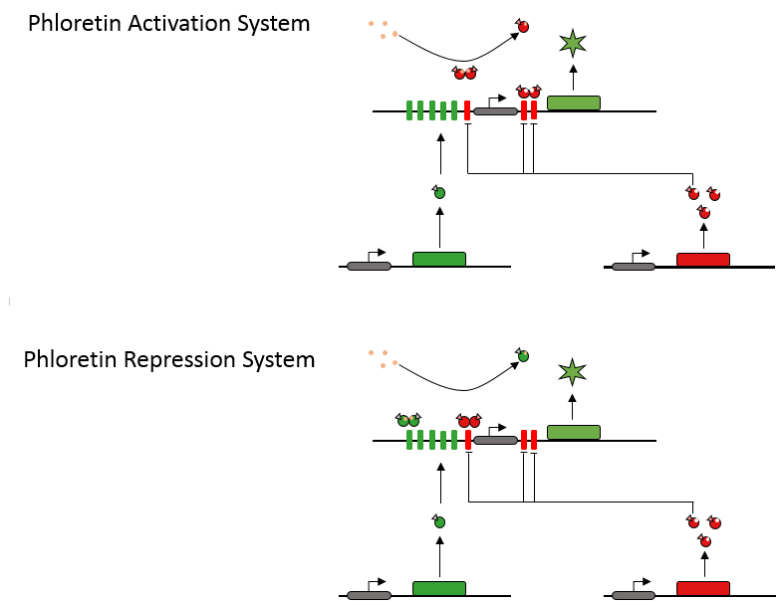


图 2-3: Phloretin 依赖的激活系统与抑制系统

在后续实验中，我使用依赖于 phloretin 的激活系统，研究报告基因随 phloretin 剂量的诱导曲线。实验所使用到的四种质粒构建示意图如下图所示：

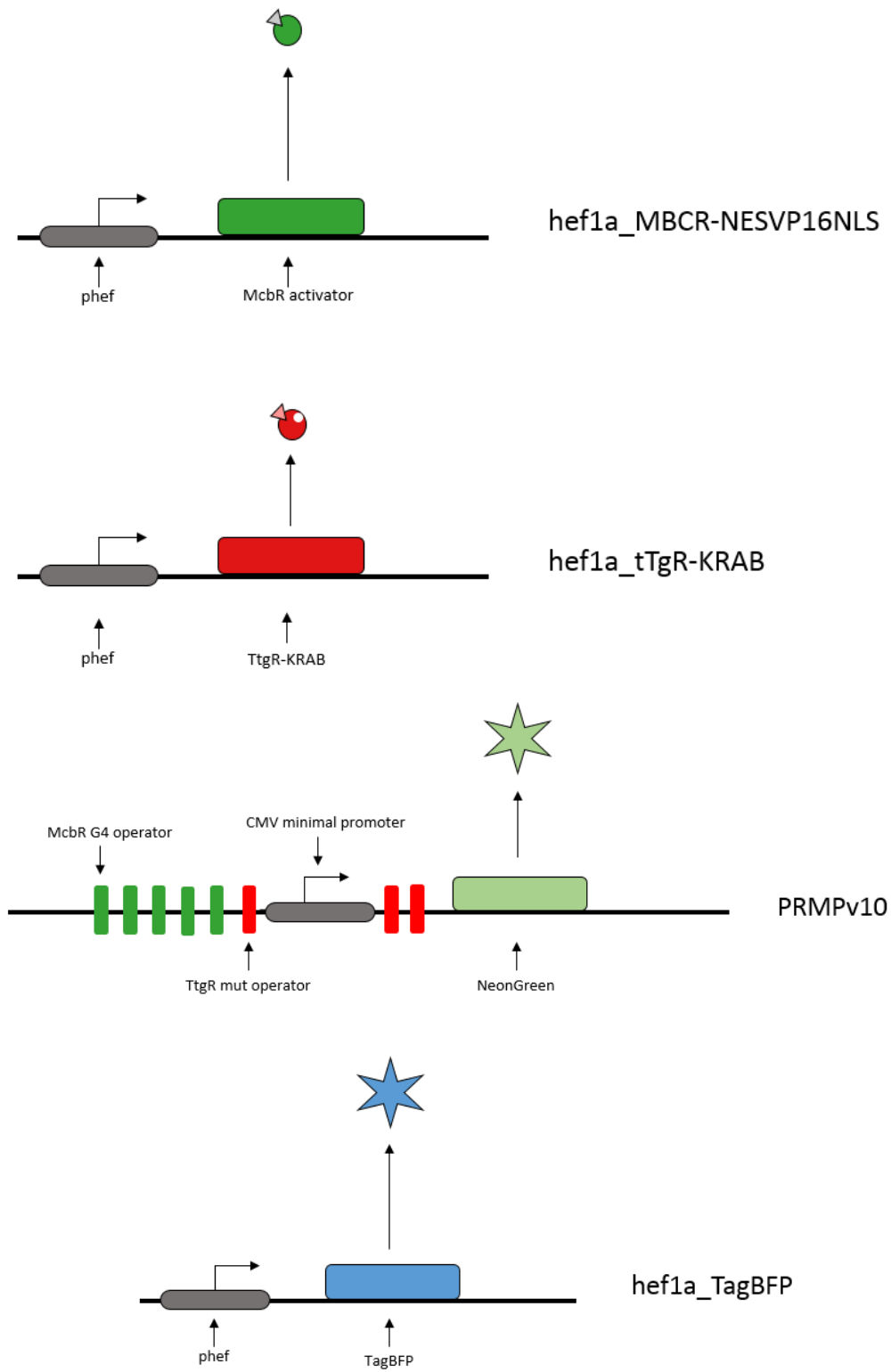


图 2-4: 激活元件 (hef1a_MBCR-NESVP16NLS), 抑制元件 (hef1a_tTgR-KRAB), phloretin 响应元件 (PRMPv10, phloretin responsive mammalian promoter version 10) 与转录标记 (hef1a_TagBFP) 质粒构建示意图

其中, hef1a 为组成型表达启动子, McbR 的存在可以与 McbR G4 operator 结合, 其 VP16 结构域可以增强下游启动子的表达, TtgR 在不存在 phloretin 的时候, 会二聚化并且与 TtgR mut operator 结合, 阻遏转录起始, 抑制基因表达。当 phloretin 分子存在时, TtgR 会从 operator 上解离, 下游基因转录起始。TagBFP 为共转染报告基因, 指示转染效率

在这里, 我对着四种质粒进行共转染, 调整激活与抑制元件的不同比例, 测试不同转染效率下的响应强度。不同比例的激活和抑制基因根据下表被共转录, 培养 24 小时后, 加入不同浓度 (0uM, 100uM) 的 Phloretin 信号分子, 继续培养 24h 后, 使用 FACs 收集细胞的荧光强度信号。

表 2-1: 共转录 DNA 实验

Total DNA	800ng	800ng	800ng	800ng	x = 100 ng
Samples	6:PRMPv10	7:PRMPv10-Pos	8:PRMPv10-Neg	9:PRMPv10-NegL	
pKK593	x	x	x	x	hef1a_TagBFP
pKK365	x	x	x	x	PRMPv10_neonnls
pKK328	x	x			hef1a_MBCR-NESVP16NLS
pKK054	x		x		hef1a_tTgR-KRAB
Filler	4x	5x	5x	6x	

Total DNA	800ng	800ng	
Samples	10:PRMPv10	11:PRMPv10-Neg	
pKK593	x	x	hef1a_TagBFP
pKK365	x	x	PRMPv10_neonnls
pKK328	x		hef1a_MBCR-NESVP16NLS
pKK054	2x	2x	hef1a_tTgR-KRAB
Filler	3x	4x	

使用 cytoflow 对收集到的数据进行分析。利用 FSC 与 SSC 筛选出单细胞群体, 利用单色对照矫正 bleedthrough, 转录效率由 TagBFP 的表达量表征, 计算不同转染效率下对应输出荧光强度的平均值与标准差, 绘制输出荧光强度对转录效率的折线图。

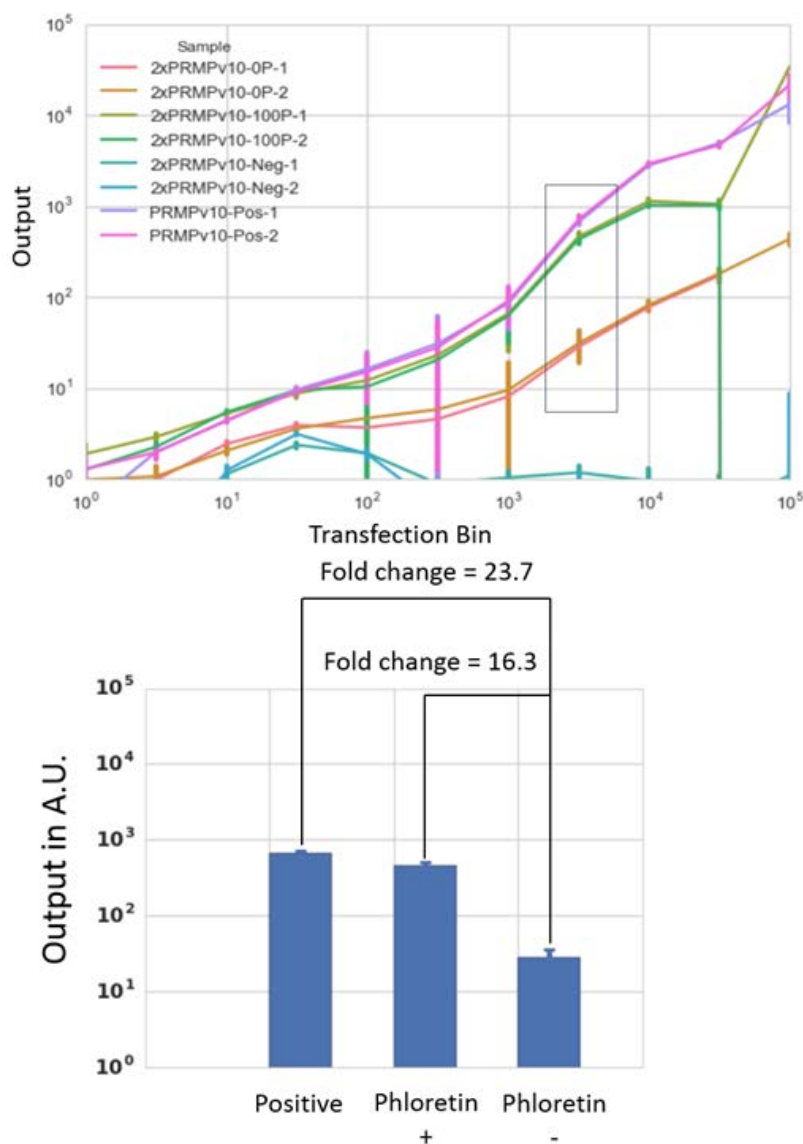


图 2-5：输出荧光蛋白表达量对转录效率折线图（2xPRMP 指两倍浓度抑制元件）（上）与最大响应倍数（下）

数据分析表明，当相对转录效率为 30000 时，相较于不加入 phloretin 而言，100uM 浓度 phloretin 的刺激能够产生高达 16.3 倍的 phloretin 响应。而阳性对照（不加入抑制元件）表明响应倍数上限约为 23.7 倍。

讨论：

Phloretin 作为信号小分子，扩散速率相当快，因此往往不容易形成有效的浓度梯度，后续相关实验表明，混合 phloretin 的生成细胞与响应细胞所形成的细胞群体，难以在空间上观测到显著的差异性蛋白质表达。根据先前的分析，一方面我们可以在之后寻找降解 Phloretin 的蛋白质来降解 Phloretin 的局部浓度，抑或是寻找相关的与 Phloretin 结合的蛋白质或物质，降低其表观扩散速率。而另一方面，在实验装置的设计上，使用具有连续液流的微流控系统，有助于稀释多于的信号小分子，协助浓度梯度的形成。

2.2.2 非稀释质粒和稀释质粒系统构建：

在一般哺乳动物细胞的瞬时转染实验中，质粒往往不携带与哺乳动物细胞复制机器相关的复制原点，因此其质粒拷贝数会随着细胞的分裂而逐渐减少。但在自然界中存在相关的哺乳动物病毒，携带有能利用哺乳动物细胞复制机器复制其复制原点的能力。SV-40 系统质粒复制系统就是一套能够在哺乳动物细胞中自主复制的质粒系统，其主要组成部分为 SV40 大 T 抗原（large T antigen）和 SV40 复制原点^[10]。大 T 抗原一方面抑制细胞周期蛋白如 p53 和 Rb，促进细胞进入细胞周期，另一方面招募 DNA 复制相关元件至 SV40 复制原点，起始质粒的复制^[11]。

SV40

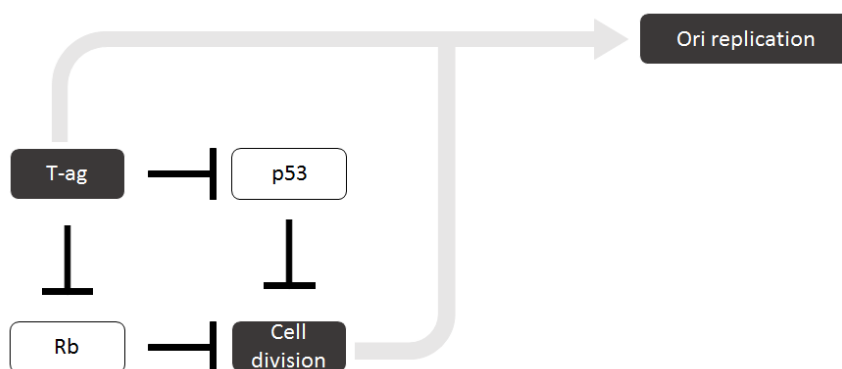


图 2-6：SV40 大 T 抗原协助质粒复制机制示意图

HEK293FT 细胞系是一种组成型表达 SV40 large T antigen 的细胞系，而此处我所使用的细胞系已将 SV40 Large T antigen 的表达系统整合入基因组。因此理论上当我们使用含有 SV40 复制原点的质粒进行瞬时转染时，该细胞系表达的大 T 抗原会协助相关质粒的复制，减少其由于细胞分裂而产生的稀释效应。如果同时使用含有 SV40 复制原点和不含 SV40 复制原点的质粒进行瞬时转染，理论上含有 SV40 复制原点的质粒的稀释速率会远小于不含 SV40 复制原点的质粒的稀释速率。

在本节中，我同时使用含有 SV40 复制原点和不含 SV40 复制原点的质粒对该细胞系进行瞬时转染，含有 SV40 复制原点的质粒携带有 CMV 驱动的 EGFP 荧光标记，而不含 SV40 复制原点的质粒则携带有 CMV 驱动的 EBFP 荧光标记。

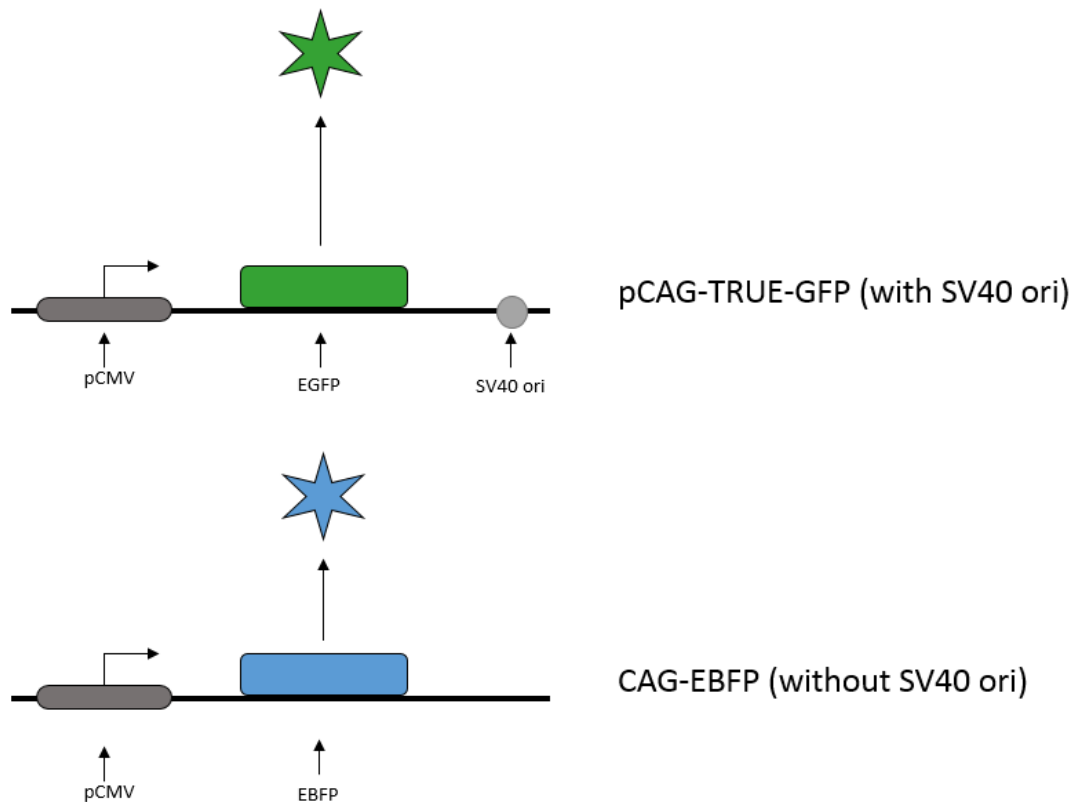


图 2-7：包含 SV40ori 的 EGFP 表达系统与不包含 SV40 ori 的 EBFP 表达系统

我们分别在 0h, 48h, 96h, 132h 在 96 孔板上对细胞进行了瞬时转染，并且在 144h 收集所有细胞，进行流式细胞仪荧光信号分析。同时分别含有这两种质粒，以及均不含两种质粒的细胞也一同被转染，作为实验的单色对照和空白对照。为了避免细胞生长过密阻碍对质粒稀释的观测，在 0h 和 48h 转染的细胞，分别在 48h 和 96h 处转移进 24 孔板上培养。

表 2-2：实验组与控制组设计

Sum	125	125	125	0	0	Plasmid Description pKK202(pCAG-TRUE-GFP) CAG-EBFP BLANK (Cag-FALSE pKK203)
Plasmid Number	1:DNA X	green	blue	Beads	Blank	
pBW361	25	25	0	0	0	
NA	25	0	25	0	0	
pBW363	75	100	100	0	125	
				Beads		

单个细胞的荧光强度通过流式细胞仪进行收集，利用 cytoflow 进行单细胞筛选和荧光泄露矫正，得到实验组和控制组的细胞数量随荧光蛋白表达强度的分布。通过统计大于一定阈值（此处设置 100 作为荧光强度的检出下限）的细胞占总细胞群体的比例来表征质粒在群体中的分布。而稀释速率则通过比较不同时间的质粒分布得出，例如，48 小时至 96 小时的稀释速率则通过以下公式计算得出：

$$48h - 96h \text{ Dilution ratio} = \frac{\frac{\text{number of cells above threshold for 48h}}{\text{number of cells for 48h}}}{\frac{\text{number of cells above threshold for 96h}}{\text{number of cells for 96h}}} \quad (2-4)$$

96h 至 144h 的稀释速率也通过相似的方式计算得出。该数值越大表明该质粒的稀释速度越快，而数值越小则表明质粒的稀释速度越慢。若稀释速率大于 1，则表明发生了质粒稀释的过程。如果稀释速率小于 1，则可能发生了质粒复制的过程。

为了探究从 96 孔板转移至 24 孔板培养细胞是否会造成质粒在群体中分布的不同，我们比较了在 96 孔板中培养了 96h，和在 96 孔板中培养 48h，再转移至 24 孔板中培养 48h 的两组实验，从数据分布可以看出，24 孔板组 EBFP 表达量在高于 1000 的细胞数量略有下降，而在 EBFP 表达小于 100 处，细胞数量呈现明显的提升，这可能是由于在 24 孔板中更多细胞，在获得了充足的营养后，细胞的分裂速度较快而造成的更快的稀释效率。分析 EGFP 高于阈值的细胞占比可以发现，在 24 孔板中的质粒分布数量呈现明显下降趋势。

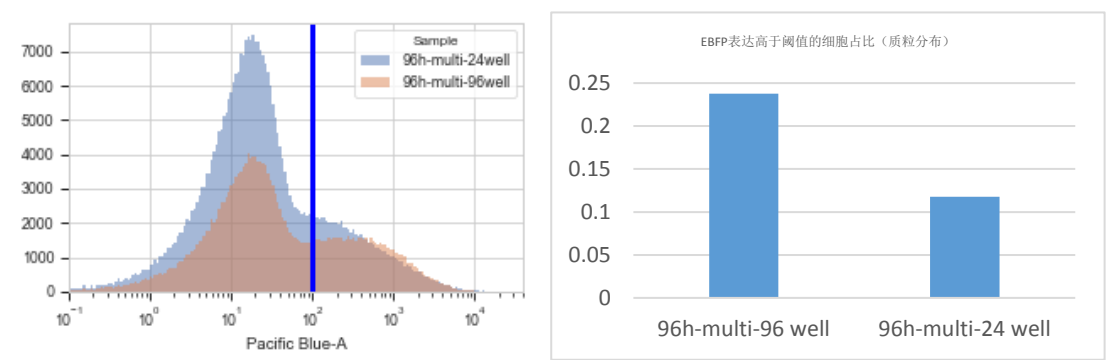


图 2-8：相同培养时间下，EBFP 表达分布

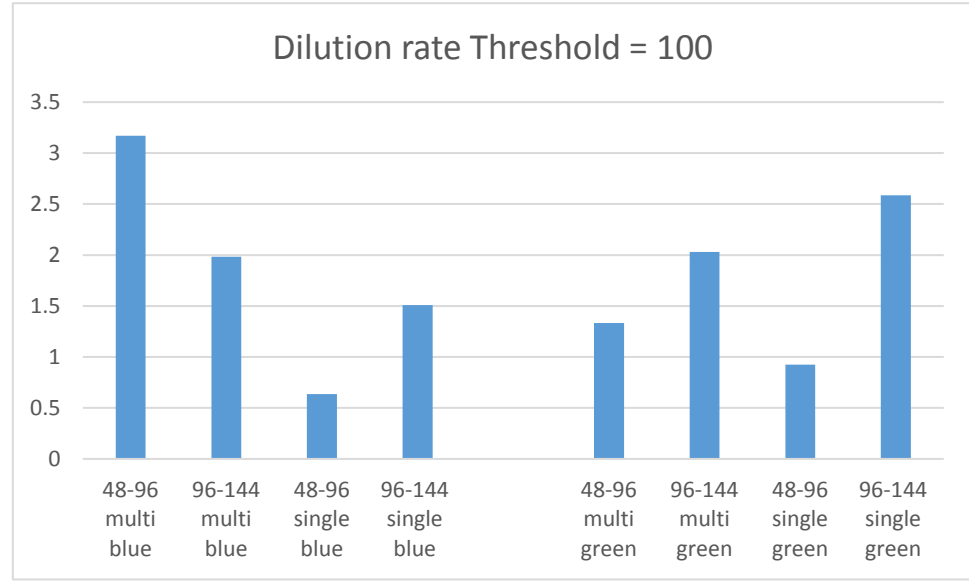


图 2-9：不同组别的稀释速率

根据式 (2-4)，我们分别计算不同组别 EGFP 和 EBFP 的稀释速率，可以发现，对于共转染质粒的实验，EGFP 的稀释速率接近与 1，而 EBFP 在共转染时有明显的质粒稀释现象，这与我们的假说相符。然而对于分别独立转染的细胞系来说，两者的质粒稀释速率都接近于 1。

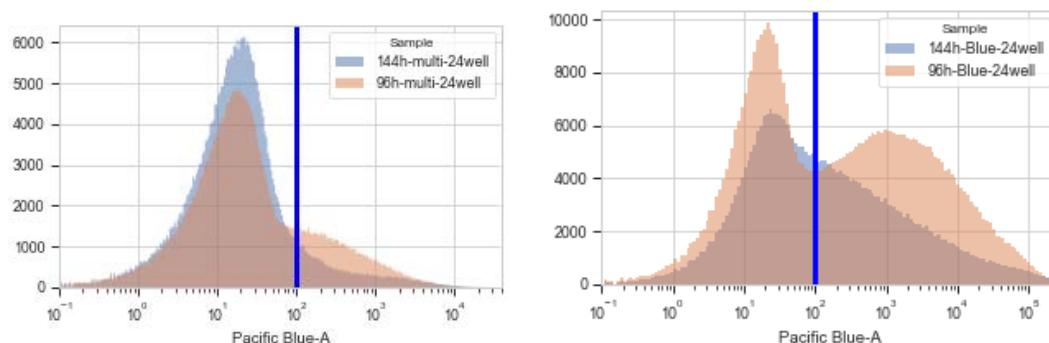


图 2-10：共转染中 EBFP 表达量比单独转染 EBFP 表达量低一至两个数量级

此外，在共转染中的 EBFP（不携带 SV40 复制原点）的表达量相较于单色对照中 EBFP 的表达量低了一至两个数量级。携带 SV40 复制起点的 EGFP 质粒似乎对不携带 SV40 的 EBFP 质粒存在某种上位效应，抑制了 EBFP 的表达量。

讨论：

该实验是通过测量荧光蛋白的表达量来反应质粒拷贝数的变化。实际上，这两者之间存在着一些微妙的不同。蛋白质的稳态浓度由蛋白质的生成和降解速率决定：

$$\text{Amount of protein in steady state} = \frac{\text{production rate}}{\text{degradation rate}} \quad (2-5)$$

而蛋白质的生成速率又由表达质粒的拷贝数，启动子的强度，转录和翻译速率等共同决定。在其他条件不变，并且转录-翻译分子机器尚未饱和的情况下，增加质粒的拷贝数往往能够增加蛋白质的稳态表达量。然而如前文所述，蛋白质的生成速率相对较快，而降解速率则相对较慢，因此在质粒拷贝数达到稳态后，还需要经过一定的时间（平均为 24h 左右，具体由蛋白质类别决定），才能够观测到蛋白质表达量达到稳态，因此在对数据在时序上进行分析的时候，需要对蛋白质表达量到达稳态的延迟加以考虑。

实验表明，质粒的蛋白表达约在 48h 左右到达稳态，而对 96h 与 144h 的实验组进行考察，可以观测到一定的质粒稀释效应。

而另一方面，稀释速率又与式（2-4）中阈值的选取有关，因为荧光蛋白表达量的分布并不是均一的，而更接近正态分布，因此在选取的阈值改变时，稀释速率的计算结果会产生变化：

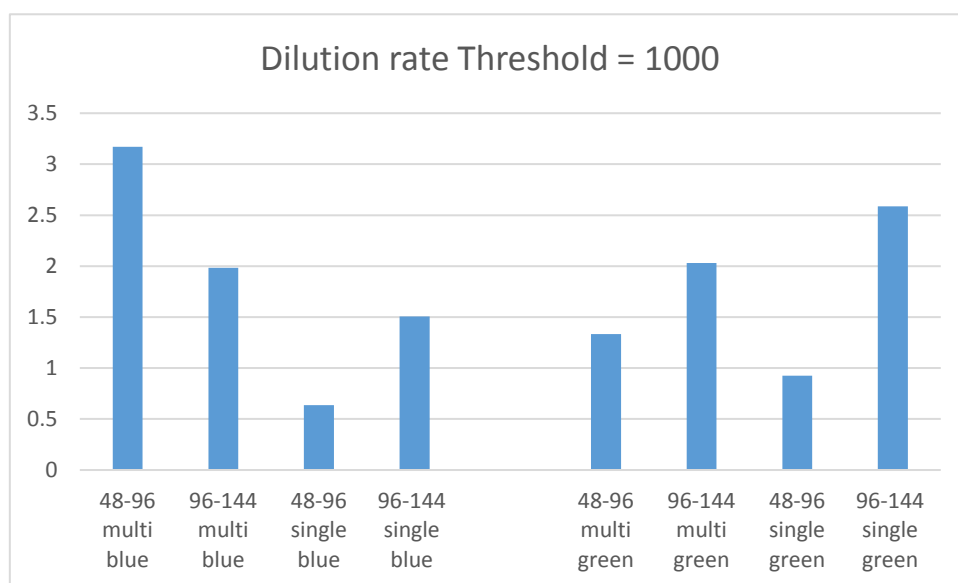


图 2-11：当选取的阈值较高时，可以在 96 小时后观测到明显的质粒稀释现象

但是 100 的阈值选择仍然是一个较为合理的值，因为小于 100 的荧光蛋白表达量相对而言已经低于检出阈值，与细胞的自发荧光处于同一数量级上。

另一方面，当选取较高阈值时，稀释速率的提高变化表明，质粒在高浓度处发生的稀释现象较为明显，倘若给与足够多的时间，其稀释效应最终将投射到低浓度阈值上来。

通过对数据的综合分析，我们可以发现，对于携带有 SV40 复制起点的质粒，我们观测不到明显的质粒复制现象。其可能的原因包括整合入基因组的 SV40 T antigen 因为某些原因被沉默了，因此后续实验可以通过免疫组化或者 qPCR 等方式鉴定该蛋白质是否在此细胞系中正常表达。表达 EGFP 的质粒对表达 EBFP 的质粒展现出的上位效应，尚不清楚是因为 SV40 复制起点造成的，还是由于两个质粒上其他的不同元件所造成的，因此后续需要在原质粒的基础上，交换其对应的复制原点重复试验，以验证是否是 SV40 复制原点造成的偏差。

第三章 反思与讨论

本文从传统的合成生物学手段出发，以构建具有一定种群数量的细胞球体作为工程案例，通过理性化的基因回路设计以及参数分析，利用 Morpheus 等相关软件对细胞的生长与死亡，信号分子的扩散等过程进行模拟，最后基于前人的工作，对细胞间长距离通讯系统以及非稀释性与稀释性质粒系统进行测试。由于时间所限，尚未能够对所有的设计细节，例如如何工程哺乳动物的程序性死亡，如何在哺乳动物中编码质粒聚集系统与正反馈回路等一一进行讨论，已经完成的工作和实验数据又提出了新的问题和挑战，但是这份工作给出了一个完整的工程化范式：如何从一个特定的工程目标出发，理性设计完成该目标所需要的各个部分，再通过建模与实验加以验证。希望能够在之后的工作中我能够有机会对该设计的细节加以完善，也希望这篇文章能够发挥一些作用，为后来的研究者提供一些参考。

参考文献：

- [1]: Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), 335.
- [2]: Toda, S., Blauch, L. R., Tang, S. K., Morsut, L., & Lim, W. A. (2018). Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science*, 361(6398), 156-162.
- [3]: Morsut, L., Roybal, K. T., Xiong, X., Gordley, R. M., Coyle, S. M., Thomson, M., & Lim, W. A. (2016). Engineering customized cell sensing and response behaviors using synthetic notch receptors. *Cell*, 164(4), 780-791.
- [4]: Din, M. O., Danino, T., Prindle, A., Skalak, M., Selimkhanov, J., Allen, K., ... & Hasty, J. (2016). Synchronized cycles of bacterial lysis for in vivo delivery. *Nature*, 536(7614), 81.
- [5]: Sara Molinari, David L. Shis, James Chappell, Oleg A. Igoshin, Matthew R. Bennett. (2018). Synthetic pluripotent bacterial stem cells. Retrieved from <https://www.biorxiv.org/content/10.1101/436535v2>
- [6]: Cell biology by the numbers, Ron Milo, Rob Philips. Retrieved from <http://book.bionumbers.org/>
- [7]: MacDonald, R. E., & Bishop, C. J. (1952). Phloretin: an antibacterial substance obtained from apple leaves. *Canadian Journal of Botany*, 30(4), 486-489.
- [8]: Terán, W., Felipe, A., Segura, A., Rojas, A., Ramos, J. L., & Gallegos, M. T. (2003). Antibiotic-dependent induction of *Pseudomonas putida* DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. *Antimicrobial agents and chemotherapy*, 47(10), 3067-3072.
- [9]: XAVIER DUPOURTET, “Developing new tools and platforms for mammalian synthetic biology: from the assembly and chromosomal integration of complex DNA circuits to the engineering of artificial intercellular communication systems” PhD diss., INRIA Rocquencourt and Massachusetts Institute of Technology, 14 November 2014.
- [10]: Mahon, M. J. (2011). Vectors bicistronically linking a gene of interest to the SV40 large T antigen in combination with the SV40 origin of replication enhance transient protein expression and luciferase reporter activity. *BioTechniques*, 51(2), 119-126.
- [11]: Fanning, E., & Zhao, K. (2009). SV40 DNA replication: from the A gene to a nanomachine. *Virology*, 384(2), 352-359.

致谢：

感谢 Ron Weiss 教授为我的毕业设计提供指导和实验空间，感谢我的 supervisor, Katherine Kiwimagi 博士，如果没有她不厌其烦的耐心指导，愿意从头开始手把手教导实验与数据分析，鼓励我在自己的路上走下去，和她相处的时光是我在 MIT 最快乐的时光之一，如果没有她的悉心教导，那么我连这篇论文中的一点点小结论都不可能取得。感谢 Jesse, Shiva, Ross 为我课题思路的设计提供的帮助，感谢沈颖，张扬和杨静老师，在我论文撰写困难时候的理解和帮助。

除此之外，我还希望能够致谢我的朋友们，感谢俞易，无论是在学术上还是生活上，能够一直理解我支持我，感谢张心瑜，张瑜，沈望尘，王蕙心，王子悦，在我失意的时候愿意倾听我诉说，帮我纾解忧愁，感谢巴方一直以来对我生活上无微不至的照顾。感谢王瑞旋，刘文基和戴鼎原，是他们带给了我无限的快乐。感谢那些所有没有被提及的，曾经在我生活无以为继时帮助过我的人们。

感谢钟桂生老师，一直以来给予我学业上的指导以及在大一就进入实验室学习的机会。感谢他对科研与学生的一腔热诚和无所求的精神，让我在科研与生活上都能有所成长。感谢廖军老师，李健老师，沈伟老师，庄敏老师，钟超老师，对于我生活，学业，科研以及 iGEM 比赛一直以来的指导，我今天的任何成绩，都离不开他们的悉心栽培。

最后我想感谢我的家人和父母，如果没有他们一直以来无私的支持和鼓励，对我无条件的爱，我绝对不可能任性地在一条自己选择的道路上毫无负担地走下去。

附录：

本文所涉及到的 Morpheus 代码与模拟视频，具体 python 代码，实验 protocol 与 notebook 皆已上传到 github，可被公开访问，部分材料如 FACs 原始数据以及质粒的详细构建图谱，由于相关工作尚未发表，暂不公开发布。

备注：Phloretin 系统的筛选由 Xavier Duportet 博士完成，而其瞬时转染质粒的构建则由 Katherine kiwimagi 博士完成