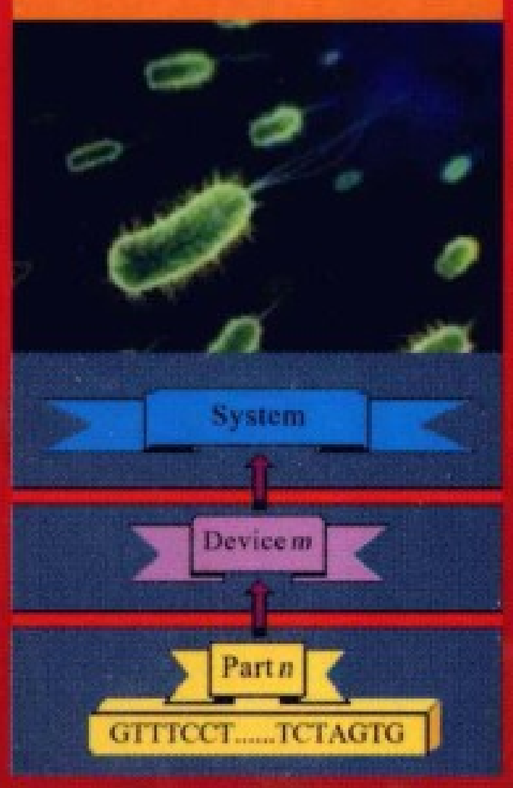
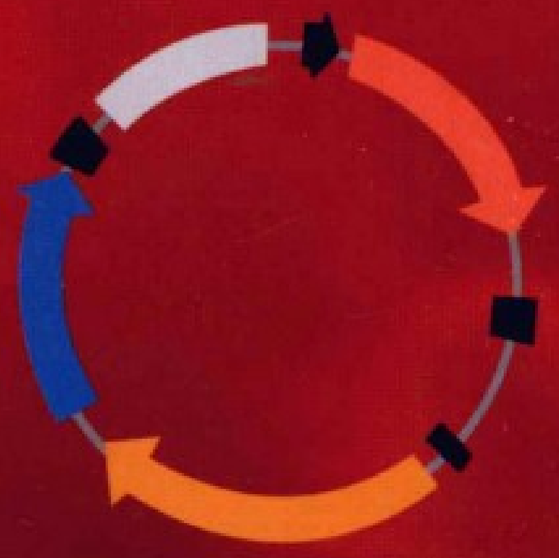
合成生物学导论

Introdution To Synthetic Biology



宋　凯 /编著
黄熙泰 /主审



科学出版社
www.sciencep.com

(Q-2438.0101)

合成生物学导论

Introdution To Synthetic Biology

高等教育出版中心
联系电话：010-64030233
http://www.sciencep.com
e-mail: biology@mail.sciencep.com

销售分类建议：自然科学/生物学/化学/信息科学

ISBN 978-7-03-026731-3
9 787030 267313 >
定 价：36.00 元

-04

合成生物学导论

宋　凯　编著

黄熙泰　主审

天津市科协自然科学学术专著基金资助出版

科学出版社

北京

Q5
5767

内 容 简 介

合成生物学是21世纪的新兴交叉科学，强调“设计”和“重设计”，其目的是通过人工设计和构建自然界中不存在的生物系统来解决能源、材料、健康和环保等问题。合成生物学的发展要以生物学、化学、物理学、数学、信息科学、工程科学、计算机科学等相关学科的发展为基础。

本书共分5章。第1章介绍合成生物学的基本概念和内容，使读者对合成生物学有一个全面的了解。第2章和第3章介绍合成生物学设计和分析的基本方法、合成生物学的层次化结构和概念及合成生物学中的常用算法。第4章和第5章介绍合成生物学在基础和实际应用方面的研究内容、成果和进展，包括经典的双稳态开关、大肠杆菌成像系统、微生物计算机、合成青蒿酸的微生物工厂等。

本书可作为具有一定生物学基础的本科生和研究生教材，也可作为相关领域的学者和研究人员的参考书。

图书在版编目(CIP)数据

合成生物学导论/宋凯编著．—北京：科学出版社，2010
ISBN 978-7-03-026731-3

Ⅰ.①合… Ⅱ.①宋… Ⅲ.①生物合成 Ⅳ.①Q5

中国版本图书馆CIP数据核字（2010）第019670号

责任编辑：单冉东 / 责任校对：李奕萱
责任印制：张克忠 / 封面设计：耕者设计工作室

科学出版社出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
北京天时彩色印刷有限公司印刷
科学出版社发行 各地新华书店经销
*
2010年2月第 一 版 开本：787×1092 1/16
2010年2月第一次印刷 印张：12
印数：1—3 000 字数：282 000

定价：36.00元

（如有印装质量问题，我社负责调换）

序　一

合成生物学是21世纪初出现的一门新兴交叉学科，在这里，生物学、工程学和数学紧密地结合在一起。合成生物学的出现是分子生物学经历了半个多世纪的发展，特别是人类基因组和模式生物基因组计划实施以来，伴随着生物信息学和系统生物学迅速发展的一个合乎逻辑的结果。在人类深入认识生命现象的基础上，合成生物学旨在设计和构建工程化的生物系统，使其能够处理信息、操作化合物、制造材料、生产能源、提供食物、保持和增强人类的健康和改善我们的环境。因此，合成生物学既是一门科学，也是一门技术，它是继遗传工程之后出现的新一代生物技术，代表了生物技术发展的一个新的制高点。2004年美国麻省理工学院主办的世界著名科技杂志 *Technology Review* 曾将合成生物学列为将改变世界的十大新出现技术之一。在2008年以合成生物学为主题的第322次香山科学会议上，与会专家一致肯定了合成生物学对我国未来国民经济发展的重大意义。

无论是我国制定的中长期科技发展规划，或者是国务院颁发的《促进生物产业加快发展的若干政策》，都对生物技术给予极为充分的重视。在中国科学院最近出版的《创新2050：科学技术与中国的未来》系列丛书之一——《科技革命与中国的现代化——关于中国面向2050年科技发展战略的思考》一书中，将人造生命和合成生物学列为22个战略性科技问题之一；同时又将人造生命和合成生物学预测为“可能出现革命性突破的4个基本科学问题”之一。在该丛书的另一分册——《中国至2050年人口健康科技发展路线图》中，将系统生物学和合成生物学技术列为2030年前后若干年间生物医学创新体系中的关键技术。鉴于合成生物学在生物科学技术中的重要地位，在大学生特别是工科大学生中普及合成生物学知识就有十分积极的意义。天津大学是一所以工程学科为主的综合性大学，从2008年开始，我们就给本科生开设了一门“合成生物学导论”的创新课程，共有150名各种专业的大学生听课，反映良好。同学们普遍感到开阔了视野，增长了合成生物学方面的新知识，感受到了最新的科技发展。2009年选修这门课程的大学生已经达到264人。经批准，这门课在天津大学将长期开下去。当前，我国大学课程存在着内容相对陈旧、不能及时反映最新科技发展的缺点，这门课程的开设是我们为改变这一状况所做的一次尝试。

为了适应教学的需要，编写一门合成生物学教材就变得十分必要了。天津大学化工学院宋凯副教授勇于承担了这项任务，经过艰苦的努力，终于撰写了这本《合成生物学导论》教材。全书由南开大学分子生物学家黄熙泰教授认真审阅、订正。就当前版本来看，还存在着篇幅过少、内容不够丰富等缺点，这也与这门学科刚刚兴起有关。我深信，随着合成生物学的迅速发展，在该书的新版本中这些问题将会逐步得到解决。最后

我要指出，由于该书内容涉及分子生物学、控制工程和数学等多门学科，而一个人的知识毕竟是有限的，书中错误之处恐怕在所难免，我衷心盼望有关专家学者不吝批评指正，帮助编著者在该书再版时加以改进。

张春霆

（中国科学院院士）

2009 年 8 月 1 日

序　二

该书是国内外第一部较为完整地描述有关“合成生物学”的书籍。

由于基因组测序及 DNA 合成技术与工具的突破性进展，催生了“组装细胞基因”的“合成生物学”的出现。最近几年合成生物学在生物科学知识产生，生物零件、装置及系统的建造，以及药物、能源、酶的生产等研究领域取得了重要进展，成为生物科学和生物经济中最前沿的学科。“合成生物学”于 2004 年被 *Technology Review* 评为将改变世界的十大新出现的技术之一（10 Emerging Technologies That Will Change Your World），短短几年的实践已证明该预见的正确性。中国在相关技术方面已经有了许多积累，在合成生物学领域与发达国家处在同一起跑线，及时的投入可以使我们在这个领域得到与发达国家同样的发展。

长期以来，我们对细胞的改造大多集中在化学物理诱变、一个到几个基因改变的基因工程或代谢路径改变的代谢工程上，这里面也包括个别基因的分子进化改造、点突变改造甚至基因片段的改造等。这些工作的进行，使我们对基因调节细胞代谢路径及代谢产物有了越来越多的认识。例如，我们可以把几个相关基因组合起来，构建一条合成某个产物的代谢路径，可以把组成这条代谢路径的基因当做一个部件，组装到同一个或另一个新的细胞中进行表达。又如，我们可以把细菌合成生物塑料聚羟基脂肪酸酯（polyhydroxy alkanoates，PHA）的三个合成基因 *phaCAB* 操纵子作为一个生物零件，组装到其他不能合成 PHA 的细胞中，从而使之也能产生 PHA 的合成。事实上，基因的调控很像电回路，比如，乳糖操纵子中，通常一个称为阻遏子（repressor）的基因处于“开”的状态，使乳糖的利用途径关闭；当只有乳糖作为唯一碳源的时候，细菌把阻遏子的基因“关”掉，这时细菌就能利用乳糖了。类似的例子在该书中都有很好的描述，相信对读者会很有启发。

把具有某个功能的几个基因（或叫操纵子）作为一个生物零件，把完成某个任务所需要的生物零件组装起来，构建一个新的细胞，就是“合成生物学”。合成生物学包括两条路线：①新的生物零件、组件和系统的设计与建造；②对现有的、天然的生物系统的重新设计。这两条路线的目的都是为了造福人类社会。近几年来，作为合成生物学技术基础的基因组测序技术及 DNA 合成技术正以指数速率增长，这正如大规模集成电路与计算机技术的发展一样。有理由认为合成生物学将会像信息技术一样得到迅速发展，并将在能源、化学品、材料、疫苗等医药领域得到广泛应用，具有巨大的社会效益及经济效益。同时，合成生物学将对人类认识生命、揭示生命的奥秘、重新设计及改造生命等方面产生重大的科学意义。在美国能源部资助下，美国生物经济研究协会（http://www.bio-era.net）于 2007 年发表了题为“基因组合成和设计未来：对美国经济的影响”的研究报告，分析了合成生物学及基因组工程支撑技术的迅速发展，从技术的发

明、创造、结合、重组、传播、推广的角度，分析了生物工程革命的经济影响范围及大小，展望了合成生物学与基因组工程的应用前景，认为在未来几年，合成生物学技术将会比重组 DNA 技术发展更快。报告预计合成生物学将在生产化学品、能源、疫苗及医药等方面有极为重要的应用前景，具有重大的经济意义及社会意义。

该书较为完整地讨论了合成生物学的诞生、定义、研究内容、意义、工程本质、合成生物学与相关生物学科的关系等内容，概要介绍了合成生物学国际会议、国际遗传机器设计大赛等背景信息，为读者全面了解合成生物学提供了较为完整的信息。

该书还描述了合成生物学的道德规范、创造人工生命所带来的问题、究竟该如何应对伦理，详细地介绍了合成生物系统的设计、数学模拟、性能分析，以及合成生物学的基础研究和应用研究等，是很值得一读的首部“合成生物学”书籍。

（陈国强，清华大学生命科学学院教授）

2009 年 10 月 31 日

前　言

合成生物学是21世纪初刚刚兴起的一门交叉学科，其工程化的理念、标准化的生物工具和大胆新颖的设计思路，一经兴起便引起了世界范围的广泛关注。诸如J. D. Keasling课题组合成青蒿酸的微生物工厂、C. A. Voigt课题组的大肠杆菌成像系统等合成生物学研究和应用方面的重大突破更是增加了世人的信心。合成生物学的发展受到多个国家政府和研究团体的高度重视，美、欧等国均投入大量资本支持合成生物学的研究与工业化。2009年，由中国科学院300多位专家经过一年多研究发布的《创新2050：科学技术与中国的未来》战略研究系列报告指出："合成生物学是可能出现革命性突破的4个基本科学问题之一"。由此可见，合成生物学的发展必将会在不久的将来给国民经济带来巨大推动。高等学校的教育必须引领科学的发展，编写一本好的合成生物学专业教材，对学生进行系统培养，无论是对于学科的发展还是对于科研的发展都至关重要。

为推动合成生物学的迅速发展，2003年，美国麻省理工学院发起了一项以本科生为主要参赛对象的国际遗传机器大赛（international Genetically Engineered Machine Competition，简称iGEM）。两年后，也就是2005年，这个最开始仅仅是麻省理工学院校内规模的比赛，迅速升级为世界范围的国际比赛。越来越多世人谙熟的著名大学——哈佛大学、剑桥大学、加州大学伯克利分校和普林斯顿大学等无不对此表现出极大的热情。2006年有32个参赛队；2007年参赛队迅速增加至54个；2008年参赛队增至84个；2009年参赛队更是多达111个，包括来自欧洲、亚洲、南美洲、北美洲数十个国家数千名参赛队员，参赛项目覆盖环境、能源、健康等方方面面。

在合成生物学方兴之时，天津大学化工学院抓住机遇，与合成生物学创始人及其他著名学者多次探讨合成生物学的发展态势。2006年，天津大学对合成生物学有兴趣的老师和同学经常在一起研讨，自然形成了一个合成生物学兴趣小组，开始筹备iGEM2007竞赛的参赛事宜，着手进行师生培训。小组成员包括张春霆、元英进、赵学明、宋凯、王庆昭、程景胜、马平生、黄鹤、班睿等指导老师和来自不同专业、不同年级的学生。2007年4月和6月天津大学化工学院承办了iGEM在亚洲区的两次研讨会，特别邀请了合成生物学专家、iGEM大赛创始人D. Endy、T. Knight、R. Rettberg、C. D. Smolke，以及Davison学院的K. A. Haynes作报告。包括澳大利亚墨尔本大学、日本千叶大学、北京大学、清华大学、中国科学技术大学在内的多所大学的教师和学生在这两次研讨会上共同学习和分享，成功促进了iGEM和合成生物学在亚澳地区的普及和推广。同年，天津大学作为中国内地首批参赛的四所高校之一，取得了金牌的优异成绩，并在2008年的iGEM竞赛中再获佳绩。

作为兴趣小组的指导教师之一，作者既被合成生物学深深吸引，也被学生的热情和

投入所感动。在两年的 iGEM 竞赛训练过程中，作者深感学习资料匮乏，深深体会到了撰写一本适合我国高校师生和研究人员使用的合成生物学参考资料的必要性。2008 年，天津大学批准开设了“合成生物学导论”创新课程，在课程的准备和讲授的过程中，作者更加感觉到编写一本合成生物学教材的迫切性。为此，作者将“合成生物学导论”的讲义编撰成书，力图将合成生物学的基本理念、研究内容和经典实例介绍给读者，为合成生物学的推广和发展略尽绵力。

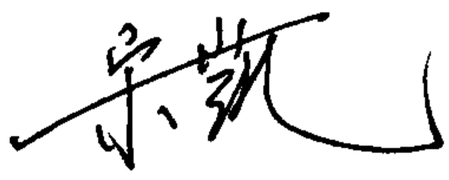
本书主要分为 5 章：第 1 章介绍合成生物学的兴起、特点和研究内容，并对合成生物学所引发的道德思考进行初步探讨，使读者对合成生物学有一个全面了解；第 2 章介绍合成生物学的层次化结构和概念，包括 BioBrick 等基本概念和合成生物系统设计中的基本逻辑拓扑结构；第 3 章介绍合成生物学中常用的建模、模式识别算法，以及生物系统的主要性能和基本分析方法；第 4 章介绍合成生物学在基础研究方面取得的研究成果和经典实例，包括基因开关、细胞群体系统脉冲发生器、细胞群体图案系统、人工转录级联线路、工程化的支架蛋白和重构 T7 噬菌体基因组等内容；第 5 章介绍合成生物学在应用方面的研究内容和成果，包括著名的合成青蒿酸的微生物工厂、大肠杆菌成像系统等。

全书由宋凯副教授主笔和定稿，由南开大学生命科学院黄熙泰教授主审。书中实例均引自期刊文献，力求内容的新颖性与准确性。部分模拟和实验结果图在原文基础上略有修改，读者可参考原始文献阅读使用。

桃李不言，下自成蹊。值本书出版之际，特对以下老师、同行的支持与帮助表示感谢：张春霆院士提出了许多宝贵的建议和修改意见；元英进教授从策划、内容安排到成书的全过程一直给予鼎力支持；赵学明教授在参与“合成生物学导论”讲授的同时给予诸多指导；此外，邹瑞阳、杨梦、陈林、李炳志、李晓倩、朱曼丽、杜瑾、侯男、张蔚然、刘夺、袁芳等在资料整理等方面也提供了许多帮助。

本书出版得到天津市科协自然科学学术专著基金、高等学校博士学科点专项科研基金新教师基金（200800561005）资助，在此一并表示感谢！

由于作者水平和研究兴趣所限，书中的缺点和不足之处在所难免，热忱欢迎广大读者批评指正。本书只是一块引“玉”的“砖”，相信随着合成生物学的发展，会有越来越多的有识之士投入其中，会有越来越完善的资料和书籍不断出现。



2009 年 12 月

专有名词英汉对照表

英文名称（缩写）	中文名称
synthetic biology	合成生物学
genetic circuit	遗传线路
gene circuit	基因线路
elementary gene circuit	基础基因线路
genetic device	遗传装置
chaperone	伴侣蛋白
essential gene	必需基因
pleiotropic	多效性
module	模块
robustness	鲁棒性
motif	模体/基元
de novo DNA synthesis	DNA 从头合成
grey goo	灰色黏质
stem-loop	茎环
hairpin structure	发夹结构
inverted repeat sequence	反向重复序列
ultrasensitivity	超敏感性
quorum sensing（QS）	群体感应
standardization	标准化
decoupling	解耦
abstraction	抽提
isopentyl pyrophosphate（IPP）	焦磷酸异戊酯
mevalonate isoprenoid（MEV）	甲羟戊酸类异戊二烯
amorpha-4，11-diene synthase（ADS）	紫穗槐-4，11-二烯合成酶
toggle switch	双稳态开关/拨动开关
international Genetically Engineered Machine competition（iGEM）	国际遗传机器大赛
BioBrick	生物积块
cohesion	内聚
coupling	耦合
Part	生物组件
Device	生物装置
System	生物系统
cascade	级联线路
[RNA] polymerase per second（PoPS）	RNA 聚合酶每秒
inverter	转换器
N-acyl-homoserine lactone（AHL）	*N*-乙酰基高丝氨酸内酯

dense overlapping regulon (DOR)	密集交盖调节网
incoherent feedforward	非一致前馈
coherent feedforward	一致前馈
sign-sensitive delay	信号敏感性延迟
elementary reaction	元反应
tree-augmented Bayesian classifier (TAN)	树扩张型朴素贝叶斯分类方法
scale-free networks	无标度网络
biphasic switch	双相开关
aptamer	适配体
linear-loop interaction	线性-环相互作用
programmed cell death (PCD)	程序性细胞死亡
apoptosis	凋亡
leucine -zipper	亮氨酸拉链
modulator	调控蛋白
recruitment	募集
decoy-zipper	诱捕拉链
refactoring	重构
chloroquine	氯喹
artemisinin	青蒿素
artemisinic acid	青蒿酸
dimethylallyl pyrophosphate (DMAPP)	二甲（基）丙烯焦磷酸酯
farnesyl pyrophosphate (FPP)	法尼基焦磷酸酯
deoxyxylulose 5-phosphate (DXP)	脱氧-5-磷酸木酮糖
artemisia annua	青蒿
biosensor	生物传感器
phytochrome	光敏色素/感光蛋白
chromophore	生色团
two-component system	双组分系统
apoprotein	脱辅基蛋白
phycocyanobilin (PCB)	藻胆青素
the burnt pancake problem (BPP)	烧焦烧饼问题
syntenic genome	同线基因组
B cell antigen receptor (BCR)	B细胞抗原受体
immunoreceptor tyrosine-based activation motif (ITAM)	酪氨酸活化模体
agammaglobulinemia	丙种球蛋白血症
ethionamide (ETH)	乙硫异烟胺
biofilm	生物膜
theophylline	茶碱
transactivator	转录反激活因子
enterocyte	肠吸收细胞
hypoxia	缺氧/低氧
recombinational enhancer (RE)	重组增强子
recombinant DNA	DNA 重组

《合成生物学导论》读者反馈表

欢迎广大专家、教师、读者与我们联系，提供宝贵意见与建议！凡使用本书作为教材的单位，可获赠教学课件一份；课件数量有限，赠完为止。本活动解释权在科学出版社。

姓名：	职称：	职务：
电话：	电邮：	
单位名称：		
通信地址：		邮编：
主要研究方向：		
所授课程（一）：		人数：
课程对象：□研究生　□本科（_____年级）　□其他_____		授课专业：
使用教材名称 / 作者 / 出版社：		
所授课程（二）：		人数：
课程对象：□研究生　□本科（_____年级）　□其他_____		授课专业：
使用教材名称 / 作者 / 出版社：		
您对《合成生物学导论》的评价及修订意见： 本书获知方式：□_____书店　□_____网络　□他人推荐　□其他_____		
贵单位（院系、研究所等）开设的相关课程有哪些？使用的教材名称/作者出版社？		
科研、教学需求的图书： 意向写作计划：		
其他建议或意见：		教学使用证明 （签章）

回执地址：北京市东黄城根北街 16 号 科学出版社高等教育出版中心（邮编 100717）
联 系 人：单冉东　　咨询电话：010-64030233　　传真：010-64034871
电子邮箱：biology@mail.sciencep.com

目　　录

序一
序二
前言
专有名词英汉对照表 …… xi

第1章　绪论 …… 1
第一部分　合成生物学概述 …… 1
1.1　合成生物学的诞生 …… 2
1.2　合成生物学的定义 …… 4
1.3　合成生物学的研究内容 …… 4
1.3.1　生物大分子的合成与模块化 …… 5
1.3.2　生物基因组的合成、简化与重构 …… 7
1.3.3　合成代谢网络 …… 8
1.3.4　遗传/基因线路的设计与构建 …… 8
1.3.5　细胞群体系统及多细胞系统研究 …… 9
1.3.6　数学模拟和功能预测 …… 9
1.4　合成生物学的意义 …… 10
1.4.1　加速合成生物系统工程化的进程 …… 10
1.4.2　验证和深化对于生物现象的理解 …… 10
1.5　合成生物学的工程本质 …… 11
1.6　合成生物学与相关生物学科 …… 13
1.6.1　与遗传工程的关系 …… 13
1.6.2　与分子生物学和细胞生物学的关系 …… 14
1.6.3　与系统生物学和功能基因组学的关系 …… 14
1.6.4　与生物信息学的关系 …… 15
1.7　合成生物学国际会议及国际遗传机器大赛 …… 16
1.7.1　合成生物学国际会议（Synthetic Biology x.0） …… 16
1.7.2　国际遗传机器大赛 …… 16
第二部分　合成生物学引发的思考 …… 17
1.8　人类≠造物主 …… 17
1.8.1　道德规范 …… 17
1.8.2　对环境无控制地排放 …… 18
1.8.3　生物武器的滥用 …… 19

1.8.4 对生物的专利与垄断 …… 20
1.8.5 创造人工生命? …… 21
1.8.6 究竟该如何应对? …… 22

第2章 合成生物系统的设计 …… 25
第一部分 合成生物系统的模块化与层次化结构 …… 25
2.1 合成生物学的模块化设计 …… 26
2.1.1 模块设计的三个特征因素 …… 26
2.1.2 细胞本底环境对模块化设计的影响 …… 27
2.1.3 无细胞合成生物系统 …… 28
2.2 标准化生物模块——生物积块 …… 28
2.2.1 BioBrick 命名规则 …… 29
2.2.2 BioBrick 的连接 …… 30
2.3 标准定量机制 …… 33
2.3.1 PoPS …… 33
2.3.2 RIPS …… 34
2.4 生物系统的层次化结构 …… 34
2.4.1 生物部件“Part” …… 34
2.4.2 生物装置“Device” …… 36
2.4.3 生物系统“System” …… 36
2.4.4 细胞群体系统及多细胞系统 …… 37
第二部分 合成生物系统的逻辑拓扑结构 …… 39
2.5 串联 …… 39
2.6 单输入 …… 40
2.7 多输入 …… 41
2.8 反馈 …… 42
2.9 前馈 …… 44

第3章 合成生物系统的数学模拟与性能分析 …… 49
第一部分 生物系统的数学模拟 …… 49
3.1 常微分方程 …… 49
3.1.1 质量作用定律 …… 50
3.1.2 米氏方程 …… 51
3.1.3 希尔方程 …… 52
3.2 Logistic 模型 …… 53
3.2.1 单种群细胞自由生长的 Logistic 模型 …… 53
3.2.2 两种细胞自由生长的 Logistic 模型 …… 54
3.3 主方程 …… 56

3.4　随机微分方程…… 57
3.5　模式识别…… 57
3.5.1　聚类分析…… 58
3.5.2　分类分析…… 60
3.6　小结…… 64
第二部分　合成生物系统性能分析 …… 64
3.7　常用性能指标…… 65
3.7.1　稳定性…… 65
3.7.2　鲁棒性…… 65
3.7.3　响应快速性…… 65
3.8　稳定性分析…… 65
3.9　鲁棒性分析/敏感性分析 …… 66
3.10　提高系统稳定性和鲁棒性的机制 …… 67
3.10.1　冗余 …… 68
3.10.2　反馈控制 …… 68
3.10.3　特异性和保真度 …… 68

第4章　合成生物学的基础研究 …… 71
第一部分　模拟逻辑功能的基础研究 …… 71
4.1　逻辑门功能遗传线路…… 72
4.2　基因调控开关…… 74
4.2.1　转换开关…… 74
4.2.2　双相开关…… 75
4.2.3　Riboswitch …… 75
4.2.4　双稳态开关…… 77
4.3　Repressilator …… 83
4.4　细胞群体系统脉冲发生器…… 88
第二部分　其他方面的合成生物学基础研究 …… 92
4.5　控制群体数量的基因线路…… 92
4.6　人工细胞群体图案系统…… 95
4.7　人工转录级联线路…… 99
4.8　工程化的支架蛋白 …… 100
4.9　重构 T7 噬菌体基因组 …… 105

第5章　合成生物学的应用研究 …… 111
5.1　合成生物学应用于维护人类健康 …… 111
5.1.1　环境控制的大肠杆菌对癌细胞的侵入系统 …… 113
5.1.2　合成青蒿素的微生物工厂 …… 117

5.2 合成生物学应用于生产生物能源 …… 121
利用工程化的大肠杆菌非发酵代谢途径生物合成高级醇…… 123
5.3 合成生物学应用于环境治理 …… 126
砷离子的检测…… 126
5.4 合成生物学应用于其他方面 …… 128
5.4.1 大肠杆菌成像系统 …… 128
5.4.2 微生物计算系统 …… 131

参考文献…… 137
主要网站…… 148
附录 A 名词解释…… 150
附录 B iGEM 参赛项目选介 …… 157
1. 幽门螺杆菌疫苗 …… 157
2. 艾滋病的治疗 …… 161
3. 芳香烃类污染物的检测 …… 164
4. 丁醇的生物合成 …… 165
5. 基因反馈调节治疗脓毒症 …… 167
6. 大肠杆菌香味剂 …… 169
名词索引…… 171
人名索引…… 173

科学绝不是也永远不会是一本写完了的书。每一项重大成就都会带来新的问题。任何一个发展随着时间的推移都会出现新的严重的困难。

——爱因斯坦

第1章 绪论

基础提要

- 细胞生物学
- 分子生物学
- 生物化学
- 细胞结构与功能
- 生物能学和代谢

合成生物学强调“设计”和“重设计”，其目的是通过人工设计和构建自然界中不存在的生物系统来解决能源、材料、健康和环保等问题。其工程化的思想和标准化的工具，一经兴起便得到全世界范围的广泛关注，2004年被美国麻省理工学院（MIT）主办的世界著名科技杂志 *Technology Review* 评为将改变世界的十大新出现的技术之一（10 Emerging Technologies That Will Change Your World）。2009年，由中国科学院300多位专家经过一年多研究发布的《创新2050：科学技术与中国的未来》战略研究系列报告中指出：“合成生物学”是可能出现革命性突破的4个基本科学问题之一。

第一部分 合成生物学概述

1953年，年仅25岁的J. D. Watson和37岁的F. Crick从DNA（脱氧核糖核酸）的X射线衍射图上解读了DNA的双螺旋结构，隐藏了几十亿年的生物密码终于渐渐露出了端倪。2003年人类基因组计划顺利完成，此后包括人类在内的各种生物基因组图谱纷纷出炉，生物遗传密码的神秘面纱正在被迅速揭开。生物科学研究由定性描述转向定量计算，从分析到设计，进而进入系统和合成生物学（synthetic biology）时代，生物科学的复杂课题成为研究的主流。如果说现代自然科学的领头学科是物理学的话，那么21世纪自然科学发展的关键带头学科就是生命科学。然而，仅仅是

解释和破译现有的生命现象已经无法完全解决人类所面临的环境、能源、健康三大危机，人类不得不探索全新的出路——人工合成全新的生物系统，借此来验证我们对生物现象的理解，满足人类对物质的无限需求，改造和更新我们生存的地球乃至宇宙。

这是一个予结束于开始的时代！自从地球上开始出现生物，遗传密码已经存在了几十亿年，现在到了人工重新排列它们的时候了。作为遗传物质的 DNA 的结构发现开启了分子生物学时代，对于其结构属性的理解和改造则令生物学进入了下一个阶段——合成生物学时代。

1.1 合成生物学的诞生

同所有其他科学学科一样，合成生物学的出现并不是偶然现象，而是历史积淀的必然。合成生物学的发展要以生物学、化学、物理学、数学、信息科学、工程科学、计算机科学等相关学科的发展为基础。分子生物学将人类对生命科学的研究从宏观世界引入到微观世界；生物信息学的出现使人类具有处理大量生物数据信息的能力；系统生物学和功能基因组学的出现使得人类具有从整体和功能角度研究和分析各种生物现象的能力。只有当这些基础学科的发展达到一定高度以后，合成生物学的出现才会成为必然。图 1-1 所示为合成生物学的学科基础及应用范围。

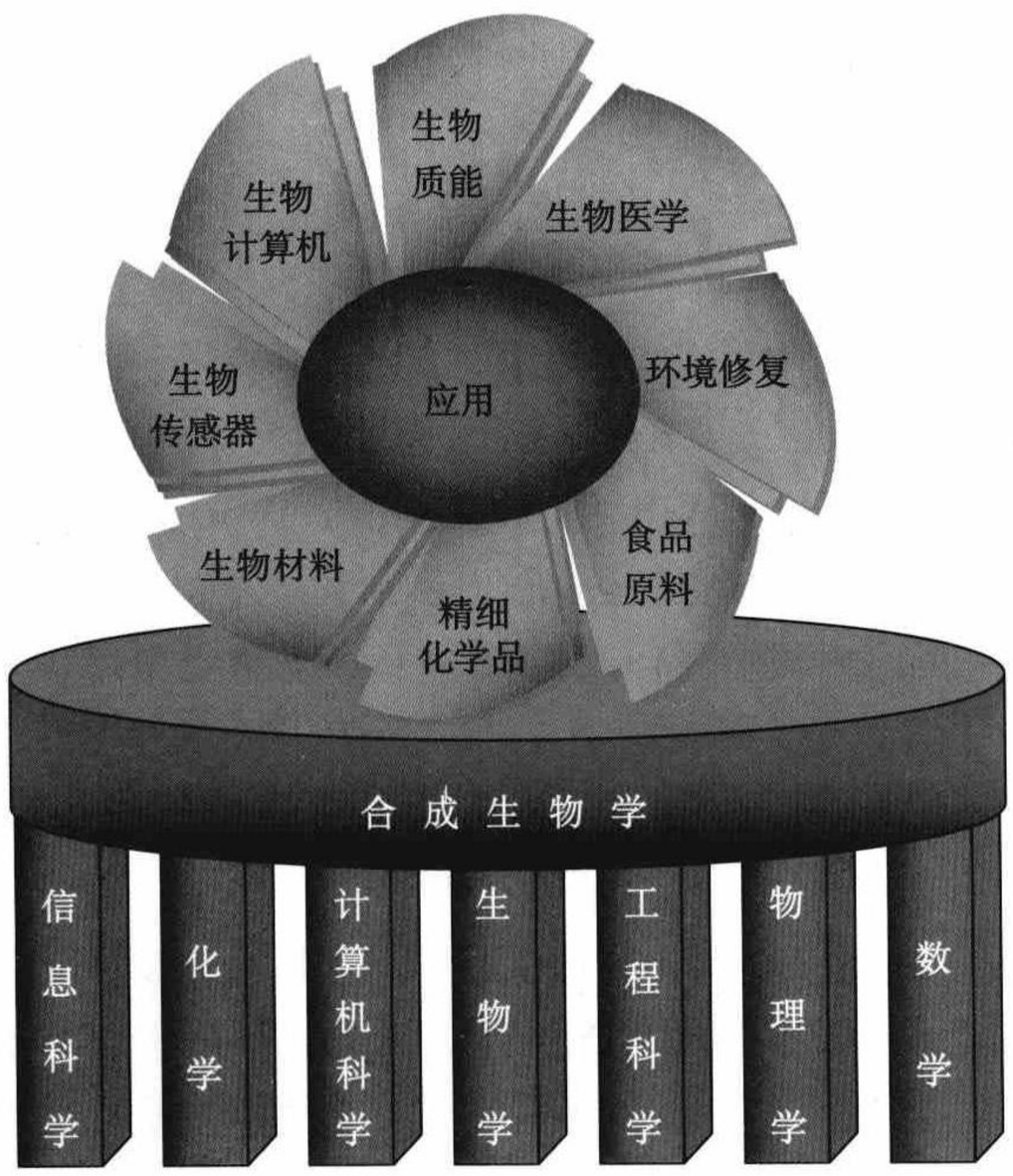


图 1-1 合成生物学的学科基础及应用范围

工程科学：将数学、化学、物理学等自然科学的原理应用到具体工农业生产部门中形成的各学科的总称。例如，水利工程、化学工程、土木建筑工程等。

1952年，S. E. Luria和M. L. Human在研究T偶数噬菌体对大肠杆菌的侵染实验中发现，细菌能将外来的DNA片段在某些专一位点上切断，从而保证自身不被外来噬菌体所侵染，而其自身的DNA由于被一种特殊的酶（甲基化酶）所修饰而受到保护，即所谓的“限制-修饰”现象[1]。1953年，G. Bertani和J. J. Weigle在λ噬菌体对大肠杆菌的侵染实验中发现了同样的“限制-修饰”现象[2]。从此人们逐渐认识到限制和修饰是存在于细菌系统中的普遍现象，作用于许多烈性（即可以导致细菌裂解）和温和（即可以导致细菌进入溶源态）的噬菌体，并存在于多种细菌中。1962年，W. Arber和D. Dussoix阐明了该现象的分子机制[3]。在发现“限制-修饰”现象的基础上，20世纪60年代末，S. Lin和W. Arber发现了执行DNA“修饰”的DNA甲基化酶和执行“限制”的DNA限制性内切核酸酶。1970年，美国约翰·霍普金斯（Johns Hopkins）大学的H. Smith进一步发现了能准确切割识别序列的DNA限制性内切核酸酶。这些能准确切开基因组特殊位点的酶被称为“分子手术刀”，它们与DNA连接酶结合应用，诞生了合成生物学常用手段之一——DNA重组（recombinant DNA）技术。同一时期，F. Jacob和J. Monod提出了调节基因表达的操纵子模型，F. Jacob认为细胞的行为是通过目的性的调控实现的，而他的这一思想被认为是合成生物学主要研究内容之一——“遗传/基因线路（genetic/gene circuit）”的雏形[4]。

1978年，波兰科学家W. Szybalski在他的一篇文章中提到为“发现了限制性内切核酸酶及其在分子生物学中的应用”的研究者所颁发的诺贝尔奖时说：“对于限制性内切核酸酶的工作不仅可以提供给我们重组DNA的工具，而且引领我们进入了一个新的‘合成生物学’的领域[5]。在此，我们不仅可以利用现存的、已经分析过的基因，并且可以创造和研究新的基因组组成的方式”，由此第一次提出了合成生物学的思想。1980年，“合成生物学”第一次作为文章的标题出现在学术期刊上——B. Hobom用“合成生物学”一词来描述通过工程化的重组基因组技术而得到的细菌，这些经过人工改造的细菌仍然能够正常存活[6]。2000年在旧金山举行的美国化学会年会上，“合成生物学”一词再次被E. T. Kool和其他学者提起用以描述生物系统中非天然存在的功能性有机分子的合成[7]。当时的合成生物学还未脱离遗传工程，与之有相当大的交叉。

2000年以后，在D. Endy、J. D. Keasling、R. Weiss、J. J. Collins和C. D. Smolke等许多人的推动下，合成生物学作为一门崭新的学科得到了飞速的发展。美国威斯康星（Wisconsin）大学的研究人员创造出了一种基因组比野生型大肠杆菌更清瘦、更精干的大肠杆菌版本[8]。这项成果表明，研究人员能够在不影响生物基本功能的情况下对其进行更精确、更大规模的改造。J. D. Keasling通过合成生物学方法在酿酒酵母中合成抗疟疾物质青蒿素的前体物质青蒿酸，这一成果被评为2006年的世界十大科技进展之一[10]。模拟各种“与”、“非”门逻辑的基因调控线路的出现、大肠杆菌成像系统的成功，所有这些实例无不证明：合成生物学作为一门独立的崭新学科步入了科学的殿堂。

作为一门迅速成长的新兴学科，目前来看，合成生物学更多的还是展望，随着相关领域的发展，在不久的将来，合成生物学将会有更广泛的发展，为人类更加深入地理解各种生物和生命现象提供有利佐证，为解决人类目前所面临的能源危机、环境危机和健康危机等提供更加有力的理论和工具。

1.2 合成生物学的定义

合成生物学（synthetic biology）长期以来被用于描述如下的生物学方法，即尝试整合不同领域的研究以建立对生命过程更加全面的理解。然而，2000 年以后该词具有了不同的含义，它标志着一个新研究领域的形成。这个新领域将科学和工程学结合在一起，目的在于设计和构建具备新的生物学功能的生物基因组和系统。

虽然 2000 年以后“合成生物学”一词在学术刊物及互联网上大量出现，但目前合成生物学的定义还处于多元化阶段[10]。

（1）美国加州大学伯克利分校化学工程教授 J. D. Keasling 认为：合成生物学正在用“生物学”进行工程化，就像用“物理学”进行“电子工程”、用“化学”进行“化学工程”一样。

（2）美国哥伦比亚癌症研究中心、测序及基因组科学中心主任 R. Holt 认为：合成生物学与传统的重组 DNA 技术之间的界限仍然是模糊的。从根本上说，合成生物学正在利用获得的“元件”进行下一层次的工作——对细胞进行实际的工程化。

（3）美国哈佛大学医学院遗传学教授、计算遗传学中心主任 G. Church 认为：合成生物学是利用我们所确信的一些“零件”进行新生物系统的工程。它在利用从系统生物学得出的最好分析法加工、制作及检验复杂的生物机器。

任何一门学科或者科学的发展都需要开放性的思想和兼容并蓄的科学态度，因此我们没有必要像背诵数学定理一样深究合成生物学定义中的每一个字眼，也没有必要强求一个完全统一的定义，我们推荐“合成生物学组织”网站* 上公布的一段描述来引导我们了解什么是“合成生物学”：

合成生物学是指按照一定的规律和已有的知识：①设计和建造新的生物部件、装置和系统；②重新设计已有的天然生物系统为人类的特殊目的服务。[Synthetic Biology is: A) the design and construction of new biological parts, devices, and systems, and B) the re-design of existing, natural biological systems for useful purposes.]

简单地说，合成生物学就是通过人工设计和构建自然界中不存在的生物系统来解决能源、材料、健康和环境等问题[11]。

合成生物学强调“设计”和“重设计”。设计、模拟、实验是合成生物学的基础。合成生物学不仅仅是实验，利用已有的生物学知识，根据实际的需要进行设计和重设计，建立数学模型对人工设计进行模拟从而指导实验的进行才是合成生物学的方法[12]。

1.3 合成生物学的研究内容

合成生物学的研究内容主要包括以下几个方面[13,14]。

（1）工程化的功能模块（module）：包括各种生物分子的合成与模块化、亚细胞模

---

* http://www.syntheticbiology.org

块、生物合成基因网络[15]、代谢途径和信号转导通路[16,17]、转运机制等。

（2）接口：调整和修改输入-输出过程、调整不同亚细胞组件间的层次化相互作用使得模块具有可拆装性。

（3）开发平台：优化生物或非生物载体，达到提高工程系统效率、降低其维护成本和要求、提供某种特殊的“鲁棒性”及对环境的兼容性等目的。

（4）调控和通讯系统：包括生物部件的反馈、前馈机制，以及行为和通讯方式的模块化。

（5）各种功能模块的仿真、预测算法和相应软件。利用计算机辅助技术是降低实验成本和复杂性的有利工具。

1.3.1 生物大分子的合成与模块化

合成生物学的研究重心是具有截然不同于自然系统特性和功能的崭新生物系统的开发和工程化，因此离不开各种标准化生物大分子的研制和开发。这方面的研究主要包括以下几部分内容。

1. 蛋白质的工程化改造与模块化

根据分子生物学的“中心法则”，生物遗传信息的主要传递过程是以 DNA 为模板指导 RNA 的合成并由 mRNA 指导蛋白质的合成。DNA 测序技术、遗传工程技术、聚合酶链反应技术（PCR 技术）、蛋白质晶体学技术等一系列生物技术的发展，使人类理论上可以进行任何蛋白质从氨基酸序列到局部超二级结构（模体，motif）的改造和组合。

合成生物学设计功能相对独立、可以被组装的蛋白模块以探索生物的可塑性和构建崭新的生物系统，因此对基本组分“工具箱”中蛋白质的需求远比自然界能够提供的种类多得多。例如，目前通过聚合酶链反应实现的基因扩增等生物过程主要是由自然界提取的酶（限制性内切核酸酶、连接酶、聚合酶）来执行的，这些酶受到底物和环境等的限制，在模块化和功能性方面无法完全满足人工生物系统的要求，因此需要开发模块化、标准化的人工蛋白模块。

这些蛋白模块可以通过蛋白自身域结构的重新组合来巧妙地实现——生物体中有两类 DNA 结合蛋白具有功能相对独立的锌指（zinc finger）结构，一类是锌指蛋白，另一类是甾类受体。蛋白质中的锌指结构由一组保守的氨基酸残基（半胱氨酸和组氨酸）的侧链和锌离子以配位键结合，在蛋白质中形成了相对独立的超二级结构，使蛋白质与 DNA 特异序列结合，像一根根手指伸向 DNA 的大沟。巧妙地利用这种结构，可以人工设计多锌指 DNA 结合蛋白，进而重组蛋白模块获得崭新的蛋白质功能[18~20]，如崭新的酶底物结合特异性等。

研究人员还可以利用计算化学指导突变，构建具有崭新功能的蛋白质或活性酶。例如，蛋白质工程研究者能够合理地重新设计和构建变化了的大肠杆菌等细菌对 TNT、L-乳酸盐等胞外配体产生响应的外周胞质；或者用复合胺替代原来的自然配体[21]，将核糖体结合蛋白（一种缺乏酶活性的感受蛋白）转化为活性蛋白，如具有丙糖磷酸异构

酶（triose-phosphate isomerase activity）活性的蛋白质[22]。

通过蛋白模块的构建和改造，不仅可以获得具有崭新功能的蛋白质，还可以利用信号蛋白能够在蛋白质-蛋白质相互作用水平进行功能重编程的特性，对信号通路进行构造和改变。例如，通过改变真核神经元 N-WASP 蛋白接收域和输出域使蛋白质具有全新的信号行为[23]；通过号称分子胶水的支架蛋白（scaffold protein）将激酶和激酶作用物紧密粘合在一起得以磷酸化从而具有特异性等[24,25]。

然而，众所周知，氨基酸彼此的相互作用使分子链发生折叠形成蛋白质的二级和三级结构，目前这一结构仅从蛋白质氨基酸序列数据中很难预测，且对蛋白质氨基酸序列进行预定义的修饰是一个高度复杂的过程。因此，可互换工程蛋白质部件的发展水平远落后于 DNA 部件的发展水平。

2. 核酸分子的人工合成

当自然界存在的 DNA 或 RNA 不具有期望属性时，合成 DNA 乃至全基因组有时是获得期望生物系统的唯一方法。合成 DNA 可以用来重设计目标基因序列、编码区域或者调控信号（报告基因、阻遏子、激活子、终止子、核糖体结合位点、启动子等），以及为适应特殊宿主或模拟生物体而改变密码子用法。合成基因允许高效构建相关的、在特殊区域有所改变的基因簇；允许目标基因的柔性设计而不需常规基因重组或克隆所需的中间步骤；允许用户选择只包括期望功能和途径的人工合成基因，以简化或切断生物进化作用所带来的影响；允许用户插入任意期望的模块，具有可扩展性；如果把合成基因组设计成只有在实验室特殊条件和培养基中才能存活的形式，则人工合成基因更具有安全性。

近来发展比较迅猛的要数 DNA 从头合成（*de novo* DNA synthesis）技术，目前此技术主要是通过短链片段的拼接，以及化学合成寡核苷酸并通过 PCR 装配成较长的 DNA 片段[26]。已经有多家公司能够提供一周内以 1 美元/碱基的低廉价格合成 2000bp 的服务了[27]。利用这项技术已经成功全合成了脊髓灰质炎病毒等较小的生物基因组[28]。然而，用标准的亚磷酸三酯化学法合成 DNA 寡核苷酸的成本和化学合成的准确性仍然是 DNA 从头合成的主要“瓶颈”[29]。虽然理论上合成长的寡核苷酸是可行的，但在实际操作中要受制于自动固相化学合成阶段平均每 250 个碱基发生一次错误的概率的限制。因此通常采用组装短链寡核苷酸（典型长度为 200 个碱基）和利用内切核酸酶及外切核酸酶移除错配的方法来将 DNA 合成的错误概率降低，目前已经能做到错误率为 1/3000bp[30,31]。

由于支撑碱基对的脱氧核糖核酸的主链相对稳定，具有非常完美的可修饰结构，对 DNA 聚合酶的轻微修饰不影响对包含除了 A、G、C、T 以外核苷的 DNA 序列进行正常的“阅读”。因此，除了上述对自然界已有基因的人工合成技术以外，某些合成生物学家还试图制造新的遗传物质，并通过实验获得了许多可以被化学合成的新型生物分子——人造基因，且证明了这些人造基因都具有繁殖和传递遗传信息的功能[32]，有力地验证和深化了人类对于生物遗传密码的理解。S. A. Benner 等人工制成了两种核苷——K 和 X，组成了被其称之为 AEGIS（an expanded genetic information system）的

系统，而且证明了核苷酸的种类可以多达12种之多[33]。美国斯坦福大学的E. T. Kool课题组在原有的4种碱基上增加苯环，形成新的4种碱基，并制成了加长的双螺旋新分子“xDNA”。通过这种方法，增加了DNA双链的间距，使DNA分子双链的氢键断裂温度提高[34]。这段双螺旋能在黑暗的环境中发光并且能在较高的温度下仍然保持稳定，有助于研究DNA的突变及其他一些生物系统性能。

1.3.2 生物基因组的合成、简化与重构

1. 人工合成生物全基因组

1979年，H. G. Khorana合成了酪氨酸阻遏tRNA基因[35]；2002年，美国纽约州立大学石溪分校E. Wimmer小组用了3年的时间合成出了脊髓灰质炎病毒（Poliovirus）的全基因组，共7500个碱基。经过实验证明，这些人工合成的病毒基因组不仅可以合成出与天然病毒蛋白完全相同的蛋白质，而且同样具有侵染宿主细胞的活力[36]。2003年，J. C. Venter的研究小组用了14天时间从头合成了噬菌体ϕX174（5386 bp）基因组[37]；2005年，美国研究人员人工合成了1918年造成全世界上千万人死亡的“西班牙流感病毒”[38,39]；2008年，J. C. Venter小组又合成了*Mycoplasma genitalium*生殖道支原体基因组（582 790 kb）[40]，这是迄今为止人类在合成生物过程中走得最远的一步[41]。虽然，J. C. Venter等合成的这个细菌基因组能否正常工作尚未报道，但J. C. Venter小组的另外一项工作却获得了成功：将关系较近的两株支原体中一株的基因组用另外一株的基因组替换，结果这株支原体仍能正常“工作”[42]。

2. 生物基因组的简化和重构

各种人工构建的生物模块难免会被植入生物体中进行扩增和培养。生物体自身的各种代谢途径、信号转导途径以及各种内源噪声对于人工模块功能的执行无疑是一种干扰。同时，天然生物基因功能的多效性和冗余性也给各种模拟算法的有效应用带来了障碍。因此，许多合成生物学家致力于生物基因组的简化和模块化，力图净化宿主细胞的代谢内环境，其中尤以“最小基因组（minimal genome）”和“必需基因（essential genes）”的研究成果最为显著[43~46]。这些基因中包括与DNA复制、RNA处理和修饰、解码遗传密码的tRNA、翻译组分和伴侣蛋白（chaperone）等能支持一个完整的生物体存活所必需的基因。现已对14种原核生物和7种真核生物基因组进行了实验，鉴别出1万余个必需基因，存放在必需基因数据库DEG 5.2中[47]。

1990年前后，美国“Institute for Biology Alternatives”的J. C. Venter小组开始研究*Mycoplasma genitalium*（生殖道支原体）细菌的最小基因组。2005年，这一小组的研究表明，该细菌大概有386个基因对DNA修复、能量代谢和其他必需过程是必不可少的。[48]。大肠杆菌的最小基因组目前已经确认的有151个基因[45]。此外，其他研究者还将T7噬菌体基因组进行重构，使其更易于细胞底物背景的量化和模拟分析[49]。本书会在第4章中对其进行较为详细的介绍。

需要注意的是，虽然生物基因组的简化和模块化具有很重要的意义，但细胞种群的

突变和进化使得模块化很难维持，无法保证工作的可重复性，因此基因组的简化应当适度，必须考虑简易性与可靠性之间的协调。

1.3.3 合成代谢网络

目前合成代谢网络主要是利用转录和翻译控制单元调控酶的表达以合成或分解代谢物。例如，将酿酒酵母中合成焦磷酸异戊酯（isopentyl pyrophosphate）因子的甲羟戊酸类异戊二烯途径（mevalonate isoprenoid）的整个代谢网络转入大肠杆菌中，通过结合插入的紫穗槐-4，11-二烯合成酶（amorpha-4，11-diene synthase）基因编码的合成酶，制造大量的抗疟疾药物——青蒿素的前体物质青蒿酸 artemisinin。

对于合成代谢网络而言，在异源宿主中，无论是最小化有毒中间代谢物的胞内累积，还是最大化目标代谢物的产量，都需要编码代谢途径的多个基因的协调与平衡[50,51]。不同来源的酶具有不同的动力学特性，很难准确预测其在异源宿主中的行为，因此对代谢网络相关基因及调节元件的合理筛选仍然是一个挑战。合成代谢网络设计中无法避免参数设计算法的错误。网络设计只能保证某些粗略的特性基本正确，而无法保证细节的正确。例如，只能够保证在代谢网络拓扑结构等特性的描述上是正确的，却无法确定许多具体的物理参数（结合常数、反应速率和扩散速率等）。此外，合成代谢途径的设计还必须充分考虑细胞的自然生长和进化过程对于网络参数稳定性的影响。这些都意味着代谢网络的设计必须与参数整定和试验相结合，通过反复的试验和模型校正进行代谢网络的完善和优化，因此仍然耗时、耗力。

1.3.4 遗传/基因线路的设计与构建

如前所述，1961 年 F. Jacob 和 J. Monod 提出的乳糖操纵子模型第一次明确提出了基因表达调控的概念[4,52]，被认为是分子生物学发展的第二个里程碑，自此越来越多的研究者致力于基因调控的研究。20 世纪 90 年代，一些研究者借鉴电磁学中描述电器件关系的“Circuit”方法，提出了“Gene Circuit”和“Genetic Circuit”概念，用于研究基因受蛋白质、mRNA 等物质调控的关系和相应的数学模型[53~56]。

遗传线路（genetic circuit），俗称“基因线路”（gene circuit）。在合成生物学中，是由各种调节元件和被调节的基因组合成的遗传装置（genetic device），可以在给定条件下可调、可定时定量地表达基因产物*。

利用转录水平、转录后水平等的控制机制，合理组合转录基元、基础基因线路、基因模块的拓扑结构，目前遗传线路的功能主要可以分为两大类：逻辑基因线路（模拟各种逻辑关系和数字元件的遗传线路）和其他功能遗传线路（具有特定生物功能的遗传线路）**。

---

* 此定义由南开大学生命科学学院生物化学和分子生物学系黄熙泰教授拟定。

** 目前的大部分文献对于“基因线路”和“遗传线路”的区分并不十分严格。就作者的理解，似乎前者侧重于基因的调控，后者侧重于基因的遗传功能。本书后面章节在讲到各种合成生物学实例时会遵照作者的原意分别采用“基因线路”和“遗传线路”的说法，其余不致引起混淆的情况下，则不会严格区分。

合成生物学中，逻辑基因线路起源于数字电路中逻辑运算的思想，主要是借鉴控制理论和逻辑电路的设计规则研究基因线路的逻辑关系与调控方法，模拟各种数字元件的功能，研究基因线路中噪声传播和响应机制等。例如，各种“与”、“或”、“非”等逻辑门关系的遗传线路[57]；各种形式的基因表达调控开关、双稳态开关[58,59]、Repressilator[60]、脉冲发生器和级联线路等。

其他功能遗传线路设计主要是利用基因模块原有的功能，设计全新的遗传线路，并利用基因重组、基因克隆等基因操作手段对现有的生物系统进行改造，使生物系统具有特定的期望功能。例如，在大肠杆菌中构造信号感知系统用以控制细胞的浓度等。

1.3.5 细胞群体系统及多细胞系统研究

基于细胞间交流的细胞群体系统及多细胞系统的开发，主要是研究细胞群体间的同步基因表达、信号交流、异步功能配合等。

目前研究者利用人工构建的群体感应机制，已经开发出了许多具有崭新功能的细胞群体系统。例如，利用SOS应急系统和群体感应的双稳态开关[61]，能够达到高或低细胞浓度的双稳态开关系统等[62]。

由于基因表达过程中内源和外源噪声的影响以及其他细胞的作用，即使同源细胞也可能具有不同表型和异步行为。而在细胞群体系统及多细胞系统的设计中，却要涉及大量完全不同的细胞，这些细胞合成和工作的可靠性必然受到多种信号组分、多种宿主细胞、多路通讯等方面的影响，因此在细胞群体及多细胞系统环境下设计通讯系统需要平衡胞内元素的敏感性，降低信号间的交叉干扰[63~65]。

1.3.6 数学模拟和功能预测

前面已经提到，用野生型装置构建模块的困难主要在于：自然进化已经将野生型生物优化到了足以适应自然环境的地步，要在人工环境里重新进行连接，这些生物装置很有可能失去原有的功能。因此合成生物学家通常需要改变装置特性以获得期望的功能。当对生物装置的机理了解比较充分时，通过合理的突变调整转录和翻译的动力学属性、操纵子（operator）结合亲和性以及转录因子结合协同性等都有助于使模块达到预期目标。

然而，即使利用合理突变优化简单模块也需要耗费大量的时间和资源。考虑到构建实际系统的高成本和耗时性，利用计算模型的辅助，通过各种数学工具抽提模拟单元的动力学特性和网络连通性，提供系统变量的描述等则是为实验提供预测信息，指导实验优化，降低实验成本的主要方法[66]。

理想情况下，模型应该能够捕捉生物系统的动态行为，在活体实验之前通过计算机模拟等方式在优化设计策略、调试等方面给出建议。虽然基于结构模型的分子动态模拟在模拟DNA结合位点的动态行为上已经显现了其丰富多彩的功能，但抽提生化反应以模拟生物装置不仅需要必要的计算工具，还需要获得相应的速度常数。到目前为止，从活体细胞中直接确定速度常数仍然是一个非常困难且不准确的过程，这也成为了限制数学模型成功应用的主要“瓶颈”。此外，其他模块的添加或者改变

都可能会改变整个系统的功能和相应参数，因此特定条件下获得的参数一般不能直接应用于其他场合[66~69]。

1.4 合成生物学的意义

合成生物学和经典生物技术的主要区别在于达到目标所用的手段不同。传统的生物技术需要反复实验，而合成生物学却尝试着结合数学、计算机方法和工程学设计和改造生物系统。合成生物学的目的是将生物系统的合成和遗传线路的连接工程化，与集成电路的制造方法一样，先设计、再制造、再设计。其最终目的是要通过这种理性设计和再设计的过程，获得人们所需要的生物功能，并通过实验得以实现。其研究的主要任务是通过工程化策略加速生物学研究和应用的进程，同时利用人工合成的生物系统验证和深化人类对于生物乃至生命的理解[70]。

1.4.1 加速合成生物系统工程化的进程

从 1978 年 W. Szybalski 提出合成生物学思想到现在已经 30 多年了，虽然遗传工程和生物学技术各自取得了令人瞩目的成绩，但生物系统的合成仍然是一项费用昂贵且不十分可靠的过程。

生物系统几十亿年的进化并不是为了人类的理解和工程化。大量单一功能组分之间无法预料的相互作用（如基因多效性）使得人类几乎无法将复杂生物系统可靠地工程化和标准化，由此造成了许多不必要的重复性劳动，使研究人员大量的时间耗费在具体的实验操作上而不是技术和科学的更新上，也因此大大限制了生物工程在生产方面的应用[71]。因此，我们需要工程化、标准化的策略和工具，将研究人员从日复一日的重复性操作中解脱出来。

合成生物学诞生的主要目的和意义之一是作为生物工程的分支，致力于工程化自组装细胞装置，制作崭新的分子和生物系统，推进疾病诊断、药物学、基因编码功能及生命起源等其他方面的研究。对于设计人员来说，合成生物学提供了验证“基因组编码自然生物系统可以被‘重写’”这一假说的手段，也提供了制造可以有效替代某些自然生物系统的工程化代用品的手段；对于工程师而言，合成生物学是一门技术，在以往遗传工程的基础上，合成生物学试图开辟更广阔的应用领域，使得设计和构建工程化生物系统更加简便易行[72]。

1.4.2 验证和深化对于生物现象的理解

回顾合成化学的发展史，我们发现：直到 19 世纪中期，小化合物分子合成的成功才使得化学家有可能提出解释原子键的结合模型；19 世纪 50 年代 M. Berthelot 就已经应用化学方法合成了苯，但是直到 1865 年 F. Kekulé 才提出了苯的环状结构式，从机理上充分阐明了 M. Berthelot 合成的苯的稳定性。这些都说明合成化学出现的先决条件绝对不是对化学原理的全面理解。更为确切地说，“合成”与“分析”是互相促进、协同进步的两种方法，它们共同推进现代化学的发展。合成通过经验学习怎样系统地操作

和干扰分子，是分子结构的最终证明，直接推动着化学结构和反应性理论的发展[33,73]。

与此极其相似的是，在现代生物学中“分析”和“合成”扮演了同样的角色。在分子生物学诞生之前，古典生物学是以个体和细胞水平的形态描述为基础的，不能了解生物发展和演化的根本原因。分子生物学的诞生使生物学建立在分子水平的定性定量分析基础之上，使现代生物学变成成熟的科学，在测量的精度和通量上取得了巨大进步。现代实验技术（基因组测序、DNA 芯片、分子结构确定及高通量的显微技术方法和体内生物传感器）代表了主要的技术进步，为我们提供了大量分子组件细节和生物系统及其行为的描述。这些进步非常类似于 19 世纪早期分析化学的进步。然而，有机化学的历史证明：为了正确理解复杂的生物系统，“合成”将是“分析”的必要补充。只有通过人工构建各种生物分子、遗传线路和代谢途径，才能更好地验证人类对于各种生物现象的理解。成功固然会帮助我们建立合成生物学的基本原则和生物系统的工程化技术；失败也同样是人类理解与自然生物本质间存在鸿沟的直接佐证，并会为我们如何更好地理解和运用源自于自然的技术提供指引。

1.5 合成生物学的工程本质

在我们前面引述的各种有关合成生物学的描述中，读者应该可以看出，虽然这些描述并不统一，却都从不同侧面提到了合成生物学的一个显著特点，也是合成生物学区别现有生物学其他学科的主要特点，即“工程化”。合成生物学家力图通过工程化方法，将复杂的人工生物系统合理简化以探索自然生物现象及其广泛的应用，并利用基因等元素设计和构建具有崭新功能的合成生物系统[72,74,75]。

合成生物学的工程化研究主要有两种策略：自上至下（逆向工程）和自下至上（前/正向工程）[75]。自上至下策略主要用于分析阶段，试图利用抽提和解耦方法降低自然生物系统的复杂性，将其层层凝练成工程化的标准模块。例如，通过敲除基因组中除复制和功能性之外非绝对必需的遗传物质，简化基因组构建，达到可模拟和预测的目的；而自下至上的策略通常是指通过工程化方法，利用标准化模块，由简单到复杂构建具有期望功能的生物系统的方法[63]。

上述两种策略都涉及最关键的三个工程化概念：对生物系统的标准化、解耦和抽提[77,78]。标准化包括建立生物功能的定义、建立识别生物部件的方法以及标准生物部件的注册登记；解耦指将复杂问题分解成简单问题、将复杂系统分解成简单的要素，在统一框架下分别设计；抽提则包括建立装置和模块的层次，允许不同层次间的分离和有限的信息交换、开发重设计的和简化的装置和模块、构建具有统一接口的部件库等。

标准化（standardization）：现实世界中有方方面面的标准，如铁路轨距、Internet 地址、度量衡单位等。生命科学中，围绕着“中心法则”这个定义大多数生物系统运行规律的理论，已经有一些被广泛应用的有效标准。例如，最具代表性的 DNA 序列数据、遗传学表征、微阵列数据、蛋白质晶体学数据、酶命名法则、系统生物学模型及限制性内切核酸酶活性等。

然而生物工程学界却尚未开发出正式的、广泛应用的各类基本的生物功能标准，由此造成了巨大的浪费。例如，一个生物工程师手中的“强”核糖体结合位点相对于其他位点而言可能只够“中等”强度；培养在 Luria 肉汤培养基中的大肠杆菌 JM2.300 菌株中起作用的一组基因开关可能在生长于合成的最低介质中的大肠杆菌 MC4100 细胞中处于关闭状态，因此被另外一套开关所取代，那么人们能否直接将这两种开关系统组合成新的“替换开关”呢？

为了实现元件的“即插即用”性能，需要规范不同部件之间的连接标准化定义，并开发各种基本生物功能（如启动子活性）、试验测量（如蛋白质浓度）、系统操作（如遗传背景、发酵液、生长速率、环境条件等）等的标准。只有这些标准规范被广泛采用，才能保证不同研究人员设计和构建的单元能够相互匹配。这些标准化工作也是支持一个活跃的、建设性的、负责任的生物工程学界所必需的。它将更加有利于加速和保护特定生物部件遗传信息的交换使用和共享以及工程化生物系统检验、证实和授权程序的顺利进行。

解耦（decoupling）：将一个复杂问题分解成许多相对简单的、可以独立处理的问题，最终整合成具有特定功能的统一整体的过程。例如，在建筑领域，将一个建筑项目解耦成设计、工程预算、建造、项目管理和监理等相对简单、可以独立处理的过程；在控制领域，设计一个解耦控制装置解除输入、输出变量间的交叉耦合，将一个多变量系统化为多个独立的单变量系统的控制问题，使系统的每个输出变量仅由一个输入变量完全控制，且不同的输出由不同的输入控制，从而实现自治控制，即互不影响的控制。同样的，生物工程中，工程师可以将复杂的“生物系统”解耦成许多套相互独立的“装置”（如标准化的细胞、标准化的核苷酸序列等），便于利用已有的标准化部件来加速开发的进度。

蓬勃发展的 DNA 合成技术大大推动了解耦方法的进程，只有在充分发展的基因和基因组合成技术的支持下，人们才有精力和能力致力于设计和构建基因组等合成生物学领域的研究。最近，该技术更是已经发展到可以将寡核苷酸和短 DNA 片段自动组装成长链 DNA 分子的水平。

抽提（abstraction）：自然界的生物系统具有难以想象的复杂性，不仅不断有新的调控细胞行为的分子机制被发现，而且不断有既有法则以外的特例出现。怎样才能推进工程化的进程使得众多生物系统组分具有期望的行为？应对复杂性的一个有力的技术就是“抽提”。分层次进行抽提是工程化常用的方法，具有许多优势，例如，系统边界概念的引入使得许多内部信息得以顺利隐藏，有利于简化复杂系统；可以分别从不同水平对生物系统的独立性和协同操作进行描述，这也是将部件组装成复杂系统的先决条件。目前来看，生物工程中有两种抽提形式值得进一步深化和推进。①利用抽象的层次模型以不同水平的复杂程度描述生物功能的信息。生物工程的抽象层次模型必须做到在每一个水平的工作不需要考虑任何其他水平的细节；而不同水平间原则上只允许有限的信息交流。②对组成生物系统的部件和装置进行重新设计和构建，使其适当简化以方便模拟和组合，如转录启动子、核糖体结合位点和开放读码框（open reading frame，ORF）[79]的重新设计和崭新组合等。

如前所述，生物系统具有高度的复杂性，远远超过非生物系统部件的整合程度，目前我们还缺乏大量的、有关这些系统是如何运作的细节信息。标准化、解耦和抽提均需要分解生物体组成部件，这虽然在概念化方面是一种有效方法，但用无生命的物质系统组成人工生物系统需要自然元件的运用和修改，而目前我们还不知道如何在生物体动态整体水平上提供生命表征。因此，生物系统的工程化与化学、物理学等自然科学领域的工程化有着本质的不同[63,77]：

（1）生物系统自身具有复制和进化的功能，而这一功能会影响许多合成系统的长期稳定性，因此需要对系统关键部件进行长期监测。合成生物学的最终目的之一是在特定宿主细胞中构建复杂系统，宿主细胞的突变和进化自然会带来很多干扰。细胞需要生长，需要摄取养分，逃避压力，还需要进化，并且在每一阶段都具有明显不同的细胞状态，细胞体积、代谢物、膜成分乃至一些全基因组范围的调节网也会随着生长环境和细胞状态的改变而改变。所有这些因素都能影响合成系统的功能。进化的作用还有可能通过许多代的变异而“破坏”合成部件。

（2）活体系统中系统边界的定义是一项非常困难的事情。细胞中信息的传输依赖于分子的扩散和传播，这导致了所有相互作用的产生；生物体为了能够在诸多环境中生存，DNA 等许多细胞元素必须同时具有多种功能，这些都增加了抽提和解耦的复杂性。因此需要人工设计完备的小规模分子-蛋白质、DNA-蛋白质以及蛋白质-蛋白质间的相互作用，以助于开发功能相对独立的生物模块。同时，研究生物基因组的最小化和重构，以期为开发崭新生物系统提供更加优化的宿主平台。

1.6 合成生物学与相关生物学科

前面已经提到，合成生物学的发展离不开工程学、生物学、物理学、化学、数学以及计算科学等相关学科的发展；而合成生物学自身的发展也对相关学科的发展产生了促进作用。互相交叉、优势互补是当今学科发展的大趋势，合成生物学与其他学科之间自然也存在着一些研究内容上的交叉与互补。例如，通过各种异源基因 DNA 从头合成整条代谢途径、优化密码子域、调整 mRNA 结构和操纵单元、允许结合位点的替换等属于合成生物学中代谢工程的研究内容；基因模块设计可以看作是用合成生物学解决遗传工程的问题，等等。然而，作为一门新兴学科，合成生物学毕竟有自己区别于其他学科的特点。下面我们来初步讨论一下合成生物学与其他相关生物学科之间的关系。

1.6.1 与遗传工程的关系

可以说遗传工程和合成生物学是生物技术发展的两个阶段，后者是建立在前者的基础之上的。

遗传工程（genetic engineering）俗称“基因工程”，是生物工程的一个重要分支，指在分子水平上对基因进行操作，将外源基因通过体外重组导入受体细胞内获得“克隆”，并使其在受体细胞内复制、转录、翻译表达的操作。

由此可知，遗传工程主要是通过自动测序技术来读取基因组 DNA 序列，并用分子生物学手段（如克隆、PCR 和 DNA 序列重组等技术）来构建 DNA 序列。而合成生物学除了上述手段以外，还要开发新的 DNA 序列自动合成技术、建立一系列标准和规则来简化人工生物系统的设计过程。具体说来两者之间至少有以下三方面的不同[41]。

（1）遗传工程是将个别外源基因转移到某生物基因组内，使之能表达有益的蛋白质（例如，抗虫棉虽然携带了抗虫基因，但它还是棉花）；而合成生物学则是从头设计和构建自然界中不存在的人工生物体系，这是两者的显著不同之处。但合成生物学也包含了对现有生物的重新设计和改造，在这点上两者又是一致的，只是在改造的深度和广度上有所不同。

（2）在遗传工程的实施过程中较少使用数学工具；而合成生物学在设计和构建人造生物体系时广泛使用各种数学工具来进行模拟，优化生物体系的设计和构建，同时达到降低实验成本和工作量的目的。

（3）遗传工程的实施过程中，由于只转移个别外源基因，一般较少或不进行细胞网络分析；合成生物学改变了“转移一个基因，表达一种蛋白质的模式”，而通常是转移一组基因，因而要在更大规模、更多层次上涉及细胞网络（如代谢网络等），所以，网络分析是合成生物学的核心内容之一。

如果说遗传工程是上一代生物技术的话，那么合成生物学则是下一代生物技术。但这并不意味着合成生物学要取代传统的遗传工程技术，而是各有各的用处，两者竞相发展。DNA 序列大规模、高精度和低成本的关键技术目前仍未解决，成为制约合成生物学发展的“瓶颈”问题。目前连合成一个高等真核生物基因组 DNA 序列都是不现实的，因此，合成生物学与遗传工程之间的差异至今为止还无法得以充分体现。

1.6.2 与分子生物学和细胞生物学的关系

分子生物学是指在分子水平上研究生物现象的科学。它通过对生物大分子（核酸、蛋白质）的结构、功能和生物合成等方面的研究来阐明各种生物现象的本质，包括遗传信息及其传递、生物进化、光合作用、发育的分子机理、神经活动的机理和癌症的发生等过程。

细胞生物学是以细胞为研究对象，从细胞的整体水平、亚显微水平、分子水平三个层次，以动态的观点研究细胞和细胞器的结构和功能、细胞的生活史和各种生物活动规律的学科。

分子生物学和细胞生物学的出现使得人类从对生物的宏观分析与观察进入到了微观领域，对 DNA、蛋白质等生物分子的研究结果是合成生物学的重要理论基础。合成生物学需要利用分子和细胞生物学的研究成果设计和重新构造新的生物分子、遗传线路乃至具有崭新功能的细胞，并在此基础上验证和检验分子和细胞生物学的理论。

1.6.3 与系统生物学和功能基因组学的关系

系统生物学就是在细胞、组织、器官和生物体整体水平上研究结构和功能各异的生物分子及其相互作用，并通过计算生物学来定量阐明和预测生物系统的功能、表型和行

为的学科。

系统生物学主要致力于研究自然生物系统整体，以仿真和建模工具作为实验信息的比较手段。随着各种高通量检测技术的发展，系统生物学已经囊括了基因组学、转录组学、蛋白质组学、代谢组学、代谢物组学等组学分析。

合成生物学与系统生物学之间具有非常亲密的关系。定量的系统生物学以工程系统和信号理论的应用为基础分析生物系统，从数学方程和复杂模型角度“定义”系统。一旦系统或系统的部件用系统生物学的方法描述，合成生物学就可以在此基础上将系统解耦成生物部件，将生物部件的功能及输入/输出以标准形式表述，再将部件存入知识库中。任何系统设计者都能够理解部件的功能特性，利用知识库中的部件组装成装置，最终组装成系统[80]。

另外，从某种角度看，合成生物学可以看作是系统生物学新出现的合理补充和验证。合成生物学的关键理念是层次性的模块化方法，即将生物系统看作具有特殊功能的、不同水平的模块（蛋白质、DNA簇、信号代谢途径、转运机制、能量供应模块以及成分、细胞器、细胞膜）间的交互系统。这种方法很具有吸引力，为系统生物学和功能基因组学提供了分子水平所无法提供的、能够将细胞看作系统的分析角度。系统生物学可以利用合成生物学手段识别和编目各种组织内和组织间的模块类型，研究模块自身的工作特性以及彼此之间的交流，并通过建立不同水平的层次结构绘制整个系统。可以说，合成生物学的研究为系统生物学和功能基因组学的研究提供了强有力的验证方法。

另外，系统生物学和功能基因组学在生物系统进化方面的洞察力使其也可以看成是基因组水平合成生物学的基础。基因组水平的合成生物学包括单细胞、多细胞及细胞群体间行为和功能的根本改变和设计。作为这一水平生物工程的先决条件，生物系统的重构和整合、相关理论的发展、数学和软件工具的完善等将会提高对这些系统模拟的准确性和加深对系统理解的程度。而这些先决条件正是系统生物学和功能基因组学追求的目标。系统生物学能够阐明网络拓扑结构的变化、生物学变化导致的生物化学参数值的变化、作为线路进化媒介的分子结构等；复杂而尖端的测量技术使得研究者在“调试”和应用合成生物学设计时更加得心应手。因此从这个角度讲，系统生物学和功能基因组学将会是基因组水平合成生物学的基础[81,82]。

1.6.4 与生物信息学的关系

随着人类基因组计划的完成，人类获得各种生物数据的能力飞速发展，如何从浩瀚的信息海洋中获得有用的知识，避免“信息爆炸-知识匮乏”的现象成为20世纪研究者面临的主要问题，生物信息学应运而生。生物信息学（bioinformatics）是一门数学、信息、计算机与生物学交叉结合的学科。其主要任务是从已经获得的数据中挖掘和抽提有用的“知识”，在人类遗传疾病与功能基因的发现与识别、基因表达与蛋白质功能的研究方面都发挥着关键的作用。其主要研究内容包括：基因组的分析和注释、表达谱分析、基因芯片、蛋白质结构、分子设计等。随着系统生物学的发展，人类已经有能力从基因组、转录组、蛋白质组、代谢组等组学水平研究各种生物功能，生物信息学的应用

也扩展到了与这些组学相应的知识获取方面。合成生物学强调生物系统的设计和优化，需要利用各种数学模型和信息处理工具模拟、预测和优化生物模块的性能，因此生物信息学的发展无疑会推动和促进合成生物学这方面的发展。

1.7 合成生物学国际会议及国际遗传机器大赛

1.7.1 合成生物学国际会议（Synthetic Biology *x*.0）

第一届合成生物学国际会议（Synthetic Biology 1.0）于2004年在美国麻省理工学院（MIT）举行。当时合成生物学刚刚引起人们的重视，还远不及现在影响深远，参会的人并不多，会议的主要问题也只是以如何推进合成生物学的发展为主。

从2006年起，合成生物学国际会议开始扩大规模，邀请全世界范围内所有从事和对合成生物学感兴趣的研究者参加。2006年的Synthetic Biology 2.0在美国的加州大学伯克利分校（UCB）举行，会议以邀请报告的形式开始，并针对合成生物学可能引发的社会问题发表声明。

2007年，为了进一步扩大合成生物学的影响，Synthetic Biology 3.0走出了美国，在瑞士苏黎世联邦理工学院（ETH）召开。

2008年10月，Synthetic Biology 4.0在中国的香港举行。全世界的合成生物学研究者以及对合成生物学感兴趣的学者云集在主办地共同商讨合成生物学方方面面的问题。

1.7.2 国际遗传机器大赛

国际遗传机器大赛（international Genetically Engineered Machine Competition，iGEM）是由美国麻省理工学院为推动合成生物学的迅速发展而于2003年发起的一项以本科生为主要参赛对象的竞赛。比赛中，参赛者被鼓励充分发挥自己的创意，以创造新的生物系统为最终目的，将简单的基因模块植入细胞体构造出复杂的生物组件、装置和系统[83]。两年后，也就是2005年，这个最开始仅仅是MIT校内规模的比赛迅速升级为世界范围的国际比赛。越来越多世人谙熟的著名大学——哈佛大学，剑桥大学、加州伯克利大学和普林斯顿大学等对此表现出极大的热情，这极大地推动了iGEM在学术界的国际影响力。同时，iGEM也不断为合成生物学吸引新的人才、注入新的力量，促使这个领域更快、更好地发展。

2006年，合成生物学在天津大学等高校的大力推动下“登陆”中国。对于这个崭新的学科来说，iGEM大赛无疑是推动其在中国发展的主要契机，也是一个与国际前沿研究方向接轨的机遇，为中国精英人才的培养和高等教育的国际化提供了更为广阔的空间。2007年11月，天津大学、清华大学、北京大学和中国科学技术大学等中国内地代表队首次赴MIT参赛，取得了令人瞩目的成绩。2008年和2009年中国内地参赛队不断增加，继续取得优异的成绩。

第二部分 合成生物学引发的思考

1.8 人类？造物主[76,84~89]

与20世纪七八十年代遗传工程的相关争论非常类似，合成生物学一经出现立刻引起了社会方方面面的广泛关注。

尽管科学家的研究初衷都是好的，防御系统也是极其严密的，然而，不可否认，生物系统的复杂性远远超出人们的预料，它可以进行自我修复、自我管理，而且能发生突变以适应环境。人们不禁担心，随着“造物技术”的不断发展和广泛应用，在生物合成监管的相关规定还没有出现以前，管理疏漏或技术偏差会不会导致新物种失去控制，演绎现实版的“生化危机”？恐怖主义分子会不会利用合成生物技术制造致命病毒或生化武器？

正如人类对于所有未知事物的原始恐惧一样，合成生物学究竟有怎样的前景？是造福人类的天使还是毁灭人类的恶魔？合成生物学的大门后面是通向造物主的后花园还是通向一直隐藏在噩梦中的地狱？人类是否应该又能否扮演好造物主的角色？这些问题恐怕是连狂热的预言家也无法想象和回答的！

所幸，合成生物学界从未对合成生物可能对人类造成的威胁避而不谈，相反，相应的道德规范和防范措施已经成为所有相关国际会议的主要议题之一，几乎任何一篇对于合成生物学的综述都指出需要道德规范、国际调控和实际行动。第二届合成生物学国际会议（Synthetic Biology 2.0）也已经发表宣言支持科学界采取维护人类安全行为的策略。

概括而言，合成生物学被科学家、非政府组织、媒体和公众所广泛关注的道德、社会、法律方面相关的问题大体可归结为：道德规范、对环境无控制的排放、生物武器的滥用、生物垄断及人工生命的创造等几方面。

1.8.1 道德规范

对于这一新兴学科道德规范的争论与20世纪七八十年代遗传工程的相关争论非常类似。合成生物学界需要自我完善、制定道德和行为规范，共同努力成为一个安全的、具有社会责任感的团体。

为此，许多政府机构和研究机构都采取了相应措施。在美国，Laurence Berkeley国家实验室和MIT都启动了包括道德规范在内的合成生物学研究项目。2005年美国Alfred Sloan基金授予MIT 570 000美元从事合成生物学可能引起的道德问题方面的研究。在欧洲，2007年，EC-FP6 Programme（欧盟第六框架计划）资助NEST（New and Emerging Science and Technology）从事“Synbiosafe：Safety and Ethical Aspects of Synthetic Biology”的项目研究，给予该项目236 000欧元的基金支持。作为对于合成生物学道德方面的回应，这一项目希望提供风险评估、安全、道德、知识产权、利润享有人和公众利益等方面的建议[90]。

此外，一些非政府组织和公众社团组织也发起了对于合成生物学的社会、安全、道德关注，其中包括比较著名的 ETC 技术监督组织（The Action Group on Erosion, Technology and Concentration）、地球之友（Friends of the Earth），以及其他诸如 Bulletin of Atomic Scientists 等致力于维护和平和安全的组织等[48]。

1.8.2 对环境无控制地排放

此方面关注的核心是合成生物体有意或无意地排放到环境中所造成的影响。

2002 年 9 月，E. Wimmer 小组合成出了脊髓灰质炎病毒（poliovirus）的全基因组；2003 年，J. C. Venter 研究小组用了 14 天时间从头合成了噬菌体 ϕX174（5386 bp）基因组；2004 年，美国等研究人员合作，人工合成了 1918 年造成全世界上千万人死亡的“西班牙流感病毒”。这些生物学“壮举”加深了世人的焦虑。

生物是进化的机器，需要通过进化和基因突变适应自然选择。合成生物体的基因组在自然选择的压力面前自然也有突变的可能，这就意味着合成生物的确有可能因为突变而产生与其他生物或者环境之间意想不到的相互作用[91]。

（1）人工合成的新物种具有非自然存在的崭新功能的 BioBrick，因而可能具有战胜自然生物体的能力，在自然选择中获得优势而无限增殖，从而彻底改变生态系统。

（2）清洁某一特殊化学物质的微生物体可能与其他微生物体相互作用，将合成基因传入自然物种中而“污染”自然界的基因库。

（3）人工生物体可能与自然界现存的物质发生作用，导致意外的副效应。

如此种种，正如“反纳米”组织和 P. Charles 等人所担心的那样，纳米技术造成的“Grey Goo”* 可以用毫无固定形态的方式杀死所有居住于银河系的生物，而 DNA 分子正是纳米级物质！

合成生物学的潜在能力甚至能把整个生化途径工程化，其导致的复杂变化使得现有的、已经建立的风险评估方法力不从心。因此，这些挑战可能还将引起评估方法和评估过程的重大变革。

现阶段也许我们还可以说，无论是逃出还是特意释放出去的合成生物，一旦脱离实验室的环境就会立刻死亡。然而如果我们最终想用合成生物进行临床和工业应用的话，我们必须考虑更多有关生物安全的问题，否则合成生物学可能无法逃脱转基因产品的厄运。

这方面关注的话题与遗传工程曾经面对的质疑非常相似。科学家们相信，只要及时制定、严格执行并随时更新相应的限制措施，这方面的危险就完全能够避免。相应的自律和监督策略可以包括以下几个方面。

---

* “纳米技术”（nanotechnology）的概念首先是由美国的未来学家德雷克斯勒（E. Drexler）提出的。德雷克斯勒也是“灰色黏质”（grey goo）这个词的创造者。1986 年，他在一本名叫《造物引擎》（*Engines of Creation*）的书中描述了能够进行自我复制的纳米尺度机器人。这种机器人可以通过移动单个原子制造出任何人们想要的东西（土豆、服装或者是一块计算机芯片）而不必使用传统的制造方式。所以德雷克斯勒担心，这样一种能够自我复制的机器人可能会失去控制：他们疯狂地复制自身，在很短的时间内就把地球变成了一大团完全由纳米机器人组成的“灰色黏质”。

• 加强和普及对合成生物学潜在风险的教育。

• 对于任何合成生物部件和生物体，制订详尽、准确的产品安全手册。

• 建立交换平台，共享对于优良习惯和紧急情况处理程序的建议。

• 加大各国及国际生物安全委员会的权限、审核范围，强化对生物安全指导方针的执行力度。

• 为合成生物学新实验和额外风险建立国家顾问团，提供全面、专业的顾问建议。

除此以外，还可以采用许多技术手段来防止合成生物学技术的滥用。例如，将合成生物体进行改造使其只有依赖于常规自然界不能存在的特殊营养物质才能生长繁殖；给合成生物体连接一个自毁装置，在菌体的密度达到特定阈值或繁殖特定代时，种群会自行毁灭。将合成基因组中某一密码子全部替换使得遗传密码中包括非天然氨基酸部分[92~94]；将质粒所有的某一遗传密码子全部替换成另一密码子，则有可能允许在质粒中非天然氨基酸的插入，从而对染色体密码子无任何竞争能力[93~95]；将全部的非同义密码子关闭，防止与自然物种的遗传物质进行功能交换，从而防止由水平基因转移导致的环境污染[96]。这些措施都可大大降低合成生物在自然界的竞争力。

1.8.3 生物武器的滥用

合成生物学在代谢途径方面取得的成果使得将特定的相关代谢基因组合植入大肠杆菌和酵母菌中继续发挥功能成为可能。这样，微生物就可以作为生物工厂来生产天然的蛋白质，自然也可能被用于生产如蛇、昆虫和蜘蛛的毒液、植物毒药甚至霍乱等疾病的细菌毒物。随着 DNA 合成和测序价格的不断降低及速度的不断提升，任何个人和组织均有可能具有合成 DNA 序列的能力。这使得人们相信生物黑客运用合成生物学技术作为武器似乎已经指日可待[97]，这无疑给生物武器的监察者带来了新的挑战。

美国国家科学顾问委员会（National Science Advisory Board，NSAB）合成基因工作组提出了一份关于合成基因组应用双重性的报告，认为这一领域的科技和知识不仅可以用于维护，也可被“错用”来危害公众健康和国家安全[84]。

2003 年，美国中央情报局（Central Intelligence Agency，CIA）对生物武器未来发展的评估是：“工程化的生物制品可能比任何人类已知的疾病都可怕”，并认为基因组革命所带来的巨大进步使得常规的监控杀伤性武器的方法已经不足以有效监控未来的武器了。为了确保情报服务仍旧能够准确掌握生物工程潜在的应用，这个报告建议与生物科学团体保持更加密切的联系。

英国国防部认为合成生物学有可能影响未来军事能力，2006 年，英国安全科学顾问委员会（Defense Science Advisory Council，DSAC）同意立项调查这一领域所造成的军事机遇和威胁[84]。

而合成生物学家则普遍认为目前主要的威胁还是来自国家水平的生物细菌战计划。他们指出：合成生物学在制造病原体方面还有着巨大的技术障碍，在目前技术水平上利用合成序列构建整个全新病原体不太可能，构建如上面讨论的脊髓灰质炎病毒等已知病原体则相对容易些。尽管这些相对简单的生物体能够被人工制造出来，但却很难成为“武器”，因为合成病毒所用的成分是常规的，且是不具危害性的。此外，合成脊髓灰质

炎病毒项目的首席科学家 E. Wimmer 即使在媒体和其他科学家的批评下仍然捍卫着这项研究，他认为这项研究不会助燃生物恐怖主义，却强调和推动了疾病疫苗的研究工作。同样，“西班牙流感病毒”的研究者们也始终强调他们的研究有助于深入理解病毒机制，帮助科学家更广泛地了解致命流感病毒。

作为这方面问题的应对，除了 1972 年生物及有毒武器公约（Biological and Toxin Weapons Convention，BTWC）严格限制了生物武器的开发、生产和储备之外，还需要从以下四个方面进行预防。

（1）基因公司和寡核苷酸生产商必须改进筛选软件，筛选并核实订单的安全性；DNA 合成者必须保证合成装置注册和申请许可证；购买试剂和相关服务的客户也必须具有许可证；科研团体应该加强在这方面的要求和规范；相关公司应该互相协作以核实危险序列是否以分割成多个订单的隐藏形式订制。

（2）建立正式规范、伦理准则、咨询顾问等专业实体。建立道德咨询委员会，通过道德热线等方式，确保人们能够在无需担心机密泄漏的情况下很方便地得到此方面的专业建议。

（3）开发和研制识别合成生物体创建者的条码。加强地方、国家、世界级范围的审核，建立国家顾问委员会监督用户和科研机构；建立道德监督机制，相关的成员能够汇报潜在的危险。

（4）确保开发和使用只有在自然界无法生存和传播的、本质安全的合成生物体。

1.8.4 对生物的专利与垄断

合成生物学知识产权方面的主要难题是专利命名范围的宽窄问题。专利范围太宽可能会限制协作并抑制这一领域的发展；而范围太窄又会使研究过程变得太复杂，利用标准化部件仅仅制造一个系统就可能涉及几百个专利。

1980 年，“戴雅蒙对查克那伯迪”（Diamond vs. Chakrabarty）的著名案例中，美国最高法院裁定一种能消化油脂的细菌有机体可以成为专利品。这次判决开创了生物可以成为专利品的先例。从此以后，生命工业（Life Industry）公司开始了对微生物、基因、细胞、器官、胚胎等基因资源和转基因动植物品种的专利控制，这就是“对生命的专利”的由来。

近 30 年来，生物学界对专利权的热衷也影响到现在的合成生物学界。根据 1980 年美国最高法院的规定：只要一个人设计的并非是自然界已有的生物就可以申请专利，没有经过改造的微生物则是不能申请专利的。

合成生物学是纳米技术、生物技术、工程学和信息技术的结合产物[98]，很容易找到合适的领域来申请专利。对知识产权的申请不仅仅会涉及生物学，同时也会涉及计算机和软件等。

引发围绕合成生物学专利与垄断问题的导火索是 J. C. Venter 关于“Synthia”的专利。他不仅在美国申请了最小活细胞的专利，还通过世界知识产权组织（World Intellectual Property Organisation，WIPO）申请了国际专利。2007 年 J. C. Venter 又提出合成基因组（UPSTO No. 20070264688）以及将其放入细胞（20070269862）的专利申请。

如果 J. C. Venter 关于最小基因组专利的申请获准，那么他很有可能垄断生物燃料的供应市场。J. C. Venter 曾对《星期日泰晤士报》（*Sunday Times*）宣称："很明显，如果我们制造出能够生产燃料的生物体，那它将是首个价值十亿或万亿美元的生物体"。

专利之争由来已久。1998 年 5 月，J. C. Venter 等建立的塞莱拉公司（Celera Genomics）介入打乱了"人类基因组计划"的原有步调。J. C. Venter 声称，要在 3 年内完成人类基因组的序列测定，目的是抢在"国际人类基因组计划"前完成，以便将人类基因组图谱申请成专利。为阻止人类基因组专利落入 J. C. Venter 之手，2000 年 3 月 14 日，当时的美国总统克林顿和英国首相布莱尔联合发表声明，宣布两国政府都支持把基因组数据向全世界免费公开，使 J. C. Venter 想垄断基因组专利的企图受挫。2000 年 6 月 26 日，时任美国总统的克林顿在白宫郑重宣布，"人类有史以来制作的最重要、最惊人的图谱——人类基因组草图完成"。站在克林顿身边的有两位科学家，其中的一位就是 J. C. Venter。

另一个著名的生物专利之争发生在 2002 年。当时瑞士先正达（Syngenta）公司（世界最大的农业化肥生产公司）写了一份长达 323 页的、关于大米基因组研究的专利申请书，在申请书中他们声称要垄断性地控制对于大米种植和其他一些农作物种植起关键作用的基因序列。这个专利的范围史无前例，不仅涉及至少 23 种主要粮食农作物，还扩展到基因序列的计算机可读序列信息方面。虽然这项申请最终在 230 个社会组织的反对下被否决，但却揭示了未来生物学专利的涵盖范围：数字化的 DNA 序列以及相应处理软件。

应对合成生物学的专利与垄断，MIT 启动了"生物积块"项目作为避开广义专利对于研究者限制的方法。2004 年仅有 100 个生物积块在网上注册，而到 2008 年已经扩展到了将近 2000 个。生物积块基金会（BioBricks Foundation，BBF）是一个非营利组织，其目的就是为了确保注册的生物积块对广大科研工作者是免费的。为此，BBF 还注册了"BioBrick"和"BioBricks"商标，通过构建网络"生物积块库"的方式实现"生物积块"的共享。"生物积块"的使用率和点击率能够向科研工作者提供一些关于"生物积块"的侧面信息，研究者还可以在网上分享"生物积块"的使用技巧和经验等。这一项目受到了全球合成生物学界的普遍支持。

此外还有人提出申请"Clickwrap"许可的办法，所有专利项目可以在拥有这项证书的会员中免费流通，避免其向非许可用户的不正当扩散。

1.8.5 创造人工生命？

合成生物学家不否认其研究目标之一是创造对人类有益的"人工生命"。显然，这激起了绝大多数人的恐惧。从世人的角度看，合成生物学家正在创造"造物主"或者"进化法则"不敢创造的生命形式。科学家要"扮演上帝"？这激发了世人关于自然生命本身及其创造过程的深层次的哲学与宗教思考。

引起这方面关注的主要是合成生物学制造的"类生命"细胞。许多科研小组以创造"类生命"细胞为主要研究课题，例如，由 13 个合作研究机构组成的"PACE"（Programmable Artificial Cell Evolution），称他们的研究目的就是拉近二元（逻辑"0"和

“1”）世界和生命世界的距离。

所幸这方面的争论并不是合成生物学特有的，许多生命科学的分支同样面临着这样的争论。正如目前“生物”和“生命”的定义本身也无法准确界定它所描述的客体一样，无论是合成生物学家还是世人，对于生命是什么、生命的起源、生命的特征以及生命的复杂性方面都非常困惑。

1.8.6 究竟该如何应对?

对于合成生物学的质疑还包括以下方面：廉价的微生物合成橡胶一旦成功是否会颠覆本就相对落后的东南亚经济？人类是否应该利用合成生物学技术重新“复活”已经灭绝的古生物？

生物工程是21世纪的核心学科，其发展依赖于现有科技发展的速度和广度。从这个意义上讲，合成生物学似乎没有任何全新的东西，分子生物学、重组DNA和遗传工程都已经讨论过类似的问题。关于生物安全、生物武器及生物工程的道德问题也都是20世纪七八十年代就开始的论题了。然而合成生物学毕竟还有一些非常重要和全新的东西，它从某方面反映了科学界从事生物工程的信心，标志着计算科学、材料科学、纳米技术等一系列有关科技的成熟。

作为合成生物学研究者应该怎样面对这些问题？

首先，必须承认维持公众认同和支持的重要性。1990年前后，欧洲对于转基因食物论战导致的公众对于基因修饰植物支持的下降就明确说明了这一点。为了避免合成生物学研究发展得太快而远远超出公众的承受能力，应该明确展现潜在的应用有利社会的一面；不应过高估计这一科技的潜在益处而造成公众的不信任，更不应造成公众的过分焦虑；应该寻求有效地增强与民间社会组织、社会科学家和伦理学家之间互相理解的方式，以便赢得公众和科学领域的支持。遗传工程的历史也告诉我们，这种争论会是漫长而艰巨的，但也是必要的，因为科学、技术、调控框架和社会含义是在任何一门学科成熟的过程中协同发展的。

其次，科学界必须在合成生物学所引起的种种争论中起到带头作用，等到实际应用已经引起广泛关注时才做出反应就已经太迟了。公众争论并不是线性或理性发展的，而是具有不可预测性的，由深层的对待自然、环境和科学的态度等多种因素所影响的，因此预先合理干涉公众关注就显得至关重要。

最后，在合成生物学应用真正实现之前必须要有强有力的监管机制。这可能需要现有监管机制的协调统一和建立新的监管体系以维护公众安全。从地方到国家再到世界范围各级别的调查局、监管委员会要真正担起责任。

过去的三十多年，生物科技发展的一个主要经验是：尽管某一科技具有很大的潜在的社会和道德危害性，但人们还是应该清醒地认识到，虽然科学的发展是迅速的，但将科学研究从实验室转化为现实世界的应用其实是要付出巨大努力的。合成生物学是推动还是危害社会和人类，应该通过对成功创造一个新的生物有多困难的现实，来进行理性判断和调和，而不应该武断地、草率地做出论断。

我们需要清醒地认识到，任何法律和道德均无法约束恐怖分子和别有用心的人，如

果合成生物学是一把双刃剑，我们能够放弃对合成生物学的研究，而让这把剑仅仅被恐怖分子和别有用心的人掌握么？

中国有句古话，“知己知彼，百战百胜”。战胜各种危机的方法，绝对不是避而不谈或不去碰触，而是尽快地完全掌握所有相关知识和技术。只有完全掌握了基因复制、重组等相关技术，才能在新的病毒和有害生物出现之后迅速利用合成生物学技术破坏其复制和侵害功能，真正提高人类抵御侵害的能力。

可以预见，未来合成生物学仍然会在人类激烈的争论中继续前进！无论合成生物学是一把多么锋利的双刃剑，这把剑能否伤人终究要取决于握着剑柄的人类……

人类的 DNA 序列是人类的真谛，这个世界上发生的一切事情，都与这一序列息息相关。

——杜尔贝科

（1975 年诺贝尔生理学或医学奖得主）

第 2 章　合成生物系统的设计

基础提要

- 遗传学
- "DNA 重组技术"
- 控制理论基础

在前面的介绍中，我们曾经多次提到“模块”这个概念。虽然“遗传工程”与“DNA 重组技术”的概念已经出现了很多年，然而大多数生物工程学家所进行的生物相关研究，却与工程技术鲜有共同之处。其中一个很重要的原因就是，现有的生物工具，在标准化和模块化方面还远没有达到与其他工程技术领域相应的水平。

众所周知，IT（information technology）行业的人员在封装、接口、标准化和模块化方面的不懈努力是推动电子工业产量和质量飞速发展的、举世公认的因素。人们在组装计算机、控制系统甚至大规模信息网络时只要设计标准化零件的逻辑关系即可，再不需耗费任何时间和精力去开发这些零件本身及零件的连接方式。与计算机革命相比，今天的合成生物学最不利的因素就是这些模块、标准化零件和接口等工程化基础的欠缺。因此，合成生物学家在各种生物元件的标准化、模块化方面做了大量的探索性工作。本章我们就来介绍这些支撑合成生物学工程化本质的标准化元件、规范和概念。

第一部分　合成生物系统的模块化与层次化结构

如果说晶体管是电子电路的基本组件，那么在生物学中，与之对应的便是基因。无论是崭新功能的蛋白质信号转导路径，还是合成崭新代谢物的代谢通路，最终都要通过对 DNA 序列的设计来实现。借鉴电子技术与程序设计的模块化和标准化设计来实现生物元件的组装、标准化接口设计等，把各种基因部件设计成标准模块，其他使用者仅仅

需要“搭积木”的技术即可以构建基因线路，将研究者从繁重的劳动中解脱出来专注于新生物体的设计和创造，这一切才是合成生物学的理念。

2.1 合成生物学的模块化设计

在合成生物学中我们称具有标准接口、功能相对独立的生物大分子、信号转导路径、基因线路等为“模块”。模块的规模可大可小，小到具体的启动子、终止子，大到单细胞、多细胞及细胞群体系统。可以说，模块化设计和构建是“标准化、解耦和抽提”三个概念的综合运用，最大限度地体现了合成生物学的思想精髓，是合成生物学的标志性内容。

所谓模块（module），必须具有以下几种基本属性：接口、功能、逻辑、状态。其中，接口、功能与状态反映模块的外部特性，逻辑反映模块的内部特性。每个模块完成一个特定的子功能，所有的模块按某种方法组装起来，成为一个整体，完成整个系统所要求的期望功能。

模块化的目的是为了降低合成生物系统设计的复杂度，使实验设计、验证和优化等操作简单化。所谓的模块化系统设计不只是简单的逐个基因的连接，而是首先利用“解耦”的思想，将整个系统按照某一标准（功能、时间等）进行分割，分割成相对独立的子系统；以“标准化”的思想定义和验证各子系统之间的输入、输出连接关系；利用逐步细化的“抽提”方法得到一系列以功能模块（子系统）为单位的算法描述。在设计好整个体系结构后，我们就已经在宏观上明确了各个模块应具有什么功能，应放在体系结构的哪个位置；再由小到大逐层验证和优化模块、子系统乃至系统的功能；最后通过试验加以实现。

保持“功能独立”是模块化设计的基本原则。只有“功能独立”的模块才可以降低开发、测试、维护等阶段的成本。但是“功能独立”并不意味着模块之间绝对的孤立。一个系统要完成某项任务，需要各个模块相互配合才能实现，此时模块之间就要进行信息交流。例如，把鼠标和键盘当作两个“功能独立”的模块，没有鼠标时，键盘照样能干活；没有键盘时，鼠标仍可以运作；但如果希望快速灵活地输入，那么就需要鼠标和键盘的配合。可以这样认为：“功能独立”的模块具有各自的作用，但只有当所有的模块完美配合，系统的功能才能以最佳状态最大限度发挥。因此在设计一个模块时不仅要考虑“这个模块提供什么样的功能”，还要考虑“这个模块应该怎样与其他模块进行信息交流”。

2.1.1 模块设计的三个特征因素

进行模块化设计和评价模块设计优劣的三个特征因素主要为：“信息隐藏”、“内聚-耦合”和“封闭性-开放性”。

1. 信息隐藏

为了尽量避免某个模块的行为干扰同一系统中的其他模块，在设计模块时就要注意信息隐藏。信息隐藏是指在设计和确定模块时，一个模块内包含的信息对于不需要这些

信息的其他模块来说是不能访问或知晓的。独立的模块彼此之间仅仅交换那些为了完成系统功能所必需的信息，而将其他一切内容“隐藏”起来。模块的信息隐藏可以通过接口设计来实现，一个模块仅提供有限个接口，充当模块与模块之间必不可少的信息交流媒介和通道。在生物系统中，接口不是电气工程里面的插头，也不是软件设计里的入口参数，而是各种信号分子、蛋白质、RNA 等生物介质。因此，生物模块的信息隐藏可以理解为尽量不产生不必要的信号物质，尽量将信号物质快速降解以避免对其他模块产生干扰，等等。

2. 内聚-耦合

内聚（cohesion）是一个模块内部各成分之间相关联程度的度量；耦合（coupling）是模块之间依赖程度的度量。内聚和耦合是密切相关的，与其他模块存在强耦合的模块通常意味着弱内聚，而强内聚的模块通常意味着与其他模块之间存在弱耦合。模块设计追求强内聚、弱耦合。通俗地说，就是要增加生物模块内部组分间的依赖性，削弱模块与模块之间的依赖性和相互作用。任何必须保留的相互作用，在理论上都要具有可预测性；对相互作用的特征化和标准化是保证模块可移植性和相对独立性的主要手段。

需要指出的是，生物系统由于自身的特性而具有区别于其他工程领域的模块化设计。合成模块和宿主细胞的内在过程具有一定的依赖性，会互相约束彼此的行为。宿主细胞过程的任何波动都会传播到模块并影响其功能；反之，模块的波动也会影响宿主细胞的过程。这给工程化生物系统的可预测性、可靠性带来了不利影响。一般来讲，模块之间以及模块与宿主细胞之间的耦合作用越少，越有利于插入模块功能的发挥。如果可以做到宿主细胞在保护模块不受胞外环境影响的同时仅仅只对过程资源产生作用，则理论上模块功能即可以独立于细胞本底环境。

3. 封闭性-开放性

如果一个模块可以作为一个独立体被应用，则称模块具有封闭性；如果一个模块可以被扩充，则称模块具有开放性。

从字面上看，让模块具有“封闭性-开放性”是矛盾的，但这种特征是客观存在的。当着手一个新问题时，我们很难一次性解决问题。应该先综观问题的一些重要方面，同时做好以后补充的准备。因此，让模块存在“开放性”并不是坏事情。然而，“封闭性”也是需要的，因为我们不能等到完全掌握解决问题的信息后再用模块。这方面的典型例证就是 IT 行业中，无论是计算机硬件的设计，还是网络通信协议等软件的设计，都会给未来的发展留有余量，即为现有模块的扩展留有余量。目前对于生物模块的研究，主要集中在“封闭性”上，要求模块对不希望的输入信号具有“绝缘”的能力，从而保证自身功能不被干扰。基于“开放性”本章后面的内容中会介绍合成生物学标准生物模块——生物积块“BioBrick”[99]，读者可以从生物积块独具匠心的前缀、后缀设计中看到合成生物学家为模块具有“开放性”的属性所做的努力。

2.1.2　细胞本底环境对模块化设计的影响

对于生物系统而言，模块无法像电容、电阻等电子器件模块那样独立存在。细胞中

模块的功能不仅取决于它的组分装置和连接性，也取决于其运行的细胞本底环境，如DNA、RNA新陈代谢、可用的氨基酸、ATP水平、蛋白质合成、细胞的周期与分裂、与外源模块相互作用的内源信号通路等。相同的合成生物系统在不同的宿主细胞中也会表现出不同的行为。另外，新模块的整合及其功能也会对宿主细胞过程产生根本性影响，从而改变细胞本底环境，而本底环境又会反过来改变模块的行为。因此，合成生物学在定义功能模块时，不仅需要对特殊生物功能模块进行标准化定义，还需要对本底细胞元素的状态进行标准化，充分考虑模块与宿主细胞本底之间的联系。

第一章所述解耦的方法，虽然在分析生物系统方面非常有效，但将复杂系统分解为许多相对简单子系统的同时也可能会丢失许多简单元素间及元素与细胞本底间的基本连接信息。将模块从细胞背景中分离，近似完备的概念化描述并不能充分定义模块的功能，宿主细胞信息的隐藏也会带来某些不利影响。因此，利用抽提和解耦方法对生物系统进行模块化分析和设计时必须充分考虑具体的细胞本底环境。只有将细胞本底背景的概念融入模块功能中，才能使模块组分发挥作用。

生物装置无法脱离模块，模块无法脱离细胞，细胞也无法脱离细胞群体。这些环境依赖是生物系统的本质特征，如果善加利用即可以变这些所谓的“生物工程障碍”为彼此的桥梁。

2.1.3 无细胞合成生物系统

为了克服细胞本底以及细胞自身内环境的影响，一些学者正在研究无细胞环境下合成生物系统的组装原则。无细胞表达系统建立在细胞提取物基础上，不仅提供了必要的替代体内蛋白质合成的方法，同时还具有许多优势：基因和聚合酶浓度的可控性、报告物测量的量化，以及可供研究的大范围参数空间，并适合于高通量方法、大范围蛋白质产量和蛋白质体外进化等。近来，已经从近100种纯化组分中再生了无细胞转录-翻译系统[100~102]。

2.2 标准化生物模块——生物积块[73,99,103]

为了克服常规基因操作中繁琐的切、连、转、筛，即DNA片段的分离、体外连接、导入受体细胞和筛选的过程，以及从成百上千种限制性内切核酸酶和底物中进行选择的困难中解脱出来，更加灵活、高效、方便地使用DNA元件，合成生物学家创造性地提出了生物积块（BioBrick）的概念，并构建了相应的DNA元件文库——iGEM Registry。

像建筑业的砖和IT行业的零件一样，生物积块也有大有小。小型的生物积块通常是具有一定功能的DNA片段，就是后面我们要讲到的组件（Part），例如，一个RBS或者一个终止子等，几十或者几百个bp；稍大一些的可以是由几个Part组成的基因调控线路，就是我们常说的Device；再大些可以是由调控线路组成的级联线路、调控网络，甚至调控系统（System）。只要经过标准化处理、具有标准的酶切位点，都可以称其为生物积块。

生物积块的标准化具有很多优点。

（1）生物积块是标准化的生物模块，种类多，相互之间容易连接，可供选择的余地大。

（2）相当多的生物积块经过遗传工程手段的改造和实验的检验，在模式菌等菌种中具有很好的生物功能，这就克服了直接从自然生物中克隆基因所必需面对的异源表达问题。

（3）标准化的酶切位点省去了寻找和优化限制性内切核酸酶、连接酶等 DNA 重组工具的繁琐工作，大大节省了时间，提高了效率。

（4）标准化的描述文件和分类的方法，为使用者迅速找到理想的模块提供了便利。

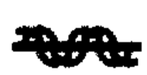
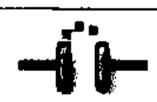
（5）标准化的动力学参数模拟、载体和宿主背景，为生物模块的功能预测奠定了参考、比较和优化的平台。

（6）iGEM Registry 提供的交流平台和文献资料中有许多使用者对于生物积块的宝贵经验共享，可以互相取长补短。

2.2.1　BioBrick 命名规则

为了方便使用，任何一个工程领域的零件库都会有完备的零件规格、功能和使用说明，向使用者提供必要的零件信息，同时也是一种零件描述的规范。作为一个新兴的工程化学科，合成生物学对于自己的零件——BioBrick，同样也有相关的定义和描述。为了方便研究者查阅 BioBrick 的功能，iGEM Registry 对其中的每一个 BioBrick 都有详细的注释，包括该片段的示意图、碱基顺序（不包括前缀和后缀）、片段的设计者对于该片段功能的阐述，以及其他使用者提供的使用经验等。表 2-1 所示为部分常用 BioBrick 的图示及功能描述。

表 2-1　部分常用 BioBrick 图示及功能

图标	说　明
	终止子部件（terminators）
	DNA 部件（DNA），包括克隆位点，SCAR 基因、重组位点、接合转座子等
	蛋白质编码部件（protein coding parts）
	报告/显示部件（reporters）
	信号接收和发送部件（receivers and senders）
	引物部件（primers）
	转换部件（inverters）
	启动子部件（promoters）

注：本表截取自 http://partsregistry.org/catalog（2009 年 11 月 4 日）。

同样的，质粒也有一套描述体系，有兴趣的读者可以参见 iGEM Registry 的开放网页 http://partsregistry.org 获得更详细的信息。

每年 iGEM Registry 都会向全世界所有参加“国际遗传机器大赛（iGEM)”的队伍免费提供最近更新的所有 BioBrick，供参赛队使用。参赛队也会免费向 iGEM Registry 上交自己队伍新创造的 BioBrick 和规范化的描述性文字。同时，这个 iGEM Registry 也得到了全世界合成生物学界的普遍支持，其他科研机构和个人也会贡献自己的力量。籍由 iGEM Registry 这个平台，使用者们互相交流经验和心得，共同为合成生物学的发展而努力。

虽然相对于所有物种庞大的基因库来说，目前已经标准化的生物积块的数量仍是九牛一毛，这个刚刚搭建的开放性的技术平台还需要海量的 DNA 资源为其充实，但我们有理由相信，随着科学家的不懈努力，这个基因库将会越来越丰富、越来越完善。

2.2.2 BioBrick 的连接

iGEM Registry 中 BioBrick 的标准化体现在每一个 DNA 模块的结构都是标准化的：除了本身的功能序列以外，它们都具有相同的前缀和后缀，每一个 BioBrick 的前缀中都包括 *Eco*RI 和 *Xba*I 两个酶切位点，后缀中包括 *Spe*I 和 *Pst*I 两个酶切位点，并且经过特殊的遗传工程手段处理，确保真正的编码序列中不含有这四个酶切位点。整个生物积块被克隆在由 iGEM 组委会提供的质粒载体上，可按照设计的需要剪切和拼接。

BioBrick 前缀的碱基序列一般为：

```
5′ GTTTCTT  C  GAATTC  GCGGCCGC  T  TCTAGA  G  [part] 3′
3′ CAAAGAA  G  CTTAAG  CGCCGGCG  A  AGATCT  C  [part] 5′
     (1)   (2)  (3)      (4)     (5)  (6)   (7)   (8)
```

下面的一条序列为设计 PCR 引物时的碱基序列，各片段的功能如下。

- 片段(1)：额外的间隔碱基序列，以利于用 *Eco*RI 酶切后进行 PCR 扩增；促进利用 *Taq* 聚合酶进行 PCR 时反向链末端 A 碱基的添加，从而保证 TA 克隆的高效性。
- 片段(2)：随机附加的间隔碱基。
- 片段(3)：*Eco*RI 识别位点。
- 片段(4)：*Not*I 识别位点。
- 片段(5)：附加碱基，防止切割/重组过程中意外出现 *Eco*BI 或 *Eco*KI 甲基化位点而抑制 BioBrick 酶的酶切作用。
- 片段(6)：*Xba*I 识别位点。
- 片段(7)：附加 G 碱基，防止意外出现 GATC 位点，GATC 位点在某些应力的作用下可发生甲基化而抑制酶切作用；防止意外出现的 ATG 起始密码子。
- 片段(8)：大约 20 bp 的、与预构建组件 5′端匹配的序列。

BioBrick 后缀的碱基序列一般为：

```
5′ [part] T ACTAGT  A  GCGGCCG  CTGCAG  G  AAGAAAC  3′
3′ [part] A TGATCA  T  CGCCGGC  GACGTC  C  TTCTTTG  5′
     (1)    (2)    (3) (4)      (5)    (6)  (7)
```

下面的一条序列为设计 PCR 引物时的碱基序列，各片段的功能如下。

- 片段(1)：大约 20 bp 的、与预构建组件 3′端匹配的序列。
- 片段(2)：*Spe*I 识别位点。
- 片段(3)：附加碱基，防止切割/重组过程中意外出现的 *Eco*BI 或 *Eco*KI 甲基化位点而抑制 BioBrick 酶的酶切作用。
- 片段(4)：*Not*I 识别位点。
- 片段(5)：*Pst*I 识别位点。
- 片段(6)：随机附加的间隔碱基。
- 片段(7)：额外的间隔碱基序列，以便用 *Pst*I 酶切后进行 PCR 扩增；促进利用 *Taq* 聚合酶进行 PCR 时反向链末端 A 碱基的添加，保证 TA 克隆的高效性。

有了上述四个标准化的酶切位点之后，需要组装的部件可以分为插入片段和载体两部分。插入片段由限制性内切核酸酶处理以后可以从载体上切割出来，通过琼脂糖凝胶电泳分离回收后可得到纯度足够高的插入片段。载体经酶切处理后可打开一个小口，并留下两个黏性末端，处理过的载体可经乙醇沉淀或琼脂糖电泳纯化。

下面我们就来具体讲一下连接方法[99]。

当我们需要将目的片段 1 插入到目的片段 2 的左侧时（图 2-1）。

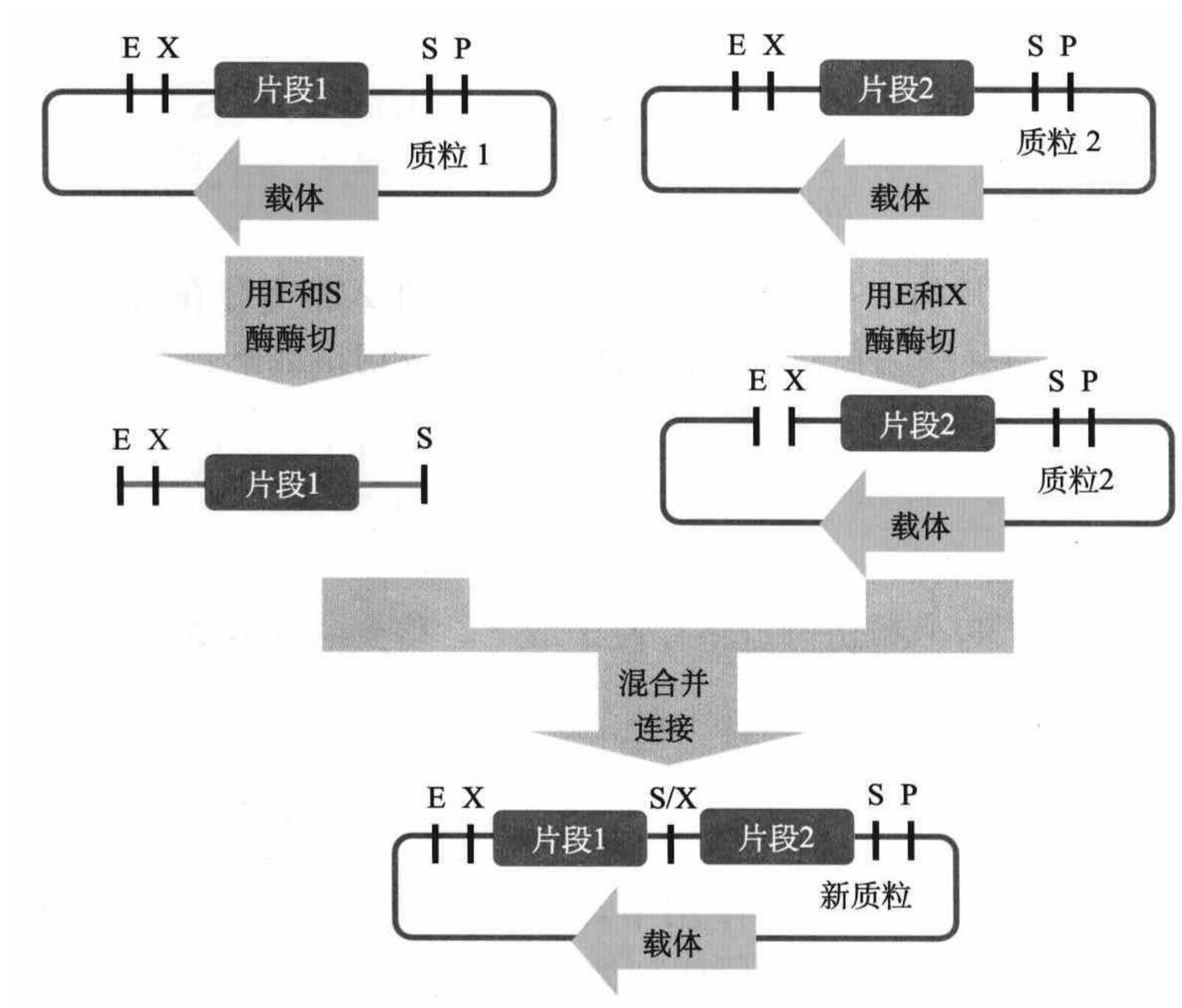


图 2-1　BioBrick 的连接方法

将目的片段 1 插入到目的片段 2 的左侧。首先利用 *Eco*RI 和 *Spe*I 两个酶切位点将片段 1 切割下来；利用 *Eco*RI 和 *Xba*I 两种酶去掉它们之间的序列，在含有片段 2 的质粒上形成一个切口。之后将两种片段混合，在连接酶的作用下，片段 1 上的 *Eco*RI 位点会与片段 2 上的 *Eco*RI 位点粘在一起，形成新的 *Eco*RI 酶切位点；而片段 1 和片段 2 上的同尾酶切位点 *Spe*I 和 *Xba*I 因留下相同的黏性末端而粘在一起。由此插入片段准确地嵌入载体上的小口中，且融合后的 S/X 混合位点不能被 *Xba*I 和 *Spe*I 识别而切开；与此同时，片段 1 又将自身的 *Xba*I 位点加入到片段 2 上，实现两个 BioBrick 的成功连接。

首先在含有已知片段 2 的质粒上用 *Eco*RI 和 *Xba*I 两种酶去掉它们之间的序列，形成一个切口如下：

```
序列：5'--gca G                    *CTAGA G--- 3'
序列：3'--cgt CTTAA*                    T C--- 5'
                 EcoRI              XbaI
```

再利用 *Eco*RI 和 *Spe*I 两个酶切位点将片段 1 切割下来，片段 1 的序列如下：

```
序列：5' *AATTC GCGGCCGC T TCTAGA G -- 片段 1 -- T A      3'
序列：3'      G CGCCGGCG A ACATCT C -- 片段 1 -- A TGATC* 5'
        EcoRI     NotI        XbaI                      SpeI
```

之后将两种片段混合。在连接酶的作用下，片段 1 上的 *Eco*RI 位点会与片段 2 上的 *Eco*RI 位点粘在一起，形成新的 *Eco*RI 酶切位点；而片段 1 上的 *Spe*I 和片段 2 上的 *Xba*I 是同尾酶切位点，因留下相同的黏性末端而粘在一起，因此插入片段可以准确嵌入载体上的小口中，且融合后的混合位点不能被 *Xba*I 和 *Spe*I 识别而切开；与此同时，片段 1 又将自身的 *Xba*I 位点加入到片段 2 上，这样就实现了两段序列的成功连接，同时又保证了连接后片段的前缀和后缀都保持不变，新片段仍然可以用这四种酶与其他片段连接。连接后的序列如下所示：

```
序列：5' --gca G  *AATTC GCGGCCGC T TCTAGA G -- 片段 1 --T A   *CTAGA G---- 3'
序列：3' --cgt CTTAA*  G CGCCGGCG A ACATCT C -- 片段 1 --A TGATC*   T C---- 5'
              EcoRI        NotI        XbaI                          混合位点
```

同样道理，如果将片段 1 插入到片段 2 的右侧，那么就需要在含有已知片段 2 的质粒上用 *Spe*I 和 *Pst*I 两种酶去掉它们之间的序列，形成切口如下：

```
序列：5' --T A                    *G gct--- 3'
序列：3' --A TGATC*            ACGTC cga--- 5'
                SpeI            PstI
```

通过 *Xba*I 和 *Pst*I 两个酶切位点将片段 1 取下，得到的序列如下：

```
序列：5' *CTAGA G-- 片段 1 -- T ACTAGT A GCGGCCG CTGCA 3'
序列：3'      T C-- 片段 1 -- A TGATCA T CGCCGGC G*     5'
         XbaI                    SpeI      NotI      PstI
```

将两种片段混合，在连接酶的作用下，两个片段连接在一起形成了新的 *Xba*I、*Spe*I 和 *Pst*I 酶切位点，连接后的片段序列如下：

```
序列：5' --T A *CTAGA G -- 片段 1 -- T ACTAGT A GCGGCCG CTGCA  *G gct--- 3'
序列：3' --A TGATC* T C -- 片段 1 -- A TGATCA T CGCCGGC G*  ACGTC cga--- 5'
          混合位点                      SpeI      NotI       PstI
```

连接后重新组合成新的 *Spe*I 和 *Pst*I 位点及一个不可分割的 *Spe*I/*Xba*I 混合位点。加上载体原有的 *Eco*RI 和 *Xba*I 位点，新的 Biobirck 仍然具有 4 个标准酶切位点。

由此可见，只要按照标准化的操作，即可以保证连接后的 BioBrick 仍然具有相同

的 4 个标准酶切位点。可以用同样的方法与其他标准片段连接。如此循环往复，即可以由简单到复杂，逐层构建更加复杂的生物系统。

2.3　标准定量机制

除了用标准化的功能模块作为承载功能的硬件之外，还需要标准化的系统量化平台和抽象的概念信号作为承载功能的软件。为此，iGEM Registry 提供了衡量和代表输入输出信号的标准——PoPS 和 RIPS。

2.3.1　PoPS[104]

PoPS（RNA polymerase per second，RNA 聚合酶每秒）用于衡量基因的被转录水平，对于每个 DNA 拷贝来讲 RNA 聚合酶分子每秒钟通过 DNA 分子上某一点的数量。从某种意义上讲，PoPS 类似于流经电线特定位置的电流流量。这个度量单位有时也被称为 PAR（polymerase arrival rate），即每秒钟到达某一特定 DNA 位点的 RNA 聚合酶的数量。

在上述的各种生物部件中，启动子可以看作是 PoPS 源（类似于电路中的电流源——电池），产生 PoPS 的稳定输出，但是没有输入。终止子相当于 PoPS 接收器或者接地的装置，即以 PoPS 作为输入，但没有输出。

基于 PoPS 的转换器（inverter）通常包含一个 RBS、阻遏蛋白（repressor）编码区域、终止子和同源启动子。此时高水平的 PoPS 输入会导致阻遏蛋白表达并与启动子结合，产生低水平输出信号；相反，低水平的 PoPS 输入时无阻遏蛋白表达，启动子被启动而产生 PoPS。RBS 的作用相当于导线，允许 PoPS 信号通过。类似的，编码区域也是导线，但却具有一定的阻抗，即其输出的 PoPS 小于它的输入，可以看作是电路中的电阻元件。

PoPS 只是一个转录水平上通用的信号载体，其提出的初衷是为了提供一个标准的衡量单位和信号描述方式，方便对基因线路规范化的表述。但 PoPS 并不是一个可以广泛使用的信号。翻译水平和代谢水平的组件不涉及 RNA 聚合酶和转录过程，因此也就无法采用此种量化方法。例如，激酶处理装置就不能采用 PoPS。

需要指出的是，PoPS 并不是我们常说的转录速率。转录速率通常是与特定转录相关的参数，衡量的是单位时间内的转录量；而 PoPS 则是指 DNA 特定位点的关键转录速率。这两者在某些情况下具有相同的物理含义，例如，编码区域下游的 PoPS 值等于编码区域的转录速率；但在某些情况下含义却是不同的，如在某些特殊位点，根本不存在转录速率的说法，但 PoPS 却仍然具有一定的含义。例如，生物工程比较关心的 RNA 聚合酶通过终止子的速率（或者说是终止子下游的 PoPS）可以用 PoPS 来衡量，却无法用转录速率来衡量。

作为一个具有一定量化作用的抽象概念，PoPS 的提出具有一定的开创性，但读者应该能够认识到它的局限性。正如 D. Endy 所说，合成生物学信号的量化和度量还是一个开放的课题，有待进一步完善。目前合成生物学在基因模块定量的标准化上还远不及

模块本身的进程那么快速普遍，许多研究者正在寻找更通用、更直观的衡量方法和更直接的测量手段。

2.3.2 RIPS

RIPS（ribosomal initiations per second）用于衡量 mRNA 的翻译水平，对于每一个 mRNA 来讲，是指核糖体分子每秒钟通过 mRNA 分子上某一点的数量。

2.4 生物系统的层次化结构

生物系统的层次化结构是合成生物学工程化本质的又一典型体现，如图 2-2 所示。具有一定功能的 DNA 序列组成的最简单的 BioBrick 称为基因部件——Part，不同功能的基因部件按照一定的逻辑和物理连接组成复杂的生物装置——Device，不同功能的 Device 协同运作组成更加复杂的生物系统——System，含有多种不同功能 System 的生物体彼此通讯互相协调组成再复杂些的多细胞或细胞群体生物系统。

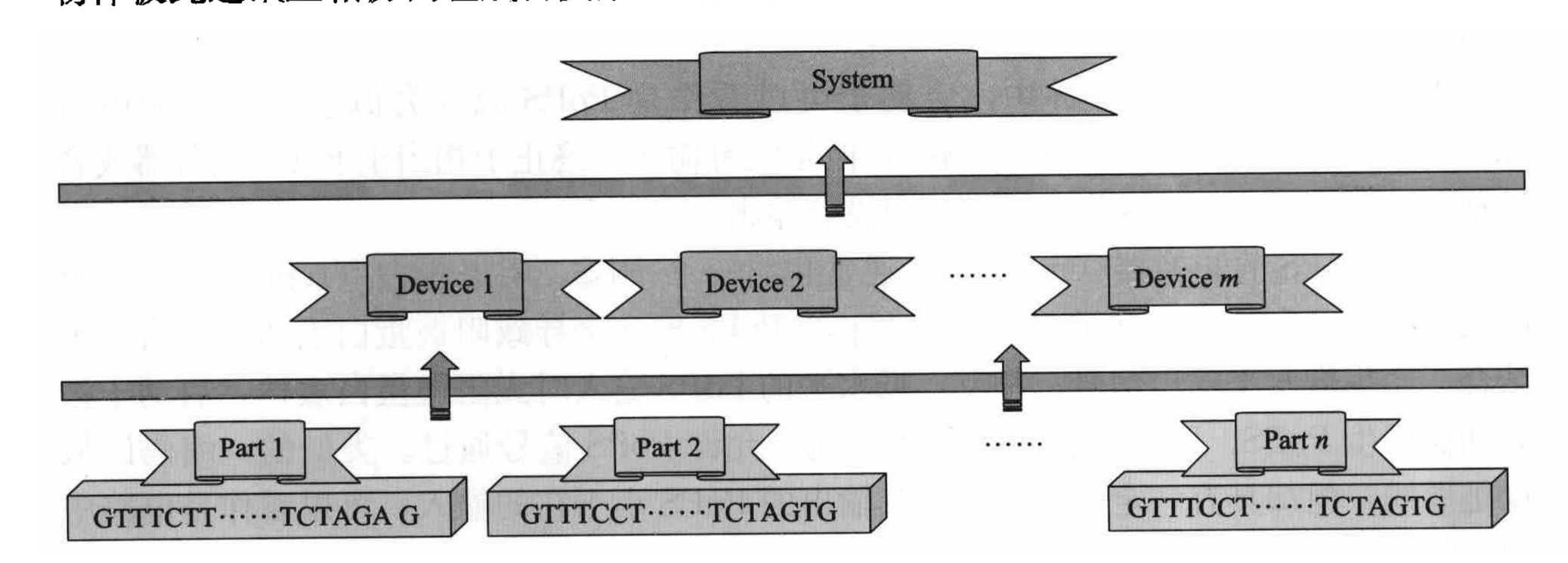


图 2-2 生物系统的层次化结构示意图

2.4.1 生物部件“Part”

遗传系统中最简单、最基本的生物积块称为生物部件——Part。Part 是指具有特定功能的核苷酸或者蛋白质序列，能够通过标准化组装方法与其他 Part 组装成具有更复杂功能的模块。

Part 按照其功能可以划分为终止子、蛋白质编码基因、报告基因、信号传递组件、引物组件、标签组件（tag）、蛋白质发生组件、转换器、启动子等类别。每一个 Part 都有一个标准的名字编码，我们可以很方便地从一块 DNA 元件的名字编码中判断出它在具体生物过程中所发挥的功能。常见的元件 Part 如下。

（1）启动子——启动子（promoter，后文简写作 P）是操纵子（operon，后文简写作 O）的一个组成部分，专一地与 RNA 聚合酶结合并决定转录从何处起始的部位，控制基因表达（转录）的起始时间和表达的程度。启动子就像“开关”，与称为转录（transcription）因子的蛋白质结合，控制基因的活动。转录因子是“起始复合物（ini-

tiation complex)”的组成成分，指导 RNA 聚合酶的转录起始。生物中有许多启动子，如大肠杆菌约有 2000 个启动子，各启动子的效率并不相同，强启动子每 2s 启动一次转录，而弱启动子每 10min 才启动一次。细菌的启动子通常具有一些为 RNA 聚合酶与启动子相结合所必需的特定的结构保守区，其碱基变化会影响 RNA 聚合酶的识别能力和结合亲和力，控制转录水平的高低。原核表达系统中通常使用的可调控的启动子有 P_{lac}（乳糖启动子）、P_{trp}（色氨酸启动子）、P_{tac}（乳糖和色氨酸的复合启动子）、T7 噬菌体启动子等。真核生物的启动子与原核生物不同，而且启动转录的活性除需启动子外，还需某些外加序列。

（2）RBS——核糖体结合位点（ribosome binding site，RBS）是指 mRNA 分子中紧靠启动子下游、起始密码子 AUG 上游的一段非翻译区序列，用于结合核糖体以便开始转录。翻译起始密码位于它的下游位置，其功能是初始化翻译。原核生物的 RBS 中有 SD 序列，长度一般为 4～9 个核苷酸，富含 G、A。该序列与核糖体 16S rRNA 的 3′端互补配对，促使核糖体结合到 mRNA 上，有利于翻译的起始。RBS 的结合强度取决于 SD 序列的结构及其与起始密码子 AUG 之间的距离，相距一般以 4～10 个核苷酸为佳，9 个核苷酸为最优。由于核苷酸的变化能够改变 mRNA 5′端的二级结构，影响核糖体 30S 亚基与 mRNA 的结合自由能，从而造成蛋白质合成效率上的差异，因此 SD 序列的微小变化往往就会导致表达效率上百倍甚至上千倍的差异。

（3）终止子——是指结束基因转录的 DNA 信号序列。在一个基因的 3′端或是一个操纵子的 3′端往往有特定的核苷酸序列，具有终止转录的功能，这一序列称之为转录终止子，简称终止子（terminator，后文简写作 T）。所有原核生物的终止子在终止点之前都有一个回文结构，它转录出来的 RNA 可以形成一个茎环式的发夹结构。终止子按其作用是否需蛋白因子的协助至少可以分为两类：一类是不依赖 ρ 因子的终止子，这类终止子在序列上有一些共同的特点，即有一段富含 GC 的反向重复序列（inverted repeat sequence），其后跟随一段富含 AT 的序列，因而转录生成的 mRNA 序列中能形成发夹式结构，后继一连串寡聚 U 序列。正是 RNA 聚合酶转录生成的这段 mRNA 的结构阻止 RNA 聚合酶继续沿 DNA 移动，并使聚合酶从 DNA 链上脱落下来，终止转录。另一类是依赖 ρ 因子的终止子，即其终止转录的作用需要 ρ 因子的协同，或至少是受 ρ 因子的影响，终止点前无寡聚 U 序列，回文对称区不富含 GC。不同终止子的作用也有强弱之分，有的终止子几乎能完全停止转录；有的则只是部分终止转录，还有一部分 RNA 聚合酶能越过这类终止序列继续沿 DNA 移动并转录。如果一串结构基因群中间有这种弱终止子的存在，则前后转录产物的量会有所不同，这也是终止子调节基因群中不同基因表达产物比例的一种方式。转录终止的机制较为复杂，结论尚不统一。在构建表达载体时，为了稳定载体系统，防止克隆的外源基因表达干扰载体的稳定性，一般都在多克隆位点的下游插入一段很强的转录终止子。

（4）操纵子——是细菌的基因表达调节装置，由启动子和其他 DNA 调节元件与串联的多个相关基因组成，由同一套调节蛋白调节。操纵子通常由 2 个以上的编码序列、启动序列、操纵序列以及其他调节序列在基因组中成簇串联组成。某些操纵序列是原核阻遏蛋白的结合位点，当操纵序列结合阻遏蛋白时会阻碍 RNA 聚合酶与启动序列的结

合，或使 RNA 聚合酶不能沿 DNA 向前移动，阻遏转录，介导负调节（negative regulation）。原核操纵子调节序列中还有一种特异 DNA 序列可结合激活蛋白，使转录激活，介导正调节（positive regulation）。

在基本部件中，被调控基因的激活或者抑制通常是通过转录调控因子与启动子操纵位点的直接作用实现的。目前合成生物学的早期研究主要依赖于这些转录单元，尤其是以 LacI-P_{lac}、cI-Pl，以及 TetR-P_{tet}对等作为复杂人工线路的“积木”。

2.4.2 生物装置“Device”

有了上述标准化的“Part”部件，就可以利用转录激活因子、转录阻遏蛋白、转录后机制（如 DNA 修饰酶）和 Riboregulator 等结合下文将要讲到的逻辑拓扑结构构建稍微复杂些的生物装置“Device”。

生物装置通过调控信息流、代谢作用、生物合成功能以及与其他装置和环境进行交流等方式处理“输入”产生“输出”。可以说，生物装置包含了一系列转录、翻译、蛋白质磷酸化、变构调节、配体/受体结合、酶反应等生化反应。不同装置各自的生物化学属性具有各自的优势和限制。

利用 iGEM Registry 提供的标准化系统量化方法，我们可以将一些生物装置进行标准化抽提，描述成如下形式：具有一定生物学功能，并且能够为外源物质所控制的一串 DNA 序列。

（1）报告基因（reporter）——其产物易于被检出的基因，在分子生物学实验中用于替换天然基因的位置，以检验其启动子及调节因子的结构组成和效率，常用的为各种荧光蛋白编码基因，如 gfp（green fluorecent protein 基因）等。

（2）转换器（inverter）——一种遗传装置，它在接收到某种信号时停止下游基因的转录，而未接收到信号时开启下游基因的转录。

（3）信号转导装置（signaling）——是指环境与细胞之间或者邻近的细胞与细胞之间接收信号和传递信号的装置。

（4）蛋白质生成装置（protein generator）——产生具有一定功能蛋白质的装置。

目前已经工程化的遗传装置还有很多，如控制基因表达的各种基因开关、切换基因表达状态的双稳态开关[58]、三种阻遏蛋白相继表达的 Repressilator[60]、模拟各种逻辑门功能的生物装置等，这些我们会在后面的章节中详细介绍。

2.4.3 生物系统“System”

为了得到更加复杂的调控行为或生物功能，可将装置以串联、反馈或者前馈等形式连接，组成更加复杂的级联线路或者调控网络，即所谓的生物系统（System）。

自然生物系统中的调控级联线路是非常普遍的，如大肠杆菌和酵母菌调控网络中包含的级联线路。级联线路调控基因的递进式表达，可以触发大肠杆菌鞭毛、酵母菌孢子的形成或者控制细菌的细胞周期。在活体细胞中，许多信号转导和蛋白激酶通路也得益于级联过程来调控其活性。在果蝇和海胆等多细胞生物体中，许多时间顺序事件通常由级联过程来调控[105,106]。同时，级联线路具有许多非常重要的特点。例如，由蛋白质控

制的级联线路对于渐变输入具有超敏感性的“全开或全关”响应，输入信号中非常小的变化即可以激励输出由低到高/由高到低的跳变[107]，如图 2-3 所示。超敏感性遗传线路对诱导物浓度在一个很窄范围内的变化具有快速响应能力，即使是微弱的输入信号，一旦达到阈值，即可快速激活遗传线路，其响应曲线非常类似于典型的阶跃响应，具有广泛的应用价值[108]。

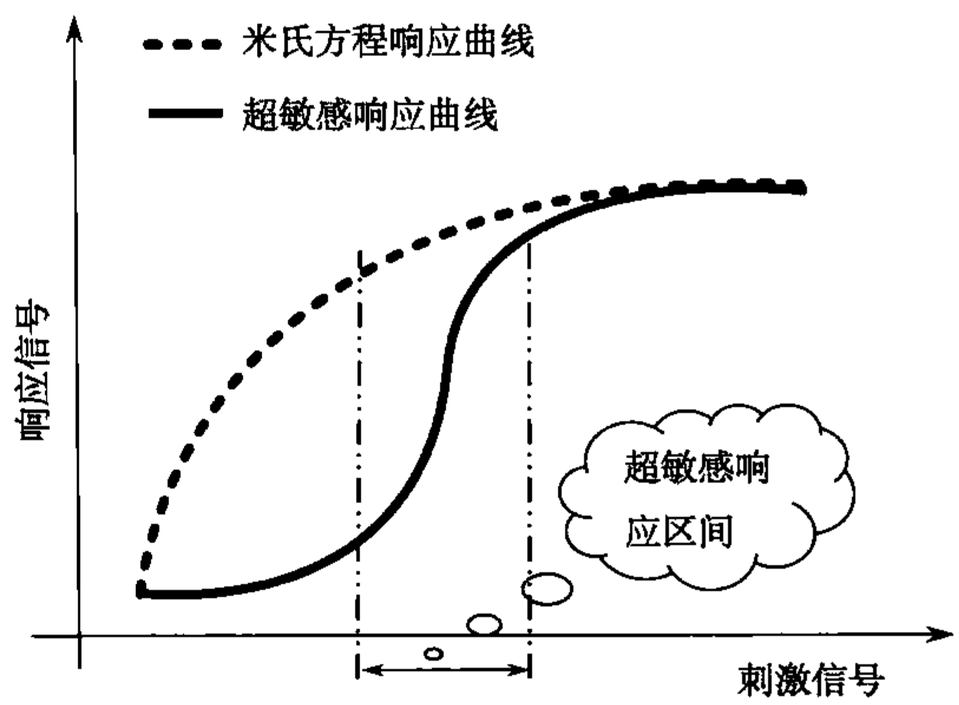


图 2-3　超敏感性响应曲线与米氏响应曲线对比图

相比于标准的米氏响应曲线，超敏感性响应对于特定区间信号的变化更敏感，对诱导物浓度在很窄范围内的变化具有快速响应能力，即使是微弱的输入信号，一旦达到阈值，即可快速激活遗传线路。

在各种级联线路和调控网络中，最简单的形式是转录水平的调控系统。核苷酸序列直接决定了相互作用的特异性，因此，相对来说控制转录和翻译以产生目的输出的装置，其搭建都比较容易且具有一定的柔性。转录控制系统具有很多其他有用的特性，包括信号放大，多个转录因子的组合控制，多个下游靶点的控制，噪声的传播、放大和衰减，以及内外因素对于表型变化的控制等[109]。对于不同长度级联线路的研究表明，在某种条件下，增加级联线路的层次深度能够增加响应的敏感性，使其输入/输出关系更加接近离散特性；同时，响应的延迟性也由线路的层次深度决定，长的级联能够起到低通滤波器的作用，对于输入噪声具有一定的鲁棒性（Robusness）[106,110,111]。一般来讲，真核细胞级联线路的长度通常要比原核细胞的长[112]。

然而，仅仅合成一种蛋白质也需要从转录到翻译的大量的生化反应才能实现；而获得能够检测的输出变化，则需要许多蛋白质合成反应的发生。因此，这些装置和系统在实现其功能时需要消耗大量细胞资源，其衡量时间是以分钟或小时为尺度的，相比于翻译水平的 RNA 调控和蛋白质水平的调控，基因线路转录输出的变化相对比较缓慢。

2.4.4　细胞群体系统及多细胞系统

由于基因表达过程中内源和外源噪声的影响以及其他细胞的作用，互不通讯的一组细胞即使起源相同也可能具有不同表型和异步行为，不可能互相协作，完成的生物功能也有限。为了更高效地实现人类期望的功能，需要多个细胞甚至多种细胞协同运作。利用细胞间通讯协调彼此的行为是目前工程化细胞群体系统的主要手段[113]。相比于单细胞而言，细胞群体系统及多细胞系统的人工构建则复杂得多，不仅要考虑细胞间的协同，还要考虑信号分子的跨膜运输、环境因素的分布梯度等[114]。

群体感应

群体感应（quorum sensing，QS）现象最早是在海洋中的一种发光弧菌（*Vibrio*

fischeri）中发现的。近年来的研究证明，微生物之间存在着信息交流，许多微生物都能合成并释放一种被称为自体诱导物（autoinducer，AI）的信号分子，胞外的 AI 浓度随微生物密度的增加而增加，当达到一个临界浓度时，AI 能启动菌体中相关基因的表达，调控微生物的生物行为，如产生毒素、形成生物膜、产生抗生素、生成孢子、产生荧光等，以适应环境的变化，这一现象被称为群体感应现象。因这一现象只有在微生物密度达到一定阈值后才会发生，所以也称为依赖细胞密度控制的基因表达（cell density dependent control of gene expression）[115]。

在发光弧菌中由 P_{lux}启动子控制群体感应表达系统，而 P_{lux}启动子是由小分子信号控制的。图 2-4 是一个利用细胞群体感应机制进行协调的例子。该荧光系统包含有向左和向右两个方向分别转录的启动子 Pr_{lux}和 Pl_{lux}。Pl_{lux}启动子能够启动 LuxR 蛋白的表达，Pr_{lux}启动子能够启动 *luxI* 和与荧光相关的基因转录，*luxI* 编码的 LuxI 能够催化小分子 AHL（*N*-乙酰基高丝氨酸内酯）的合成。当细胞浓度较低时，细胞内 *luxI* 基因的产物很少，细胞及其周围介质中的自诱导物 AHL 的浓度很低。随着菌体数量的增加，AHL 的浓度逐渐升高，当达到某一临界值时，该种小分子就会与 LuxR 蛋白形成一种复合物，该种复合物能够强烈促进 Pl_{lux}启动子，从而启动与荧光相关基因的转录以及 *luxI*、*luxR* 基因的表达，产生更多的 AHL，形成正反馈调控。

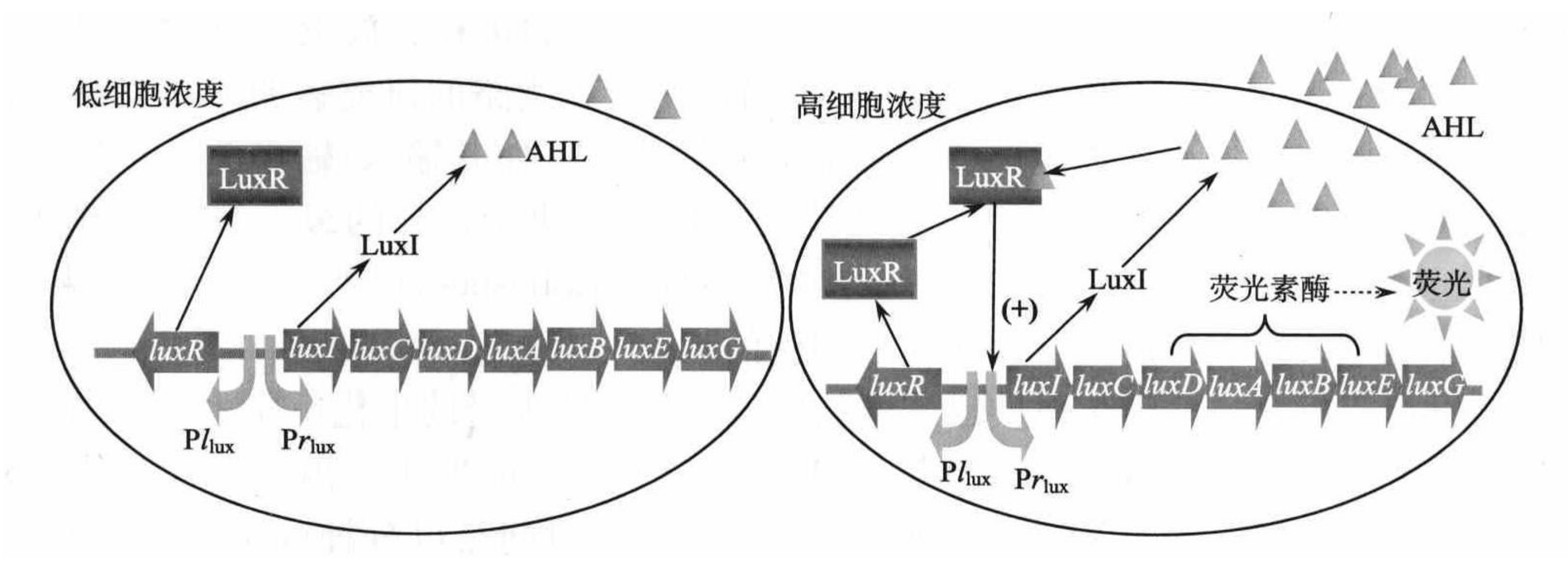


图 2-4 细胞群体感应机制示意图

图中的三角形代表信号分子 AHLO 系统中，Pl_{lux}启动子能够启动 LuxR 蛋白的表达，Pr_{lux}启动子能够启动 LuxI 和与荧光相关的基因转录。当细胞浓度较低时，细胞内 *luxI* 基因的产物很少，其催化产生的自诱导物 AHL 的浓度很低。随着菌体数量的增加，AHL 的浓度逐渐升高，当达到或超过某一临界值时，AHL 小分子与 LuxR 蛋白形成一种复合物，强烈促进 Pl_{lux}启动子启动与荧光相关基因的转录以及 *luxI*、*luxR* 基因的表达，系统中不仅有荧光产生，且产生更多的 AHL，形成正反馈调控。

群体感应一般用于系统中群体细胞之间的同步响应，它提供了一种依赖细胞浓度调控基因表达的有效手段。合成生物学在此方面的初期尝试就是在大肠杆菌中开发细胞间的通讯模块以协调细胞群体行为。将费氏弧菌（*Vibrio fischeri*）中响应群体感应的基因分别植入独立的发生细胞和接收细胞中。发生细胞表达 LuxI 催化 AHL 的合成。AHL 信号分子自由扩散到接收细胞中和 LuxR 蛋白组成复合物激活目的基因转录，从而通过 AHL 信号分子的扩散实现细胞群体间的通讯和协作[116]。

除了研究比较普遍的在大肠杆菌中构建群体感应机制获得目的功能外，R. Weiss 等

通过在酿酒酵母中构建人工群体感应系统而将其研究扩展到了真核细胞中。从拟南芥（*Arabidopsis thaliana*）中获取的信号系统与天然的酵母磷酸化级联线路相连接控制工程化的基因开关[117]。此系统实现了高度的模块化设计，考虑到对系统响应的测量和调控的需要，还对信号进行了必要的放大。

第二部分　合成生物系统的逻辑拓扑结构

生物系统的模块化设计中小到 DNA 片段、大到调控网络，各个模块的逻辑结构都有一些共性。较为简单的基因调控单元为调控网络基元和基础基因线路。这些简单模块可以利用逻辑拓扑结构中的前馈、反馈等的合理组合连接成具有一定功能的遗传线路，同样，遗传线路又可连接成调控网络乃至生物系统。

调控网络基元（motif）是转录因子和靶基因之间相互调控关系的特定小规模组合，通常由一组相关基因和调节元件按照一定的拓扑结构构成。

基础基因线路（elementary gene circuit）也是比较简单的基因线路形式，如果一个基因的表达受单一转录因子的调节并在给定条件下对一种信号分子作出反应，则可称为“基础基因线路”[118]。

原核生物基因表达调控主要发生在转录水平，因此我们以转录水平的调控为主来介绍遗传线路的各种逻辑拓扑结构。转录调控的基本单元是操纵子，在转录水平上对基因表达进行调控取决于 DNA 的结构、RNA 聚合酶的功能、蛋白因子及其他小分子配基的相互作用。

对于这些拓扑结构的分类方式有很多种，大体可以归纳为串联、反馈和前馈等类型。其中，反馈可分为正反馈和负反馈；前馈可分为一致前馈和非一致前馈等。这些拓扑结构并不是生物系统所特有的，在电子线路、控制理论、计算机图论等领域中均有类似的拓扑结构，只是因为生物系统是由生物元件组成，具有不同于电子线路、控制回路等的特点，因此读者需要留心这方面的特性和共性。

2.5　串联

从概念上讲，串联是指上游模块的输出信号作为下游模块的输入信号的连接形式，如图 2-5 所示。对于生物模块来讲，这种信号可以是小分子代谢物、蛋白质、mRNA 等。由于这种结构比较简单，这里我们就不详细介绍了。

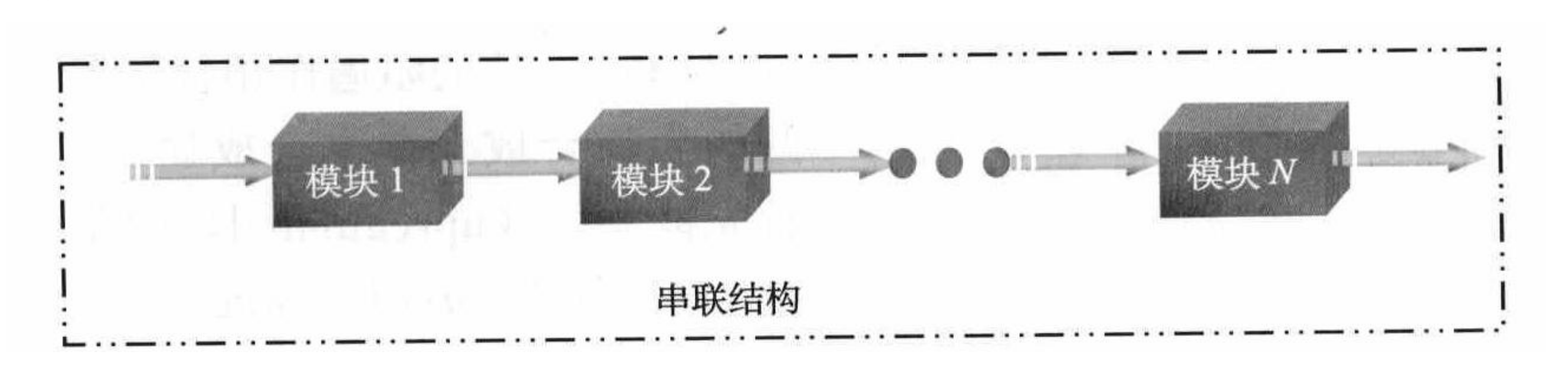


图 2-5　模块串联结构示意图

2.6 单输入

图 2-6 为单输入模块（single input module，SIM）结构示意图，主模块的输出作为下层分模块的输入。在单输入结构中，只有一个主模块作为下层模块的输入。

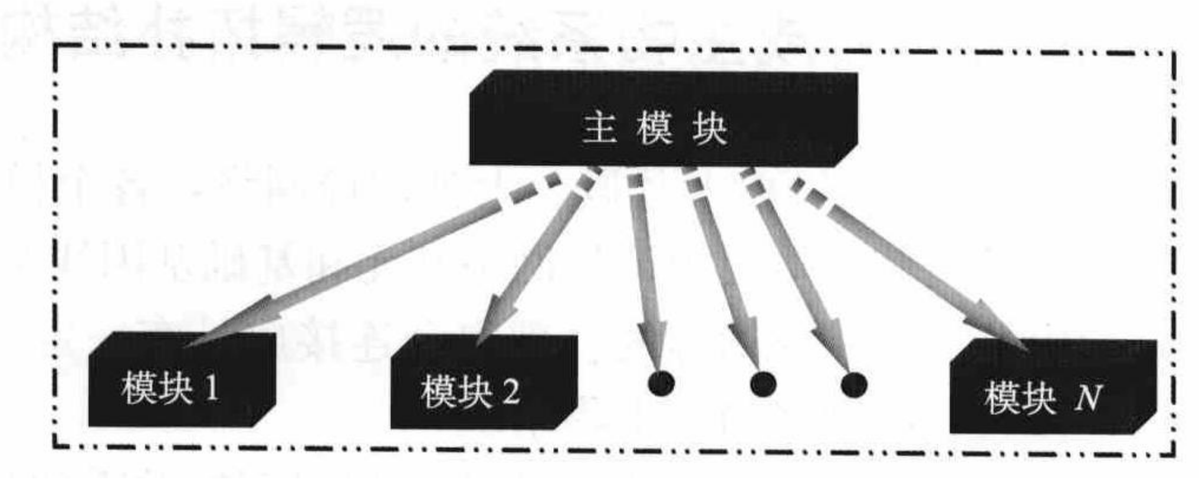


图 2-6 单输入模块结构示意图

这种结构在自然生物的转录基元（motif）中很常见，其主要功能是实现一组基因的共表达。此外还可以实现模块时序表达的功能，即模块 1 到模块 N 按预定义顺序表达：如果主模块代表激活因子，激活下层启动子模块，启动子序列和结合位点不同，导致各个启动子激活阈值不同，当主模块的活性缓慢递增时，阈值最低的启动子先启动，依此类推，阈值最高的启动子最后启动，实现基因表达的时序控制。同理，当主模块代表阻遏子基因时，可以实现类似的时序控制。通过高时间分辨率的检测手段可以观测到在大肠杆菌中有许多系统具有 SIM 结构。这一结构具有一个很重要的属性——时间序列与基因功能序列吻合，即在一个多基因过程中，越早需要的蛋白质/代谢物，它的基因越先被激活。这种时序控制可以防止蛋白质产物在需要之前过早生成，实现能源的节约，避免不必要的交叉干扰。

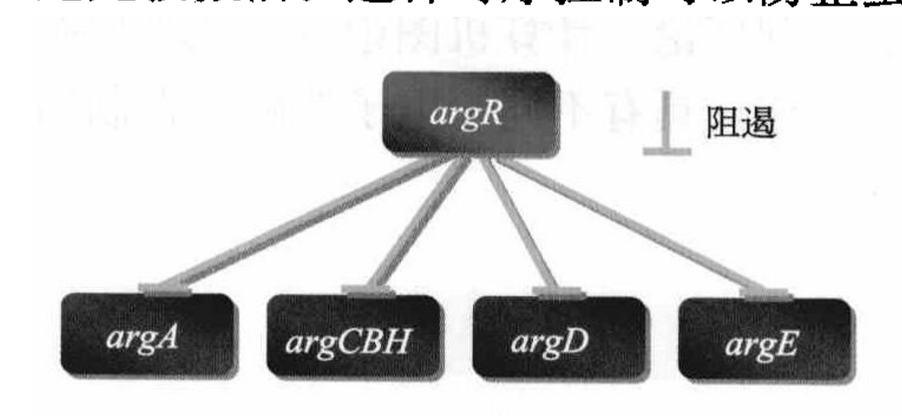


图 2-7 大肠杆菌精氨酸生物合成中的 SIM 拓扑结构

在精氨酸合成的主通路中，阻遏蛋白 ArgR 同时调控 *argA*、*argCBH*，*argD* 和 *argE*。当细胞内缺乏精氨酸时，阻遏蛋白 ArgR 对 *argA*、*argCBH*、*argD* 和 *argE* 的抑制作用被解除。由于阻遏作用阈值的不同，与谷氨酸酯转化成鸟氨酸所需的酶 ArgA、ArgBC、ArgD 以及 ArgE 的顺序一致，*argA* 启动子首先上调，接下来依次是 *argCBH*、*argD*、*argE*。

例如，在大肠杆菌的精氨酸生物合成系统中，阻遏蛋白 ArgR 调控多个编码精氨酸生物合成过程中所需要的酶的操纵子，如图 2-7 所示。精氨酸合成的主要途径是利用谷氨酸盐作为前体的。此外还需要谷氨酰胺和天冬氨酸，其产物通过另外两条无支路的途径合成，并最终会合于主路径。当细胞内缺乏精氨酸时，ArgR 对 *argA*、*argCBH*、*argD* 和 *argE* 的阻遏作用被解除，开启精氨酸的生物合成。在主合成途径中，*argA* 是首先被上调（upregulated）的基因，接下来是 *argCBH*、*argD*、*argE*。这与 ArgA、

ArgBC、ArgD 以及 ArgE 将谷氨酸酯转化成鸟氨酸* 的顺序一致。在精氨酸合成的另外两条无支路的代谢途径中，基因的时序表达与酶的需求顺序仍然吻合：在将谷氨酸转化为瓜氨酸的代谢途径中先是 *argAB*，之后依次是 *argF* 和 *argI*；在瓜氨酸转化成精氨酸的过程中，*argG* 先于 *argH* 基因表达。需要注意的是，尽管 ArgG-ArgH 途径是 ArgA-ArgE 途径的终产物，但 ArgG-ArgH 途径具有与 ArgA-ArgE 途径相似的时间顺序[119]。

此外，大肠杆菌的 SOS 应急反应中的 DNA 修复系统、大肠杆菌和新月柄杆菌（*Caulobacter crescentus*）的鞭毛系统以及许多生物的细胞周期基因系统中都有类似的拓扑结构。

2.7　多输入

多输入模块（multi input module，MIM）结构也称为“密集交盖调节网”（dense overlapping regulons，DOR）。我们还是以转录水平的调控为例，这种结构表示一组调控因子共同控制一组基因。大肠杆菌中有很多 DOR 结构，负责碳的利用、厌氧环境下生长、胁迫响应等广泛的生物功能。类似的结构在酵母菌中也有很多。

图 2-8 所示为 MIM 结构示意图。由图中可以看出，MIM 可以看作是一个逻辑门阵列，多路输入进行组合运算以后控制下游模块。因此，为了充分理解模块之间的调控机制，仅用箭头来表示信号的传递是不够的。图 2-9 为相关生物功能的 MIM 示意图。在 8 个子图中，只有子图 C、G 中两个调控因子对被调控基因的作用是完全一致的。读者应该能够试着画出其余的调控组合形式。

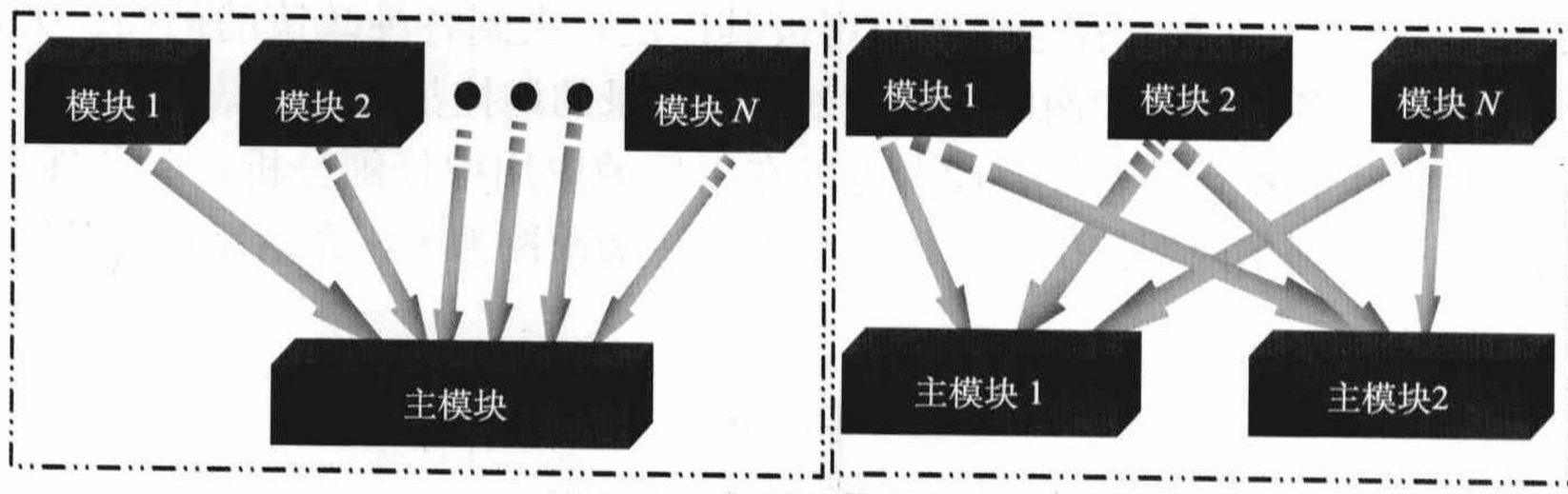


图 2-8　MIM 结构示意图

目前大部分转录水平的 MIM 的具体功能细节还不完全清楚，转录网络、代谢网络等不同水平的调控互相影响，很难准确界定每个 MIM 的规模。第 3 章将要讲到的聚类算法是目前识别 MIM 的主要手段，但具体聚类算法的不同，可能会导致识别出的 MIM 规模和功能略有不同，需要结合背景知识仔细分析。

除了转录水平的 MIM 之外，还有很多其他水平的 MIM 结构。在著名的神经网络算法中，这种结构非常常见，有兴趣的读者可以参考相关方面的书籍。

---

* 精氨酸合成的其中一条途径是：鸟氨酸转化成瓜氨酸，瓜氨酸再通过精氨基琥珀酸合成酶（argininosuccinate synthetase，ASS）及精氨基琥珀酸裂解酶（argininosuccinate lyase，ASL）合成精氨酸。

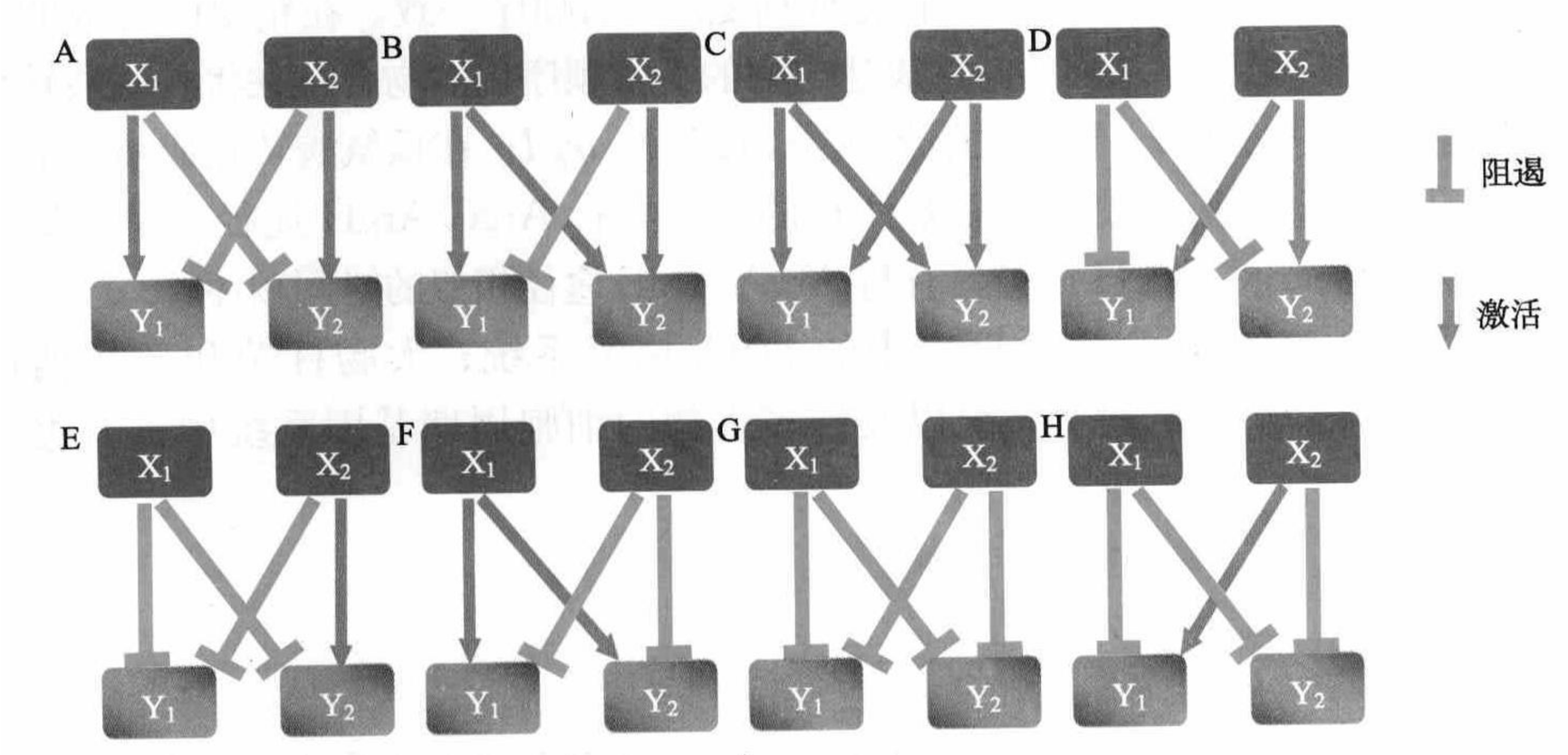


图 2-9 相关生物功能的 MIM 示意图

2.8 反馈

反馈是控制理论中最重要的概念之一，其控制系统如图 2-10 所示。从抽象的信号层面讲，反馈是指系统的输出会反过来影响系统的输入，进而影响自身的一种控制机制。这种输出对于输入的影响可能会导致最终输出的降低，即所谓的负反馈；也可能会导致最终输出的增加，即所谓的正反馈。反馈的正负特性由具体的网络动力学特性决定。由于从概念层面上讲，信号的传递形成了一个闭合的回路，因此也常称此结构为“反馈回路”。“反馈”在现实社会中的应用极其广泛，大到各种政策的调控、大规模化工厂的流程控制、各种生物代谢网络，小到基因转录的调控都有反馈结构存在。它们之间有一些共性，也有各自的特性，读者可以在以后的应用中仔细分析这些共性和差异。本章，我们还是回到基因调控中来，以最基本的转录调控为主，介绍反馈在基因调控中的应用。

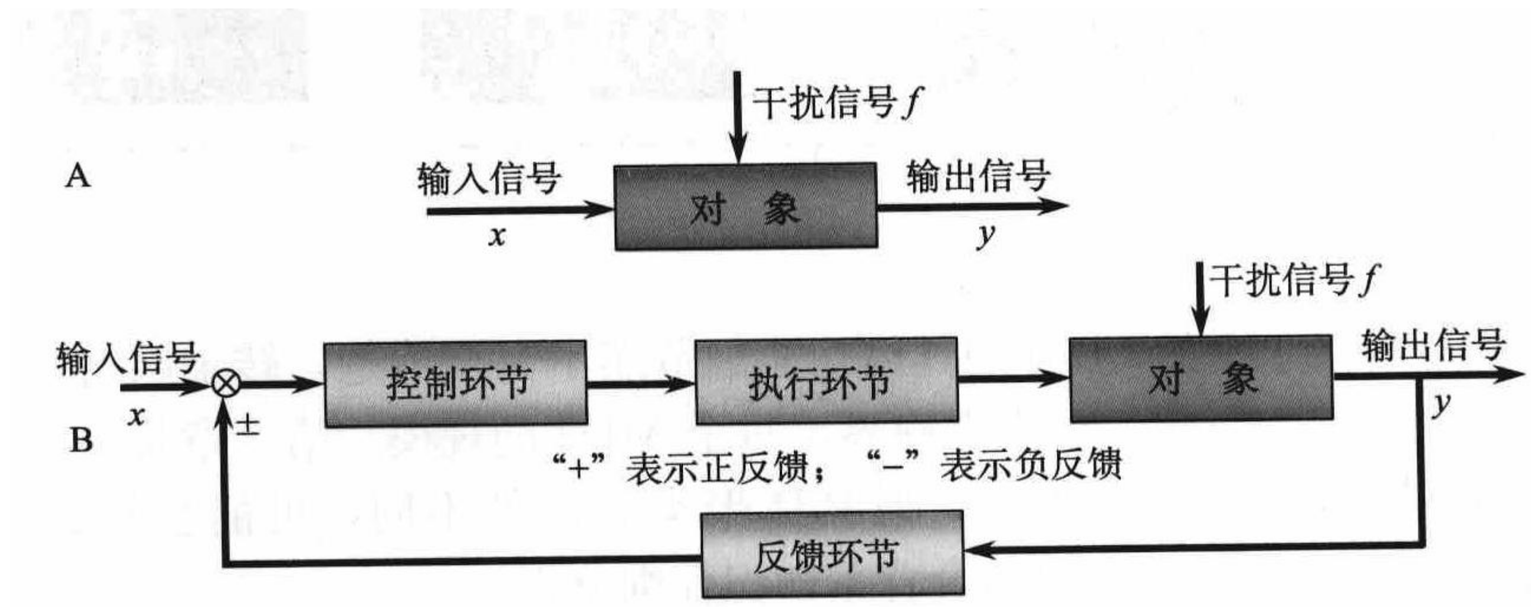


图 2-10 经典开环和反馈的控制系统方框图

A. 开环；B. 反馈

在最早的遗传线路研究中，人们发现基因编码的蛋白质可能会反过来增加或者阻遏上游基因的开启。因此，人们借鉴控制领域中的相应理论和概念来描述这种现象，称其

为“反馈”。遗传线路中也存在着两种基本反馈方式，即负反馈和正反馈，如图 2-11 所示。M. A. Savageau 经过研究，较早提出了“负反馈的调控比开环或者正反馈的调控要更稳定”的论断[120]。后来 A. Becskei 等在实验中证明了这两种基因反馈的存在[121]。A. Becskei 等同时还证明了正反馈对于生物系统的双稳态行为具有重要作用[122]。

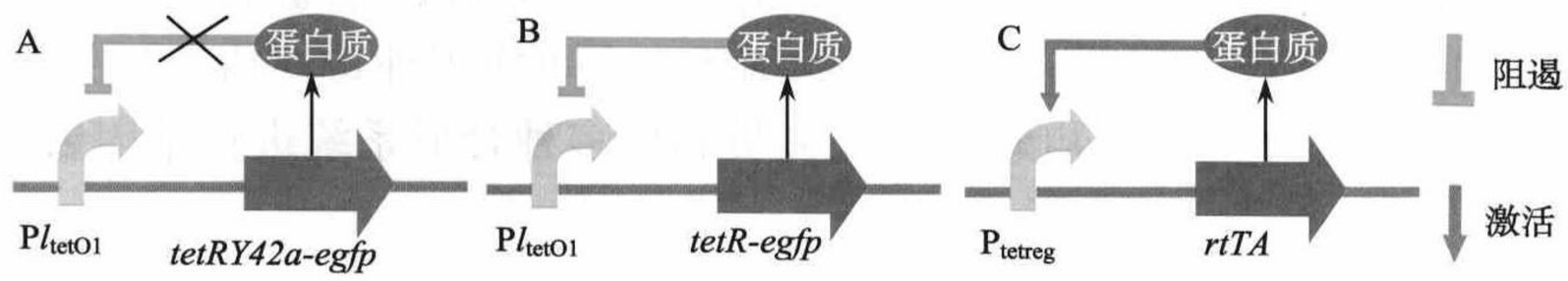


图 2-11　基因线路中的三种基因反馈

A. 开环；B. 负反馈；C. 正反馈

图 A，Pl_{tetO1}启动子启动 *tetRY42a-egfp* 的表达，翻译产生的蛋白质对于 Pl_{tetO1}没有任何影响，属于开环结构。图 B，四环素阻遏蛋白（TetR）阻遏启动子 Pl_{tetO1}，进而阻遏 *tetR-egfp* 表达蛋白和荧光蛋白，形成负反馈。可以利用荧光显微镜通过观察 TetR 与 EGFP 融合蛋白来观察线路的状态（即细胞中 TetR 分子的数量）。图 C，P_{tetreg}启动子启动四环素响应转移活化因子 *rtTA* 的表达，而翻译产生的 RtTA 激活 P_{tetreg}启动子，最终促进自身的表达，形成正反馈回路。

正反馈：基因表达的产物会反过来促进相关基因表达。

负反馈：基因表达的产物会反过来抑制相关基因表达。

图 2-11A、B 为 A. Becskei 等构建的开环和负反馈基因线路图，图 2-11A 中 Pl_{tetO1}启动子启动 *tetRY42a-egfp* 的表达，翻译产生的蛋白质对于 Pl_{tetO1}没有任何影响，属于开环结构。图 2-11B 中利用一个能被四环素阻遏蛋白（TetR）关闭的启动子 Pl_{tetO1}控制 TetR 的产生。当 Pl_{tetO1}启动子开启以后激活 *tetR* 的转录，产生的 TetR 阻遏蛋白能够关闭 Pl_{tetO1}启动子，从而关闭 *tetR* 的转录，属于负反馈机制。通过把编码荧光蛋白的基因与 *tetR* 基因相融合，利用产生的 EGFP 与 TetR 融合蛋白可以在荧光显微镜下观察线路的状态（即细胞中 TetR 分子的量）。大量细胞取样后产生一个荧光强度分布，稳态的稳定性与分布的宽度成反比，较窄的分布暗示了更强的稳定性。数学模拟的结果表明，就像早期的分析一样，负反馈线路比未受调控的线路更稳定，实验结果也证实了这个预测。实验中三种调控最终荧光蛋白的表达都会到达一个稳态，但负反馈中各个细胞的状态分布比正反馈或者开环的时候更集中，说明负反馈的稳定性更强[65, 121]。

图 2-11C 为 A. Becskei 等构建的正反馈基因线路。P_{tetreg}启动子启动四环素响应，转移活化因子 *rtTA* 的表达，而翻译产生的 RtTA 激活 P_{tetreg}启动子，最终促进自身的表达，属于正反馈结构。通常，当环境使得细胞对某种物质的需求量较高时，细胞会采用这种基因表达调控方式[65,122]。

如果基因的表达量随着信号分子的加入而增加，这种遗传线路就被定义为可诱导的；如果基因的表达量随着信号分子的加入而减少，这种遗传线路就被定义为可阻遏的。与分解代谢有关的遗传线路倾向于被代谢中间产物所诱导，即代谢中间产物对于分解代谢具有正反馈作用。与之相反，合成遗传线路倾向于被一种代谢途径的终产物抑制，即终产物通过负反馈抑制自身合成，否则将会造成合成酶的过度浪费。

反馈的引入增加了基因线路调控的精确性，线路的稳态输出不再是输入的简单函数。例如，负自反馈机制能够降低基因表达的噪声；单组分（single-element）负反馈能够缩短启动子受到诱导后达到稳态最高水平的响应时间[123]；在大肠杆菌中引入由两个表达产物互相抑制的启动子组成的反馈系统可以组成双稳态开关[58]，通过对外源刺激（干预两个阻遏蛋白之一的 DNA 结合亲和性）的暂态响应实现网络稳定状态间的切换；利用类似于反馈驱动的网络在哺乳动物细胞中也可实现外特性切换开关[124]；更为重要的是，振荡器表型的变化启发了研究人员利用多种合成系统进行基因表达噪声的研究。

2.9 前馈

前馈控制也称预先控制或提前控制，其基本原理就是测取进入过程的扰动量（包括外界扰动和设定值变化），并按照其信号产生合适的控制作用去改变控制量，使被控制变量维持在设定值上。前馈控制与反馈控制的最大区别就在于，它是在偏差出现之前就采取控制措施，而后者则是在偏差出现之后才采取纠正措施。相应的前馈控制框图如图 2-12 所示。前馈结构的特点是干扰信号进入系统后分成干扰通路和补偿通路两条不同途径影响最终变量。从概念层面讲，单纯前馈结构中信号的传递并未形成一个闭合的回路，因此尽管很多生物相关的文献中都将前馈结构缩写成 FFL（feed forward loop），但本书并不推荐这种说法，以免造成概念上的混淆。

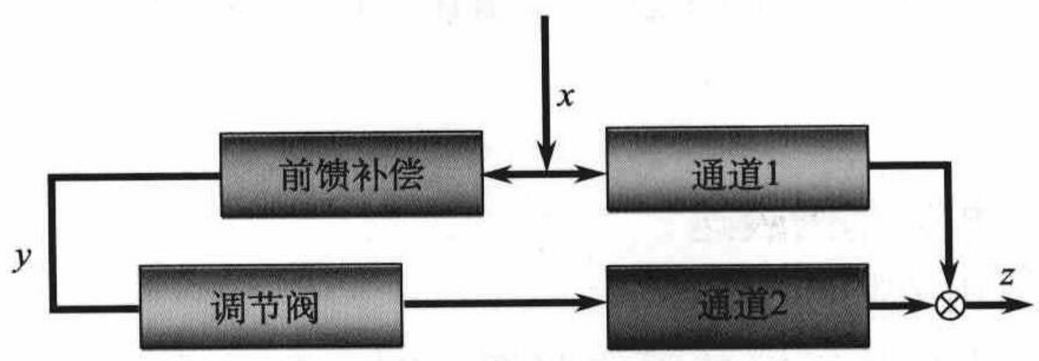


图 2-12 控制理论中的前馈控制方框图

基因线路中利用“前馈”来表示上游基因通过两条不同途径影响下游基因的控制结构。根据这两条通路对最终基因的影响效果是否一致，前馈结构还可进一步划分为“一致”前馈（coherent）和“非一致”前馈（incoherent）两类。“一致前馈”（coherent feedforward）模块中直接和间接调控途径对输出模块的作用相同；“非一致前馈”模块（incoherent feedforward modules）中直接和间接调控对输出模块的作用相反[125~127]。

最简单的前馈转录基元（motif）由三个基因组成：基因 X 调控基因 Y 和 Z，基因 Z 同时受到基因 X 和 Y 的调控。如图 2-13 所示，X、Y、Z 基因之间彼此的促进和抑制关系不同，因此三个元素组成的前馈具有 8 种可能的结构类型。图 2-13A 所示为 4 种“一致前馈”类型，即 X 对 Z 的直接调控促进 X 通过 Y 对 Z 的间接调控，两条支路对 Z 的调控效果是一致的。图 2-13B 所示为 4 种“非一致前馈”类型，即 X 对 Z 的直接调控抑制 X 通过 Y 对 Z 的间接调控，两条支路对 Z 的调控效果是不一致的[119]。

理解前馈结构必须很好地理解基因 X 和 Y 对 Z 的共同作用。我们可以用两种逻辑门来模拟 X 和 Y 对 Z 的调控作用：“与”门表示 Z 的激活需要 X 和 Y 同时满足要求，如图 2-13C 所示；“或”门表示 X 和 Y 中只要有一个满足要求即可激活 Z，如图 2-13D 所示。此外，在鞭毛系统中还存在 X 和 Y 的“加和”作用，以及 *lac* 启动子中“与”、

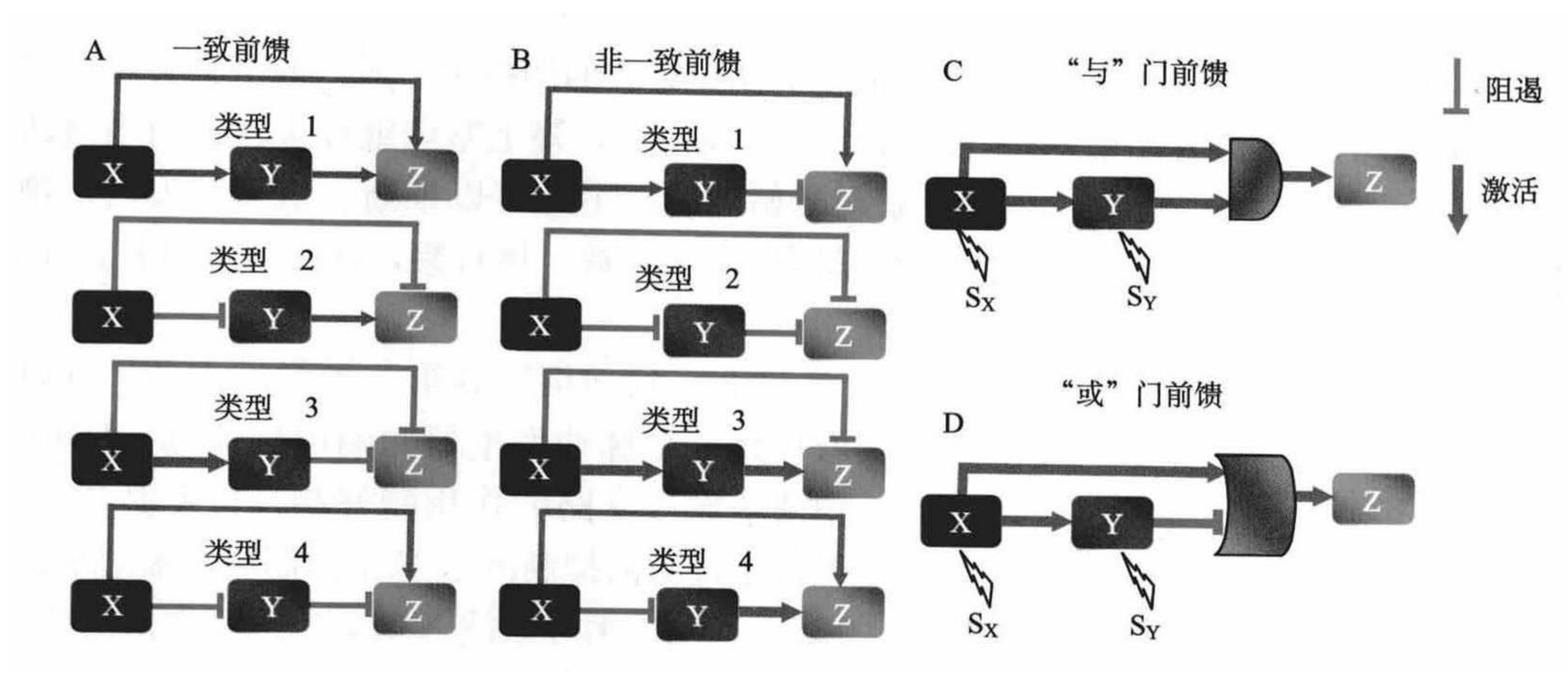


图 2-13　前馈回路 motif 的种类

最简单的前馈转录基元（motif）由三个基因组成：基因 X 调控基因 Y 和 Z，基因 Z 同时受到基因 X 和 Y 的调控。X、Y、Z 基因之间彼此的促进和抑制关系不同，导致前馈具有八种可能的结构类型。A 图所示为四种“一致前馈”类型，即 X 对 Z 的直接调控作用与 X 通过 Y 对 Z 的间接调控效果“一致”。B 图所示为四种“非一致前馈”类型，即 X 对 Z 的直接调控与 X 通过 Y 对 Z 的间接调控效果“不一致”。C 图“与”门表示 Z 的激活需要 X 和 Y 同时满足要求，如图 2-13C 所示；D 图“或”门表示 X 和 Y 中只要有一个满足要求即可激活 Z。

“或”的混合作用等。

鉴于“与”、“或”逻辑是最广泛的作用形式，下面我们就以“与”、“或”作用来解释前馈的具体运作。在研究最为透彻的大肠杆菌和酵母菌转录调控网络中，图 2-14“一致”和“非一致”的类型 1 是最广泛的前馈结构，这里简写为 C1-FF 和 I1-FF。

C1-FF 可以看作是一个“信号敏感性延迟（sign-sensitive delay)”单元和持续性检测单元。X 和 Y 是转录激活因子，假设 X 和 Y 对 Z 的调控作用是“与”的关系，则 C1-FF 结构对于上升刺激的响应具有一定的延迟，但对下降刺激的响应则无延迟特性。其机理如下：信号 S_X 出现，X 变为激活状态时，迅速与下游启动子结合，启动 Y，Y（为简单起见，在不会产生歧义时，我们用同样一个字母代表基因编码的产物）开始累积。因为 X 和 Y 对 Z 的调控作用是“与”的关系，只有当 Y 的浓度达到甚至超过阈值之后，Z 才开始启动。因此，从结果上看，Z 的表达滞后于信号 S_X 的出现。相反，移除信号 S_X，X 迅速失去活性，同样是由于 X 和 Y 对 Z 的调控作用是“与”的关系，此时，无论 Y 的状态如何，Z 的表达立刻被“关闭”。这种动态行为常被称为“信号敏感性延迟”，即对于信号 S_X 输入的滞后响应[128]。滞后时间由 Y 调节因子的生化参数决定。Y 对于 Z 的激活阈值越高，滞后时间越长。通常情况下，由于短时间内出现的脉冲噪声无法使 Y 的量积累到超过阈值限制，不会激活 Z 的响应，只有持续时间足够长的信号才能激活 Z 表达，因此这种滞后能够滤除瞬时脉冲噪声干扰。在大肠杆菌阿拉伯糖利用系统中就存在这种前馈结构。此系统对于 cAMP 信号的输入具有延迟效应，对其移除则没有延迟。延迟时间大概为 20min，略大于生长条件改变时欺骗脉冲的时间长度。

当 Z 启动子的两个输入是“或”逻辑时，C1-FF 对于信号的响应刚好和“与”逻辑

相反。在大肠杆菌的鞭毛系统中就有这种“或”逻辑的 C1-FF 结构。鞭毛马达基因相当于 Z，FlhDC（相当于 X）和 FliA（相当于 Y）对 Z 的调控为“或”的逻辑关系。当激活 FlhDC 的输入信号结束以后，FliA 不会立即消失，鞭毛基因继续表达；而当激活 FlhDC 的信号输入时，鞭毛马达基因立刻开始响应。由此可以推断，信号消失后，鞭毛马达基因表达延迟的时间取决于 FliA 基因的生化参数。据观察，这种延迟可长达 1h 左右。

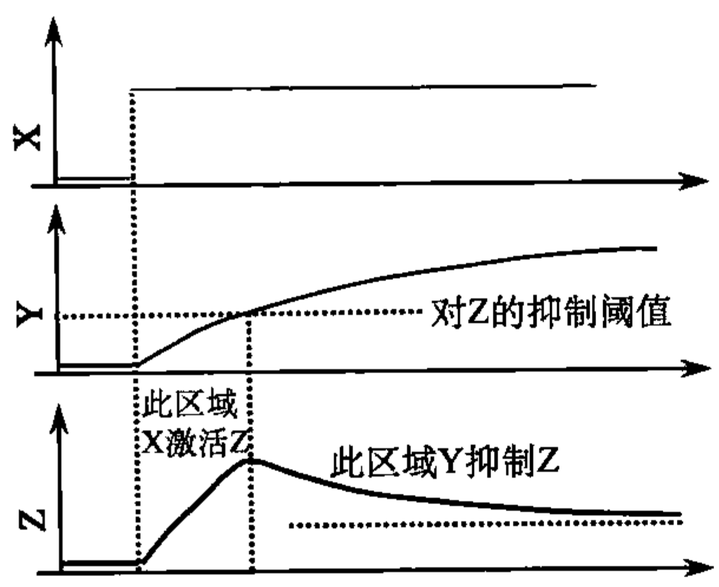


图 2-14 I1-FF 中 Z 的脉冲响应曲线

“非一致前馈”I1-FF 中，X 直接激活 Z，但也通过激活抑制因子 Y 来间接抑制 Z（参见图 2-13B）。当输入信号激活 X 时，X 对 Z 的直接激活作用迅速作出响应；而 X 对 Z 的间接抑制路径中，X 首先要激活 Y，一段时间以后，只有当 Y 的量超过对 Z 的抑制阈值时，Z 的产量才开始逐渐减少，浓度下降，因此 Z 的响应曲线类似于脉冲的动态特性。

“非一致前馈”的第一种类型 I1-FF 可以看作是一个脉冲发生器和响应加速器。I1-FF 中两条输入支路的作用刚好相反：X 激活 Z，但也通过激活抑制因子 Y 来抑制 Z（参见图 2-13）。当输入信号激活 X 时，Z 迅速作出响应，然而，一段时间以后，由于 X 同时也激活 Y，当 Y 的量超过对 Z 的抑制阈值时，Z 的产量开始减少，浓度下降，因此 Z 的响应曲线类似于脉冲的动态特性，如图 2-14 所示。极端情况下，当 Y 能够完全抑制 Z 时，脉冲会降到零点。除了类似脉冲响应，在 Y 的抑制作用产生之前的时间里，如果 Z 强烈的起始表达使得 Z 的浓度达到一定值，当 Y 不会完全抑制 Z，Z 的浓度达到一个非零稳态水平时，I1-FF 还可以看作具有响应加速器的功能。负自反馈和 I1-FF 都可以加速响应，负自反馈仅作用于转录因子（或者与转录因子在同一操纵子上的基因），而 I1-FF 结构却可以作用于任意目标基因 Z。大肠杆菌的半乳糖利用系统中就具有这种 I1-FF 结构。当缺乏半乳糖时，葡萄糖匮乏会快速诱导半乳糖代谢基因至适当的表达水平。这种结构的响应时间比单因素系统（*lac* 系统）快三倍。在哺乳动物细胞中，MYC（X）、*E2F1*（Z）和一个相当于 Y 的 miRNA 同样可以组成 I1-FF 结构。在秀丽隐杆线虫（*Caenorhabditis elegans*）中也发现了多种 miRNA 组成的前馈结构[119]。

将上述基本拓扑结构合理组合，即可以实现很多复杂调控结构，如前馈-反馈、反馈-单输入、多层反馈，等等，如图 2-15 所示。

除了上述基本的逻辑拓扑结构外，还有许多其他拓扑结构，如前一章中讲到的在代谢网络中非常普遍的无标度网络等。鉴于篇幅有限，这里就不一一讲解了。对一个复杂系统的研究，往往是从其基本逻辑拓扑结构开始的。例如，人工级联线路清晰的模块化层次性结构使其成为噪声传播、干扰和抑制等方面机理研究的有利工具[129~132]。当前，人们还远远没有完全理解生物系统网络中包含的所有调控机理。生物学家正在从网络的结构、组织方式和相关的功能分析等方面对生物调控网络进行研究。虽然有机活体的基因调节网络在速度和复杂度上受到许多实际的物质限制，但是理论上，任何人工的生物调控机制均可以通过上述结构（级联、前馈和反馈）的合理组合来实现。

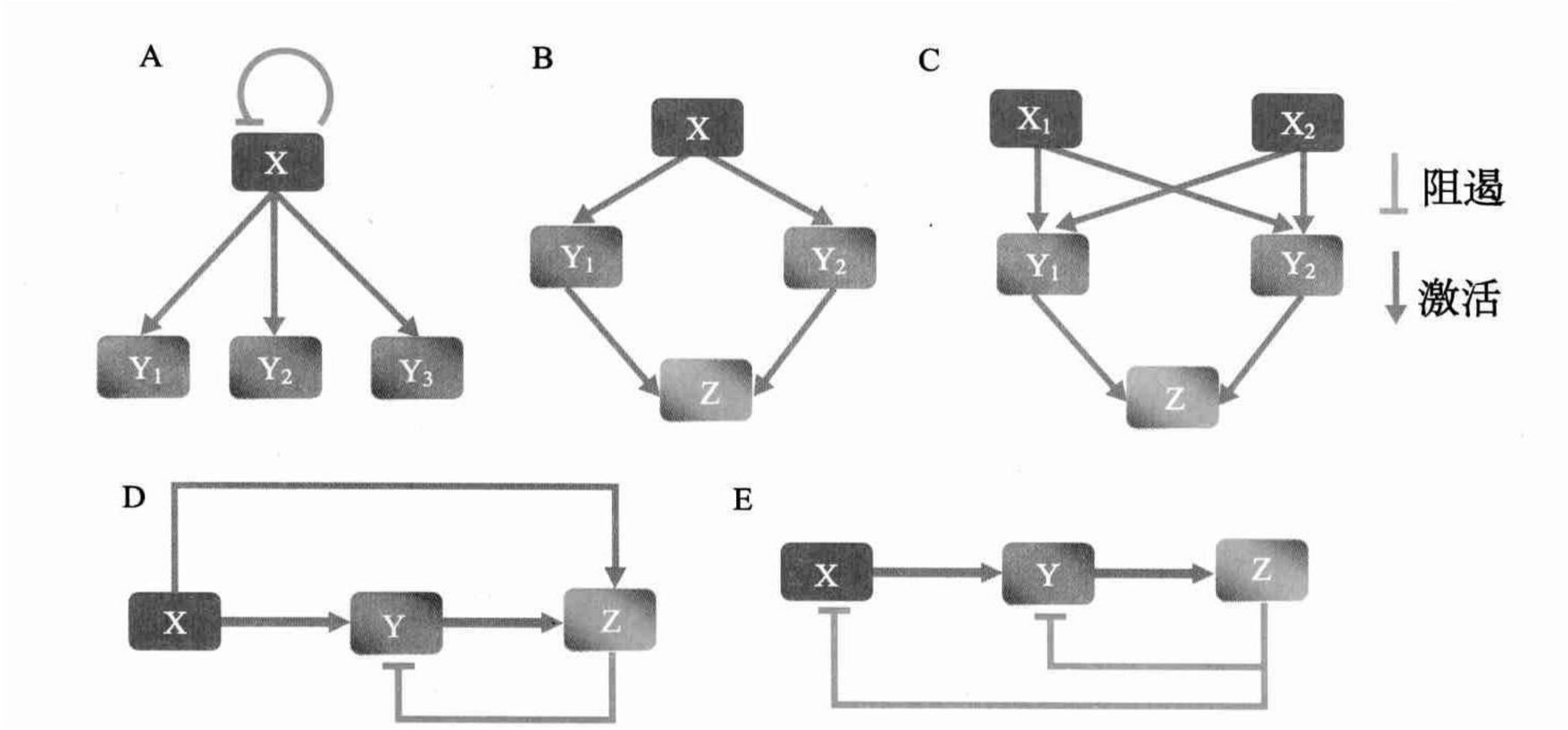


图 2-15　复杂逻辑拓扑结构

A. 自反馈＋SIM；B. 钻石型；C. 感知机型；D. 前馈＋反馈；E. 多层反馈

一门科学，只有当它成功地运用数学时，才能达到真正完善的地步。

——马克思

第3章　合成生物系统的数学模拟与性能分析

基础提要

- 微积分基本知识
- 多元统计分析
- 酶动力学

合成生物学强调“设计”和“重设计”。大量借助计算机科学、信息科学、数学和物理学原理，利用已有的生物学知识，建立数学模型，对合成生物系统进行模拟和性能分析，指导和优化实验设计，是合成生物学的重要研究手段。

第一部分　生物系统的数学模拟

随着基因操作技术的飞速发展，人们已经能够构建出一些非常复杂的遗传线路，并将其导入不同的生物体中进行表达，获得目标生物功能。然而，理论的发展并没有跟上实验发展的脚步。基因的表达受到不同水平因素的影响，如启动子的启动强度、催化因子的活性、终止子的终止能力等。此外，在特定条件下，DNA 和 RNA 拓扑结构产生的序列单元激活水平、宿主细胞的代谢物、蛋白质加工和折叠系统的能力、营养物质的丰富程度和环境压力等均是影响基因表达的重要因素。目前，人们还不能准确获得生物系统基因表达过程中的所有参数，也就无法准确预测生物系统的响应行为。因此，有效结合实验验证和算法开发，利用实验验证和优化模型，通过模型来指导实验是行之有效的方法之一。

3.1　常微分方程

目前有许多建模方法可用于描述生物系统，且每种建模方法都有自己的特点。其中，利用生物系统的动力学特性建立的各种常微分方程（ordinary differential equation，

ODE）方法，由于其连续性和相对简单性成为最常用的方法。尽管这种方法有一定的缺陷（如方程成立的理论机理与真实机理存在差异、大量参数目前很难测定等），但仍具有很大的应用价值[133,134]。

凡是表示未知函数的导数以及自变量之间关系的方程，都可以称为微分方程。当需要知道物质在一定条件下的运动变化规律时，常常需要用到微分方程。其中应用比较普遍的一类就是常微分方程。

常微分方程（ordinary differential equation，ODE）是指如果在一个微分方程中出现的未知函数只含一个自变量，这个方程就叫做常微分方程。

例如，在各种生化反应中，通常人们最关心的是在某些影响下反应物浓度随时间的变化。对于反应 $B+C \xrightarrow{p} A$，由质量守恒定律能够推导出平衡反应方程[135~137]：

$$\frac{d[A]}{dt} = [B] \times [C] \times p \tag{3-1}$$

其中，A 是 B 和 C 反应的产物；［*］表示物质的浓度；p 是这些分子的亲和力常数。B 和 C 在形成 A 反应过程中的竞争抑制关系可采用下式来描述：

$$\frac{d[A]}{dt} = \frac{1}{1+[B] \times [C]} \times p \tag{3-2}$$

其中，p 同样表示反应分子亲和力常数。上述两式中常数 p 的数值很难确定，通常要用实验的方法才能获得近似值。

下面我们就来介绍几种常用的常微分方程。

3.1.1 质量作用定律

细胞中的新陈代谢可以看作是各种各样的化学反应，因此化学反应中的许多基本定律对于生物系统建模仍然非常有用。这其中较著名的一个就是质量作用定律（the law of mass action），并且可基于此构建化学平衡反应、米氏动力学方程等来描述生物系统中的相互作用。

19 世纪中期，挪威的克利斯蒂安娜（如今的奥斯陆）大学应用数学教授古德贝格（C. M. Guldberg，1836～1902）和化学教授瓦格（P. Waage，1833～1900）提出：化学反应速率与反应物的有效质量成正比。此即质量作用定律，其中的有效质量实际是指浓度。近代实验证明，质量作用定律只适用于元反应，因此该定律可以更严格且更完整地表述为：元反应的反应速率与各反应物浓度的幂的乘积成正比，其中各反应物浓度的幂的指数即为元反应方程式中该反应物化学计量数的绝对值。例如，反应 $aA+bB \longrightarrow gG+hH$，它的反应速率公式可表示为 $v=k[A]^m[B]^n$。其中，［A］、［B］分别表示反应物 A 和 B 的浓度；m、n 是反应的级数；k 是反应速率常数，在数值上等于单位浓度时的反应速率。不同反应的 k 值不同，同一反应的 k 值与反应温度、催化剂等因素有关，而与反应物浓度、压强等无关[138]。

质量作用定律成立的假设是：理想情况下分子的运动是自由的和彼此独立的，两个分子碰撞的概率和发生反应的概率可以表示为占据反应空间的分子体积乘以确定常数的乘积。

虽然质量作用定律只是理想气体和晶体的经验法则，但在实际应用中却可用于推导非理想系统的行为。在细胞中进行的化学反应不是在理想气体中进行的，而是在非理想的液体中进行的；分子不是单调分布也不能自由移动，但是可以证明，在相对稳定条件下，仍然能使用上述质量作用定律。

由反应物分子（或离子、原子、自由基等）直接作用而生成新产物的反应，叫做元反应。简单反应只包含一种元反应，可直接应用质量作用定律。在复合反应中，不是所有反应物分子都直接作用，所以质量作用定律不能直接应用于复合反应，必须将其分解成一步一步的元反应，质量作用定律才适用。因此，进行实际系统描述时，通常需几十个甚至成千上万个质量作用定律方程来描述，计算量和所需的存储空间非常巨大，极大地限制了质量作用定律的应用。

3.1.2　米氏方程

米氏方程（Michaelis-Menten equation）是在上述质量作用定律的基础上应用较普遍的一个描述酶促反应起始速度 V 与底物浓度 [S] 之间关系的速度方程，是在假定存在一个稳态反应条件下推导出来的。对于简单的酶促反应，反应速率可用式（3-3）所示的米氏方程来描述[139]。

$$V = \frac{V_{max}[S]}{K_m + [S]} \tag{3-3}$$

其中，K_m 值为米氏常数，由酶的性质决定，而与酶的浓度无关，因此可用 K_m 的值来鉴别不同的酶；V_{max} 是酶被底物饱和时的反应速度，单位一般为 mol/L。米氏方程的模拟图如图 3-1 所示。

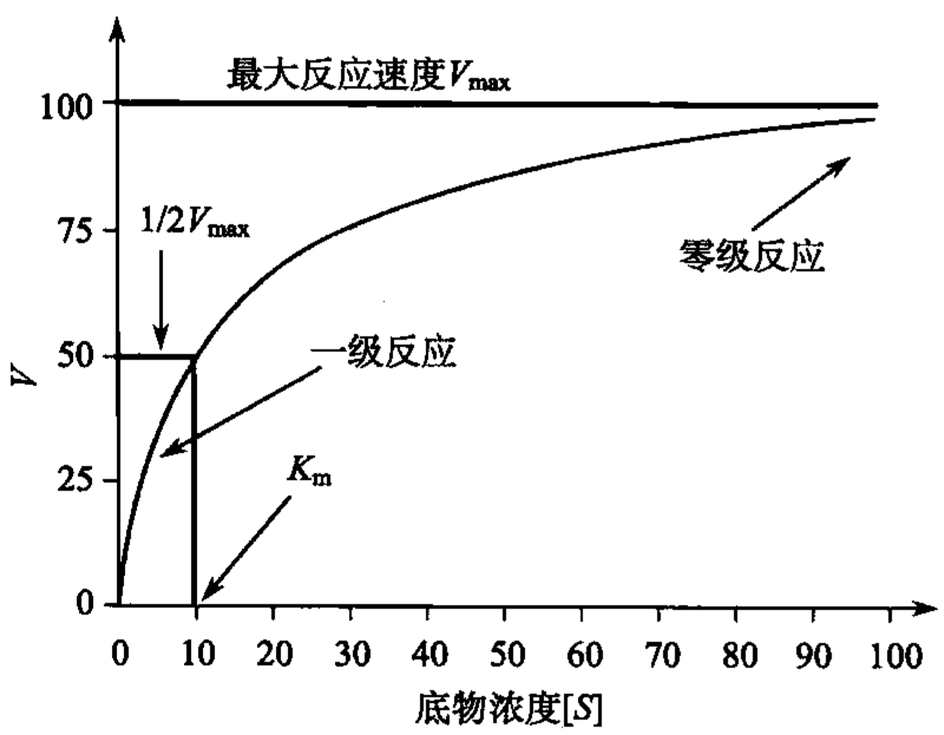


图 3-1　米氏方程模拟图

低底物浓度时，反应速度近似于底物浓度的一级反应。当底物浓度非常大时，反应相对于底物浓度是零级反应。当底物浓度处于中间状态时，反应由一级反应向零级反应过渡，可以看作是混合级反应，此时速度 V 相对于 [S] 的曲线为双曲线。

在低底物浓度时，分母中的 [S] 相对于 K_m 可以忽略不计，反应速度近似于底物浓度的一级反应（first order reaction）。当底物浓度非常大时，分母中的 K_m 相对于 [S] 可以忽略不计，反应速度接近于 V_{max}，此时酶几乎被底物饱和，再增加底物对反应速度几乎没有什么影响，此时称反应相对于底物 S 是零级反应（zero order reaction）。当底物浓度处于中间状态时，反应由一级反应向零级反应过渡，反应速度相对于底物浓度可以看作是混合级反应（mixed order reaction），此时速度 V 相对于 [S] 的曲线为双曲线。

在用米氏方程进行建模时，通常把遗传线路中的一些反应，如 RNA 聚合酶与 DNA 的结合反应等看成为可逆酶促反应，利用米氏方程对这些反应步骤进行建模，便可得到该线路的动力学模型。

3.1.3 希尔方程

希尔方程（Hill equation）是合成生物学中关于遗传线路调控的常用方程，最初由 A. V. Hill 于 1910 年提出，描述 O_2 与血红素的结合。在不考虑启动子的渗漏表达时，方程的形式如式（3-4）所示，其中，α 为启动子最大启动能力（设为 1）；A 为被调控基因的表达速率；X 为调控因子量；$m>0$ 为 HILL 系数；X_M 为模型变化趋势改变时的 X 值。做不同 m、X_M 取值下 HILL 方程模拟基因调控图，如图 3-2 所示。图中横坐标表示调控因子的量，纵坐标表示被调控基因的相对表达速率。m 值越大，调控因子对于基因的调控行为越接近于阶跃的开关行为，在 X_M 处转变[140]。可以看出，HILL 方程是当 $m=1$ 时的米氏方程[141]。

$$\begin{aligned} \frac{dA}{dt} &= \frac{\alpha X^m}{X^m + X_M^m} && \text{正调控} \\ \frac{dA}{dt} &= \frac{\alpha}{1+(X/X_M)^m} && \text{负调控} \end{aligned} \tag{3-4}$$

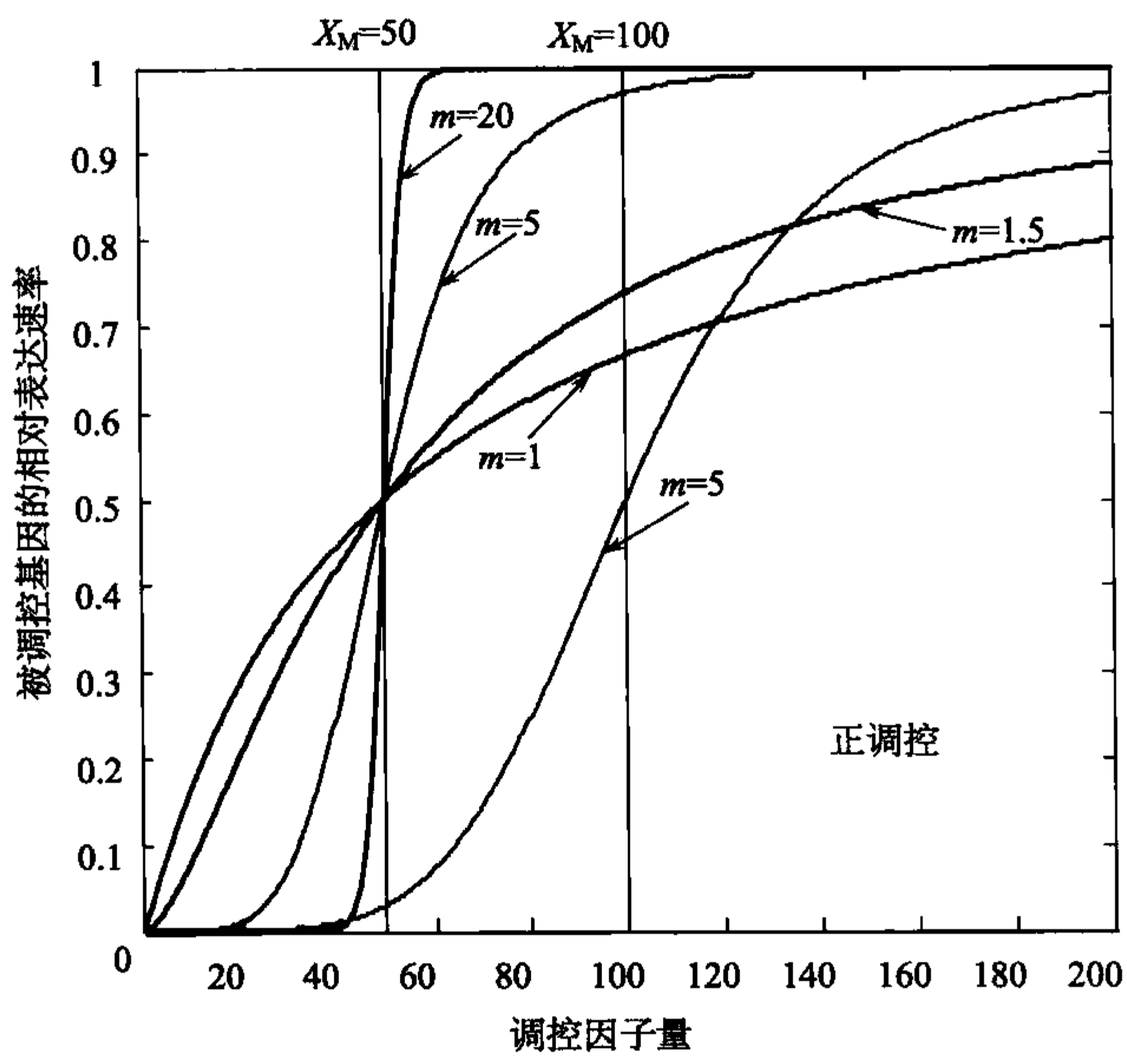


图 3-2 不同参数取值下 HILL 方程模拟基因调控图

图中横坐标表示调控因子的量，纵坐标表示被调控基因的相对表达速率。当 m 值增加的时候，调控因子对于这个基因的调控行为接近于阶跃的开关行为，在 X_M 处转变。

如果考虑启动子渗漏表达，设 ξ 为启动子的渗漏量，与 α 单位一致。此时希尔方程可表述为

$$\begin{aligned} \frac{dA}{dt} &= \frac{(\alpha-\xi)X^m}{X^m + X_M^m} + \xi && \text{正调控} \\ \frac{dA}{dt} &= \frac{\alpha-\xi}{1+(X/X_M)^m} + \xi && \text{负调控} \end{aligned} \tag{3-5}$$

如果同时考虑基因表达产物的降解，则可用下面的方程描述（其中 δ 表示降解速率）

$$\frac{\mathrm{d}A}{\mathrm{d}t}=\frac{(\alpha-\xi)X^m}{X^m+X_{\mathrm{M}}^m}+\xi-\delta A \qquad \text{正调控}$$
$$\frac{\mathrm{d}A}{\mathrm{d}t}=\frac{\alpha-\xi}{1+(X/X_{\mathrm{M}})^m}+\xi-\delta A \qquad \text{负调控} \tag{3-6}$$

3.2　Logistic 模型

Logistic 模型是目前常用的描述种群生长状态的模型，是在米氏方程基础上衍化而来的。

如果假定细胞在无限空间、无限营养来源和无约束条件下生长或繁殖，则其生长量或繁殖量 N 将是与种群密度无关的、以时间 t 为指数的函数，为

$$\frac{\mathrm{d}N}{\mathrm{d}t}=rN,\quad r=b-d \tag{3-7}$$

其中，b 为新增率；d 为死亡率。其时域的解为：

$$N(t)=N_0\mathrm{e}^{r(t-t_0)} \tag{3-8}$$

其中，$N_0=N(t_0)$ 为初始时刻 t_0 的值。式（3-8）即为种群增长的米氏方程。例如，设 $t=0$ 时刻有 3 个细胞，它们每经过 1h 便分裂 1 次，则经 xh 后，细胞总数 N 为：

$$N=3\times2^x=3\times(\mathrm{e}^{0.693164})^x \tag{3-9}$$

由上式可得，6h 后的细胞数为：$N=3\times2^6=192$，10h 后的细胞数为：$N=3\times2^{10}=3072$。

但在自然情况下，上述关系通常只能在生长或繁殖初始的瞬间成立。当细胞生长量增大后，环境恶化、养料不足、衰老加快、繁殖减慢、死亡增多，致使上述指数增长关系式不能保持下去，物种的相对生长率必然随着 N 的增大而减小，由此可以引入一个竞争项，模拟生物间的反馈作用等约束条件，此即为 Logistic 模型或生物总数增长的统计筹算律。Logistic 模型是由数学生物学家弗赫斯特（P. F. Verhulst）首先提出的，且已经有大量实验资料表明用 Logistic 模型来描述种群的增长，效果还是相当不错的。

3.2.1　单种群细胞自由生长的 Logistic 模型

在环境总承受量一定的情况下，由于空间和资源都是有限的，不可能供养无限增长的种群个体，当个体数量过多时，由于平均资源占有率的下降和环境恶化等原因，新增率降低而死亡率却会提高。设环境能供养的种群个体数量的上界为 N_{m}（近似地将 N_{m} 看成常数），N 表示当前的种群个体数量，$N_{\mathrm{m}}-N$ 则为环境还能供养的种群个体数量，个体增长率与两者的乘积成正比。此时细胞生长的 Logistic 模型可以通过反馈抑制模型推导得来。设 X 表示种群中的一个个体。该模型可以简化为两个反应，分别是 X 的复制和 X 对自身增长的抑制，见式(3-10)。

$$X \xrightarrow{k} 2X \quad \text{复制(产生一个 } X\text{)}$$
$$2X \xrightarrow{k/N_m} X \quad \text{自抑制(消耗一个 } X\text{)} \tag{3-10}$$

根据前述质量作用定律，可以推导出物种浓度增长速率，如式（3-11）所示。

$$\mathrm{d}N/\mathrm{d}t = k \cdot N \cdot (1 - N/N_{\mathrm{m}}) \tag{3-11}$$

上两式中 N 代表种群个体个数；N_{m} 代表系统营养所能供养的最大种群个体数量；k 代表细胞生长速率。

Logistic 模型见图 3-3 所示。图中横坐标表示时间，单位是小时（h）。模型中取 $N_{\mathrm{m}}=5\times10^8\mathrm{CFU/ml}$，$k=0.9\mathrm{h}^{-1}$。实线表示细胞的浓度，虚线表示细胞增长速率。可以看出，细胞的增长速率随着细胞浓度的增加先增加再减少，细胞的最终浓度等于特定营养条件下的最大浓度，符合实际中微生物的生长情况。

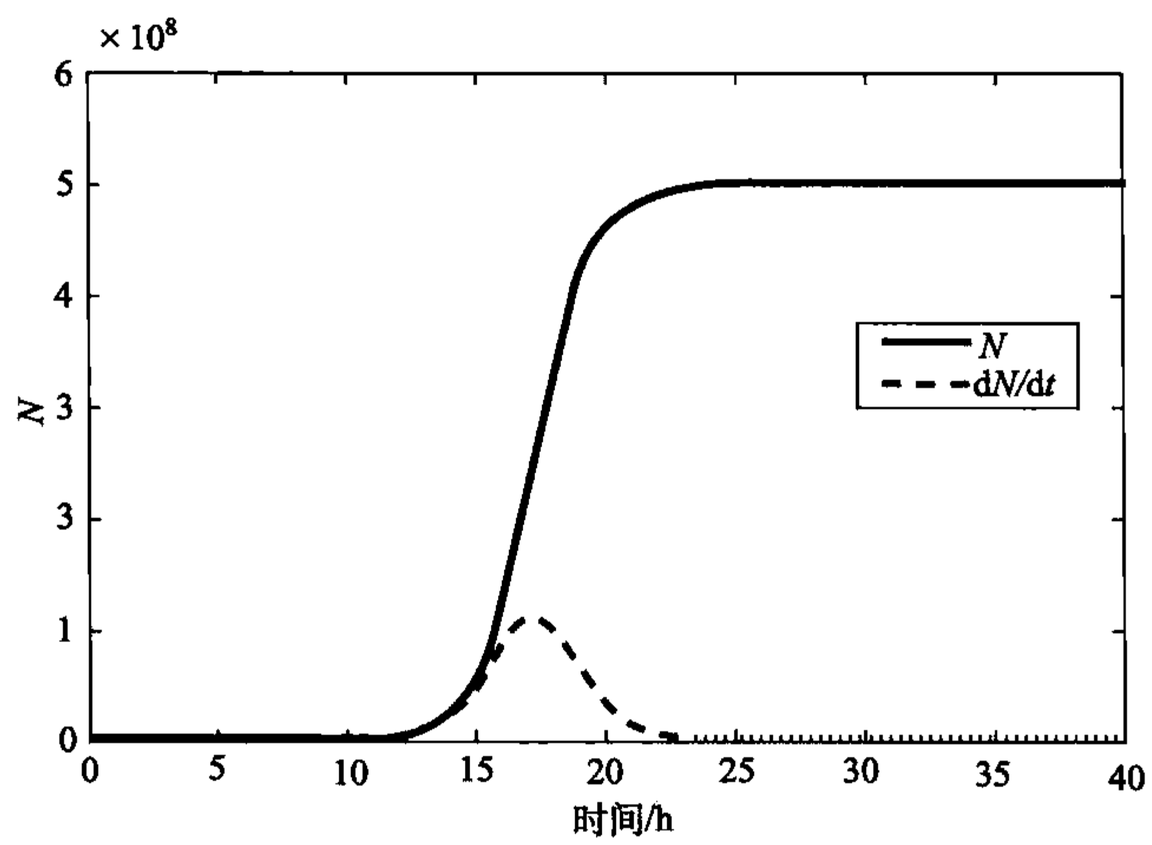


图 3-3 Logistic 模型的示意图

3.2.2 两种细胞自由生长的 Logistic 模型

当培养基中有两种细胞共同生长的时候，情况则会相对复杂。如果不考虑两种细胞之间直接的相互竞争（例如，分泌化合物或者改变培养基的 pH 等影响对方的生长），仅考虑营养限制条件下两种细胞自身的复制、对于自身增长的抑制作用、对于另一种细胞增长的抑制作用，则可以把两种细胞共同生长的模型简化为式（3-12）所示。

$$X_1 \xrightarrow{k_1} 2X_1 \quad X_1 \text{ 复制(产生一个 } X_1\text{)}$$
$$X_2 \xrightarrow{k_2} 2X_2 \quad X_2 \text{ 复制(产生一个 } X_2\text{)}$$
$$2X_1 \xrightarrow{k_1/N_{\mathrm{m}}} X_1 \quad \text{自抑制(消耗一个 } X_1\text{)} \tag{3-12}$$
$$2X_2 \xrightarrow{k_2/N_{\mathrm{m}}} X_2 \quad \text{自抑制(消耗一个 } X_2\text{)}$$
$$X_1 + X_2 \xrightarrow{k_3/N_{\mathrm{m}}} \quad \text{相互抑制(消耗一个 } X_1\text{,一个 } X_2\text{)}$$

同样，根据质量作用定律，可以得到两种细胞的增长速率公式，见式（3-13）。其中，k_1、k_2 为两个物种自身的生长速率，k_3 为种群间相互抑制对方增长的参数。

$$\begin{aligned} dN_1/dt &= k_1 \cdot N_1 \cdot (1-(N_1+(k_3/k_1)\cdot N_2)/N_m) \\ dN_2/dt &= k_2 \cdot N_2 \cdot (1-(N_2+(k_3/k_2)\cdot N_1)/N_m) \end{aligned} \tag{3-13}$$

其中，细胞增长速率 k 和最大细胞浓度的取值与式（3-11）相同，下文 k 和 N_m 的取值也与此同。如果设计的系统都是相同菌种，仅是具有的功能不同而已，例如，具有不同功能质粒的大肠杆菌菌种，其生长速率可简化为：$k=k_1=k_2=k_3$。将其代入式（3-13）中可以得到两种不同功能的大肠杆菌共同生长时各自生长速率的表达式，见式（3-14）。将式（3-14）的两个式子加和得式（3-15）。当 $N=N_1+N_2$ 时，式（3-15）就还原为式（3-11），即单种群的自由生长模型。

具有两种不同功能质粒的大肠杆菌共同生长的情况如图 3-4 所示。横坐标表示时间，单位是小时（h）。实线、虚线分别表示两种不同功能的细胞。计算时取初始浓度分别为 2000 cfu/ml（N_1，实线）与 1000 cfu/ml（N_2，虚线）。从图中可以观察到，两种细胞分别不停地生长，随着浓度和的增加，各自的生长速率逐渐减小，当浓度和到达最大承受浓度时，同时停止生长，保持稳定。细胞初始浓度的不同，决定了系统到达最终浓度时不同功能细胞能够到达的各自最终浓度的不同。

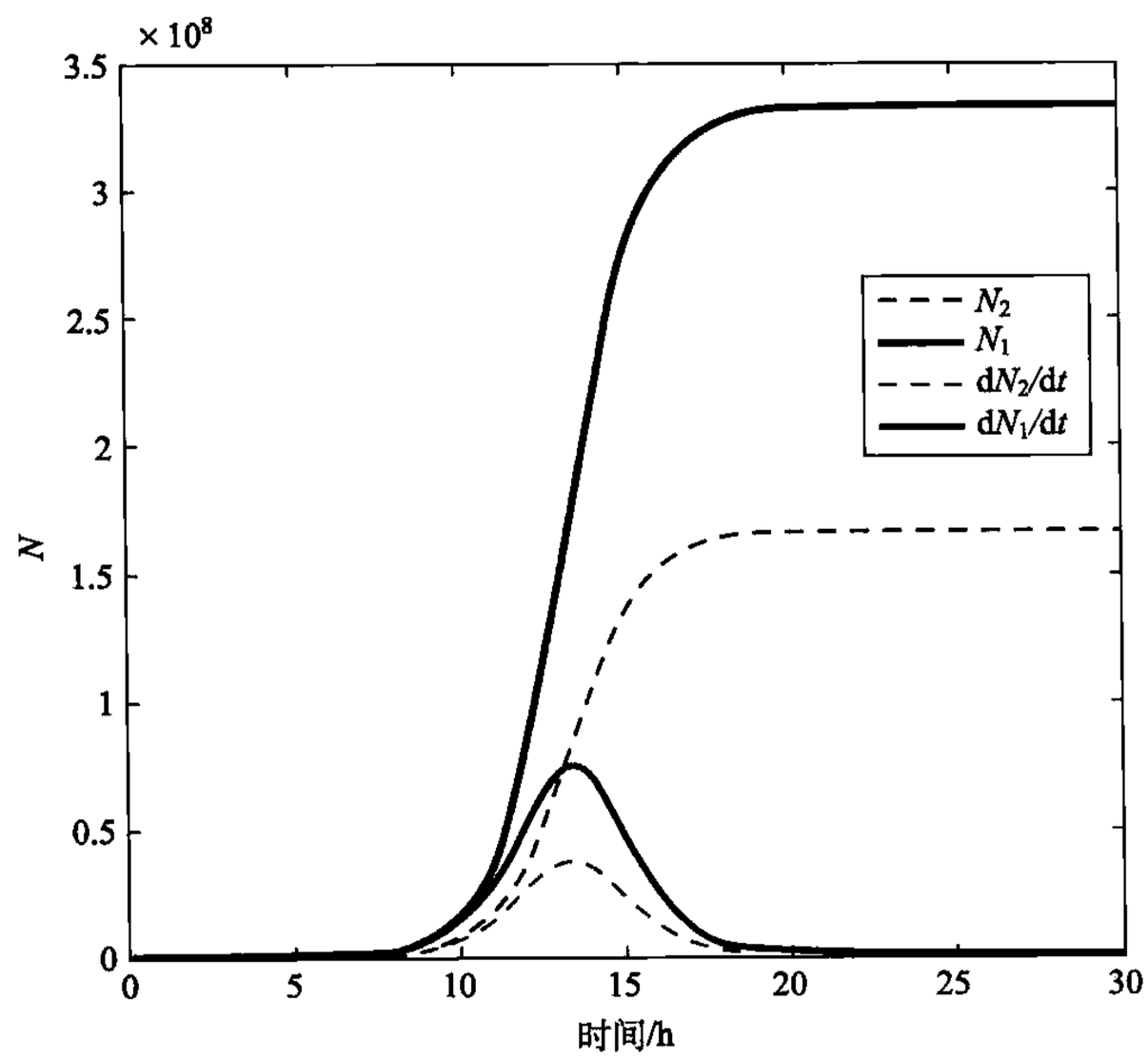


图 3-4　两种具有不同功能质粒的大肠杆菌细胞共同生长示意图

计算时取初始浓度分别为 2000 CFU/ml（N_1，实线）与 1000 CFU/ml（N_2，虚线）。两种细胞分别不停地生长，随着浓度和的增加，各自的生长速率逐渐减小，当浓度和到达最大承受浓度时，同时停止生长，保持稳定。细胞初始浓度的不同，决定了系统到达最终浓度时不同功能细胞能够到达的各自最终浓度的不同。

$$dN_1/dt = k \cdot N_1 \cdot (1-(N_1+N_2)/N_m)$$
$$dN_2/dt = k \cdot N_2 \cdot (1-(N_1+N_2)/N_m) \quad (3\text{-}14)$$
$$d(N_1+N_2)/dt = k \cdot (N_1+N_2) \cdot (1-(N_1+N_2)/N_m) \quad (3\text{-}15)$$

3.3 主方程

前面讨论的模型都是确定性的模型，不考虑基因表达时的随机波动。当基因高速表达、反应物浓度比较大的时候，环境变化对于基因表达的影响可以忽略不计，此时可以用确定性模型来表述基因表达的性状。但是当表达速率相对较低、反应物的量相对较少的时候，各种随机因素造成的影响无法忽略，基因的表达就会在一定范围内出现随机波动的现象，此时上述确定性模型就无法准确反映被描述系统的真实情况，必须采用随机模型[142]。

主方程（master equation）[136]是描述离散状态随时间变化的随机概率模型，通常用于研究基因个体行为。由中心法则可以知道，基因激活时转录为 mRNA，继而 mRNA 翻译为蛋白质。因此，主方程模型的建立主要基于下列假设：非活性基因 I 可以被转变为激活基因 A，这是一个速率分别为 b 和 c 的可逆反应

$$I \underset{c}{\overset{b}{\rightleftharpoons}} A \quad (3\text{-}16)$$

激活基因以 $H \cdot G(A)$ 的速率转录为 mRNA，其中，H 是转录速率；$G(A)$ 表示基因的开关状态 $[G(A)=1;\ G(I)=0]$；ϕ 代表 mRNA 的降解产物。

$$A \xrightarrow{H \cdot G(A)} \text{mRNA} \longrightarrow \phi \quad (3\text{-}17)$$

mRNA 以速率 K 翻译成蛋白质，蛋白质以速率 r 降解

$$\text{mRNA} \xrightarrow{K} \text{蛋白质} \xrightarrow{r} \phi \quad (3\text{-}18)$$

如果用 x 和 y 代表 mRNA 和蛋白质的量，G 为表示基因是否激活的随机二进制数，则可用三元函数（x；y；G）描述系统。x 和 y 的联合分布可表示为下列概率函数：

$$f_{xy}(t) = P[\#\text{mRNA} = x, \#\text{protein} = y, G = 0];$$
$$g_{xy}(t) = P[\#\text{mRNA} = x, \#\text{protein} = y, G = 1] \quad (3\text{-}19)$$

式（3-19）表示细胞中当基因激活或未激活时 mRNA x 和蛋白质 y 的量。基于此分布函数，结合化学反应速率，系统随时间的变化可用主方程描述为：

$$\frac{df_{xy}}{dt} = bg_{xy} - cf_{xy} + (x+1)f_{x+1,y} - xf_{xy} + Kxf_{x,y-1} + r(y+1)f_{x,y+1} - (Kx+ry)f_{xy} \quad (3\text{-}20)$$

$$\frac{dg_{xy}}{dt} = -bg_{xy} + cf_{xy} + G(A)Hg_{x-1,y} + (x+1)g_{x+1,y} - (HG(A)+x)g_{xy} + Kxg_{x,y-1} + r(y+1)g_{x,y+1} - (Kx+ry)g_{xy} \quad (3\text{-}21)$$

其中，x，$y \in \mathbf{N}_0$。前两项（$bg_{xy} - cf_{xy}$）表示启动子激活导致的概率变化。激活基因的

数目越多，非激活基因的概率倾向越大，反之亦然。

当分子数量较少、大量 x 和 y 基因的存在可以忽略其概率影响时，系统可用有限数量的 ODE 方程描述，这些方程可通过数值解法求解。然而，随着分子数目的增加，ODE 的数量以二次函数的形式增长。对于真核细胞，仅一个过程就会涉及上百个 mRNA 分子和几十万个蛋白质，通常要由 10^7 个 ODE 方程来描述，几乎不存在 ODE 建模的可行性。即使是少量分子组成的系统，严格意义上讲，系统仍然要用无数的方程来描述。因此，尽管主方程理论上能够获得系统的精确描述，但在实际应用中用其描述真实系统仍然是不太可行的。从实用角度讲，需要对主方程进行简化，用随机微分方程来描述系统。

3.4　随机微分方程

上述反应系统中，基因的状态可用随机过程 $G(t)$ 来描述：

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -x + H \cdot G(t) \tag{3-22}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = KX - ry \tag{3-23}$$

其中，x 和 y 分别代表 mRNA 和蛋白质水平。方程（3-22）此时为依赖于组分 $G(t)$ 的随机微分方程。蛋白质产生和降解仍然采用 ODE 方程。因为 y 依赖于 x，进而间接依赖 $G(t)$，所以 x 和 y 均可看作是连续的随机变量。不考虑蛋白质/mRNA 产生的噪声干扰，上式可以看作仅基因的活性是随机组分的主方程的近似模型。为了从这个模型中得到有用解，可用期望值和方差来表征其稳态属性[143]。

当处理含有大量蛋白质的系统时，上述模型仍过于复杂。此时可以考虑忽略随机性来简化系统模型，由此得到的解虽然不是精确解，但却是可行解。

随机模型的模拟还有很多其他方法，如蒙特卡罗（Monte-Carlo）模拟和 Gillespi 模拟[144]等。这些方法的计算量较大，速度也较慢，对于计算机的要求比较高。为此，D. T. Gillespie 提出了一种新的“τ-算法”，以牺牲一定精确度为代价来提高运算速度。这种算法首先自动选取一个时间间隔 τ，并假设在这个时间间隔内各个反应发生概率的误差在可接受范围内，之后计算所有反应在这段时间内发生的次数，最后根据各个反应发生的次数计算系统分子数的改变量。这种算法的计算速度比原始的 Gillespie 算法提高了几个数量级[145]。但是 Gillespie 算法和 τ-算法用同样的方法计算所有反应，并未将反应分类。因此，E. L. Haseltine 等和 H. Salis 等提出将反应按照发生的速度快慢分为快速反应与慢速反应，用常微分方程计算快速反应，用随机方法计算慢速反应[146,147]。M. Griffith 等根据以前的工作提出了将反应分为连续和离散反应的计算方法等[148]。以上这些算法都在一定程度上提高了计算效率。

3.5　模式识别

在基因表达预测、基因数据识别、代谢网络分析等多个领域常常会用到模式识别方

法。模式识别有时也称作知识发现，是目前人类从大量数据中获取有用信息的主要手段。

按照广义的定义，所谓模式是指一些供模拟用的、完美无缺的样本。存在于时间和空间中可观察的事物，如果我们可以区别它们是否相同或相似，都可以称之为模式。但模式不是事物本身，而是我们从事务中获得的信息。模式识别就是识别出特定客体所模拟的样本，其作用和目的在于面对某一具体事物时将其正确地归入某一类别。这里的客体可以是人类能用感官直接或间接接收的外界信息，如声音、图像、文字；也可以是各种检测数据，如心电图、脑电图、地震波。在合成生物学中，基因的表达水平、基因的调控模式[149]、代谢网络规模、蛋白质和基因等大分子的种类及相互关系等都可以看作模式。

模式识别的方法主要依靠通过对事物属性的度量值进行计算来对事物进行分类。从广义上说，模式识别属于人工智能的范畴，但由于历史的原因，它已经形成了独立的学科，有其自身的理论和方法。目前模式识别主要研究对事物的识别，俗称分类。例如，在疾病诊断中将样本分为有病、无病两类，进而寻找致病基因或者可进行疾病诊断的生命表征；在代谢网络识别中，根据连接度不同将节点分为中心节点、边界节点等，或者根据连通性将大规模网络拆分成几个子网络；在代谢组学中，将样本分为高产和非高产两类，并找出导致高产的关键因素等。

模式识别中有许多具体算法，根据分类信息是否已知，可将这些算法分为监督学习方法和非监督学习方法两大类。在进行分类器设计时，假定每个样本的类别都是已知的，即每个样本均被“标记”了类别属性，利用已知标记类别的样本集进行分类器设计的方法称为监督学习方法或有导师学习方法，也称为分类分析。但在大多数实际应用中，无法预先知道样本的类别，只能根据无样本类别标记的样本集进行分类器设计，这就是通常说的非监督学习方法或无导师学习方法，也称为聚类分析。

（1）非监督学习方法也称聚类分析，用途很广，用来分析数据的内在规律。一般聚类强调数据自然聚集状况分析，没有检验或者测试样本。当人类没有任何先验知识可以利用时，可以采用非监督学习方法对数据进行初步分析，再结合具体背景分析数据特性。例如，可以通过解释网络连通性提供系统结构信息，此时网络的边界可以代表直接的因果关系、间接的相互作用或者常规因素，还可以分析特定过程的微阵列数据来解释基因之间的关系[150]。

（2）监督学习方法也称分类分析，其用途很明确，就是要对样本进行分类。训练样本集中给出不同类别样本的实例，从这些实例中找出区分不同类样本的方法，从而在特征空间中划定决策域，寻找导致类别不同的关键因素。例如，寻找致病基因或者导致某次生代谢产物高产（例如广谱抗癌物质紫杉醇）的关键催化酶等。

下面我们来分别介绍这两大类方法中常用的一些算法。

3.5.1 聚类分析[151]

如前所述，聚类与分类和预测不同，通常是在没有任何先验知识的情况下对数据进行分析处理。聚类前通常并不知道将要划分的类别的数量和类型，也不知道根据哪些属

性来定义类别。聚类的原则是根据一定的度量函数在尽量扩大类间距、缩小类内距的原则下将数据划分到不同的组中。

常用聚类算法概述

1）划分方法（partitioning method，PAM）

划分方法将数据集划分为 k 个子集（划分），每个子集代表一个聚类（$k \leqslant n$，n 为样本数），这些划分必须满足：①每个划分至少应包含一个对象；②每个对象必须只能属于某一划分。

后一个要求在一些模糊划分方法中可以放宽。具体分析时，在事先给定需要划分的个数 k 的情况下，创建一个初始划分；利用循环再定位技术，即通过移动不同划分中的对象来改变划分内容，在一定指标函数下对划分进行优化。

大多数划分算法通常采用以下两种启发式方法。

（1）k-means 算法：首先从 n 个数据对象中任意选择 k 个对象作为初始聚类中心；对于其余对象，则根据它们与这些聚类中心的相似度（距离），分别将它们分配给与其最相似的（聚类中心所代表的）聚类；然后再计算每个所获新聚类的聚类中心（该聚类中所有对象的均值）；不断重复这一过程直到标准测度函数开始收敛为止。一般都采用均方差作为标准测度函数。

（2）k-medoids 算法：k-means 算法用对象的均值作为聚类中心，对异常值较敏感。为此，在 k-medoids 算法中对 k-means 算法进行改进，采用相应聚类中离聚类中心最近的对象来表示。分割算法的原则是使得每个个体与其参照点的非相似性之和最小。当数据中存在噪声和奇异值时，k-medoids 算法比 k-means 算法具有更高的鲁棒性，因为一个类中具有代表性的中心点比该类中所有个体的均值更不易受异常值的影响。

上述两个算法均需穷举所有可能的对象划分直到得出最优划分或者满足某一终止条件为止，在分析中小规模数据集以发现圆形或球状聚类时效果很好；但当样本数量很大时（如各种组学数据），计算速度和效果均不十分理想。

2）层次方法（hierarchical method）

层次方法可以分为自上而下（分解）和自下而上（合并）两种操作方式。

自下而上的层次方法从每个对象均为一个单独的组开始，逐步将相近的组进行合并，直到所有数据分为一个大组或满足终止条件为止。

自上而下层次方法则初始认为所有样本均属于一个大组，每一次循环将其分解为更小的组，直到每个对象构成一组或满足终止条件为止。

层次方法存在的缺陷是进行组分解或合并后，无法回溯，因此也就无法纠正错误决策。为弥补分解与合并的不足，层次方法经常要与其他方法相结合，其中较有效的方法是将层次方法与循环再定位算法相结合，先用层次方法对样本进行层次划分，再利用循环再定位技术对结果进行调整。当然，层次方法也有一定的优点，即在进行分解或合并时无须考虑不同选择所造成的组合爆炸问题。

3）基于密度的方法

大多数划分方法是基于对象间距离进行聚类的，仅能发现圆形或球状分布的聚类而较难发现其他形状分布的聚类。相反，基于密度概念的聚类方法利用样本分布密度的阈值控制对样本进行聚类，不仅能够发现任意形状的聚类，还可以在一定程度上克服噪声的影响。OPTICS 就是一个典型的基于密度的方法，该方法并不明确产生一个聚类，而是计算出一个增强聚类顺序以便用户根据具体背景知识和特殊要求进行自动或交互式数据分析。

4）基于网格方法

基于网格的方法首先将对象空间划分为确定个数的网格结构，以网格为单位，根据网格内所含数据的特性和网格边界决定不同网格的分类情况。这种方法的主要优点是处理时间与数据对象个数无关而仅与网格数有关，在处理海量密集样本时具有明显优势。

3.5.2 分类分析

分类是模式识别的另一主要领域，用于在类别标记属性已知的情况下描述重要数据类的模型或预测未来的数据趋势。其本质是通过分析数据，为每个类做出准确的描述，并在此基础上预测新样本的属性或将之上升为分类规则。数据分类过程主要包含两个步骤：第一步，建立描述已知数据集类别或概念的模型、生成分类规则；第二步，利用所获模型进行分类操作，评估分类准确率。如果准确率满足要求，则使用第一步所建立的模型对类别属性未知的新样本或对象进行分类识别。

1. 常用分类算法概述

用于分类的算法有很多，涉及的技术面在模式识别各种方法中也最广泛。从采用的技术来看，可以分为以下几大类：

（1）基于信息论（主要包括决策树系列算法）。

（2）基于概率统计（主要包括贝叶斯/贝叶斯网络、回归算法）。

（3）基于要求（主要包括最近邻分类、基于案例的推理）。

（4）基于人工智能（主要包括神经网络、遗传算法等）。

1）k-最近邻（k-nearest neighbors，KNN）算法[152,153]

KNN 分类基于类比学习，训练样本用 n 维数值属性描述，每个样本代表 n 维空间的一个点，由此将所有训练样本都存放在 n 维模式空间中。对于给定的一个未知样本，KNN 分类法搜索模式空间，找出最接近未知样本的 k 个训练样本，确定其分类，这 k 个训练样本就是未知样本的 k 个“近邻”。

KNN 分类方法是懒散的学习算法，即它首先存放所有的训练样本，只有当有新的样本需要被分类时才从样本库中搜索所有的与其最邻近的 k 个训练样本，对未知样本进行分类。这与诸如判定树归纳和前向传播之类的急切学习法形成鲜明的对比，后者在接收待分类的新样本之前先建立一个模型，而前者是在接收新样本后才进行搜索的。

虽然 KNN 方法有许多优点，如容易更新知识库，且比较直观、易于理解等，但是该方法仍然有一定的缺点，如当训练样本的数据量很大时将导致计算开销过大，因此需

要有效的索引技术来提高搜索效率。此外，不需要清楚的背景理论知识是 KNN 算法的优点，这使得此算法可以用于许多场合；但这同时也是 KNN 的缺点，因为缺乏清楚的应用背景知识，KNN 不考虑属性的背景特性，认为每个属性对分类的贡献相同，当数据中存在许多不相关属性时，可能会引起分类混乱和精确度下降。

2）判定树归纳分类

判定树是一个类似流程图的树型结构，其中每个内部节点表示在一个属性上的测试，每个分枝代表一个测试输出，而每个树叶节点代表类或类分布。判定树归纳是构造判定树的基本算法。在判定树构造时，有些分枝可能反映的是训练数据中的噪声或孤立点。此时，可用树剪枝方法检测和剪去这类分枝，提高对未知数据分类的准确性。

判定树归纳以自顶向下递归方式构造判定树。该算法的基本策略如下：

（1）从代表训练样本的单个节点开始。

（2）若一个节点的样本均为同一类别，则该节点成为树叶，并用该类标记。

（3）否则，算法使用信息增益度量作为启发信息，选择能够最好地将样本分类的属性作为该节点的“测试”属性。对测试属性的每个已知的值，创建一个分枝，并据此划分样本。

（4）算法使用同样的过程，递归地划分操作仅当下列条件之一成立时停止：①给定节点的所有样本属于同一类。②没有剩余属性可用来进一步划分样本。此时，使用多数表决将当前节点强制转换为树叶，并用训练样本中类别个数最多的类标记它。③测试属性的一个分枝没有样本。在这种情况下，以训练样本中的多数类创建一个树叶。

判定树归纳算法无需相关领域知识，归纳的学习与分类识别的操作处理速度都非常快。尤其是对于具有细长条分布性质的数据集合来讲，判定树归纳算法相应的分类准确率非常高。

3）神经网络算法[154]

神经网络模型最早是由心理学家和神经生物学家提出的，旨在开发和测试神经的计算模拟。粗略地说，神经网络是模拟人类神经系统中节点与中枢神经之间信息的传递与处理关系而建立的输入/输出单元的连接，其中每个连接都有一个权函数，表示连接的强度。在学习阶段，通过不断调整权函数，逐步提高网络对于已知样本的拟合精度，直到达到预定精度或有限次迭代次数为止。此时的这组权函数即是通过神经网络学习获得的“知识”或“模型”。当需要对新样本进行预测或分类时即可利用这组权函数计算输出结果，进行分类判定。

神经网络的优点包括其对噪声数据的高承受能力和对未经训练的数据进行正确分类的能力。虽然神经网络的训练耗时长、占用资源多、所需样本数大，且可能存在不收敛现象，但其强大的非线性处理能力使其仍然是目前模式分类方法的主流技术。

2. 贝叶斯分类算法介绍[136]

贝叶斯分类器具有理论上的完备性，是目前许多被广泛采用的实用分类算法的基础，因此本书将对其理论和算法应用进行较为详细的介绍。

1）贝叶斯定理

设 X 是类标号未知的数据样本（如基因芯片数据、组学数据等）。设 H 为某种假定，如数据样本 X 属于某特定的类 C。对于分类问题，我们希望确定给定观测数据样本 X，假定 H 成立的概率，即 $P(H \mid X)$。

$P(H \mid X)$ 称为后验概率，或称条件 X 下 H 的后验概率。例如，在疾病诊断中，假定数据样本域由患病和未患病人体的体征数据组成（如白细胞含量等），X 为这些体征数据，H 表示已患病，则 $P(H \mid X)$ 反映白细胞含量高于某一阈值时，判定所属人体已经患病的可信程度。作为对比，$P(H)$ 是先验概率，或称 H 的先验概率，独立于 X。在上述例子中，它是任意给定样本的数据样本中为患者的概率，即常说的患病率。由此可见，后验概率 $P(H \mid X)$ 比先验概率 $P(H)$ 基于更多的背景知识。类似地，$P(X \mid H)$ 是条件 H 下 X 的后验概率，即已知 X 是患者，X 的白细胞含量高于阈值的概率。$P(X)$ 是 X 的先验概率，指从数据集中随机取出一个数据样本，其白细胞含量高于阈值的概率。

由上述的定义可知，对于特定的样本集，$P(X)$、$P(H)$ 和 $P(X \mid H)$ 可以由给定的样本数据计算。贝叶斯定理提供了由 $P(X)$、$P(H)$ 和 $P(X \mid H)$ 计算后验概率 $P(H \mid X)$ 的方法，即根据白细胞高于阈值的概率、患病率和患病样本中白细胞高于阈值的概率来推算白细胞高于阈值时该样本是患病样本的概率。如果这一概率大于某一特定值，则可用白细胞含量来作为患病与否的判定标准，当有人来医院就医时，我们就可以根据其白细胞含量来判定他是否患病。

贝叶斯定理：

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)} \tag{3-24}$$

通常，判定样本类别需要同时考虑样本的多个属性，如上述的例子中，我们很难仅仅通过白细胞的含量就判定就医的人是否患病，还需要同时考虑其他症状，例如，体温、血小板量等等。

多个属性的贝叶斯分类器的工作过程可描述如下。

(1) 每个数据样本用一个 n 维特征向量 $X=\{x_1, x_2, \cdots, x_n\}$ 表示，分别描述对 n 个属性样本 $A_1, A_2, \cdots, A_n$ 的 n 个度量。

(2) 假定有 m 个类 $C_1, C_2, \cdots, C_m$。给定一个未知的数据样本 X（即没有类标号），朴素贝叶斯分类将未知的样本分配给类 C_i，当且仅当

$$P(C_i \mid X) > P(C_j \mid X), \quad 1 \leqslant j \leqslant m, j \neq i \tag{3-25}$$

根据贝叶斯定理 $P(H \mid X)=\frac{P(X \mid H)\ P(H)}{P(X)}$ 可知，最大化 $P(C_i \mid X)$ 即可进行分类。其 $P(C_i \mid X)$ 最大的类 C_i 称为最大后验假定。$P(X)$ 代表属性集 $A_1, A_2, \cdots, A_n$ 取值为 $x_1, x_2, \cdots, x_n$ 时的联合概率，为常数，所以最大化时只需对 $P(C_i \mid X)P(C_i)$ 最大即可。类的先验概率可以用 $P(C_i)=S_i/S$ 计算，其中，S_i 是类 C_i 中的训练样本数，而 S 是训练样本总数。

(3) 当需要对未知样本 X 进行分类预测时，可分别针对每个类别 C_i 计算 $P(X \mid C_i)\ P(C_i)$。当且仅当 $P(X|C_i)P(C_i)>P(X|C_j)P(C_j)$，$1\leqslant j\leqslant m$，$j\neq i$ 时，样本 X 被指派到类 C_i，即样本被指派到相应的 $P(X \mid C_i)\ P(C_i)$ 最大的类 C_i。

2) 朴素贝叶斯分类器

在上述的分类过程中，所需考虑的属性越多、属性间的联系越复杂，则计算 $P(X \mid C_i)$ 即$P(A_1=x_1\cdots,\ A_n=x_n \mid C_i)$ 的开销就越大，判别样本类别时所需计算量和存储空间就越大。朴素贝叶斯分类就是为了简化所需计算而假定一个属性值对给定类的影响独立于其他属性的值，即假定属性值相互条件独立，在属性间不存在依赖关系。这一假定称作类条件独立。此时可得

$$P(X \mid C_i)=\prod_{k=1}^{n} p(x_k \mid C_i) \tag{3-26}$$

概率 $P(X_1 \mid C_i)$，$P(X_2 \mid C_i)$，…，$P(X_n \mid C_i)$ 可以由训练样本计算。

(1) 如果 A_k 是离散属性（例如，将基因的开启与否看作离散量“1”和“0”），$P(x_k|C_i)=\dfrac{N(A_k=x_k,\ C=C_i)}{N(C=C_i)}$，其中，$N(C=C_i)$ 是样本集中属于类 C_i 的样本个数；$N(A_k=x_k,\ C=C_i)$ 是样本集中属于类型 C_i 且属性 A_k 取值为 x_k 的样本个数。

(2) 若 A_k 是连续值属性（如各种组学数据），常用的处理方法有两种：一是对其离散化，按照离散值处理；二是假定这一属性服从某一分布，在概率分布的基础上进行计算。

3) 贝叶斯网络

虽然朴素贝叶斯分类方法是一种非常实用的学习分类技术，与决策树学习、基于规则的学习和神经网络学习方法相比具有许多优势，如对噪声的健壮性、学习过程简单、不需要搜索等。但朴素贝叶斯方法是以一个很强的简单假设（数据中的属性相对于类别标识相互独立）为基础的。这个假设条件在实际样本中很少能够得到满足。因此，许多研究人员着手研究贝叶斯网络。在贝叶斯网络中，每一节点代表一个属性，节点间用有向连接线连接，网络中无环状闭合途径。其工作原理如下。

(1) 基于统计学中的条件独立，即给定父辈点属性，每个节点对于其祖辈、曾祖辈等都是条件独立的。

(2) 根据概率论中的链式规则，n 个属性 a_i 的联合概率如下：

$$P[a_1,a_2,\cdots,a_n]=\prod_{i=1}^{n} p[a_i \mid a_{i-1},\cdots,a_1] \tag{3-27}$$

贝叶斯网络是一种无环图，可以对网络节点进行排序，使节点 a_i的所有先辈节点序号小于 i，再根据条件独立假设推导出下式：

$$P[a_1,a_2,\cdots,a_n]=\prod_{i=1}^{n} P[a_i \mid a_{i-1},\cdots,a_1]=\prod_{i=1}^{n} P[a_i \mid a_i\text{ 的父节点}] \tag{3-28}$$

由此可将研究的重点转移至怎样根据大量的有效数据来建立贝叶斯网络的学习过程。但 N. Friedman 等将朴素贝叶斯方法和贝叶斯网络进行比较后发现，在某些领域使

用非限制性贝叶斯网络通常并不能提高精确性，甚至会降低精确性。为此，N. Friedman 提出了一种折中的办法，称之为树扩张型朴素贝叶斯方法（tree-augmented bayesian classifiers，TAN）。其基本思路是放松朴素贝叶斯的独立性假设条件，借鉴贝叶斯网络表示依赖关系的方法，扩展朴素贝叶斯的结构，使其能容纳属性间存在的依赖关系。其方法是在朴素贝叶斯分类器上添加连线，类属性是朴素贝叶斯网络每一节点的单一父辈节点，TAN 为每个节点增加第二个父辈节点。如果排除类节点和其相应的所有连线，假设只有一个节点没有增加第二个父辈节点，则结果分类器包含以无父辈节点为根节点的树结构。TAN 是当前贝叶斯网络分类器学习算法中性能较好的一种算法[155]。

3.6 小结

目前人们对基因的转录、翻译等过程细节还不是十分了解，还是以定性描述为主。各种检测手段的匮乏，使得人们还没办法准确获得基因转录和翻译等生物过程的许多关键参数，很多模型尚无法得到实验的充分检验。人们无法断定哪种生物系统肯定适用某种算法或者肯定不能用某种算法，只能通过各种尝试寻找这方面的规律。定量的建模、预测和优化方法还不能作为主要手段，只能为研究人员提供一些辅助参考信息。为了改变这种现状，充分借助计算机的高性能计算指导合成生物学的理性设计和实验优化，许多研究者正在不断努力尝试将现有算法进行改进用于生物系统模拟，或者开发各种新算法。因此，生物系统的建模和分类方法呈现多元化趋势。本书只是简单介绍了几种目前应用比较广泛的算法，读者大可不必拘泥于此。合成生物学的发展急需建模算法和各种参数检测手段的开发，本书会在后面几章结合经典实例介绍几种目前尚被认可的模型。读者不妨以这些经典实例模型为例，自己动手寻找一下更适用的算法，这也符合合成生物学开放、探索的思想。

第二部分 合成生物系统性能分析

设计合成生物系统的目的是为了获得目的产物或者崭新的生物功能，因此生物模块作为一个功能单元难免被植入到异源或同源生物体中，且必然会受到生物体中诸多信号的影响；合成生物学强调生物模块的“即插即用”性，要保证设计好的生物模块被方便地整合入其他功能模块中，生物模块就必须具有很好的稳定性和抵抗各种外界干扰的鲁棒性。此外，由于各种蛋白质、激酶自身的降解，为了维持宿主细胞内环境中各种蛋白质、代谢产物的平衡，生物模块必须具有很好的快速响应特性，能够以足够快的速度接收输入信号做出反应。因此，分析生物模块的稳定性、鲁棒性和快速响应特性等属性，总结和抽提设计的普遍原则，对于生物模块的功能扩展和生物系统的工程化至关重要[118,156]。

3.7　常用性能指标

3.7.1　稳定性

稳定性（stability）反映在暂态扰动消失后的时间响应性质上，对于稳定性的分析，有助于人们了解系统能够承受的最大干扰的幅值和频率，从而对实验设计有一定的指导意义。可以这样来定义：系统在任何足够小的初始偏差的作用下，对于暂态扰动的时间响应随着时间的推移而逐渐衰减并趋向于零，则该系统是稳定的；否则，该系统是不稳定的。稳定性是相对而言的，再稳定的系统所能承受的外界干扰也是有限的。如果外界干扰幅值过大，系统会因无法承受而崩溃，或者发生结构改变。

3.7.2　鲁棒性

鲁棒性（robustness）就是系统的健壮性。它是在异常和危险情况下系统生存的关键。所谓“鲁棒性”，是指系统在一定（结构、大小）的参数摄动下，维持某些性能的特性。例如，计算机软件在输入错误、磁盘故障、网络过载或有意攻击情况下，能否不死机、不崩溃，就是该软件的鲁棒性。

鲁棒性和稳定性是从两个不同角度衡量合成系统抵御宿主细胞内外环境扰动的能力。鲁棒性侧重扰动导致模块自身结构参数变化时维持原来状态的能力；而稳定性则侧重模块受到外界扰动后维持稳定的能力，模块也许会从原来的稳定状态过渡到新的稳定状态，但仍然维持稳定的能力。

3.7.3　响应快速性

所谓的响应快速性就是处于稳定状态的系统对于外部环境改变的快速反应能力，也是系统在受到扰动后迅速进入稳态的能力。对于生物模块来讲，任何输入信号，无论是信号分子、蛋白质，还是代谢物都会降解、排出或者转化成其他物质。生物模块必须要在这些输入信号消失之前做出响应，才能确保自身功能的发挥。

响应快速性可以通过改变具体的基因序列或者生物大分子构象等多种方法来改进。而且，响应的快速性是建立在系统的鲁棒性和稳定性基础上的，只有保证模块或者系统不被分解、不发散，快速性才有意义。因此，对于生物模块的性能分析还是以稳定性和鲁棒性为主。

3.8　稳定性分析

目前对于系统的稳定性分析还是主要建立在系统模型的基础之上，对于模型的精度具有一定的依赖性。一般可以采用两种方法来分析其稳定性：解析法和数值法。解析法分析通常需要繁琐的运算，对于没有解析解的模型来说甚至根本行不通，因此通常采用数值解法来分析系统的稳定性。

相位面分析是数值分析的常用方法。通过绘制模型解的近似曲线，相位面分析能够

判断系统的稳定性。在相位面分析中通常用模型的雅可比矩阵（Jacobian matrix）来分析稳定性。在向量微积分中，雅可比矩阵是一阶偏导数以一定方式排列成的矩阵，其行列式称为雅可比行列式。雅可比矩阵的重要性在于它体现了一个可微方程与给定点的最优线性逼近。因此，雅可比矩阵类似于多元函数的导数。通过雅可比矩阵的特征值可以判断系统的稳定性。通常正实数特征值说明系统不稳定，负实数特征值说明系统稳定，虚特征值说明系统处于临界状态。

因为微分是系统中偏导数的线性变换，因此方程中每一个线性独立的部分可以看成是特征函数 $h(t)$，满足约束条件$\frac{dh}{dt}=\lambda h$，其中，λ 为特征值。负的特征值代表的特性分量会随着时间而衰减，属于稳态分量。正特征值代表的特性分量则会随着时间的增加而发散，属于不稳定分量。虚根代表的特性分量处于临界状态，需根据具体问题进行具体分析。

分析系统的稳定性可以通过引入参数摄动和改变测量行为的方式进行。可以随机选择起始参数并考察状态的预定变化。在控制理论中，对于线性和非线性系统均有多种分析方法。由于线性系统模型简单实用，相应的控制理论比较完备，因此对于任意一个系统，人们还是喜欢先将其进行线性化后，再利用线性控制理论来分析。只有当线性控制理论实在无能为力时才会考虑各种非线性控制方法。随着计算机计算和存储能力的增加，新的控制方法层出不穷，有兴趣的读者可以参阅相关资料。

3.9 鲁棒性分析/敏感性分析

鲁棒性分析关心的是系统参数发生摄动时，系统能否保持原有的性能和状态，因此需要分析参数变动对于状态的影响，即参数敏感性分析。敏感性分析并不能得出真实的速率常数，只是辅助确定哪个参数更重要。在知识不完备情况下进行系统设计时，确定对系统输出具有最显著影响的参数是至关重要的：如果需要维持系统性能，则应尽量维持最敏感的参数不变；如果需要改变系统性能，则改变最敏感参数可能是最便捷的途径，因为系统输出通常只对少数几个参数敏感，相应突变的候选反应数量则完全可能缩小到实验可以处理的范围内[157]。

参数状态敏感性是相对于单独的参数扰动的状态稳定性而言的，如式（3-29）所示。以 $x(t)$ 为状态向量，ρ 为时变参数向量，$f(x(t),\rho)$ 为具有导数 $F(x(t),\rho)$ 的系统动态矩阵，S_{ij} 代表状态 i 对参数 j 的敏感性，有

$$S_{ij}=\frac{dx_i(t)}{d\rho_j} \tag{3-29}$$

根据链式法则可以得出所有状态的参数敏感性：

$$\frac{dS(t)}{d\rho}=\frac{\partial F(x(t),\rho)}{\partial x}S(t)+\frac{\partial F(x(t),\rho)}{\partial \rho} \tag{3-30}$$

其中，$S(t)$ 代表特定时间 t 所有的状态敏感性（status sensitivity）。由此可以衡量时间

周期 τ 内的敏感性，作为振荡频率或者周期的无偏估计。从 Shape Sensitivity 度量 $S^c(t)=\left(\frac{\delta x(t)}{\delta \rho}\right)_\tau$ 的定义可以看出，状态敏感性 $S(t)$ 可以看作是 Shape Sensitivity 和 Period Sensitivity 的组合：

$$S(t) = S^c(t) - \frac{t}{\tau}F(x(t),\rho)\frac{\mathrm{d}\tau}{\mathrm{d}\rho} \tag{3-31}$$

其中，Shape Sensitivity 独立于 Period Sensitivity。利用仿真工具即可获得微弱扰动对系统特定部分的影响。

敏感性变化独立于响应类型（振荡、稳态或混沌）。参数与真实系统具有直接联系，它们的敏感性能够暗示它们所代表的物理对应物的重要性。一般可将参数分为四大类：转录和翻译控制参数、mRNA 和蛋白质的降解参数、转运反应参数、蛋白磷酸化和去磷酸化参数。同一类参数的敏感性具有一定的相似性。

为了全面衡量系统的鲁棒性，必须同时考虑单参数和多参数的摄动。通常复杂系统中，单个参数的摄动会被其他分支减弱，因此复杂系统对于单参数的摄动具有更高的鲁棒性。而多参数系统复杂性的变化会比单回路模型更易导致周期偏离。在设计合成网络时，可以利用这些原则提高系统的鲁棒性，即为了获得一定的鲁棒性，可以适当提高系统的复杂性和加深层次结构。

由上面的介绍可知，稳定性和敏感性分析需要建立系统模型，而生物系统的复杂性使得研究者很多时候无法获得满意的模型。此时，装置的定向进化就成为另外一种优化装置功能的策略。定向进化的主要任务是确定系统进化的特点和行为，应用合适的筛选方法或选择性压力实现复杂行为的进化。如果对装置的内在特性不是十分了解，或者蛋白质特殊残基突变对装置行为的影响不清楚时，定向进化也许能够改善模块功能。例如，通过对阻遏蛋白和相应核糖体结合位点的突变，定向进化能够将无功能的二阶转录级联装置转变成可工作的模块[158]。

综上所述，综合运用推理设计、参数估计、敏感性分析以及定向进化等方法是优化装置功能的有效途径。

3.10　提高系统稳定性和鲁棒性的机制

生物系统各种各样的内源和外源噪声都能影响合成生物模块在活体内发挥期望的功能。基因表达中内源因素引起的噪声主要是指单个基因水平上生化反应的随机性；而外源噪声则包括代谢物、核糖体以及聚合酶等细胞组分的波动。有些研究表明，基因表达内源噪声引起的干扰一般要快于细胞周期，而外源噪声引起的干扰则具有较长的时间尺度且对合成模块的功能具有更显著的影响[111,159]。

噪声对于具体合成模块的影响不仅取决于噪声本身的属性，也取决于模块的属性。在进行装置及模块功能模拟和仿真时考虑噪声的影响对于生物系统设计按照预期功能工作非常重要。我们可以利用某些机制来提高系统的稳定性和鲁棒性。

3.10.1 冗余

生物系统确保预期行为的最简单方法即是冗余性[160]。这方面的例子有很多，例如，胰岛素信号转导通路[161]和酵母菌大部分代谢途径都具有一定的冗余性。当用在连通性能完好的节点上时，冗余性能提高无标度网络的稳定性。细胞的代谢系统大多数属于无标度网络，因此这一机制非常有用。

3.10.2 反馈控制

许多网络都具有一些防止系统输出达到无法控制情况的机制。正如单回路控制系统比开环控制系统具有更强的鲁棒性一样，转录级联线路比单基因转录系统更能抵抗外源噪声的影响。某些情况下，反馈能够增加系统的超敏感性，而生物模块因超敏感性而具有的近似离散的响应，使得构建的复杂系统具有一定的离散性和鲁棒性。

负反馈回路能够抑制噪声干扰并提高系统的稳定性[162]。正反馈通常具有较差的稳定性，但由于能够提高系统的敏感性，因此对于系统的鲁棒性也有至关重要的影响。一个非常微弱的信号能够通过正反馈环放大并启动期望行为，导致快速或者局部化的切换/开关。正反馈系统不能改善稳定特性，但能够造成系统“爆炸”。因此，当不再需要正反馈的行为时，必须有机制能够及时断开正反馈回路。当将转录速率限制在特定水平时，正反馈能够有效消除抑制转录速率的扰动，使其稳定在有限水平[163]。

3.10.3 特异性和保真度

合成网络设计中，模块性和特异性非常重要。通过模块和特异性设计可以降低不希望的部件间相互作用并提供易于扩展和调整的“友好”接口。

将细胞分解为各个功能模块，采用孤立错误行为的方法，能够阻止故障在整个系统中传播扩散，提高系统鲁棒性。尤其是在无标度环境中，模块化更能提高系统的鲁棒性，防止网络焦点被系统其他部分的错误所干扰。为此，必须精密设计模块对信号的特异性。“局域化”是这方面的主要手段。自然系统的局域化划分主要通过将细胞中不同空间区域的不同通路确定下来，排除信号分子通过组分边界的相互作用来完成。支架蛋白的运用可以看作是利用支架作为组分特定反应区间的“微局域化”的一种形式。如果将支架结合到某个通路中，则信号转导和特异性也就实现了“微局域化”。

信号通路的特异性和保真度能够帮助我们衡量系统各组分间相互作用的程度。

信号通路 X 的特异性定义为可信输出与不可信输出之间的比率。X_{in}和 X_{out}代表可信输入和输出，Y_{out}代表不可信输出：

$$S_X = \frac{X_{out}/X_{in}}{Y_{out}/X_{in}} \tag{3-32}$$

当没有不可信输出时，通路具有完备的特异性，如图 3-5 所示。图 3-5A 中启动子 P 启动，基因 A 和 B 均编码各自的蛋白质，其中蛋白质 B 是不期望的副产物，此时系统的特异性取决于 A 和 B 产量的比值。图 3-5B 中，P_1 是期望的启动子，而 P_2 是不期望的

启动子，两个启动子都可以启动下游基因的编码，因此系统的绝缘性和保真度不是很高[164]。

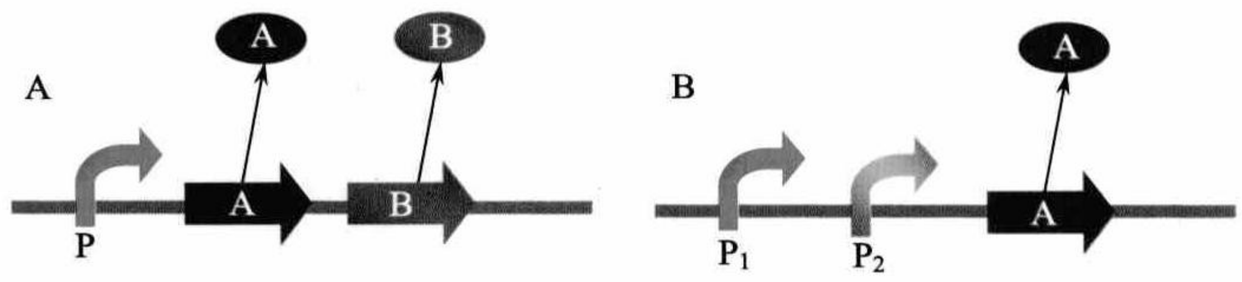


图 3-5　遗传线路特异性和保真度示意图

图 A 中启动子 P 启动，目的基因 A 和副产物基因 B 均编码各自的蛋白质。系统特异性取决于 A 和 B 产量的比值。图 B 中，P_1 启动子是期望的启动子，而 P_2 是不期望的启动子，两个启动子都可以启动下游基因的编码，因此系统的绝缘性和保真度不是很高。

科学的探讨和研究，其本身就含有至美，其本身给人的愉快就是酬报。

——居里夫人

第 4 章　合成生物学的基础研究

基础提要

- 乳糖操纵子模型
- 基因的表达与调控
- 分子病毒学

合成生物学的研究和操作对象是各种生物，而不是常规工程领域的无生命物质。生物的生长和繁殖有其特定的自然规律，需要特殊的培养环境，而进化和突变使得生物系统的行为具有明显的不确定性，因此无法对其进行精确的控制和预测，生物组件之间的接口也不可能像其他工业组件一样用简单的尺寸和标准即可规范。

这些研究对象的特殊性导致了合成生物学研究的特殊性。合成生物学的研究必须有选择性地借鉴而不能完全照搬信息系统、控制系统等其他领域的成果和理论。在后续章节的内容中，读者应该仔细留意其他工程领域知识对于合成生物学的借鉴和带动作用，更应该留意合成生物学区别于常规工程领域的特殊之处。

第一部分　模拟逻辑功能的基础研究

计算机系统中各种令人叹为观止的功能最终都可以简化为数字“0”和“1”之间的运算；组成复杂生物系统的各基因部件之间的协同运作，在逻辑上几乎均可简化为彼此之间的“开启”（对应逻辑 1）和“关闭”（对应逻辑 0）。生物系统的四编码（A、T、C、G）是所有生物现象的终极载体，信息系统的二编码（0、1）也是所有信息领域的终极载体；生物系统中发现了大量的“正”、“负”反馈调控结构，物理系统中也普遍存在着“正”、“负”反馈控制，这些惊人的相似和纷繁复杂的联系，启发合成生物学家大量借鉴各种逻辑运算和控制机制来研究和开发生物系统，即模拟逻辑功能的合成生物学设计和基础研究。

4.1 逻辑门功能遗传线路

逻辑门是数字电子线路的基本内容，是各种现代化高精尖数字仪器的基础部件，也是最基本的运算单元，从模拟逻辑功能角度出发的合成生物学设计和研究自然要从模拟各种逻辑门功能开始。下面我们就以转录水平遗传线路的构建来说明如何利用生物部件实现逻辑门运算功能。

图 4-1 所示为 R. Weiss 等设计的利用激活因子和诱导物实现逻辑“与”门功能的遗传线路。其工作机理为[57]：RNA_P（RNA 聚合酶）与启动子结合的亲和力很低，基础转录相对较小，在没有激活因子和诱导物输入时，系统输出为低水平。如果仅有激活因子而没有诱导物输入时，激活因子也因为亲和力较低而无法与启动子结合，系统输出仍然为低水平。只有当激活因子和诱导物同时存在时，诱导物与激活因子结合改变激活因子构象，激活因子/诱导物复合物与操纵子结合并帮助 RNA_P 与启动子结合起始转录，系统的输出才为高水平。

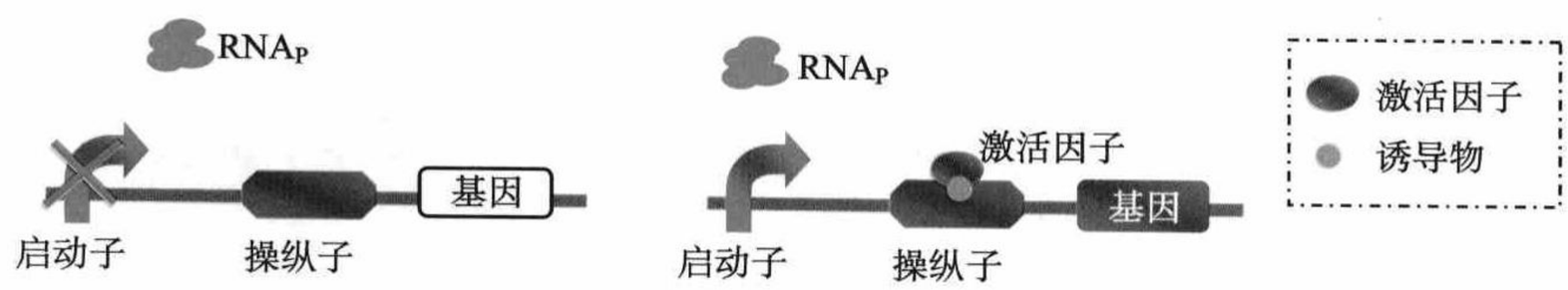


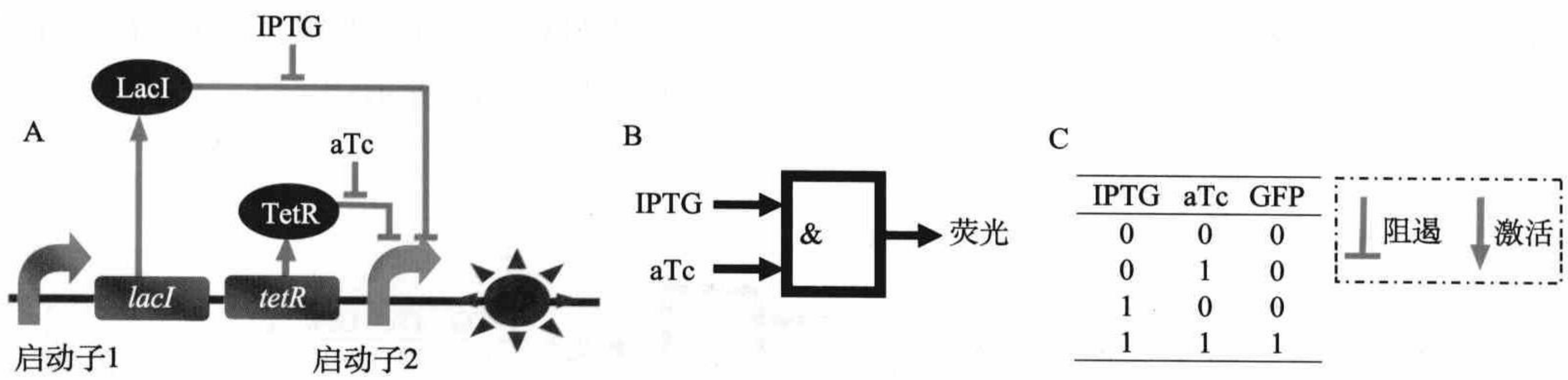
图 4-1 利用激活因子和诱导物实现逻辑“与”门功能

RNA_P（RNA 聚合酶）与启动子结合的亲和力很低，基础转录相对较小，只有当激活因子和诱导物同时存在时，诱导物与激活因子结合改变激活因子构象，形成的激活因子/诱导物复合物与操纵子结合并帮助 RNA_P 与启动子结合起始转录，系统的输出才为高水平。

例如，对于阿拉伯糖启动子，由 AraC 二聚体和阿拉伯糖组成的复合物诱导。如果将 AraC 蛋白和阿拉伯糖看做系统输入，只有当 AraC 蛋白和阿拉伯糖同时存在时，阿拉伯糖启动子才能开启，否则处于关闭状态。

图 4-2A 所示为另外一种“与”门功能的遗传线路。启动子 1 组成型表达 *lac* 和 *tet* 基因的多顺反子，工程化的启动子 2 能够被 LacI 或 TetR 蛋白抑制。当 LacI 或 TetR 存在时，与启动子 2 结合阻止其启动，*gfp* 基因的表达被关闭。只有当 IPTG（异丙基-β-D 硫代半乳糖苷，isopropyl β-D-1-thiogalactopyranoside）和脱水四环素（anhydrotetracycline，aTc）同时存在时，才能阻止 LacI 和 TetR 与启动子 2 结合，启动子 2 顺利启动，*gfp* 基因的表达才能被开启，系统有 GFP 蛋白表达。如果将 IPTG、aTC 看做系统输入，GFP 看做系统输出，则此遗传线路即可看做双输入“与”门系统，其逻辑图和真值表如图 4-2B 和图 4-2C 所示[65]。

构建遗传线路的一个有效途径就是由简单到复杂。在上述“与”门功能遗传线路的基础上，稍作改变，即可以构建带记忆功能的“与”门遗传线路。例如，在启动子 2 下游加入一个其产物能够抑制启动子 1 的基因，如图 4-3 所示。新插入的 *cI* 基因在系统处于开启状态时，表达的 CI 蛋白能够抑制启动子 1，从而抑制 *lac* 和 *tet* 基因表达相应



IPTG	aTc	GFP
0	0	0
0	1	0
1	0	0
1	1	1

图 4-2　逻辑“与”门功能遗传线路

A. 遗传线路图；B. “与”门逻辑图；C. “与”门真值表

图 A 为“与”门的遗传线路。启动子 1 组成型表达 *lac* 和 *tet* 基因的多顺反子，工程化的启动子 2 能够被 LacI 或 TetR 蛋白抑制，其下游 *gfp* 基因无法表达。只有当 IPTG 和 aTc 同时存在，同时阻止 LacI 和 TetR 与启动子 2 结合，启动子 2 才能顺利启动，*gfp* 基因的表达才能被开启，系统有 GFP 蛋白，从而实现 IPTG 和 aTc 为输入，GFP 为输出的逻辑“与”门功能。图 B 和 C 为相应的逻辑图和真值表（表中“1”表示系统中有该物质，“0”表示系统中无该物质）。

蛋白质，相当于解除了 LacI 和 TetR 对启动子 2 的抑制作用。当外界条件发生变化时（例如，IPTG 或 aTc 输入不同时存在时），“与”门本应被关闭，但 CI 蛋白的存在会抑制启动子 1，进而抑制 LacI 和 TetR 蛋白的产生，启动子 2 得以继续表达，系统仍然会维持开启状态不变，荧光蛋白还是会在一定时间内持续存在。即系统的输入发生改变时，系统的输出仍然会在一定时间内维持（“记忆”）原状态，具有“记忆”功能[65]。此时若采用对温度敏感的 CI 阻遏蛋白，则可通过环境温度的改变，将系统转换至“与”门关闭的状态。

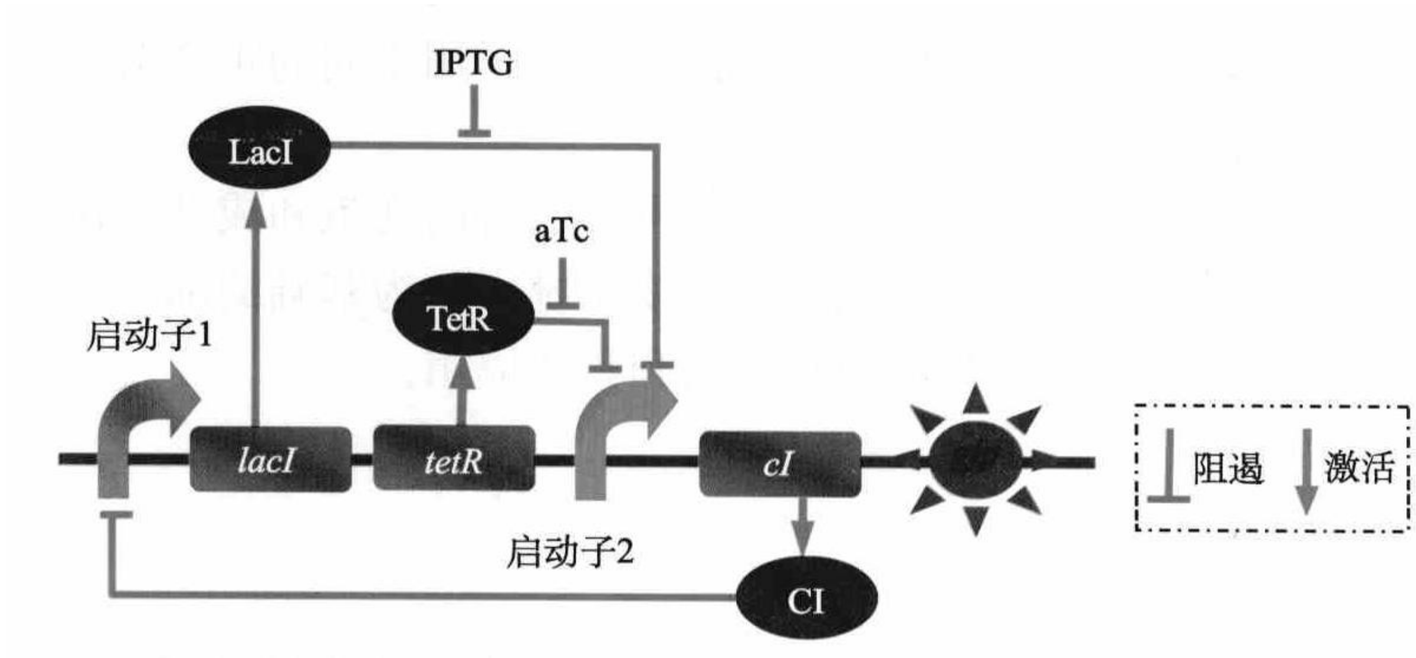


图 4-3　带有“记忆”功能的“与”门遗传线路

在系统处于开启状态时，在图 4-2A 基础上新插入的 *cI* 基因表达的 CI 蛋白能够抑制启动子 1，从而抑制 *lac* 和 *tet* 基因表达相应蛋白，相当于解除了 LacI 和 TetR 对启动子 2 的抑制作用。当外界条件发生变化，IPTG 或 aTc 输入不同时存在时，CI 蛋白仍然会维持对启动子 1 的抑制，抑制 LacI 和 TetR 蛋白的产生，启动子 2 得以继续表达，系统仍然会维持开启状态不变，具有“记忆”功能。

图 4-4A 为模拟逻辑“或”门功能的遗传线路。LacI 抑制启动子 2，TetR 抑制启动子 3。IPTG 可以解除 LacI 对启动子 2 的抑制，开启其后的 *gfp* 基因；aTc 可以解除 TetR 对启动子 3 的抑制，开启其后的 *gfp* 基因。因此，当 IPTG 或 aTc 中的任何一个

存在时，均可以导致 *gfp* 基因的表达，从而实现以 IPTG 和 aTc 为输入、GFP 为输出的逻辑“或”门功能。其逻辑图和真值表如图 4-4B 和图 4-4C 所示[65]。

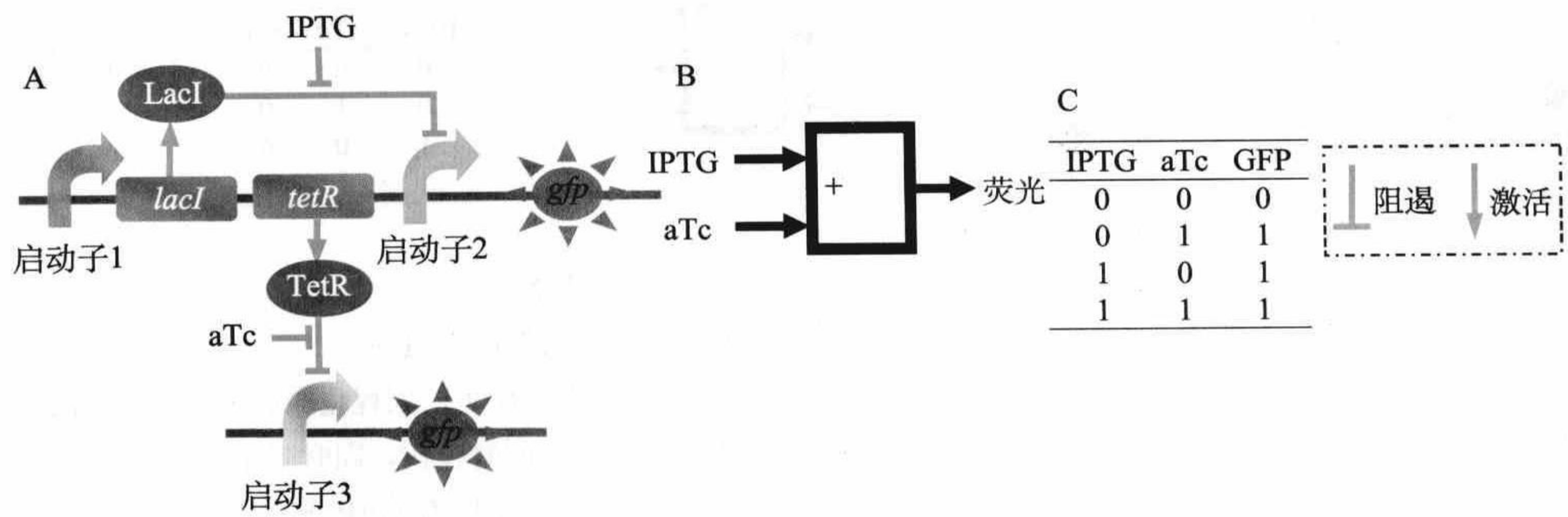


图 4-4 逻辑“或”门功能遗传线路

A. 遗传线路图；B. “或”门逻辑图；C. “或”门真值表

当系统中有 IPIG 输入时，可以解除 LacI 对启动子 2 的抑制，使其后的 *gfp* 基因表达；当系统中有 aTc 输入时，aTc 可以解除 TetR 对启动子 3 的抑制，开启其后的 *gfp* 基因。从而实现以 IPTG 和 aTc 为输入，GFP 为输出的逻辑“或”门功能。图 B 和 C 为相应的逻辑图和真值表（表中“1”表示系统中有该物质，“0”表示系统中无该物质）。

除了上述基本的转录水平的逻辑门遗传线路之外，对于逻辑门功能的模拟已经扩展到了转录后水平：O. Rackham 和 J. W. Chin 利用筛选法获得了正交的 rRNA-mRNA 对，构建了基于合成 rRNA 和 mRNA 相互作用的双输入逻辑“与”门，只有当两者都表达时逻辑门才能产生功能蛋白，而正交对内源配对物间不会发生相互作用，因此这一正交对理论上是非常理想的底物。同时，他们还证明了用不同的正交对拓扑结构可以构建逻辑“或”等功能的逻辑门[165]。

逻辑门功能生物模块的构建方式还有很多种，已有的文献和成功实例也很多，随着合成生物学的飞速发展，逻辑门已经成为了生物系统中较为基础的部件，在本书后续内容中还会介绍逻辑门生物模块的其他构建方法和具体应用。

4.2 基因调控开关

除了逻辑门遗传线路以外，最基本的基因表达调控部件即是控制基因表达的各种“基因开关”。基因开关是指某种化学诱导物存在或缺乏时，或者在两个独立的外源刺激作用下，基因处于两种可能状态中的一种的系统[122]。

4.2.1 转换开关

转换开关（inverter）因其输出是输入的转换函数而得名，即输入为低时输出是高，反之亦然。天然系统中有许多基因调控的转换开关存在，其中大家比较熟悉的正控阻遏系统（即效应物分子的存在使激活蛋白处于非活性状态，转录不进行）即可以看做是天然的转换开关：效应分子输入为高时，基因不表达，系统输出为低；效应分子输入为低

时，基因表达，系统输出为高[166]。

4.2.2　双相开关[167,168]

λ噬菌体中通过一个双向操纵子 O_R 来维持溶源状态，这个操纵子由阻遏蛋白 CI 和反阻遏蛋白 Cro（分别由 *cI* 基因和 *cro* 基因编码）结合到相邻的三个结合位点（O_{R1}、O_{R2}和 O_{R3}）来调控，如图 4-5 所示。CI 蛋白对 *cro* 基因的转录起着负调控作用，对 *cI* 基因本身的转录则既有正调控作用又有负调控作用，即我们所说的双相开关（biphasic switch）。其机理可描述如下：O_{R1}结合位点具有较高亲合力，较低浓度的 CI 蛋白首先与之结合并促进与 O_{R2}的结合和自身的转录。当 CI 蛋白浓度逐渐增高时，O_{R1}、O_{R2}和 O_{R3}三个位点均被结合，这时 *cI* 基因的转录反而被抑制，CI 蛋白浓度逐渐降低。由此可知，这一调控开关中，CI 蛋白对于 *cI* 基因的调控作用分为两个状态：CI 蛋白低浓度时的正调控状态（即开启状态）和高浓度时的负调控状态（即阻遏状态）。

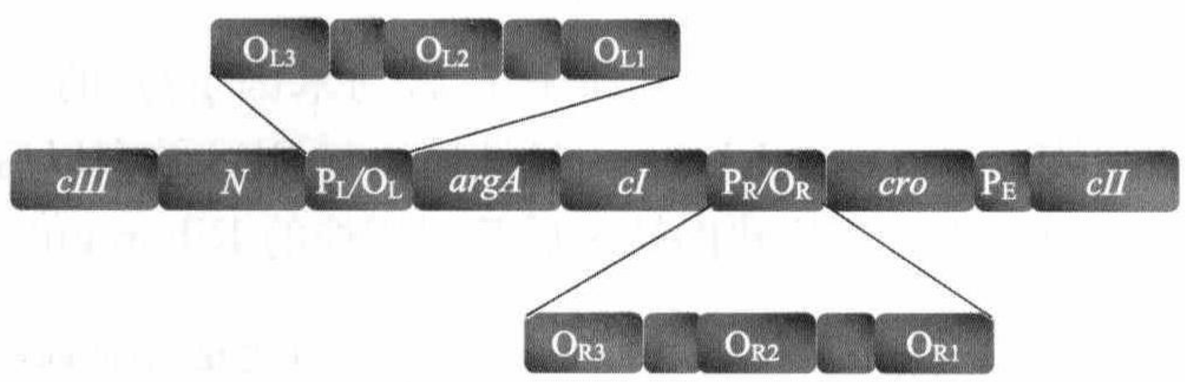


图 4-5　λ噬菌体中溶源 / 裂解基因调控示意图

cI 基因和 *cro* 基因编码λ噬菌体转录的主要调节蛋白 CI 和 Cro。O_{R1}结合位点具有较高亲合力，较低浓度的 CI 蛋白首先与之结合并促进与 O_{R2}的结合和自身的转录；当 CI 蛋白浓度逐渐增高时，O_{R1}、O_{R2}和 O_{R3}三个位点均被结合，这时 *cI* 基因的转录反而被抑制，CI 蛋白浓度逐渐降低。由此可知，CI 蛋白对 *cI* 基因本身的转录既有正调控又有负调控，即我们所说的双相开关。

4.2.3　Riboswitch

生物体能够通过各种精巧的机制控制基因表达的时间和数量，几年前发现的 Riboswitch 就是这些精巧机制中的一种。Riboswitch 主要是通过核糖核酸（RNA）构象的改变来实现“开关”的功能，阻止或开启目的蛋白的生成。大部分 Riboswitch 只有一个识别靶向配体的结合位点或适配体（aptamer）。适配体一般位于基因表达区域附近，当其与代谢产物结合时，会改变自身结构，在转录或翻译水平上行使基因调控的功能。Riboswitch 可被一种调节 sRNA（此处指小片段的调节 RNA，非编码 RNA 的一种，对于真核和原核生物的调节作用非常重要）“打开”，重新激活转录或翻译[169]。

Riboswitch 在进化上可能是 RNA 世界遗留的分子化石，广泛存在于革兰氏阴性菌和革兰氏阳性菌的代谢相关基因中，在真菌和植物中也有发现。Riboswitch 调节基因表达不需要任何蛋白因子作为中介，主要参与调节维生素、氨基酸和核苷酸等基础物质的代谢过程，因此可用于研究基因功能、开发新型药物及进行基因治疗。

为了研究 Riboswitch 的调节机制，J. J. Collins 课题组的 F. J. Isaacs 等构建了模块化的 Riboswitch。他们首先在 DNA 上游引入一个短的、与 RBS 互补的核苷序列，命名

为 *cis*-抑制序列，位于启动子下游 5′-UTR 和 RBS 序列之间。*cis* 序列的引入并不改变目标基因的编码框，也不影响转录速率。*cis* 序列由两部分组成：由互补于 RBS 的 19 个核苷酸残基组成的发夹序列和由 6 个核苷酸残基组成的环域。转录产生的 crRNA（*cis*-repressed RNA）的 5′-UTR 能够发生折叠而掩盖 RBS，从而阻止功能蛋白的翻译[170]。

为了能够在必要时重新激活基因表达，F. J. Isaacs 等还构建了另外一个启动子，引入一个模块内线性-环相互作用（linear-loop interaction）定向 taRNA-crRNA 结合的基元，这个基元的 YUNR（pYrimidine Uracil Nucleotide puRine）序列，是许多内源系统模块内 RNA 复合物的经典靶向物，具有一定的适应性。转录产生的 taRNA（*trans*-activating RNA）具有与 *cis* 抑制序列互补的核苷酸序列。taRNA 能够以高特异性优先与 crRNA 杂交，产生的 RNA 二聚体导致 crRNA 发生构象变化，发夹结构解开，RBS 暴露在外允许翻译继续进行。模块内 RNA 相互作用的特异性由特异的 crRNA 序列决定，从而保证了模块的特异性。

图 4-6A 为基因表达的基本步骤。启动子 P 启动基因 *gfp* 的表达。转录形成的 mRNA 具有一个能被核糖体识别的 RBS。一旦核糖体与 RBS 结合以后，mRNA 即可以翻译成功能蛋白。图 4-6B 为人工构建的转录后基因调控的 Riboregulator 系统。在这个

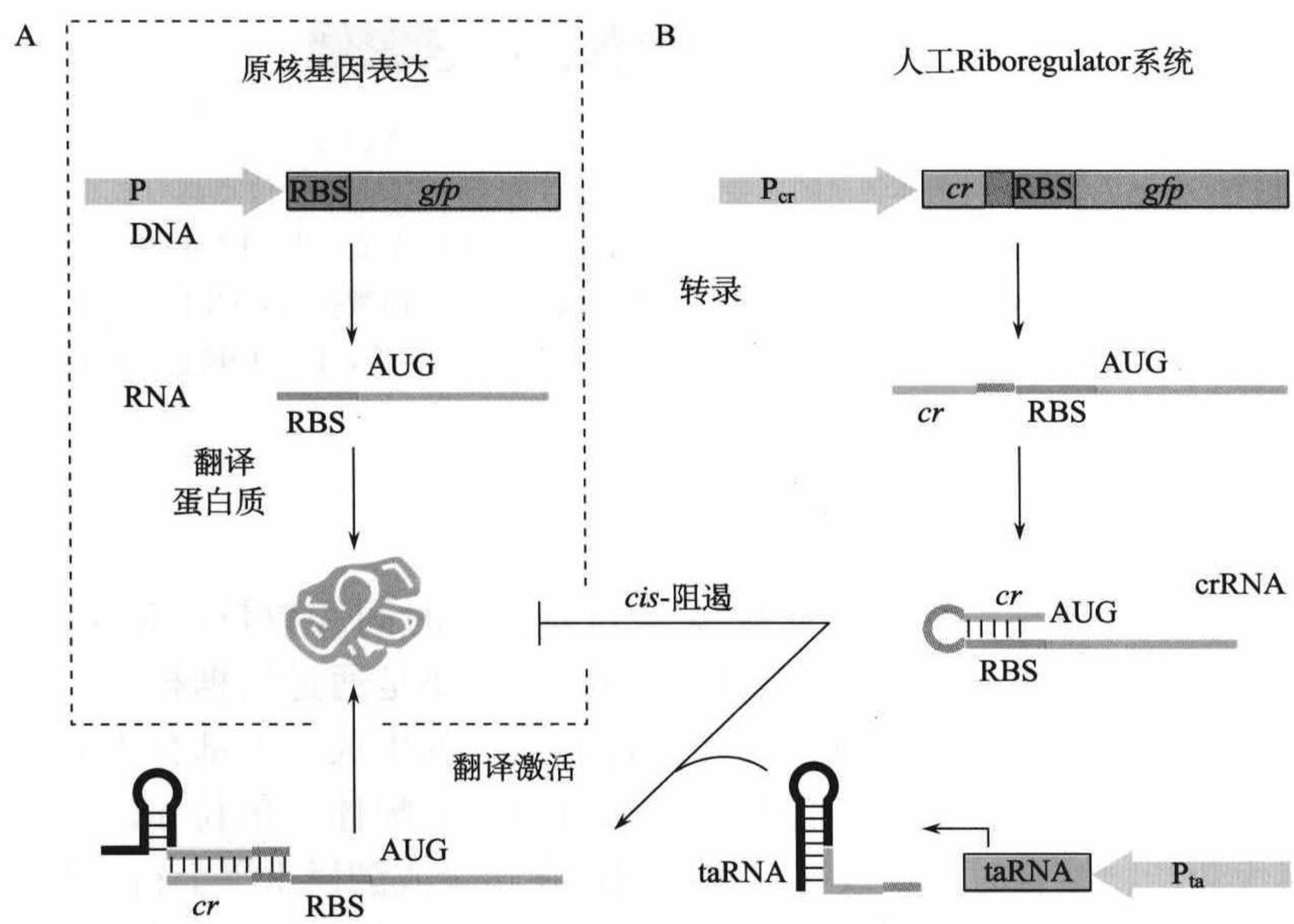


图 4-6 用于控制转录后基因调控的人工 Riboswitch 系统[170]

图 A 为基因表达的基本步骤。启动子 P 控制 *gfp* 的表达。转录形成的mRNA 的RBS 暴露在外，一旦核糖体识别并与之结合以后，mRNA 即可以翻译成蛋白质。图 B 为人工构建的 Riboregulator 系统的调控过程示意图。此系统中，在启动子 P_{cr}和 RBS 中间人工插入了一小段与 RBS 互补的 *cr* 序列。转录后 mRNA 的 5′端，互补的 *cr* 与 RBS 结合形成发夹结构，将 RBS 掩盖阻止核糖体与之结合，从而阻止翻译的进行。另外一个启动子 P_{ta}启动后表达一个短链非编码 RNA——taRNA。taRNA 能够以高特异性优先定位于 crRNA，共同经过线性-环相互作用将 crRNA 的发夹环解开，taRNA 的尾部与 crRNA 互补配对。此时 crRNA 的 RBS 位点重新暴露在外，核糖体可与之结合重新激活翻译。

人工 Riboregulator 中，启动子 P_{cr}和 RBS 中间插入了一小段与 RBS 互补的 *cr* 序列。在转录后 mRNA 的 5′端，互补的 *cr* 与 RBS 就会形成发夹结构，阻止核糖体结合从而阻止翻译的进行。另外一个启动子 P_{ta}启动后表达一个短链非编码 RNA——taRNA。taRNA 能够以高特异性优先定位于 crRNA。taRNA 和 crRNA 共同经过线性-环相互作用将 crRNA 的发夹环解开，taRNA 的尾部与 crRNA 互补配对。此时 crRNA 的 RBS 位点重新暴露在外，激活翻译。

图 4-7 为线性-环相互作用示意图。整个相互作用共分为三个步骤：①taRNA 的 5′端线性区域（灰色部分）识别 crRNA 环形域中的 YUNR（灰色部分）；②不稳定的 taRNA 尾部与发夹的环状区域互补核苷酸序列进行配对，使得掩盖 RBS 而阻止核糖体识别的发夹结构失去稳定性；③形成稳定的 RNA 双链结构，并将 crRNA 的 RBS 位点暴露在外，允许翻译的进行。

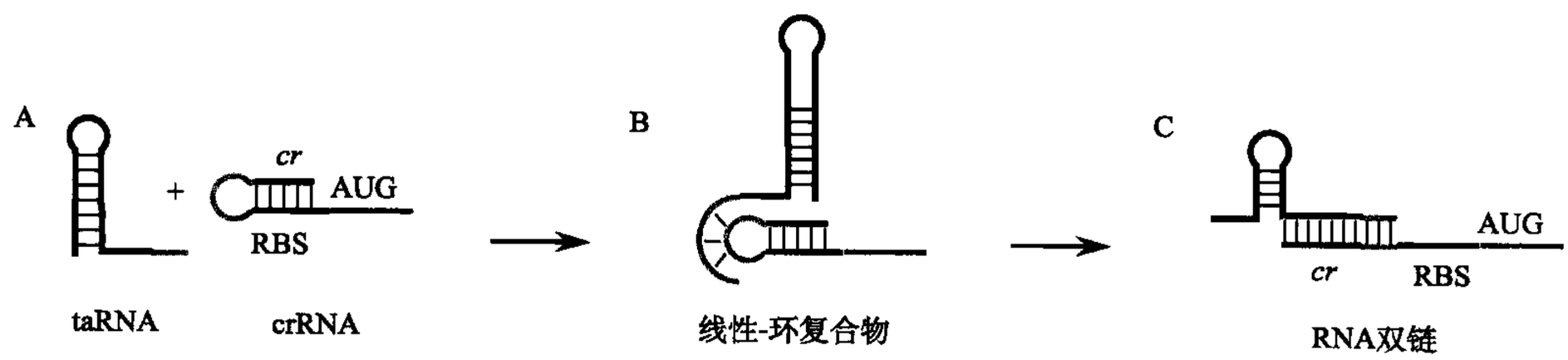


图 4-7　线性-环相互作用示意图[170]

图 A taRNA 的 5′端线性区域（灰色部分）识别 crRNA 环形域中的 YUNR（灰色部分）；图 B 不稳定的 taRNA 尾部与 YUNR 序列进行配对，使得掩盖 RBS 而阻止核糖体识别的发夹结构失去稳定性；图 C 形成稳定的 RNA 双链结构，将 crRNA 的 RBS 位点暴露在外，允许翻译的进行。

调控基因表达的代谢产物浓度是由配体结合的速度及 RNA 聚合酶作用的速度决定的。细胞可以通过使 Riboswitch 的适配体或蛋白受体发生突变，加强或减弱与配体结合的亲合力，从而改变代谢产物的浓度。由于有了 Riboswitch，细胞就可以通过发生分子的突变来调节相应的基因表达，这些突变可以改变 Riboswitch 折叠构象或 RNA 聚合酶催化 Riboswitch 合成过程的速度。相比于转录水平上的调控，这种转录后调控机制具有更加快速的特点。

与此有异曲同工之妙的是 T. S. Bayer 和 C. D. Smolke 等设计的短链 RNA 基因调控系统。如图4-8 所示，系统中短链 RNA 片段经过人工改造，能够自身与自身配对，不会对 mRNA 片段造成影响。当加入特定小分子效应物后，小分子与短链 RNA 结合改变其构象，使短链 RNA 与自身配对的一部分暴露，并且和 mRNA 片段头部的特定部位结合，阻止蛋白质的合成[171]。

此外，还有许多 RNA 水平的基因调控设计，包括反义 RNA 调控、与 RNA 酶结合的调控、小段 RNA 调控、微 RNA 调控等[172~174]。

4.2.4　双稳态开关

双稳态开关也称“拨动开关”，可通过人为调控实现基因线路在两种不同稳定状态

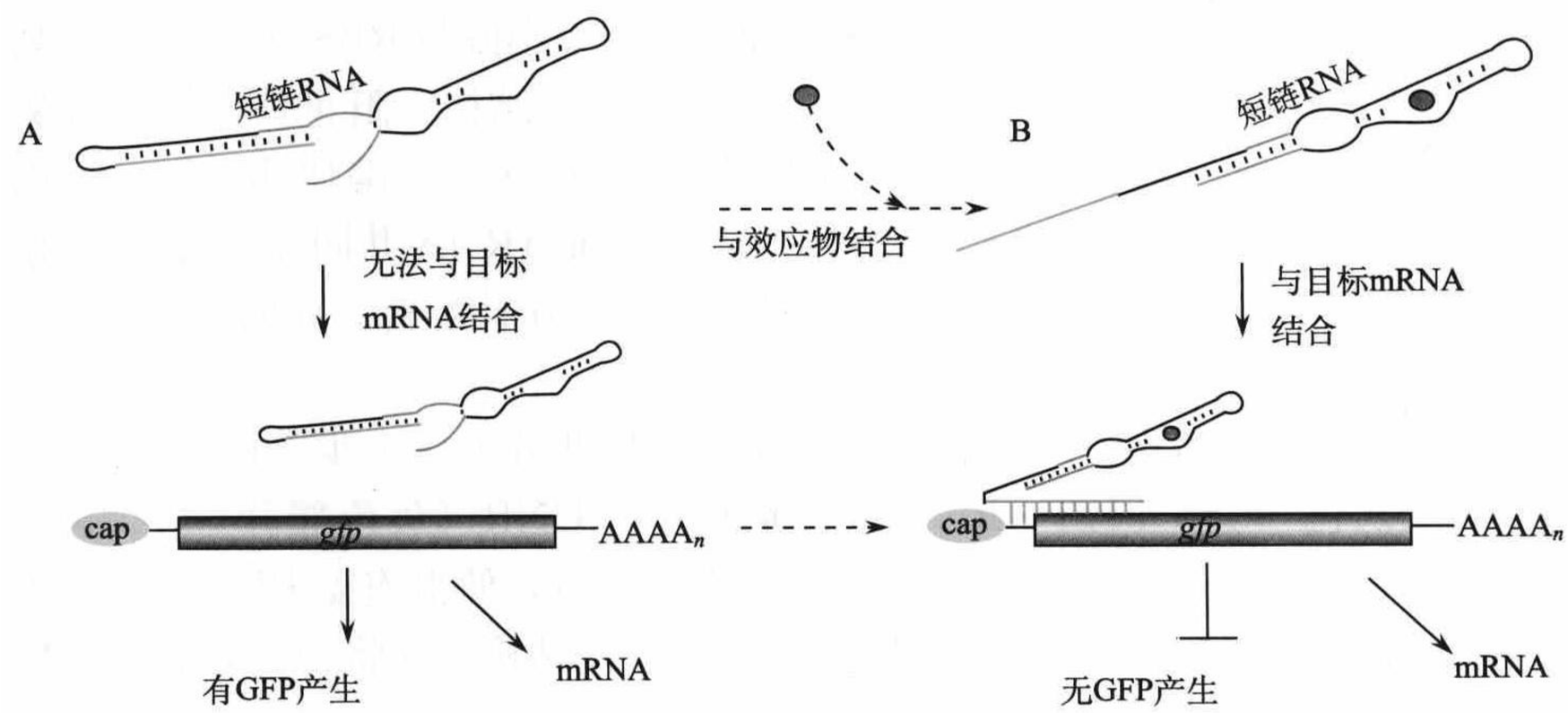


图 4-8 程序化控制的真核基因表达的小 Riboregulators[171]

图 A. 系统中短链 RNA 经过人工改造，能够自身与自身配对，不会对 mRNA 片段造成影响，*gfp* 正常翻译。图 B. 当加入特定小分子效应物后，效应物与短链 RNA 结合改变短链 RNA 构象，使其与自身配对的一部分暴露，并且和 mRNA 片段头部的特定部位结合，阻止 GFP 蛋白质的合成。

间的切换。经典的转录水平双稳态开关由两个启动子组成，每个启动子的“开”和“关”状态之间具有明显的界限，只有通过瞬态的诱导因素变化才能切换基因拨动开关至其中的一个稳态，且在移除其输入激励后仍会维持原来的状态。

双稳态开关对于输入信号在一个较宽范围内的变化不敏感，因此对信号扰动具有一定的鲁棒性。通常由一种状态切换到另一种状态时，需要时间来进行转换，因此双稳态开关起始状态为“开”或“关”时会得到不同的响应曲线。这一特性使双稳态开关具有“记忆”的功能，可能会导致细胞状态的不可逆转性。

利用合成生物学方法构建的双稳态开关有很多，在原核、真核乃至哺乳动物细胞中均实现了人工构建双稳态开关[175]。下面我们以 J. J. Collins 课题组构建的经典的双稳态开关为例，讲解其构建和调控机理。

在大肠杆菌中构建双稳态开关[58]

1）基因线路设计

基于特定蛋白质对特定基因表达的开启和抑制作用，即正、负调节作用，J. J. Collins 课题组的 T. S. Gardner 等设计了一种转录水平的双稳态开关，其逻辑线路设计如图 4-9 所示。

图 4-9 中有两种不同的启动子——启动子 1 和启动子 2，它们分别控制着阻遏子 2 和阻遏子 1 基因序列的表达状态；而阻遏子 2 和阻遏子 1 基因表达的产物又会分别抑制启动子 2 和启动子 1 的启动。阻遏子 1 下游报告基因的表达产物可以作为输出信号表征系统目前所处的状态。由于上下游基因的启动和相互抑制，整个系统共有下面两种稳定的输出状态。

状态 A：当启动子 1 启动的时候，阻遏子 2 基因被表达，其产物导致启动子 2 被抑制，阻遏子 1 不表达，无法对启动子 1 产生抑制作用，系统稳定于启动子 1 启动而启动

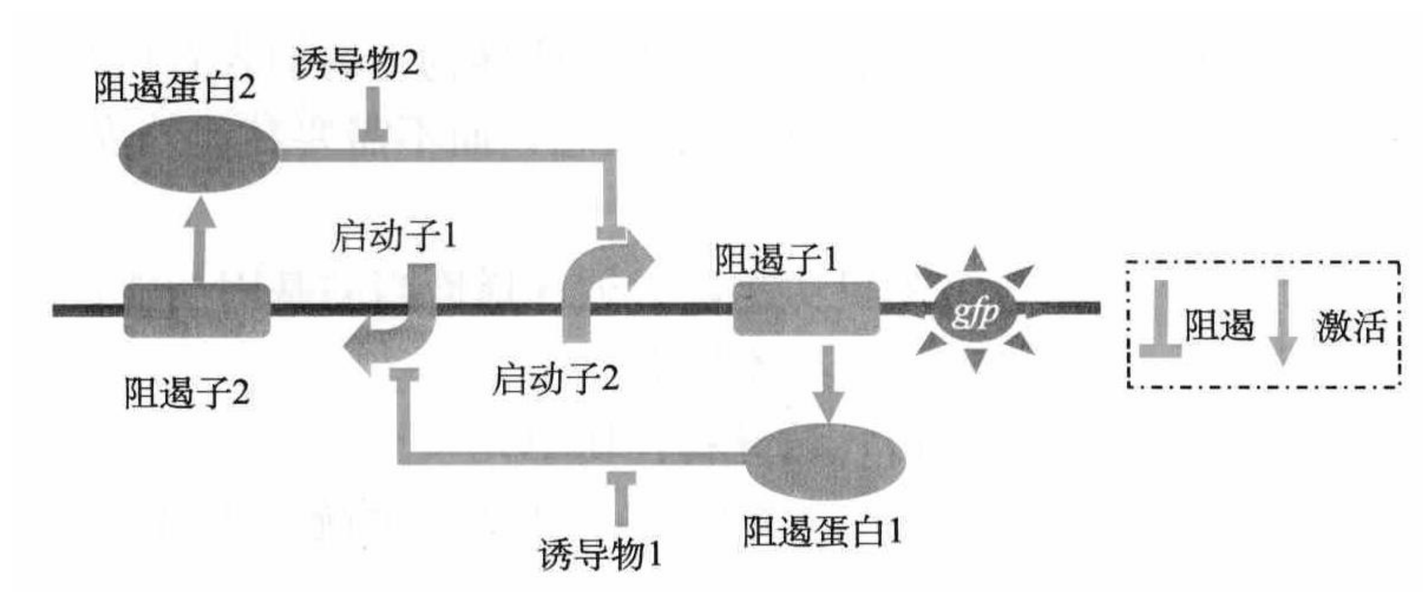


图 4-9　双稳态开关的基因线路设计

由于上下游基因的启动和相互抑制，整个系统共有两种稳定的输出状态：①当启动子 1 启动时，阻遏子 2 基因表达，其产物导致启动子 2 被抑制，阻遏子 1 不表达，无法对启动子 1 产生抑制作用，系统稳定于启动子 1 启动而启动子 2 关闭的状态，*gfp* 基因无法表达，系统中无荧光蛋白；此时若加入诱导物 2，解除阻遏子 2 对于启动子 2 的抑制作用，使得阻遏子 1 可以表达，抑制启动子 1，终止阻遏子 2 基因的表达，一段时间以后系统从 A 状态转换至 B 状态；②当启动子 2 启动时，阻遏子 1 基因被表达，启动子 1 被抑制，阻遏子 2 不表达，不能抑制启动子 2 的启动。系统稳定于启动子 2 启动而启动子 1 被关闭的状态。启动子 2 下游的报告基因同时被开启，可以检测到报告基因的产物。同理，加入诱导物 1 作用一段时间，也可以使系统从状态 B 切换到状态 A。

子 2 关闭的状态。启动子 2 下游的报告基因也无法开启，系统无特定的输出产物。

状态 B：当启动子 2 启动的时候，阻遏子 1 基因被表达，启动子 1 被抑制，阻遏子 2 不表达，不能抑制启动子 2 的启动，系统稳定于启动子 2 启动而启动子 1 被关闭的状态。启动子 2 下游的报告基因同时被开启，可以检测到报告基因的产物。

表 4-1 所示为状态 A 和 B 所对应的相关基因状态图，“1”表示开启/表达，“0”表示关闭/不表达。

表 4-1　双稳态开关状态表

状态	启动子 1	启动子 2	阻遏子 1	阻遏子 2	荧光蛋白
A	开启	关闭	不表达	表达	无
B	关闭	开启	表达	不表达	有

系统的两个外部输入信号——诱导物 1 和诱导物 2，分别能解除阻遏子 1 基因的表达产物对启动子 1 的抑制作用和阻遏子 2 基因的表达产物对启动子 2 的抑制作用。它们的加入可以实现系统在两种状态之间的转换，是人工控制状态切换的手段。如图 4-10 所示，当系统稳定于状态 A 时，通过加入诱导物 2，可以解除阻遏子 2 对启动子 2 的抑制作用，使得阻遏子 1 表达，从而抑制启动子 1，终止阻遏子 2 的表达。一段时间以后，系统从状态 A 转换至状态 B，而即使解除诱导物 2 的输入，系统依然会稳定于状态 B。同理，加入诱导物 1 作用一段时间，也可以

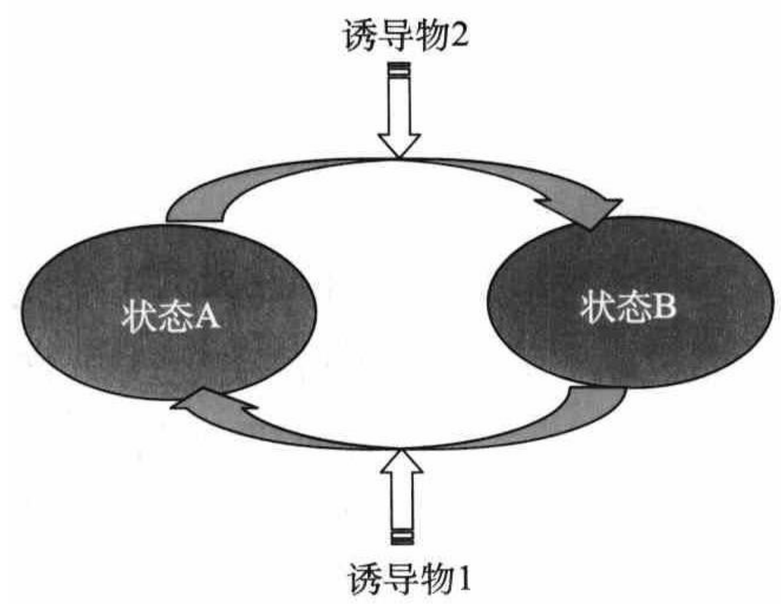


图 4-10　双稳态开关状态转换示意图

使系统从状态 B 切换到状态 A。如果将诱导物 1 和诱导物 2 的加入看作是对开关的“拨动”，那么每次拨动后，系统都会稳定于相应的状态，而不需要其他外界信号来维持。

2）基因线路实现

为了实现上述设计的基因线路逻辑功能，必须选择恰当的基因功能模块。作者共选取了三种启动子-阻遏子-诱导物方案，具体如下。

A. 启动子：$P_{trc\text{-}2}$；阻遏子：*lacI*；诱导物：IPTG。

B. 启动子：Pl_{s1con}；阻遏子：*cIts*；诱导物：温度（为统一起见，此处仍称其为“诱导物”）。

C. 启动子：Pl_{tetO1}；阻遏子：*tetR*；诱导物：aTc。

方案 A 中，外加乳糖类似物 IPTG 作为诱导物以解除 LacI 对于乳糖操纵子的抑制作用。方案 B 中，使用 λ 噬菌体中对于温度敏感的 CI 阻遏蛋白抑制启动子的表达，因此环境温度的升高可以解除这种抑制作用。方案 C 中，利用四环素启动子受 TetR 蛋白抑制的机制，加入四环素类似物 aTc 以解除相应的抑制作用。

为了能够形象地显示系统的状态，选取绿色荧光蛋白作为系统的报告蛋白。通过检测基因 *GFPmut*3 表达出来的绿色荧光蛋白的荧光值表征系统状态。

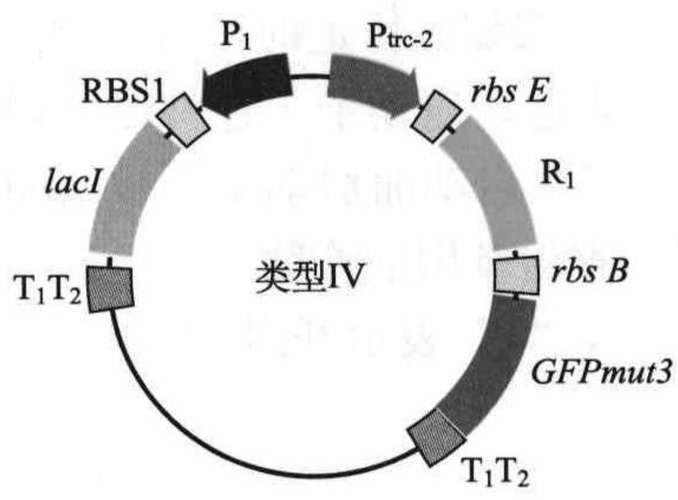


图 4-11　双稳态开关质粒结构[58]

研究者利用不同的载体，分别构建了两类不同的双稳态开关。一类是利用方案 A 和方案 B 在 pTAK 载体上构建双稳态开关，称为 pTAK 类；另一类是利用方案 A 和方案 C 在 pIKE 载体上构建双稳态开关，称为 pIKE 类。完整的质粒如图 4-11 所示。其中，P_1、R_1 由于类别的不同会有所不同；T_1、T_2 分别为两个终止子；RBS1、*rbsE*、*rbsB* 分别为不同的核糖体结合位点。

通过改变核糖体结合位点，可以改变基因的表达速率。因此，通过选用不同的核糖体结合位点，能够构造出 4 种 pTAK 载体的双稳态开关（pTAK117、pTAK130、pTAK131、pTAK132）和 2 种 pIKE 载体的双稳态开关（pIKE107、pIKE105）。

3）数学模型的建立

为了更好地研究双稳态开关系统的性能，预测和指导实验，研究者建立了数学模型模拟系统的行为。

该数学模型由一组微分方程组成，如式（4-1）。其中，U、V 分别表示两种阻遏蛋白的量；α_1、α_2 为两种启动子（包含核糖体结合位点共同作用）在没有阻遏蛋白时的表达速率；$-U$、$-V$ 表示两种阻遏蛋白的自然降解速率；而 β、γ 为启动子的抑制参数，数值越大，表示阻遏蛋白对启动子的抑制作用越强。

$$\begin{aligned}\frac{\mathrm{d}U}{\mathrm{d}t}&=\frac{\alpha_1}{1+V^{\beta}}-U\\ \frac{\mathrm{d}V}{\mathrm{d}t}&=\frac{\alpha_2}{1+U^{\gamma}}-V\end{aligned}\tag{4-1}$$

U、V 的变化速率为 0 时，说明两种阻遏蛋白的含量处于相对稳定的状态，即系统处于稳态。当两种启动子的启动能力比较相近，即 α_1 和 α_2 的数值比较接近时，双稳态开关的仿真模拟如图 4-12A 中两条实线所示。两条线共有三个交点：当系统的状态位于两条实线在 45°对角线上的交点时，任何一个微小的波动，都有可能会导致系统从一种状态进入另外一种状态，因此这个交点为系统的相对稳态点；而另外两个点为稳定的状态点，当系统状态发生微小的波动以后，系统会恢复到原来的状态。所以，整个系统有两个稳定状态，分别对应 V 含量相对较多的状态 1 和对应 U 含量相对较多的状态 2。

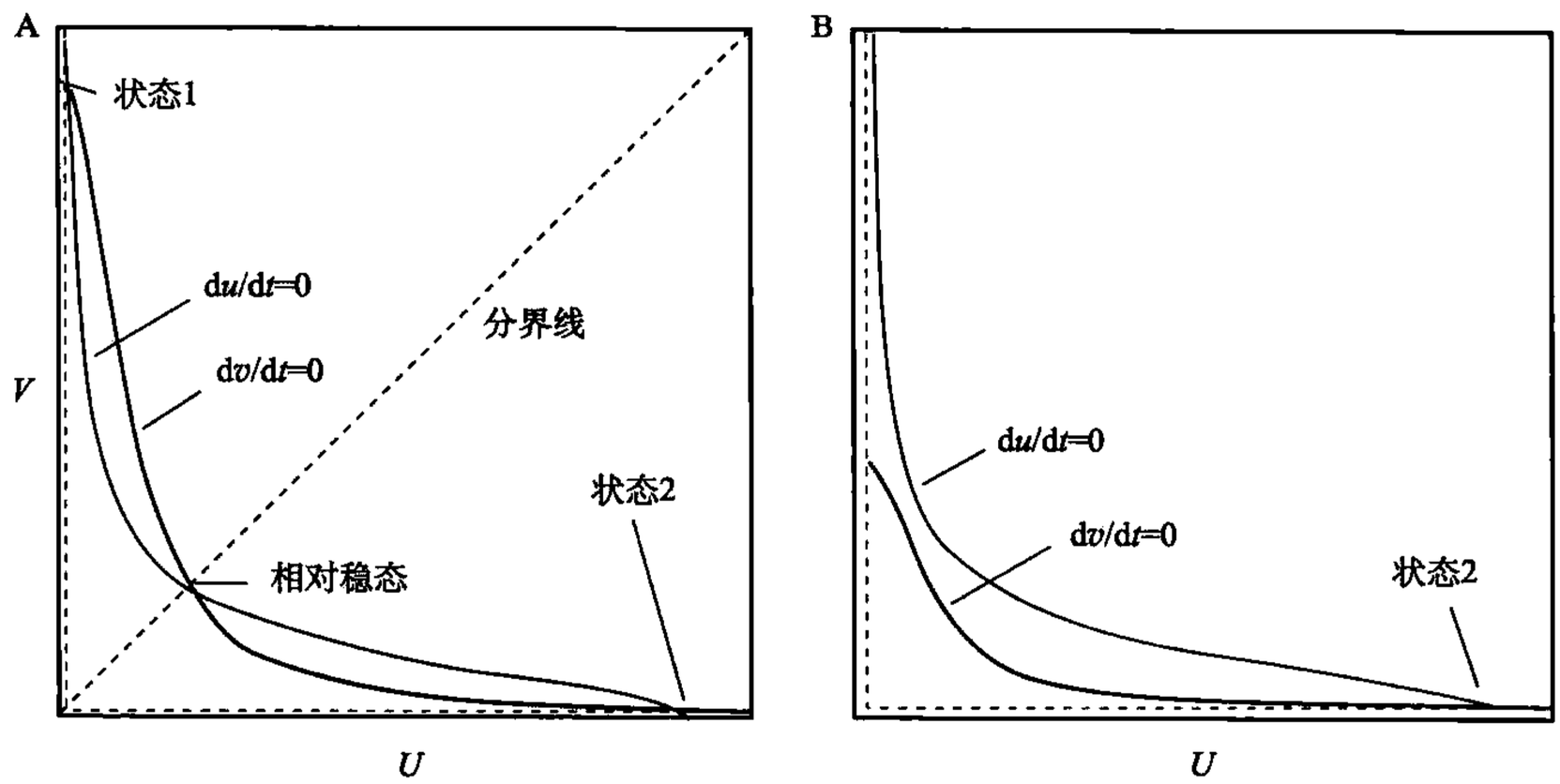


图 4-12　双稳态开关状态图[58]

A. $\alpha_1 \approx \alpha_2$；B. $\alpha_1 \gg \alpha_2$

图 A 中当两种启动子的启动能力 α_1 和 α_2 的数值比较接近时，双稳态开关的仿真模拟如图中两条实线所示。两条线共有三个交点，当系统的状态位于两条实线在 45°对角线上的交点时，为系统的相对稳态点。另外两个点为稳定的状态点，当系统状态发生微小的波动以后，系统会恢复到原来的状态。图 B 中当两种启动子的能力相差比较大时（如 $\alpha_1 \gg \alpha_2$）系统模拟的状态图，此时系统只有一个稳态。

当两种启动子的能力（包括核糖体结合位点）相差比较大的时候，如 $\alpha_1 \gg \alpha_2$，系统模拟如图 4-12B 所示。由图可知，系统只有一个稳定状态，不存在双稳态情况。

图 4-13A 模拟了 α_1 和 α_2 取值不同的情况下，系统的状态变化趋势；图 4-13B 则模拟了抑制作用不同（β、γ 不同）时系统的状态。可以看出，阻遏蛋白的抑制作用越强，系统具有双稳态的参数取值区间越大。因此，为了获得较稳定的双稳态开关，可以选择抑制作用强的阻遏蛋白编码基因。

4）功能检测

所有的双稳态开关都被转入大肠杆菌 JM2.300 菌种中验证其功能。带有双稳态开关质粒的大肠杆菌在培养基中培养 23.5h，分别用不同的诱导物诱导一段时间后再转入新的培养基中培养。通过检测不同时刻系统中荧光的量来检测双稳态开关的功能。

为了验证结果的可靠性，研究者还分别构建了受 IPTG 控制产生荧光蛋白的 pTAK102 质粒、受温度控制产生荧光蛋白的 pTAK106 质粒以及受 aTc 控制产生荧光

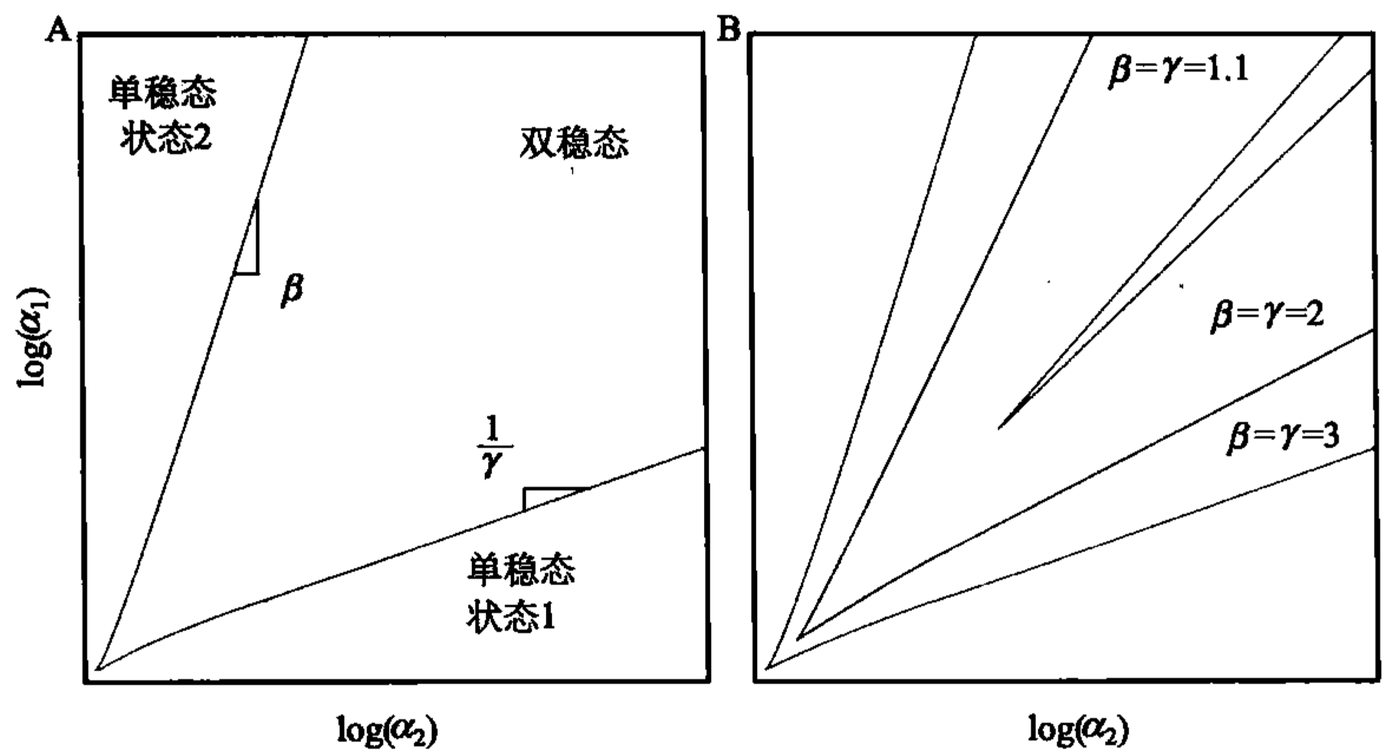


图 4-13 不同启动能力和抑制作用下双稳态开关的状态图[58]

A. 启动能力不同；B. 抑制能力不同

图 A 模拟了 α_1 和 α_2 取值不同的情况下，系统的状态变化趋势；图 B 则模拟了抑制作用不同（β，γ 不同）时系统的状态。可以看出，阻遏蛋白的抑制作用越强，系统具有双稳态的参数取值区间越大。

蛋白的 pIKE108 质粒作为实验对照质粒。所有的 6 种双稳态开关（4 种 pTAK 载体和 2 种 pIKE 载体的双稳态开关）以及 3 种对照质粒实验的结果如图 4-14 所示。以加入 IPTG 的初始时刻为 0 时刻点。从图中可以看出，在加入 IPTG 以前，系统中的荧光量较少；在加入 IPTG 6h 以后，所有的双稳态开关以及对照质粒 pTAK102 的系统中荧光值都升至了较高水平。当重新放置于新的培养基中 5h 以后，除了开关 pIKE105 以及对

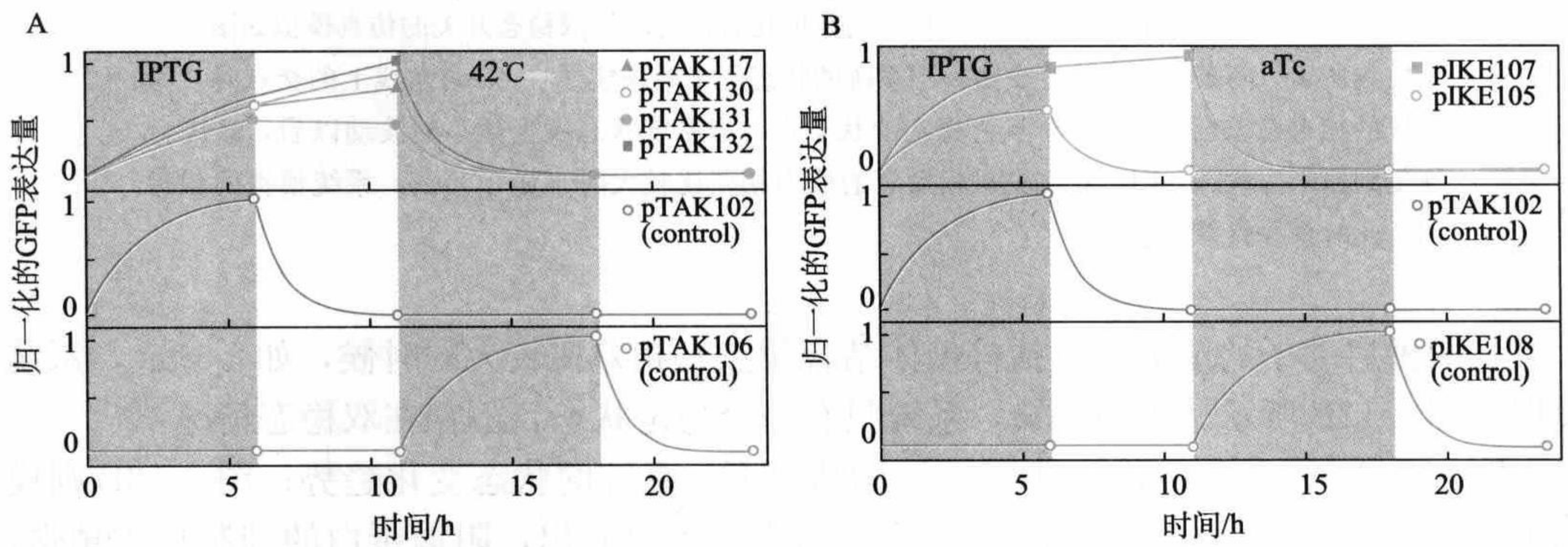


图 4-14 双稳态开关实验结果图[58]

4 种 pTAK 载体、2 种 pIKE 载体的双稳态开关以及 3 种对照质粒的实验结果如图所示。以加入 IPTG 的初始时刻为 0 时刻点。A、B 图中可以看出，在加入 IPTG 以前，系统中的荧光量较少，在加入 IPTG 6h 以后，所有的双稳态开关以及对照质粒 pTAK102 的系统中荧光值都升至了较高水平。当重新放置于新的培养基中 5h 以后，除了开关 pIKE105 以及对照质粒 pTAK102 以外，其他开关的荧光量都基本维持较高水平，说明系统从初始对应于荧光量较低的状态转换并维持在另一种对应于荧光量较高的稳定状态。在接下来的 7h，在另一种诱导物（升高温度或者引入 aTc）的作用下，荧光值处于高水平的开关全部降到了低水平，并且作为对照实验的 pTAK106、pIKE108 质粒荧光值升到了高水平。在放入新的培养基以后，所有的细胞荧光值都基本维持稳定，说明系统又转换回对应于荧光值较低的稳态。由此可见双稳态开关的功能已经基本实现。

照质粒 pTAK102 以外，其他开关的荧光量都基本维持较高水平，说明系统从初始对应于荧光量较低的状态转换并维持在另一种对应于荧光量较高的稳定状态。接下来的 7h 内，在另一种诱导物（升高温度或者引入 aTc）的作用下，荧光值处于高水平的开关全部降到了低水平，并且作为对照实验的 pTAK106、pIKE108 质粒荧光值升到了高水平。在放入新的培养基以后，所有的细胞荧光值都基本维持稳定，说明系统又转换回对应于荧光值较低的稳态。由此可见双稳态开关的功能已经基本实现。

4.3　Repressilator[60]

基因振荡是一种基因调控机制，由振荡的幅度和周期来决定基因表达的时间。这种基因表达的时间控制可以实现在大规模基因网络中仅利用少量的调控因子即可调控相对较多的基因，从而实现对复杂细胞行为的调控。

在转录水平上，如果将三个表达产物相互抑制的基因模块串联成一个环状结构，利用基因模块间的彼此抑制和解抑制即可实现振荡器的功能。这就是 M. B. Elowitz 等构建的基因振荡器。该 Repressilator 实现了在没有外加输入作用下，生物系统中基因表达按照特定的周期变化的功能。下面我们就来讲解 Repressilator 的构建和实验结果。

1. 遗传线路设计

由 M. B. Elowitz 等设计完成的 Repressilator 的遗传线路如图 4-15 所示。图中采用了三个启动子，每个启动子所控制的基因的表达产物均能抑制下游最近邻一个启动子的启动：启动子 Pl_{lacO1} 控制基因 *tetR-lite* 的转录，产生的四环素阻遏蛋白 TetR 抑制下游 Pl_{tetO1} 启动子；Pl_{tetO1} 控制 *cI-lite* 的转录，其产物 CI 蛋白抑制启动子 P*r*；P*r* 控制 *lacI-lite* 的转录，产生的乳糖阻遏蛋白 LacI 抑制 Pl_{lacO1}。三个启动子环环相扣形成一个完整的

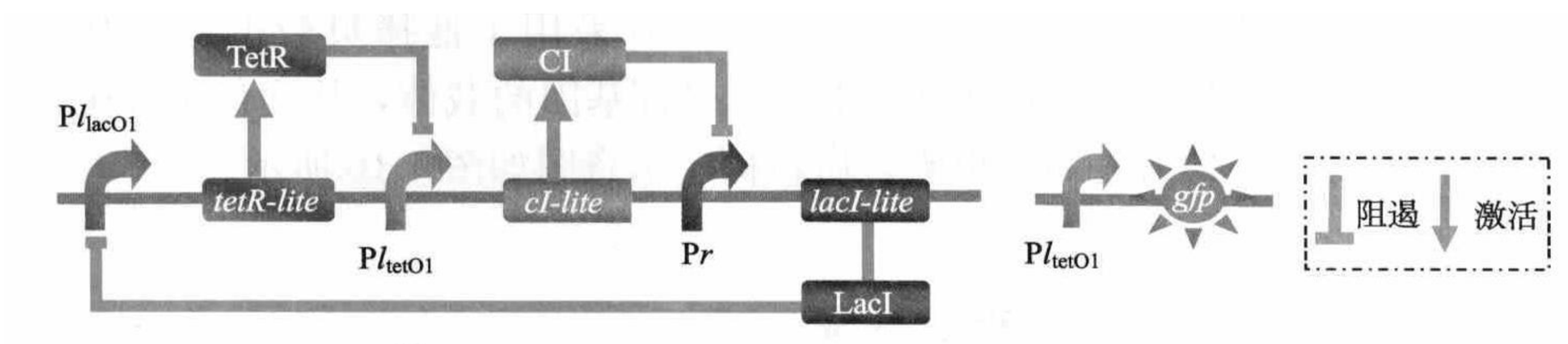


图 4-15　Repressilator 遗传线路图

启动子 Pl_{lacO1} 控制基因 *tetR-lite* 的转录，产生的四环素阻遏蛋白 TetR 抑制下游 Pl_{tetO1} 启动子；因 Pl_{tetO1} 控制 *cI-lite* 的转录，CI-lite 蛋白无法产生，P*r* 启动子的抑制被解除；P*r* 下游的 *LacI-lite* 基因顺利表达，产生 LacI-lite 蛋白并抑制 Pl_{lacO1} 启动子，Pl_{lacO1} 启动子的状态从“开启”转换为“关闭”；下游基因停止产生 TetR-lite 蛋白，Pl_{tetO1} 启动子的抑制被解除，状态由“关闭”转换为“开启”；再下游基因产生 CI-lite 蛋白抑制 P*r* 启动子，导致 P*r* 启动子的状态从“开启”转换为“关闭”；P*r* 启动子的被抑制使得 LacI-lite 蛋白无法继续产生，Pl_{lacO1} 启动子的抑制被消除，其状态由“关闭”转换为“开启”，下游基因表达 TetR-lite 蛋白抑制 Pl_{tetO1} 启动子。如此往复，实现了整个 Repressilator 在三种阻遏子基因轮流表达、阻遏蛋白相继降解的过程中周而复始的振荡。作为输出信号，当系统状态在三个阻遏子基因顺序开启的过程中振荡的时候，蛋白 TetR-lite 的浓度也会发生振荡，其对启动子 Pl_{tetO1} 抑制作用的强弱出现相应的振荡，导致下游绿色荧光蛋白 GFP 的表达量也随着 TetR 浓度的变化而变化，即随着 Repressilator 的周期变化而变化。

循环抑制 Repressilator，在三种阻遏子基因轮流表达、阻遏蛋白相继降解的过程中实现自身及其调控基因开启状态的顺序转变，从而实现周而复始的振荡。

上述 Repressilator 的状态转换可具体表述如下：当 Pl_{lacO1} 启动子得以成功启动，产生的 TetR-lite 蛋白会抑制下游 Pl_{tetO1} 启动子，CI-lite 蛋白无法产生，P*r* 启动子的抑制被解除，P*r* 下游的基因顺利表达，产生 LacI-lite 蛋白并抑制下游 Pl_{lacO1} 启动子，Pl_{lacO1} 启动子的状态从“开启”转换为“关闭”，下游基因停止产生 TetR-lite 蛋白，Pl_{tetO1} 启动子的抑制被解除，状态由“关闭”转换为“开启”，下游基因产生 CI-lite 蛋白去抑制 P*r* 启动子，导致 P*r* 启动子的状态从“开启”转换为“关闭”；P*r* 启动子的被抑制使得 LacI-lite 蛋白无法继续产生，Pl_{lacO1} 启动子的抑制被消除，其状态由“关闭”转换为“开启”，下游基因表达 TetR-lite 蛋白抑制 Pl_{tetO1} 启动子。如此往复，实现了整个 Repressilator 的状态转换。

作为输出信号，荧光蛋白的表达量可以表征系统的状态。当系统状态在三个阻遏子基因顺序开启的过程中振荡的时候，蛋白 TetR 的浓度也会发生振荡，其对启动子 Pl_{tetO1} 抑制作用的强弱出现相应的振荡，导致下游绿色荧光蛋白 GFP 的表达量也随着 TetR 浓度的变化而变化，即随着 Repressilator 的周期变化而变化。

通过加强启动子的启动能力（包括改变核糖体结合位点）调节相应蛋白质对于启动子的抑制作用以及 mRNA 和蛋白质的降解速率，可以更好地促使系统进入振荡状态。在系统中，通过操纵子序列与启动子的结合，可以使系统中的启动子启动能力增强，并且可以使阻遏子基因对启动子的调控更加完全。此外，在三种阻遏子基因以及荧光蛋白基因的末端加上能够被大肠杆菌蛋白酶识别的标记，可以达到加快蛋白质降解的目的，使阻遏蛋白降解半衰期从大于 60min 减小到 4min 左右，更加接近于 mRNA 2min 的降解半衰期，同时使荧光蛋白的降解半衰期降低到 30～40min，以便快速表征系统状态的变化。

为了使系统的振荡状态更好地被表征，研究者采用了低拷贝数的载体作为 Repressilator 的载体，而相对高拷贝数的载体作为报告基因的载体，从而使 Repressilator 状态的变化可以被更好地放大表征出来，质粒构建示意图如图 4-16 所示。

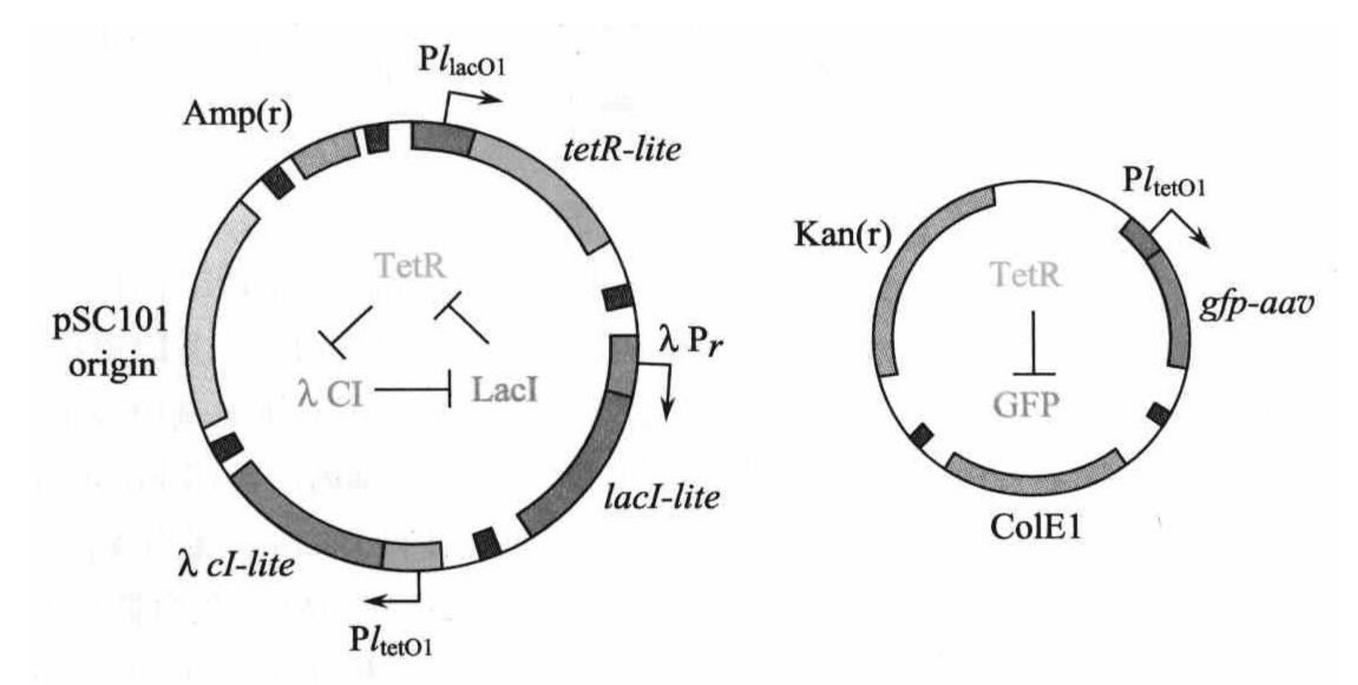


图 4-16 阻遏子调控模型质粒构建示意图[60]

左图为 Repressilator 质粒。其中黑色方块代表转录终止子。后缀‘*lite*’表示具有快速降解标签的蛋白。右图为携带荧光基因的质粒。

2. 数学模型模拟

为了进一步分析系统行为，研究者建立了连续的数学模型，如下式：

$$\begin{aligned}\frac{\mathrm{d}m_i}{\mathrm{d}t} &= -m_i + \frac{\alpha}{(1+p_j^n)} + \alpha_0 \\ \frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\beta(p_i - m_i)\end{aligned} \quad \begin{pmatrix} i = lacI, tetR, cI \\ j = cI, lacI, tetR \end{pmatrix} \tag{4-2}$$

其中，p_j 为三种阻遏蛋白的浓度；m_i 为它们对应的 mRNA 的浓度；α_0 为启动子的渗漏启动速率；$\alpha+\alpha_0$ 为启动子的正常启动速率；β 为蛋白质降解速率与 mRNA 降解速率的比值；n 为希尔系数。模型中认为三种启动子的最大转录量、渗漏表达量、阻遏子基因的调控作用以及蛋白质的翻译速度均相同。

通过模拟可以得到在不同参数下系统的状态图，如图 4-17 所示。Repressilator 会出现两种状态。

I：系统处于稳定的状态，各种阻遏子基因的表达量、阻遏蛋白的量不随时间变化。

II：系统处于周期变化的状态。

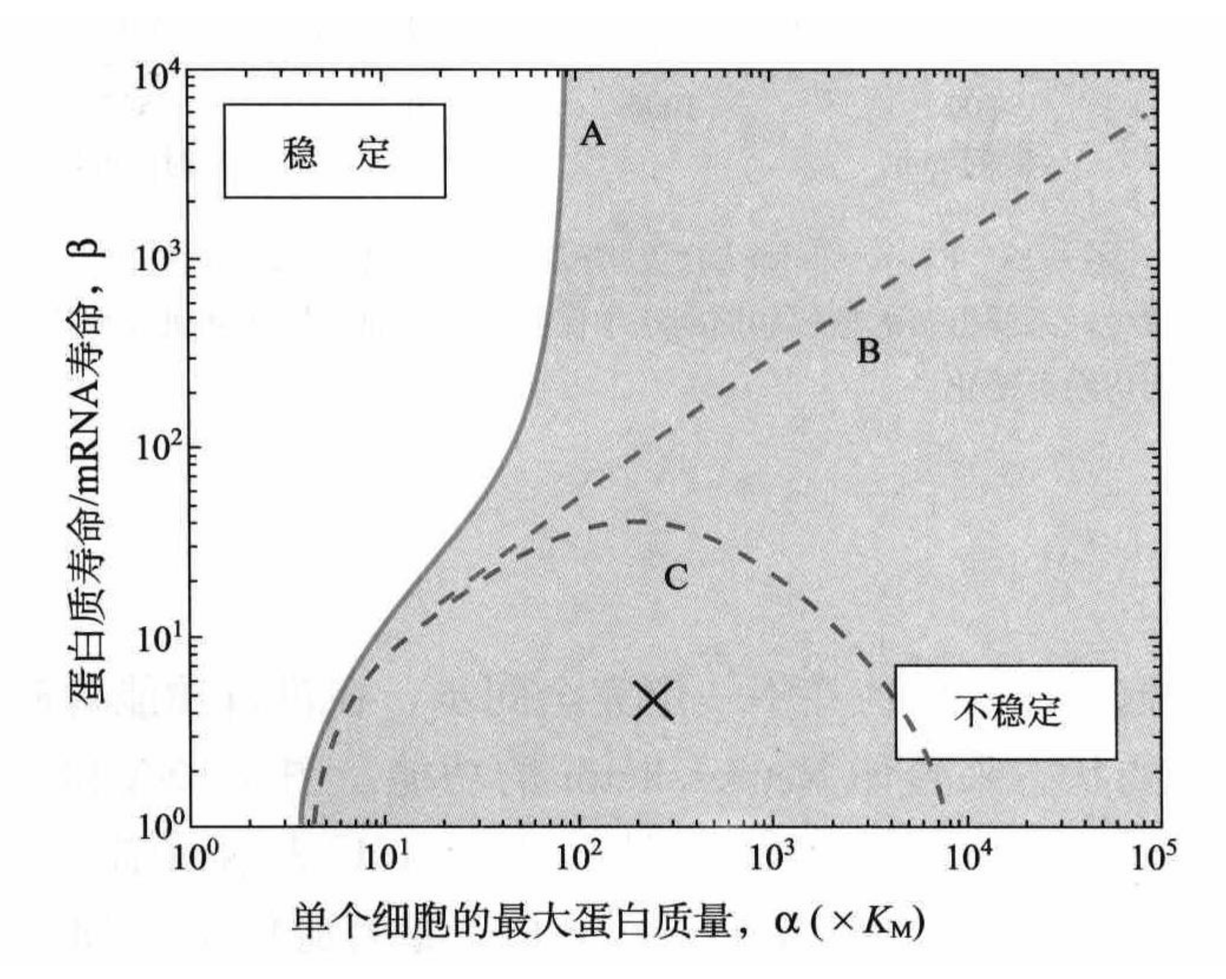


图 4-17　Repressilator 稳定性随参数变化的状态图[60]

参数空间被分为左上角的稳定区域和右下角的不稳定区域。曲线 A、B 和 C 分别为不同参数值时稳定区和不稳定区的分界线。其中 A：$n=2.1$，$\alpha_0=0$；B：$n=2$，$\alpha_0=0$；C：$n=2$，$\alpha_0/a=10^{-3}$。

从图 4-17 中可以看出，参数空间被分为两个区域，即左上角的稳定区和右下角的不稳定区域。曲线 A、B 和 C 分别为不同参数值时稳定区和不稳定区的分界线。其中，A：$n=2.1$，$\alpha_0=0$；B：$n=2$，$\alpha_0=0$；C：$n=2$，$\alpha_0/a=10^{-3}$。可以明显看出，参数的变化是导致系统状态及状态边界变化的主要原因。希尔系数 n 的增大和渗漏启动速率 α_0 的减小都会增大系统的不稳定性。

基因表达过程中会由于以下原因不可避免地带有随机性：①阻遏蛋白与启动子不同位点的结合；②阻遏蛋白从启动子上第一个位点脱落；③阻遏蛋白从启动子上第二个位点脱落；④从未结合阻遏蛋白的启动子上进行转录；⑤从被蛋白阻遏的启动子上进行转录、翻译；⑥蛋白质以及 mRNA 的降解等。

为了更好地分析系统的随机行为，研究者建立随机模型对系统的随机因素进行模拟。确定模型和随机模型模拟系统的结果如图 4-18 所示。可以看到，确定性模型中（A 图），系统振荡的振幅和周期基本保持不变；而在随机模型模拟下（B 图），系统振荡的周期和振幅均发生变化。后续实验结果表明，随机模型的模拟与实验验证结果能够更好地吻合，更真实地反映系统状态。

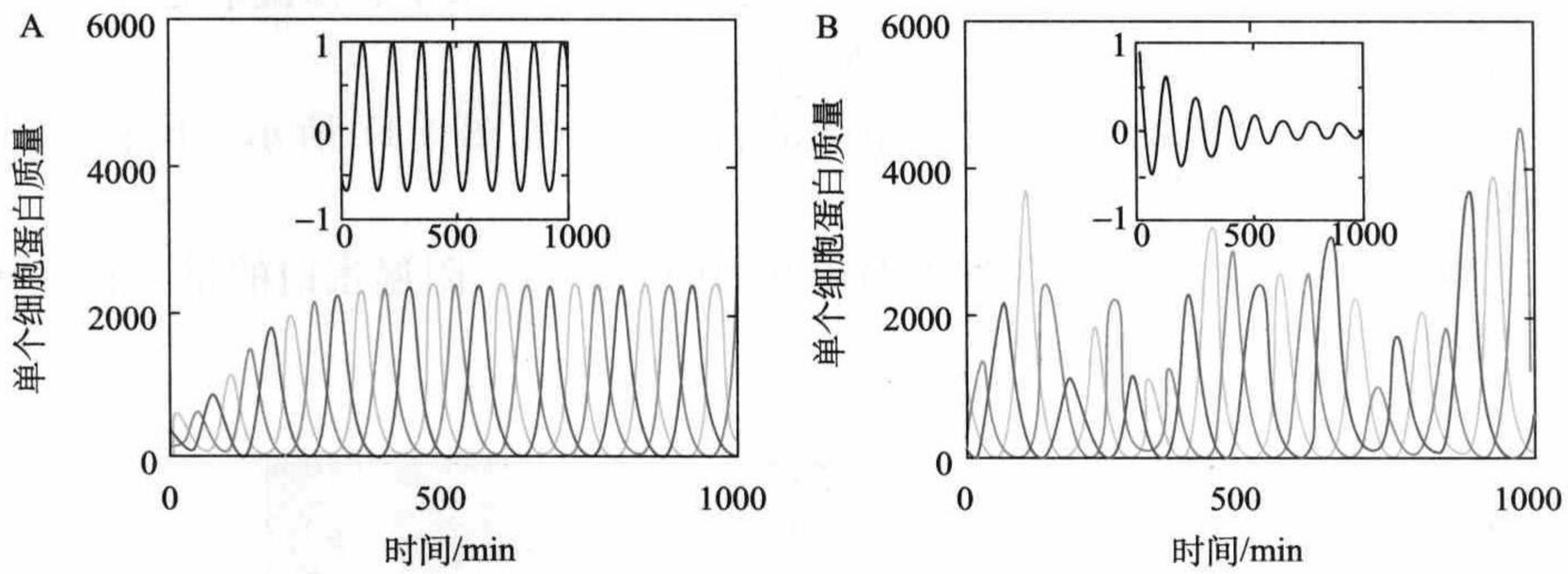


图 4-18 Repressilator 确定性模型和随机模型仿真结果图[60]

A 图确定性模型中，系统振荡的振幅和周期基本保持不变；而 B 图在随机模型模拟下，系统振荡的周期和振幅均发生变化。

3. 功能检测

考虑到各个细胞之间的振荡并不一定完全同步，在进行功能检测时，研究者通过跟踪观察一个细胞的状态来验证 Repressilator 的功能。图 4-19A 和图 4-19B 为单个细胞不同时刻在荧光和可见光下观察的结果图，图 4-19C 表示该细胞的荧光值随时间的周期振荡变化。当细胞分裂时，Repressilator 的功能会遗传至子细胞，子细胞的荧光也会发生周期性变化。可以看到，细胞的荧光值按照大约 150min 的周期进行振荡。统计得到 Repressilator 的平均振荡周期为（160±40）min，不同于细胞分裂的周期，说明两者之间并无耦合。由于荧光蛋白的降解周期大于 Repressilator 的振荡周期，荧光蛋白还未完全降解时系统即进入下一振荡循环，因此系统的荧光值呈现振荡上升趋势。

考虑细胞的分裂因素后，荧光值随时间的变化见图 4-20。红色为母细胞的荧光检测结果，绿色和蓝色为细胞分裂以后两个子细胞的荧光值变化曲线。可以看出，细胞分裂以后，子细胞传承了母细胞振荡的性状继续进行振荡。统计得到，细胞分裂以后两个子细胞之间荧光值协同变化到相互独立变化的过程为（95±10）min。这个时间比细胞分裂的周期明显要长，说明细胞分裂以后系统的状态可以很好地传承

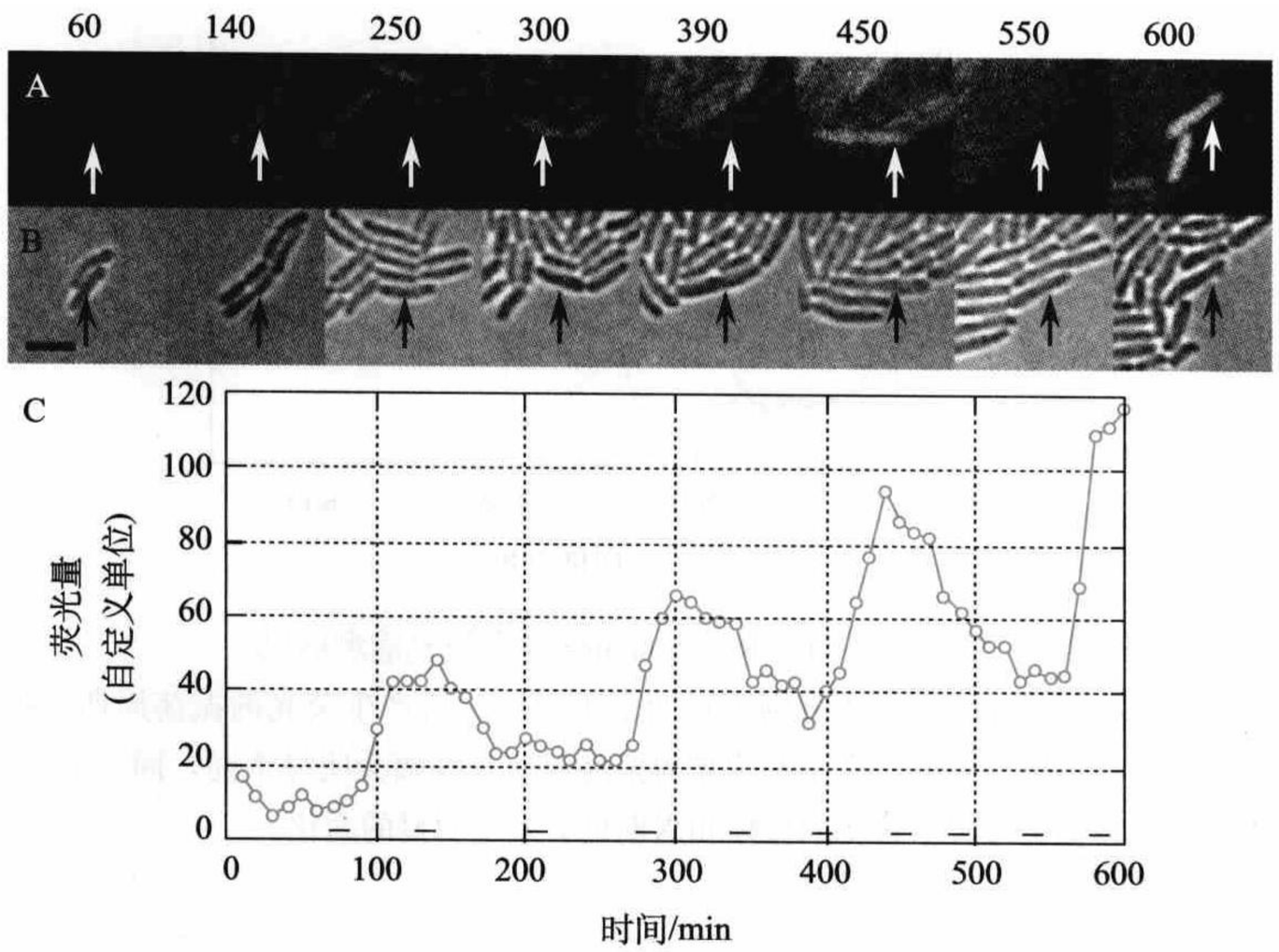


图 4-19　Repressilator 荧光检测结果（不考虑细胞分裂因素）[60]

图中下层的条柱显示的是不同时间细胞的分裂。可以看出，振荡的周期比细胞分裂周期长，并且子代仍然具有周期振荡现象。由于荧光蛋白的降解周期大于 Repressilator 的振荡周期，荧光蛋白还未完全降解时系统即进入下一振荡循环，因此系统的荧光值呈现振荡上升趋势。

到子细胞中去。

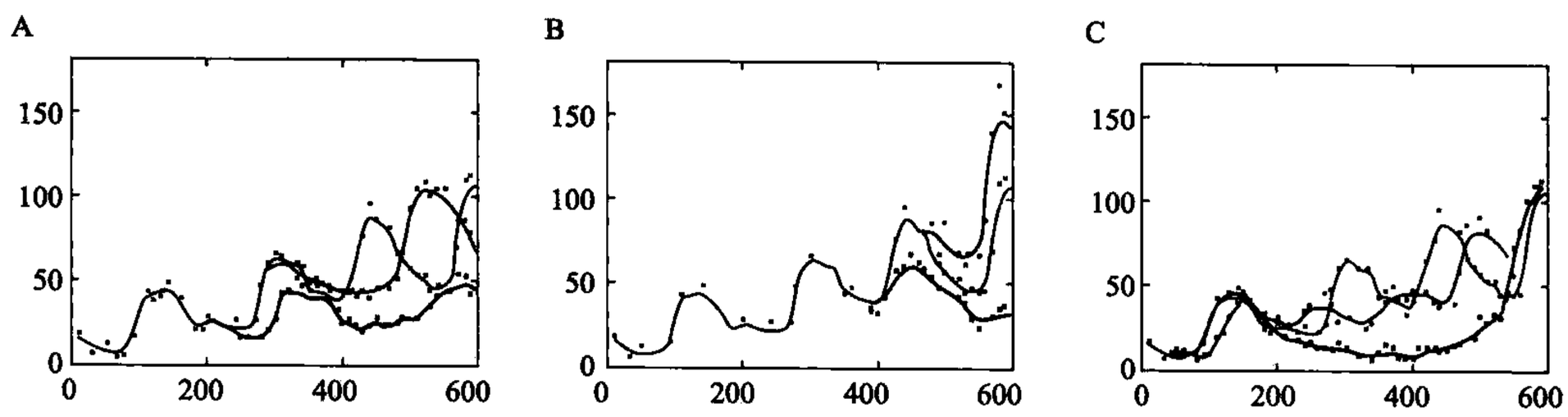


图 4-20　Repressilator 荧光检测结果（考虑细胞分裂因素）[60]

图中红色为母细胞的荧光值变化曲线，绿色与蓝色为细胞分裂以后两个子细胞的荧光值变化曲线。可以看出，细胞分裂以后，子细胞传承了母细胞振荡的状态继续进行振荡。

多个 Repressilator 细胞的荧光检测结果如图 4-21 所示。从图中可以看出，不同细胞所产生荧光的振荡周期和振幅明显不同，且同一个细胞所产生的荧光值在振荡过程中自身的振幅和周期也会发生明显的变化。

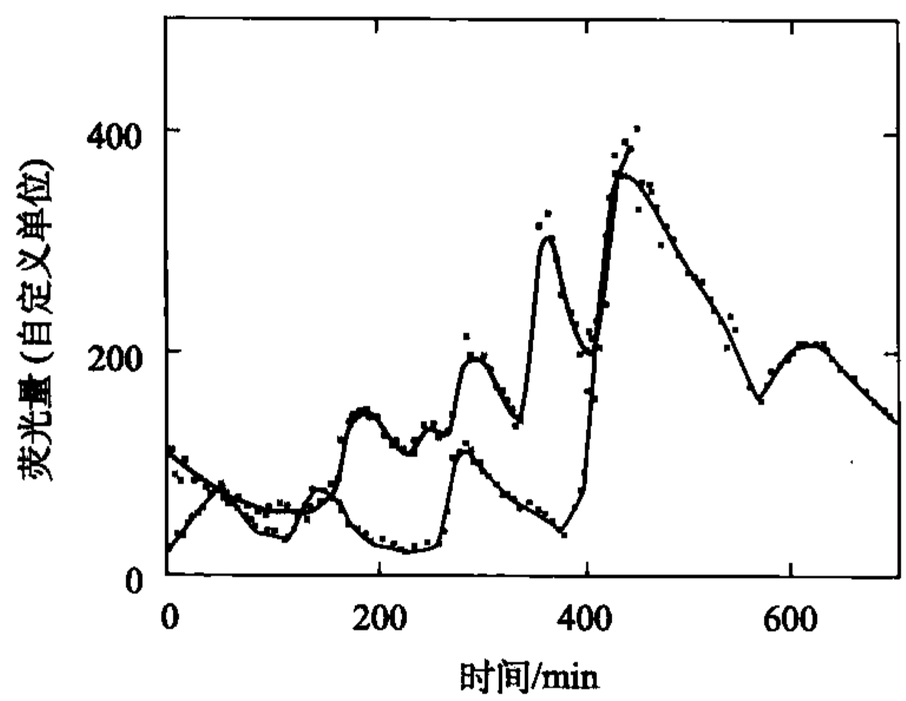


图 4-21 两个 Repressilator 细胞荧光检测结果

图中红色和蓝色采样点和曲线分别表示两个不同细胞所产生荧光的振荡周期和振幅。可以看出，不仅不同细胞所产生的荧光的振荡周期和振幅会明显不同，同一个细胞所产生的荧光在振荡过程中自身的振幅和周期也会发生明显的变化。

4.4 细胞群体系统脉冲发生器[176]

前面讲述的经典实例都是在单细胞内构建遗传线路，不涉及细胞群体间的信号交流和通讯。然而，细胞协同操作共同完成特定任务是工业化的必然趋势和要求，构建协调细胞群体操作的时空信号系统则显得尤为重要。如同在大规模和超大规模电子线路中一样，在生物系统中，脉冲信号仍然是重要的同步时钟信号。因此合成生物学家开发了许多生物脉冲发生器，其中不乏各种精巧的单细胞系统和具有特殊功能的细胞群体系统。下面我们以细胞群体系统脉冲发生器为例，讲解脉冲发生器遗传线路的设计和构建。

S. Basu 等设计了细胞群体系统脉冲发生器，如图 4-22 所示。该系统由“信号发生细胞”和“脉冲发生细胞”两种细胞构成。图 4-22A 表示初始状态下，信号发生细胞中由于 TetR 蛋白抑制 *luxI* 上游的启动子，无 LuxI 蛋白催化 AHL 信号分子的产生，两种细胞间无信息交流。图 4-22B 表示当向“信号发生细胞”中加入 aTc 时，解除了 TetR 对 *luxI* 启动子的抑制作用，使 LuxI 蛋白得以表达，催化产生 AHL 信号分子。AHL 迅速扩散到细胞外，其中一部分进入脉冲发生细胞，并与 *luxR* 基因的产物 LuxR 结合，其复合物能够激活 cI^*（此处的 * 表示人工改造过的基因）和 *gfp* 上游启动子，启动 cI^* 和 *gfp* 基因的表达。图 4-22C 表示 cI^* 基因的产物 CI 蛋白能够抑制荧光蛋白的表达，当 CI 的浓度积累到一定数量时，会完全抑制 *gfp* 的表达，细胞中的 GFP 蛋白快速降解。因此，从脉冲发生细胞接收到信号分子产生 GFP 开始到 GFP 完全降解为止，在这段时间间隔内，脉冲发生细胞内荧光蛋白的浓度曲线类似于脉冲信号。

回顾第二章中的内容，读者应该能够注意到，脉冲发生器细胞中遗传线路的设计正

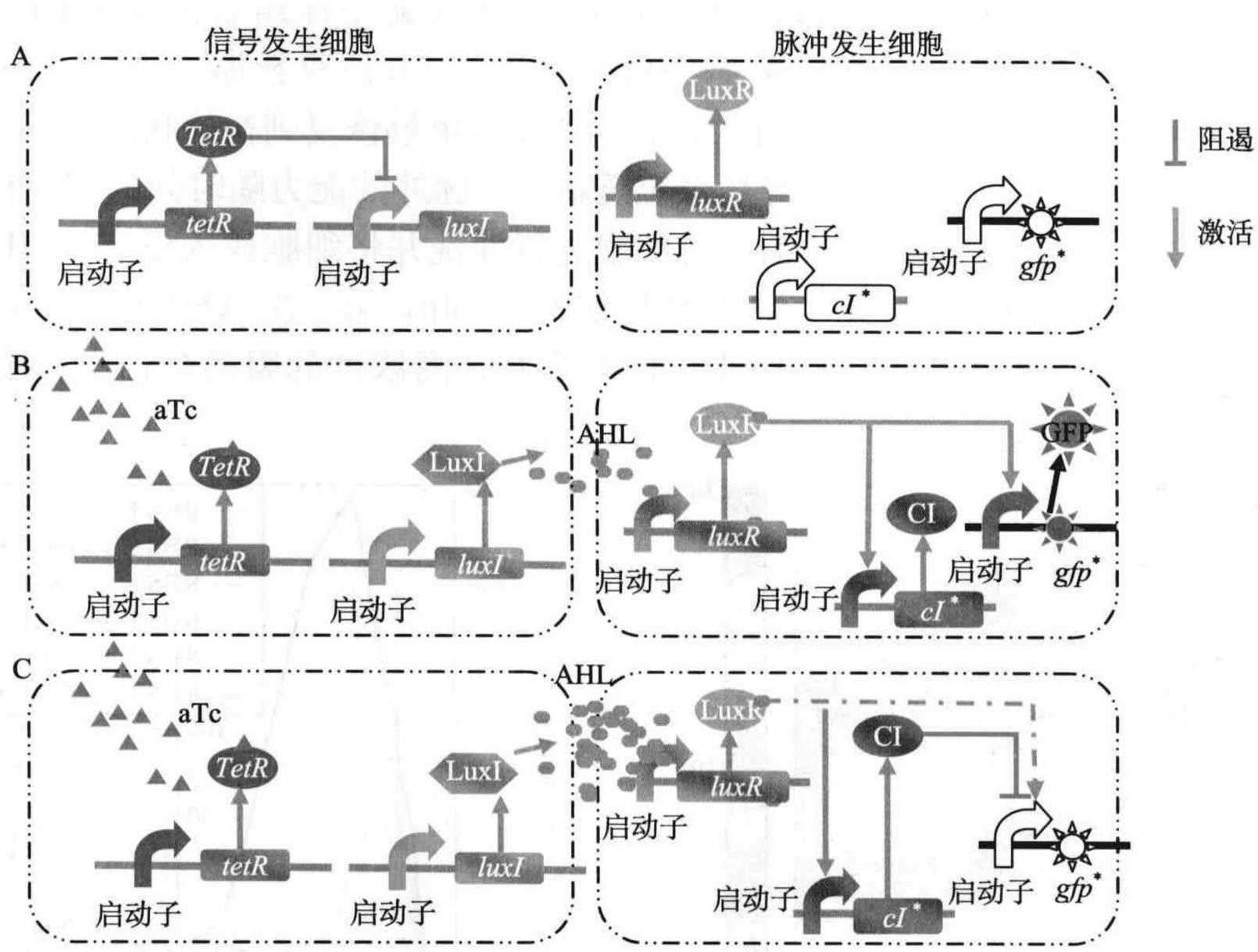


图 4-22　细胞群体系统脉冲发生器遗传线路图

系统由“信号发生细胞”和“脉冲发生细胞”两种细胞构成。图 A：初始状态下，信号发生细胞中，TetR 蛋白抑制 *luxI* 上游的启动子，无 LuxI 蛋白催化 AHL 信号分子的产生，两种细胞间无信息交流。图 B：向“信号发生细胞”中加入 aTc 时，aTc 解除 TetR 对 *luxI* 启动子的抑制作用，表达 LuxI 蛋白质，催化产生 AHL 信号分子。AHL 迅速扩散到细胞外，其中一部分进入脉冲发生细胞并与 *luxR* 基因的产物 LuxR 结合，启动 *cI** 和 *gfp* 基因的表达。图 C：*cI** 基因的产物 CI 蛋白抑制 *gfp* 基因上游的启动子，当 CI 的浓度积累到一定程度时，会完全抑制 *gfp* 的表达，细胞中的 GFP 蛋白快速降解。

是利用了“非一致前馈”的调控机制来实现其功能的。将 *luxR* 基因及其启动子看做 X、*cI** 基因及其启动子看做 Y、*gfp* 基因及其启动子看做 Z，读者应该不难画出所对应的“非一致”前馈逻辑拓扑图。

图 4-23A 模拟了脉冲发生器细胞中 CI 翻译效率和阻遏蛋白——操纵子结合亲和力对脉冲增益★的影响。在高 CI 翻译效率和操纵子结合亲和力时 GFP 的表达水平不会明显提高。此时即使没有 AHL，渗漏表达的 CI 蛋白仍然能够完全阻遏控制 GFP 表达的 $Pr_{luxcI\ OR1}$ 启动子。当模型参数取值较低时，响应于 AHL 的 GFP 水平较高，后续抑制较低。

图 4-23B 显示不同 RBS 强度和不同操纵子位点组合时的 GFP 响应。具有较强 RBS 结合能力（RBS II）和原始 CI OR1 操纵位点的脉冲发生器并没有产生显著的脉冲响应。而较弱 RBS 结合能力（RBS H）和 OR1 单残基突变体的组合能够获得更高的脉冲幅值和更长的响应时间。实验结果也证明这类动力学参数的改变显著影响着脉冲强度和

★ 脉冲增益：GFP 表达量的最大值和最终稳态值之间的差值除以 GFP 最终稳态值所得的比值。

宽度。图 4-23B 插入的小图为 RBS D 和 OR1 单残基突变体组合的实验图。在加入 AHL 4h 后，更换培养基，新的培养基中没有 AHL，6h 后重新加入 AHL，可以看出第二个脉冲响应的水平与第一个脉冲类似，说明系统能够恢复到初始状态。

图 4-23C 显示第一次 AHL 诱导后系统重新产生脉冲的能力随时间的变化行为。利用 140 nmol/L AHL 诱导细胞 6h 后，对细胞进行冲洗并将细胞放入没有 AHL 的新培养基中，之后重新用 140 nmol/L AHL 诱导细胞 20min，第二次 AHL 输入后，系统恢复初始脉冲浓度的时间间隔取决于 CI 蛋白的衰减，高浓度的阻遏蛋白会减弱对后续

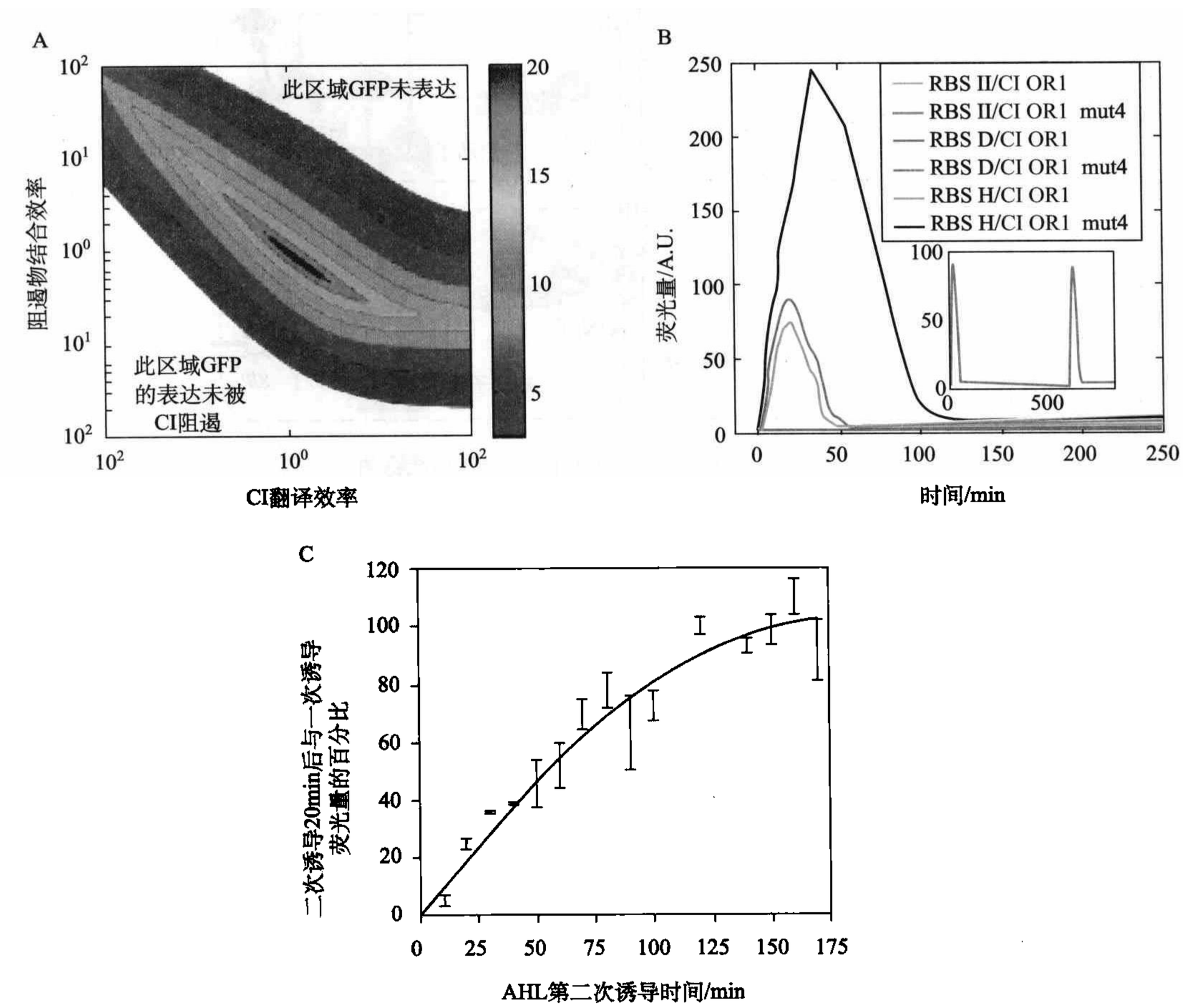


图 4-23 脉冲发生细胞系统参数变化对于脉冲响应的影响[176]

图 A 模拟了 CI 翻译效率和阻遏蛋白-操纵子结合亲和力对脉冲增益的影响。在高 CI 翻译效率和操纵子结合亲和力时 GFP 的表达没有被激活。当模型参数取值较低时，GFP 并未被 CI 蛋白抑制。图 B 显示不同 RBS 结合强度和不同操纵子位点组合时的 GFP 响应。具有较强 RBS 结合能力（RBS II）和原始 CI OR1 操纵位点的脉冲发生器并没有产生显著的脉冲响应。而较弱 RBS（RBS H）和 OR1 单残基突变体的组合能够获得更高的脉冲幅值和更长的响应时间。图 B 插入的小图为 RBS D 和 OR1 单残基突变体组合的实验图。在加入 AHL4h 后，更换新的无 AHL 的培养基，6h 后重新加入 AHL，可以看出第二个脉冲响应的水平与第一个脉冲类似。说明系统能够恢复到初始状态。图 C 显示第一次 AHL 诱导后系统重新产生脉冲的能力随时间的变化行为。利用 140 nmol/L AHL 诱导细胞 6h 后，对细胞进行冲洗并将细胞放入没有 AHL 的新培养基中，之后重新用 140 nmol/L AHL 诱导细胞 20min，第二次 AHL 输入后的 GFP 响应。

GFP 的激活。同时，脉冲响应也受到 LuxR-AHL 复合物（激活 Pr_{lux}表达 CI）衰减的影响，而 CI 和 LuxR-AHL 复合物的浓度由 AHL 水平决定。因此，初始 AHL 诱导物浓度也会影响恢复周期。

图 4-24 为不同输入浓度和速率的脉冲响应曲线。图 4-24A 为不同 AHL 浓度时的脉冲曲线，可以看出，当 AHL 浓度为 140 nmol/L 时，脉冲响应的峰值最高；当输入浓度超过 140 nmol/L 时，由于 CI 抑制的增强，脉冲响应的峰值反而下降。图 4-24B 和图 4-24C 显示了 AHL 输入增加速率对脉冲响应影响的模拟和实验结果。无论是模拟还是实验结果均表明，AHL 输入增加速率较低时会降低脉冲峰值和延迟脉冲峰值到来的时间。当 AHL 增加速率较高时，初始 GFP 和 CI 累积量较高，此后 CI 快速关闭 $Pr_{luxcI\ OR1}$启动子的活性。然而，此过程中 GFP 已被大量表达，导致脉冲的高幅值和短宽度。相反，当 AHL 增加速度较低时，初始 GFP 和 CI 相应较低，CI 需用较长时间关闭 $Pr_{luxcI\ OR1}$启动子的活性。此时，尽管脉冲宽度加大了，由于转录速率较低，CI 转录和翻译占据了大部分时间，二聚作用和操纵子结合则只占据了很小一部分，总的 GFP 表达量仍会降低。

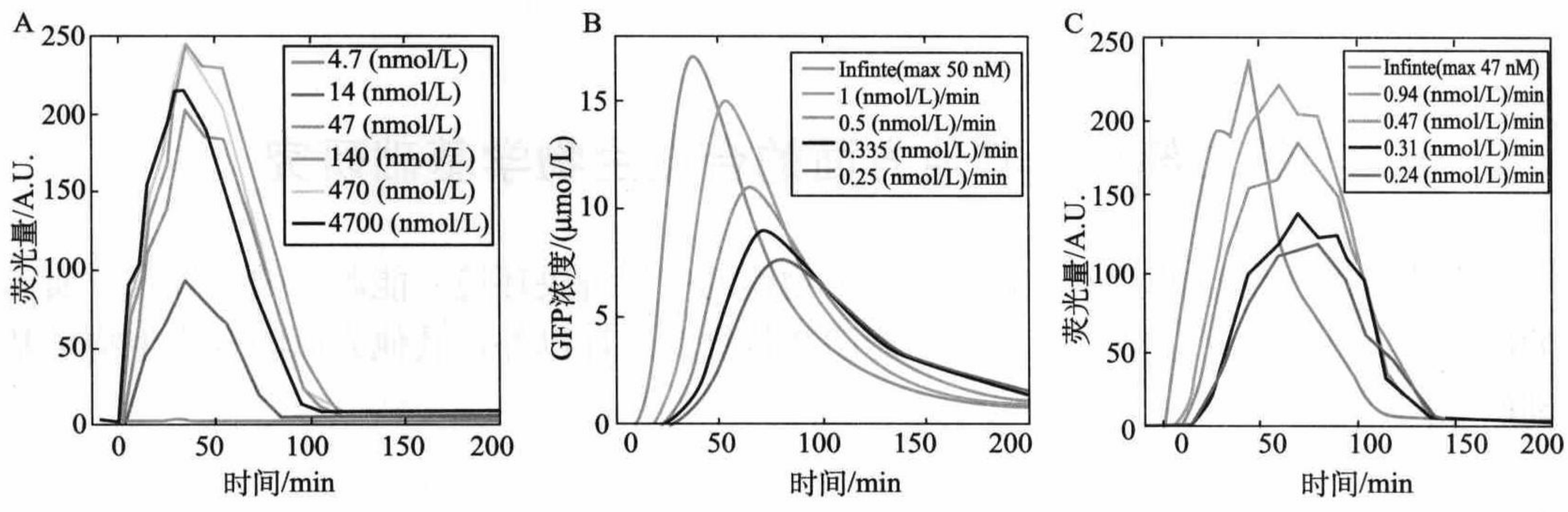


图 4-24　不同输入浓度和速率对脉冲响应的影响[176]

图 A 为不同 AHL 浓度时的脉冲曲线，可以看出当 AHL 浓度为 140 nmol/L 时，脉冲响应的峰值最高，当输入浓度超过 140 nmol/L 时，由于 CI 抑制的增强，脉冲响应的峰值反而下降。图 B 和 C 显示了 AHL 输入增加速率对脉冲响应影响的模拟和实验结果图。无论是模拟还是实验结果均表明 AHL 输入增加速率较低时会降低脉冲峰值和延迟脉冲峰值到来的时间。

图 4-25 为脉冲发生器的时空行为。图 4-25A 为信号发生器细胞和脉冲发生器细胞相距 2.5mm 时的时间序列状态，图 4-25B 为相距 3mm 时的时间序列状态。可以看出随着两种细胞距离变远，信号分子扩散到脉冲发生器细胞的时间变长，脉冲达到最高值的时间便会错后。由图 4-25 也可以看出，图 4-25A 脉冲达到最大峰值的时间为 32min，而图 4-25B 中脉冲达到最大峰值的时间却延迟至 36min。

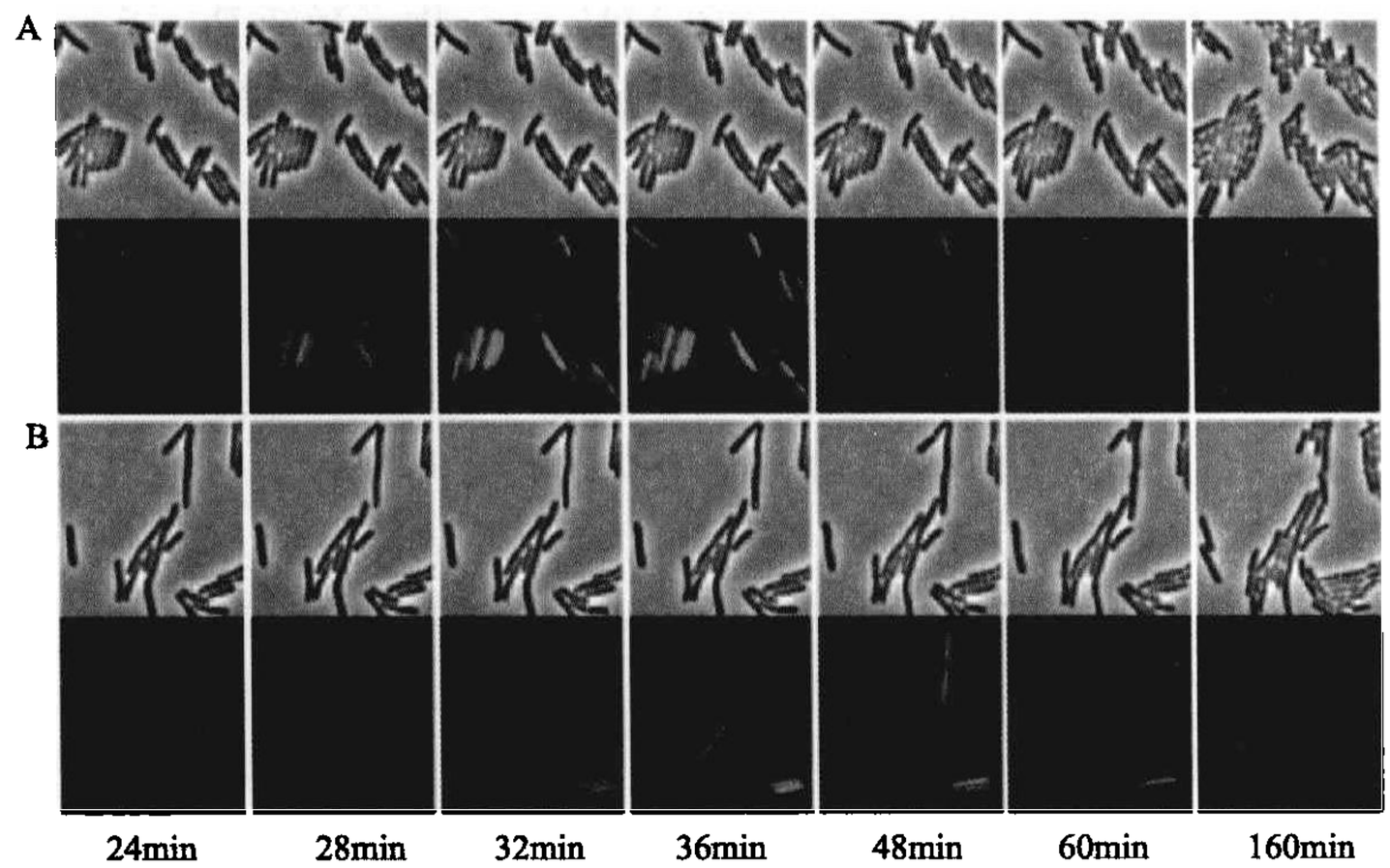


图 4-25 脉冲发生器细胞状态随时间变化图（M9 琼脂切片）[176]

图 A 为信号发生器细胞和脉冲发生器细胞相距 2.5mm 时的时间序列状态图，脉冲达到最大峰值的时间为 32min。图 B 为相距 3mm 时的时间序列状态图，脉冲达到最大峰值的时间为 36min。

第二部分 其他方面的合成生物学基础研究

合成生物学的目的是通过各种人工生物体的应用解决环境、能源、健康等方方面面的问题。因此，除了上述模拟逻辑功能的遗传线路设计以外，其他方面的合成生物学基础研究同样备受关注。

4.5 控制群体数量的基因线路[177]

如第二章所述，在许多微生物中存在着群体感应（quorum-sensing）现象，在生物体的生命活动中发挥了重要的作用。当这些生物形成的群体达到一定数量时，群体中细胞分泌的信号分子也会积累到一定浓度，此时，群体感应系统会对信号分子作出相应的响应，调节微生物群体行为。L. You 等正是基于上述机理，在大肠杆菌中构造了利用信号感应系统调控细胞凋亡来控制细胞群体浓度的基因线路。此研究将单个细胞基因线路的作用结果扩散至细胞群体中，通过以群体行为的统计学结果作为基因线路的表征，实现对细胞群体密度的人工调控。

细胞凋亡又称程序性细胞死亡（programmed cell death，PCD），是一种高度调节的生理性细胞死亡过程。与坏死不同，细胞凋亡是指细胞接受某种信号或受到某些因素刺激后，会发生主动的、由基因调控的消亡过程。由于细胞凋亡过程通常伴随有重要的生物现象，能有效调节细胞的生长、发育、癌变、免疫缺陷等过程，研究其行为具有积极的生物学意义，因此已成为当今生命科学研究的热点之一。

控制群体数量的基因线路和质粒构建如图 4-26 所示。整个系统由以下两部分组成。

（1）由 IPTG 控制的信号分子发生装置。IPTG 控制的 $P_{lac/ara-1}$启动子控制 *luxI* 基因和 *luxR* 基因的表达。*luxI* 表达的产物按一定速率催化合成信号分子 AHL 并迅速扩散出细胞外。

（2）由 P_{luxI}启动子控制的检测和程序性凋亡装置。当环境中的信号分子浓度达到或超过某一阈值后，*luxR* 表达的产物 LuxR 与信号分子结合启动启动子 P_{luxI}。P_{luxI}之后的 *lacZα-ccdB* 融合基因表达出 LacZ-CcdB 融合蛋白，其中的 LacZ 部分用于检验融合蛋白的表达量，而 CcdB 部分可使 DNA 解旋酶失活，引起细胞程序性死亡。

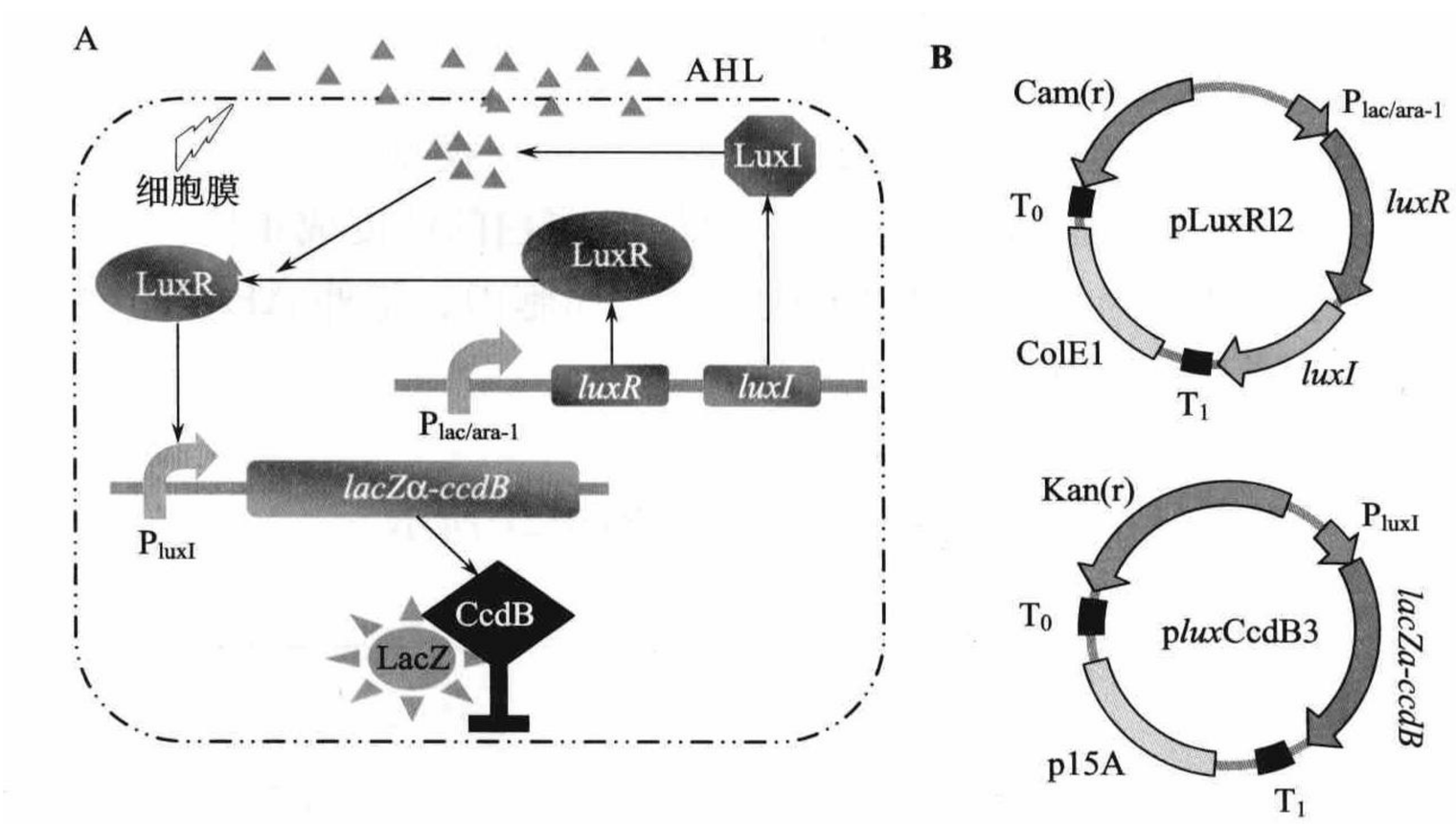


图 4-26　控制群体数量的基因线路

图 A 控制群体数量的基因线路图。由 IPTG 控制的 $P_{lac/ara-1}$启动子控制 *luxI* 基因和 *luxR* 基因的表达。*luxI* 表达的产物按一定速率催化合成信号分子 AHL 并迅速扩散出细胞外。当环境中的信号分子浓度达到或超过某一阈值后，*luxR* 表达的产物 LuxR 与信号分子结合启动启动子 P_{luxI}。P_{luxI}之后的 *lacZα-ccdB* 融合基因表达出 LacZ-CcdB 融合蛋白，其中的 LacZ 部分用于检验融合蛋白的表达量，而 CcdB 部分可使 DNA 解旋酶失活，引起细胞程序性死亡。图 B 为质粒结构图。质粒 pLuxR12 包含由 IPTG 控制的 $P_{lac/ara-1}$启动子、催化信号分子生成酶的编码基因 *luxI*、其产物接收环境中信号分子并与之结合控制下游基因表达的 *luxR* 基因。质粒 pluxCcdB3 包含有受启动子 P_{luxI}调控的 *lacZα-ccdB* 融合基因。

整个系统中，随着细胞浓度的增加，环境中的信号分子浓度也增加。而信号分子的浓度越高，*ccdB* 基因的表达量越多，凋亡细胞的数量就越多。当细胞的增长与凋亡达到一个动态平衡时，菌体浓度进入一个相对稳定的状态。

R. Weiss 等给出了这个基因线路的数学模型，见式（4-3）。

$$\begin{aligned}\frac{\mathrm{d}N}{\mathrm{d}t} &= kN(1-N/N_{\mathrm{m}})-dEN\\ \frac{\mathrm{d}E}{\mathrm{d}t} &= k_E A-d_E E\\ \frac{\mathrm{d}A}{\mathrm{d}t} &= \upsilon_A N-d_A A\end{aligned} \tag{4-3}$$

其中，N 为细胞密度；N_{m} 为此环境下可能达到的最大细胞密度；E 为 CcdB 蛋白的浓

度；A 为信号分子 AHL 的浓度；k 为细胞生长速率常数；d 为细胞死亡速率常数；k_E 为 CcdB 的生成速率常数；d_E 为 CcdB 的降解速率常数；υ_A 为 A 的生成速率常数；d_A 为 A 的降解速率常数。

由上式可知，N 的变化由两部分组成：第一项是细胞自然生长的 Logistic 模型，N_m 为自然生长条件下细胞的最大数量，k 为细胞增长速率；第二项是细胞凋亡的速率。CcdB 的变化也由两部分组成：第一项是信号分子促使凋亡基因合成的速率；第二项是凋亡基因自身的降解速率。A 的变化同样由两部分组成：第一项是细胞产生信号分子的速率；第二项是信号分子自身降解的速率。

这里用到了 5 个前提假设：

(1) 无 IPTG 诱导时，细胞密度的变化遵循 Logistic 模型。

(2) 有 IPTG 诱导时，细胞的死亡速率与 CcdB 蛋白的浓度成正比。

(3) CcdB 的生成速率与 AHL 的浓度成正比，而胞内、胞外 AHL 的浓度相等。

(4) AHL 的生成速率与细胞密度成正比。

(5) CcdB 和 AHL 的降解为一阶反应。

实验结果（pH 7 条件下）和模型模拟结果如图 4-27 所示。

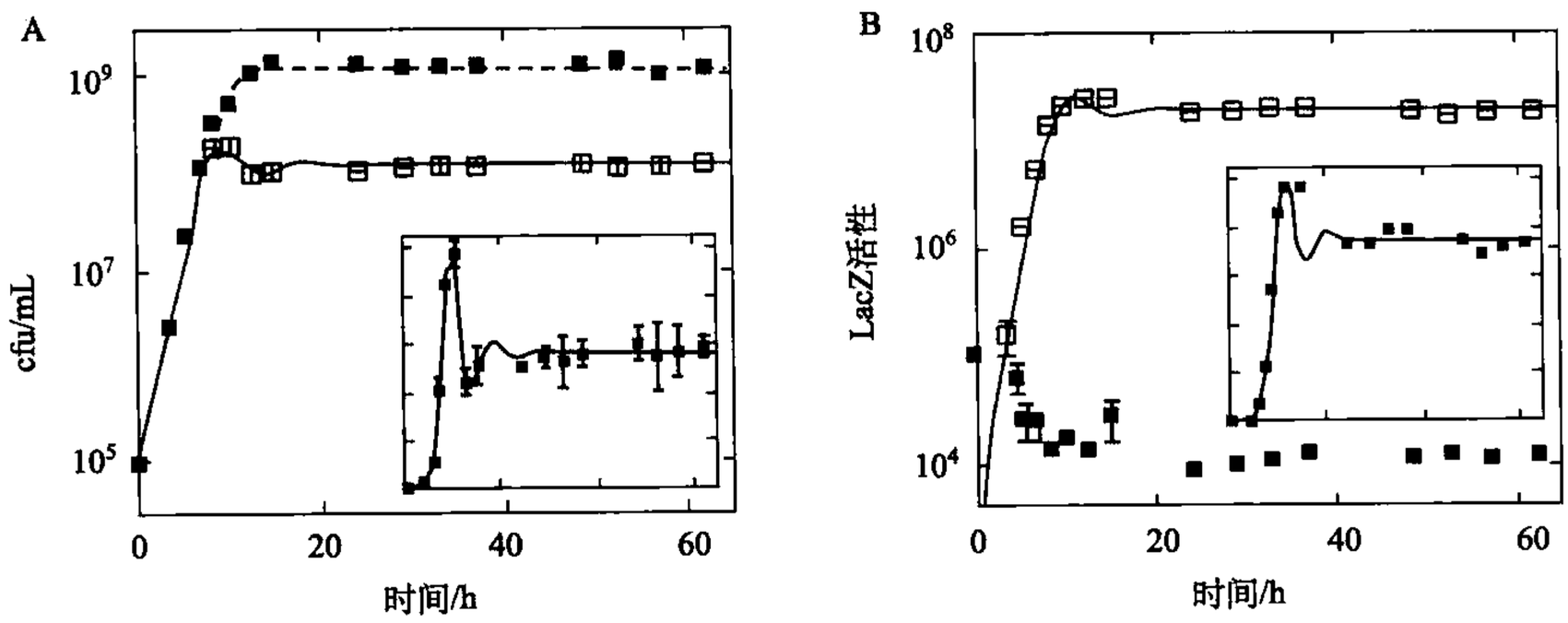


图 4-27 控制群体数量的基因线路的模拟与实验结果图[177]

图 A 中实心点代表没有加入 IPTG 下自然生长的细胞密度，空心点为 IPTG 诱导开启细胞凋亡线路时的细胞密度，蓝色的实线为模型模拟的结果。无 IPTG 诱导时，细胞自然增长并稳定于营养能够供给的最大值。有 IPTG 诱导时，细胞生长初期与自然生长情况相同，达到最大值后衰减振荡至新的稳定值，这与模型的预测完全吻合。细胞的最终稳定密度为自然生长状态下最大密度的 1/10。图 B 为 LacZ 检验的结果。自然生长的细胞中 LacZ 基本没有活性。与模型的预测一样，在 IPTG 诱导以后，LacZ 的活性先增加，衰减振荡后达到新的稳定值。

图 4-27A 中实心点代表没有加入 IPTG 时自然生长的细胞密度，空心点为 IPTG 诱导细胞开启凋亡基因线路时的细胞密度，蓝色的实线为模型模拟的结果。无 IPTG 诱导时，细胞自然生长，随着时间的增加，由于营养的限制逐渐达到稳定。而 IPTG 诱导下细胞增长初期与自然生长情况一样，当细胞达到一定密度以后，出现一个衰减振荡后达到新的稳定值，这与模型的预测完全吻合。细胞的最终稳定密度为自然生长状态下最大浓度的 1/10。由此可见，凋亡基因线路的转入很好地控制了细胞的密度。

图 4-27B 为 LacZ 检验的结果。自然生长的细胞中 LacZ 基本没有活性。可以看到，与模型的预测一样，在 IPTG 诱导以后，LacZ 的活性先增加，衰减振荡后达到新的稳定值。

改造过的大肠杆菌在凋亡基因的控制下，其细胞的最终浓度比自然生长浓度要小很多，因此模型中 N/N_m 的值明显小于 1，可将其忽略以简化模型。简化后的模型，细胞稳定状态下的各个量如式（4-4）：

$$
\begin{aligned}
N_s &= \frac{d_A d_E k}{v_A k_E d} \\
E_s &= k/d \\
A_s &= \frac{d_E k}{k_E d}
\end{aligned}
\tag{4-4}
$$

由上式可以看出，细胞的最终浓度与信号分子的降解速率成正比，而环境中 pH 的变化与信号分子降解速率的变化满足线性关系，可以方便地实现细胞密度的人工调控。同时，根据模型，系统稳定状态下 $1/d=E_s/(kN(1-N/N_m))$ 为定值，而凋亡基因的表达量可以用 LacZ 活性来表示。

4.6　人工细胞群体图案系统[178]

本章前面讲述的细胞群体脉冲发生系统利用的是遗传线路对信号浓度随时间变化的响应，即遗传线路的时间响应。而本节要讲述的细胞群体图案系统，则是利用遗传线路对信号浓度随距离变化的响应，即遗传线路的空间响应。读者在学习本节内容时，可以参照 4.4 节的内容，比较这两个实例之间的异同。

为了深入研究和模拟信号分子的介导机制，S. Basu 等设计了可以在中心细胞一定范围内形成发光带的人工细胞群体图案系统。该系统由两类细胞组成：发生器细胞和接收器细胞。由于信号分子的扩散作用，发生器产生的信号分子浓度自发生器细胞向周围成梯度下降分布，位于不同区域的接收器细胞对不同浓度的信号分子做出不同响应，表达不同颜色的荧光蛋白，从而形成不同颜色、不同形状的发光带。

图 4-28 为发光带系统遗传线路图。信号发生器细胞在接收到外界信号以后，开启 *luxI* 基因表达，催化信号分子 AHL 的生成。这种信号分子可以穿过细胞膜，扩散到环境中。由于扩散形成浓度梯度，故距离发生器越远的地方，AHL 信号分子的浓度越低。

信号接收器细胞表达 LuxR 蛋白，这种蛋白质能与 AHL 信号分子结合，开启下游 λ 噬菌体 *cI* 基因以及修饰过的乳糖操纵子 *lacIM1* 基因的表达。当接收器距离发生器细胞过近时，由于 AHL 信号分子浓度较高，导致 CI 蛋白和 LacIM1 蛋白过量表达，LacIM1 蛋白抑制荧光蛋白的启动子，此区域内无荧光蛋白产生。相反，当接收器离发生器过远时，由于信号分子过少，CI 蛋白和 LacIM1 蛋白都只有极少量表达，无法抑制下游基因，导致野生型的 *lacI* 基因顺利表达，其产生的 LacI 蛋白抑制荧光蛋白的表达，此区域内仍无荧光蛋白产生。只有当接收器离发生器距离适当的时候，适量的信号分子浓度引起适量的 CI 蛋白和 LacIM1 蛋白的表达，由于 CI 对于下游启动子的抑制作用明显比 LacIM1 要强烈，所以野生型 *lacI* 基因基本不能表达，同时 *lacIM1* 基因表达

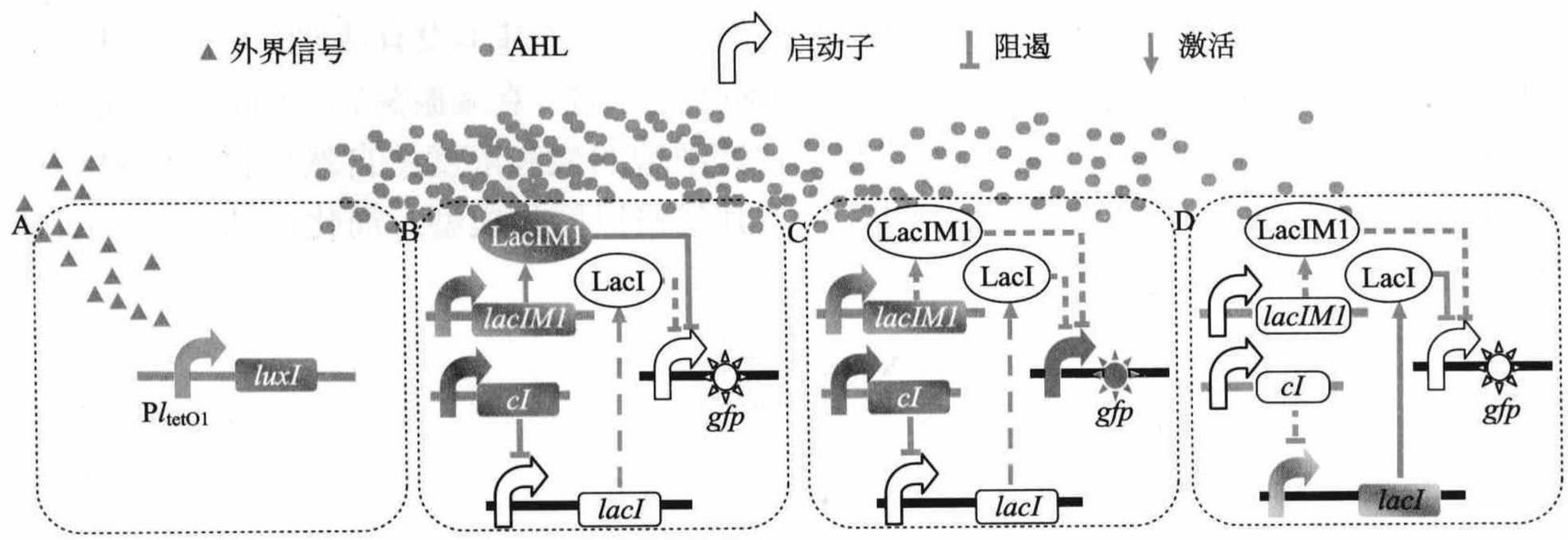


图 4-28 人工细胞群体图案系统遗传线路图

A. 信号发生细胞；B. 高浓度区域接收细胞的响应；C. 中浓度区域接收细胞的响应；D. 低浓度区域接收细胞的响应

图 A 信号发生器细胞在接收到外界信号（绿色三角形）以后，开启 *luxI* 基因表达，催化信号分子 AHL（蓝色实心圆）的生成。AHL 能够穿越细胞膜，扩散到环境中，其浓度以发生器细胞向周围成梯度下降分布。图 B 当接收器距离发生器细胞过近时，过高浓度的 AHL 过表达 CI 蛋白和 LacIM1 蛋白，LacIM1 蛋白抑制荧光蛋白的启动子，此区域内无荧光蛋白产生。图 C 当接收器离发生器距离适当的时候，适量的 AHL 引起适量的 CI 蛋白和 LacIM1 蛋白的表达。CI 对下游启动子的抑制作用强使野生型 *lacI* 基因基本不表达，而 *lacIMl* 基因表达的量不足以抑制下游荧光蛋白的启动子，荧光蛋白顺利表达，并在这个特定区域内形成环状荧光带。图 D 当接收器距发生器过远时，过少的 AHL 无法开启足量的 CI 蛋白和 LacIM1 蛋白，因此无法抑制下游基因，导致野生型的 *lacI* 基因顺利表达，其产生的 LacI 蛋白抑制荧光蛋白的表达，此区域内也无荧光蛋白产生。

的量也不足以抑制下游荧光蛋白的表达，因此会在这个特定区域内形成环状荧光带。

从上面的分析可以看出，浓度过高时 *lacIM1* 基因表达，产物抑制荧光蛋白的产生，浓度过低时 *lacI* 基因表达，产物抑制荧光蛋白的产生。只有当信号分子浓度低于某一上限高于某一下限值时荧光蛋白才会表达。因此，要有一个浓度上限控制装置控制细胞响应浓度的上限，以及一个浓度下限控制装置控制细胞响应浓度的下限。由此研究者构建了相应的高浓度反应质粒（包含 *lacIM1* 基因的 pHD 质粒）和低浓度反应质粒（包含 *lacI* 基因的 pLD 质粒），如图 4-29 所示。为了使图案更具多样性，研究者开发了

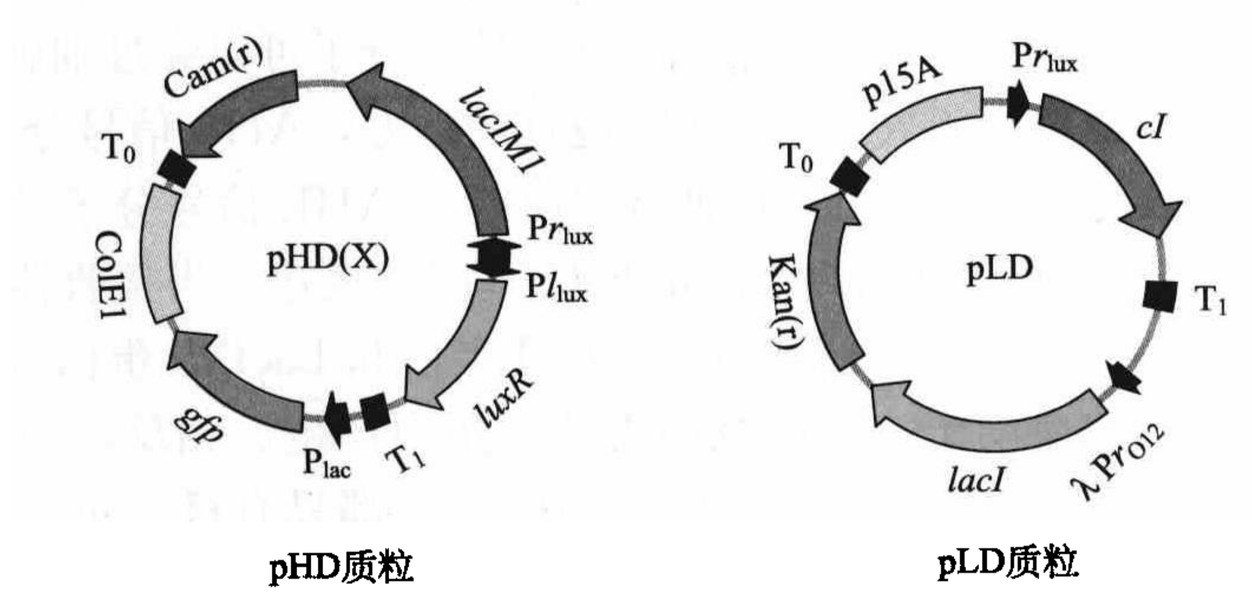


图 4-29 响应 AHL 信号的质粒遗传图[178]

浓度上限控制装置控制细胞能够响应的 AHL 浓度上限，命名为 pHD 质粒（包含 *lacIM1* 基因）；浓度下限控制装置控制细胞能够响应的 AHL 浓度下限，命名为 pLD 质粒（包含 *lacI* 基因）。

3 种浓度上限控制装置。

（1）HD1：将 *luxR* 基因进行突变改造，改造后的 *luxR* 基因产生的 LuxR 蛋白敏感性较高，能感知相对较低浓度的 AHL 分子。

（2）HD2：野生型 *luxR* 基因，感知中等浓度的 AHL 分子。

（3）HD3：将 *luxR* 基因转入 *E. coli* 中的低拷贝数质粒，感知相对较低浓度的 AHL 分子。

图 4-30A 为利用数学模型模拟的、含有浓度上限控制装置的接收器细胞在不同 AHL 浓度作用下荧光蛋白的表达量。从图中可以看出，三类细胞能够感知的 AHL 浓度上限的关系为：HD1 < HD2 <HD3。图 4-30B 图为实验检测结果图，可以看出，虽然在数量上有所区别，但数学模型对三类细胞趋势的预测相当准确。

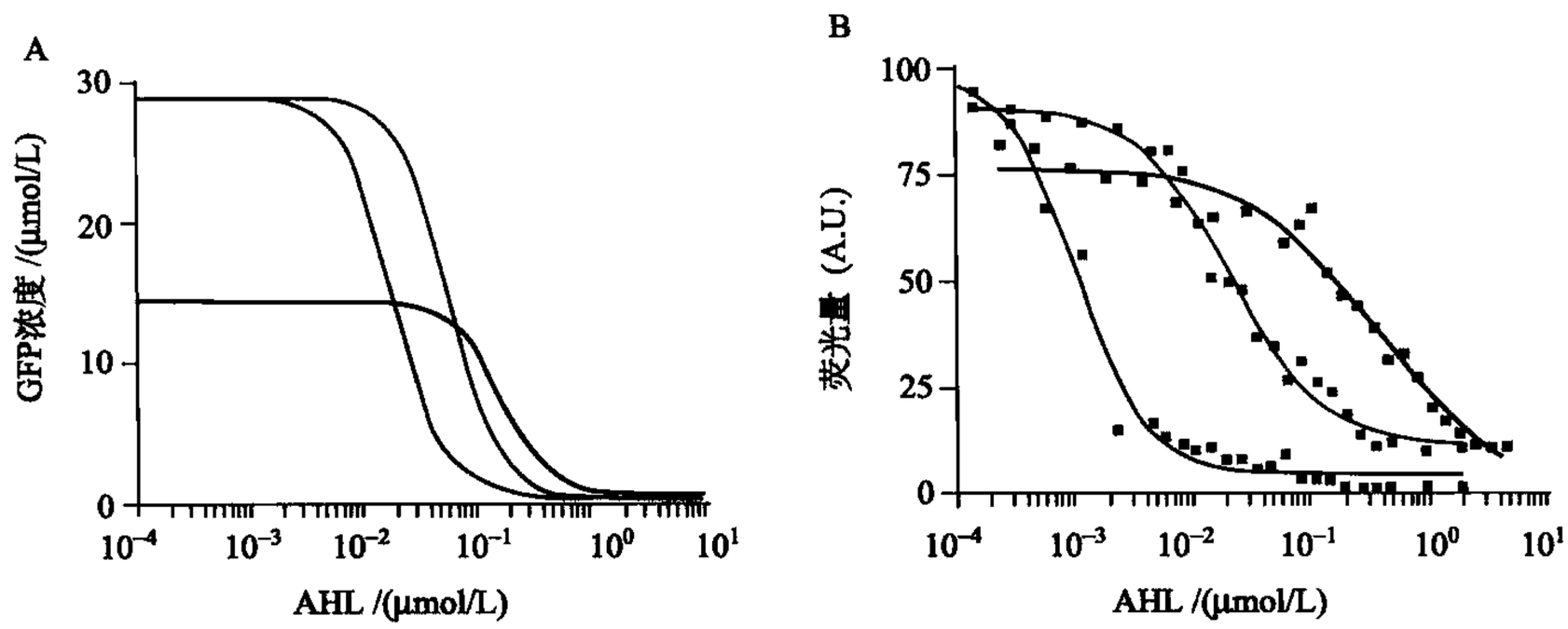


图 4-30　信号接收细胞荧光蛋白在不同 AHL 浓度作用下表达量的仿真和实验图[178]

A. 模型仿真图；B. 实验检测图

（蓝色-HD1 细胞；红色-HD2 细胞；黑色-HD3 细胞）

将这三种浓度上限控制装置与浓度下限控制装置进行组合，研究者共构建了三种不同的信号接收细胞，将 HD1 感应器和浓度下限控制装置组成的接收细胞称为 BD1，相应于 HD2 的为 BD2 细胞，相应于 HD3 的为 BD3 细胞。图 4-31A、图 4-31B 分别为这三种细胞的性能模拟和实验结果图。可以看出，环形条带之间会有小部分重合，与模型结果吻合。

三种不同的信号接收细胞会在不同距离范围内形成发光的条带。当荧光蛋白出现峰值时，三种细胞响应的 AHL 浓度大小关系为：BD1 < BD2 <BD3，相应的，其距中心发生细胞的距离远近关系为：BD3 < BD2<BD1。

如果将 3 种细胞中的荧光蛋白颜色进行调整并按设定顺序涂抹在琼脂固体培养基上，可得图 4-32 所示实验结果。其中，图 4-32A 为可见光下的信号发生细胞图；图 4-32B 为BD2 接红色荧光蛋白、BD3 接绿色荧光蛋白时在荧光显微镜观察下的放大图；图 4-32C 为 BD2 接红色荧光蛋白、BD1 接绿色荧光蛋白时在荧光显微镜观察下的放大图。

如果在多块区域同时放置信号发生细胞，就可以形成如图 4-33 中多种不同形状的颜色区域。图 4-33A 为有两个信号发生器区域的实验结果图；图 4-33B 为两个信号发生

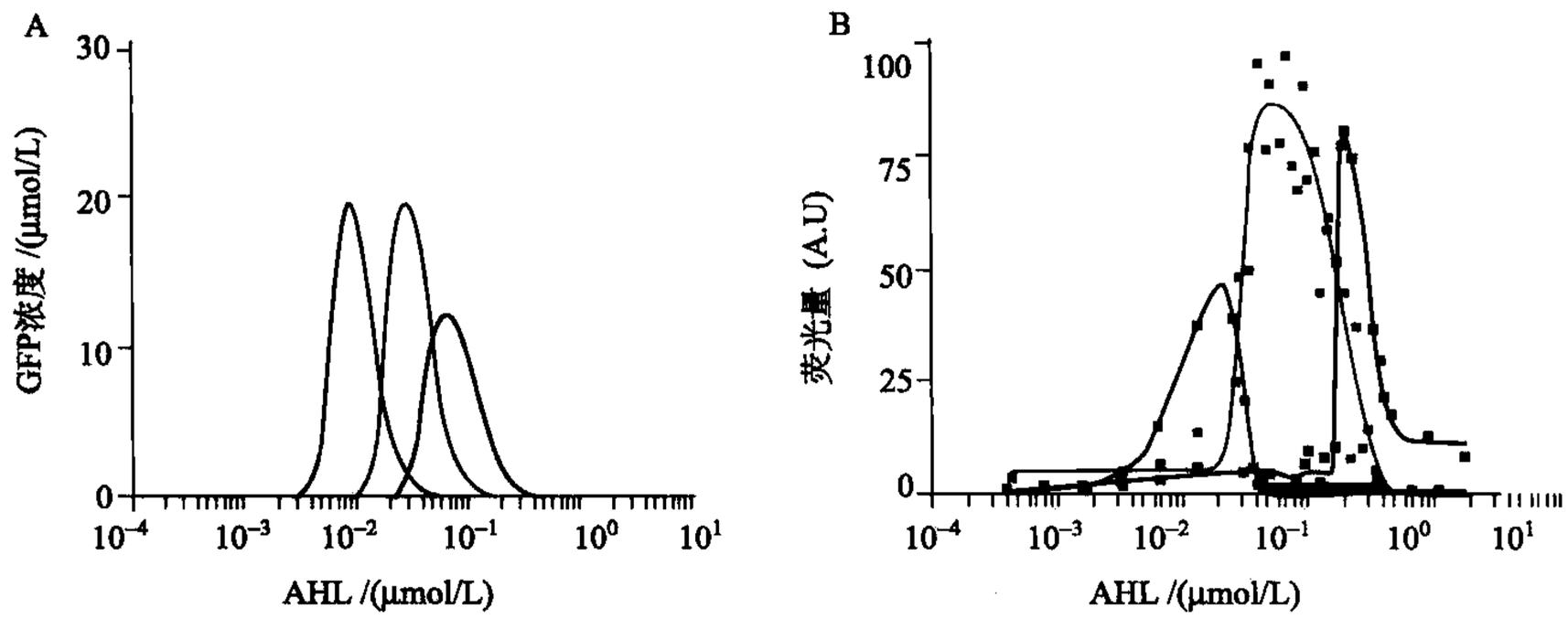


图 4-31 三类接收细胞模拟及实验结果[178]

A. 模型仿真图；B. 实验检测图

（红色-BD2 细胞；蓝色-BD1 细胞；黑色-BD3 细胞）

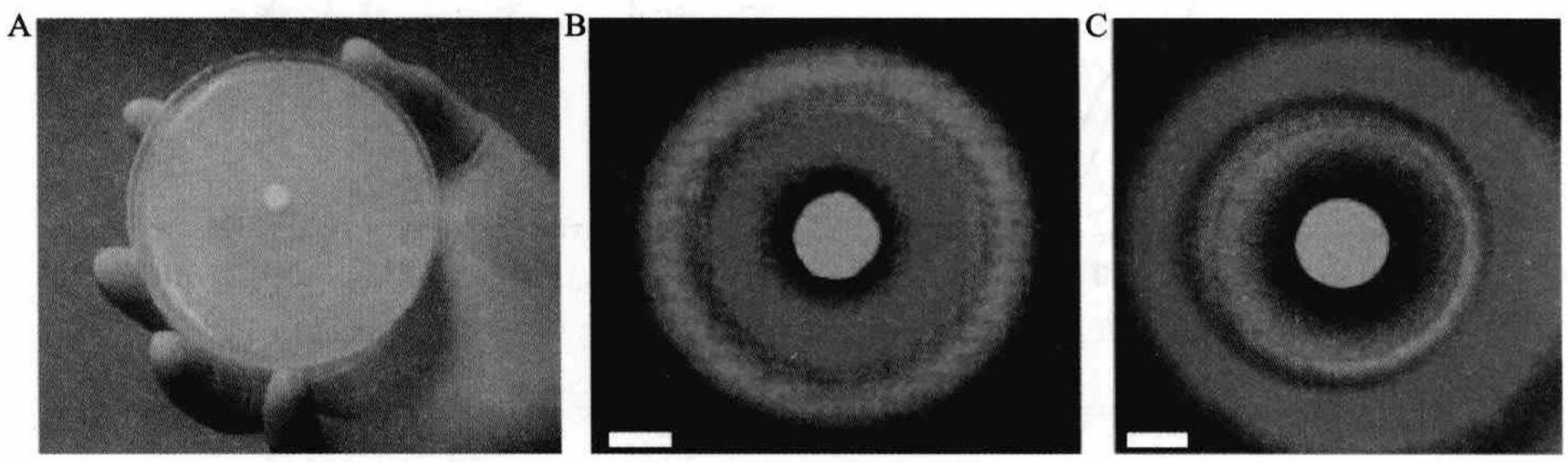


图 4-32 发光系统实验结果图[178]

图 A 为可见光下的信号发生细胞培养皿图；图 B 为 BD2 接红色荧光蛋白、BD3 接绿色荧光蛋白时在荧光显微镜观察下的放大图；图 C 为 BD2 接红色荧光蛋白、BD1 接绿色荧光蛋白时在荧光显微镜观察下的放大图。环形条带之间会有小部分重合，与模型结果吻合。

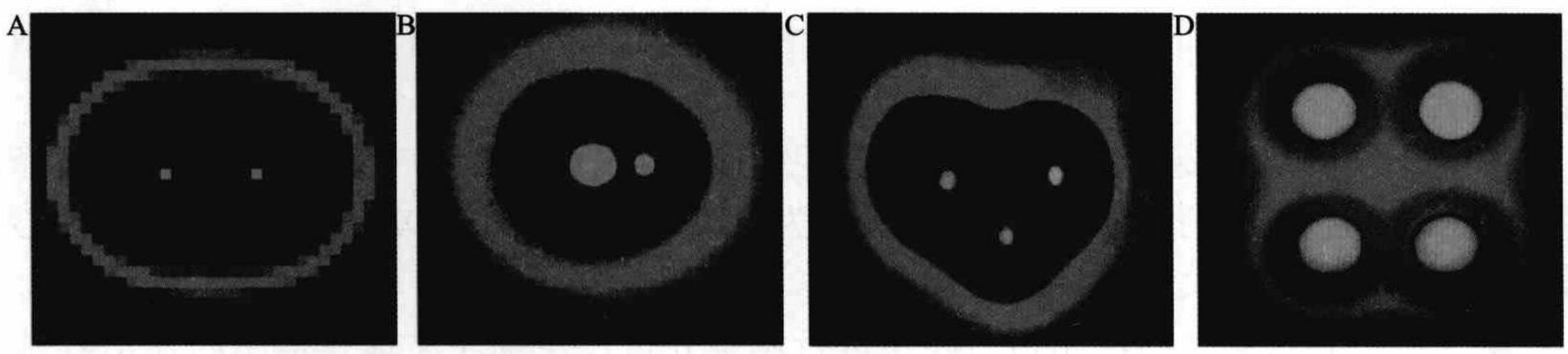


图 4-33 发光系统具有多个信号发生区域时的实验结果图

器区域的浓度不一致时的实验结果图，左侧浓度高的信号发生器区域相应的环状荧光带的半径要大些；图 4-33C 为有三个信号发生器区域的实验结果图；图 4-33D 为有 4 个高浓度信号发生器区域的实验结果图。

4.7　人工转录级联线路[179]

为了研究级联线路的超敏感性和噪声传播等特性，S. Hooshangi 等构建了人工三阶转录级联线路。图 4-34 中的 A、B、C 子图分别表示一阶、二阶和三阶级联线路的基因调控情况。

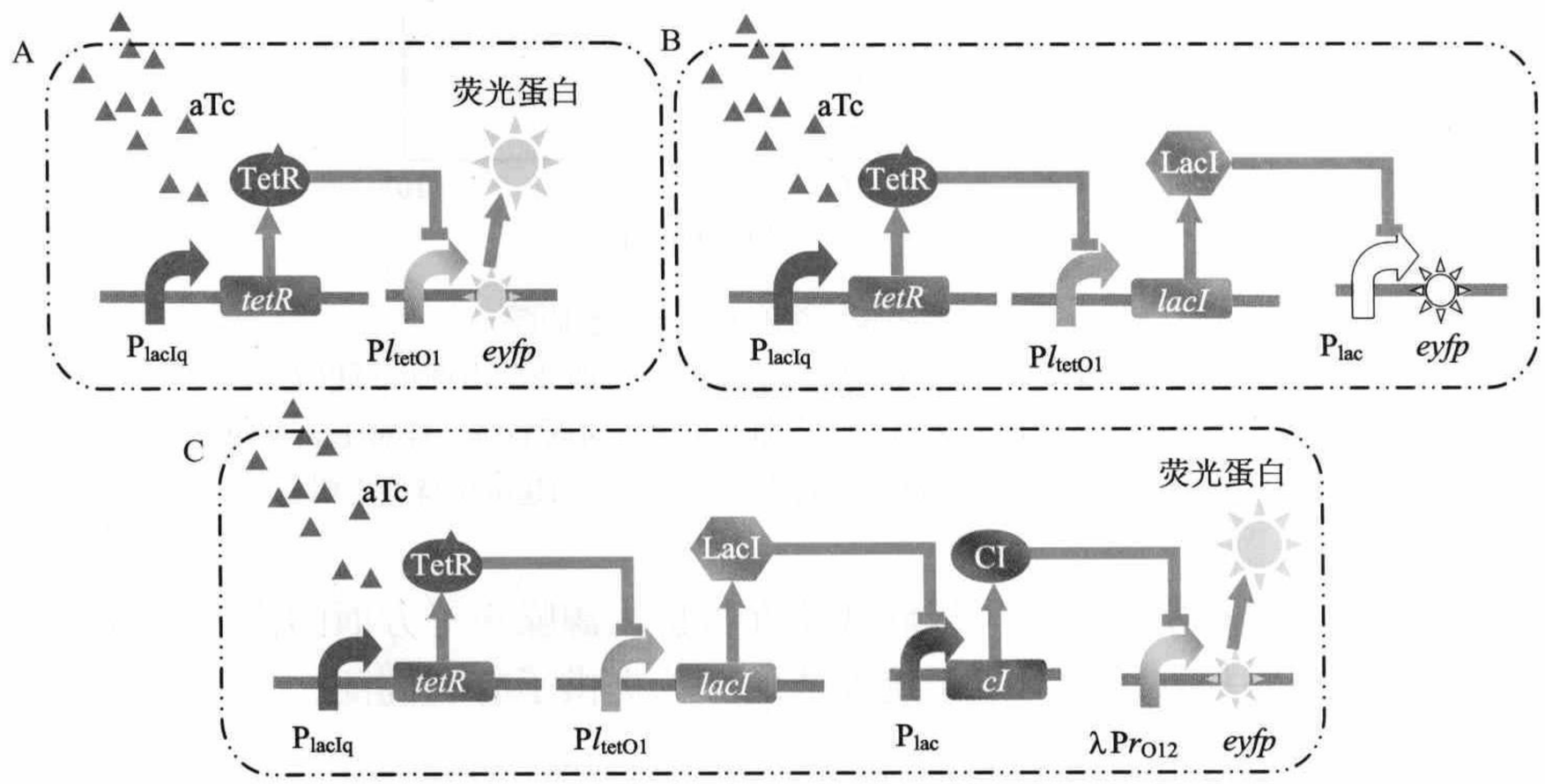


图 4-34　转录级联线路设计

在所有线路中，P_{lacIq}启动子控制 TetR 的表达，aTc 能够顺利扩散进入细胞，与 TetR 结合抑制 Pl_{tetO1}启动子的启动。图 A 为一阶级联线路，由 TetR 阻遏蛋白控制 *eyfp* 基因的表达。图 B 为二阶级联线路图，LacI 蛋白控制 *eyfp* 基因的表达。而 LacI 蛋白的表达与否则由 TetR 阻遏蛋白控制。图 C 为三阶级联线路图，*eyfp* 基因的表达受 CI 阻遏蛋白控制，*cI* 基因的表达受 LacI 蛋白的控制，而 LacI 蛋白的表达与否则由 TetR 阻遏蛋白控制。

整个系统的逻辑关系为：

（1）当系统仅由一阶级联线路组成时，aTc 信号的加入会与 TetR 蛋白结合而解除其对 Pl_{tetO1}启动子的抑制，荧光蛋白得以表达。

（2）当系统由二阶级联线路组成时，aTc 信号的加入解除了对 Pl_{tetO1}启动子的抑制，启动 *lacI* 基因表达产生 LacI 蛋白抑制 P_{lac}启动子，荧光蛋白无法表达。

（3）当系统由三阶级联线路级联组成时，aTc 信号的加入会解除对 Pl_{tetO1}启动子的抑制，启动 *lacI* 基因表达产生 LacI 蛋白使得 P_{lac}启动子被抑制，*cI* 基因无法表达，CI 蛋白无法产生，从而使被 CI 蛋白抑制的 λPr_{O12}基因得以启动，荧光蛋白表达。

由上面的分析可以看出，一阶级联线路的功能相当于基因开关中的转换开关，aTc 信号控制荧光蛋白的表达；二阶级联线路相当于在其后串联一个转换开关；三阶级联线路相当于串联两个转换开关。一至三阶级联线路对于不同 aTc 输入浓度的响应曲线如图 4-35 所示。从图中可以看出，三阶级联线路比一阶级联线路的响应曲线更接近阶跃响应，且对于 aTc 输入浓度的响应范围比一阶级联线路要窄，由此可以得出：级联线路的阶数越高，响应曲线的超敏感性越明显。

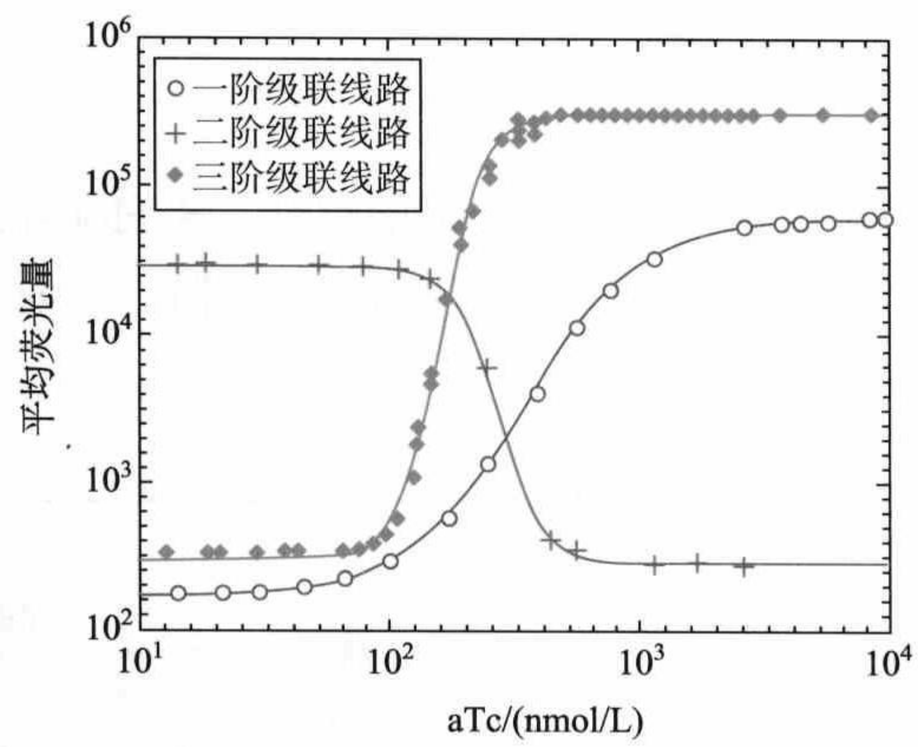


图 4-35 级联线路的 aTc 输入响应图

一至三阶级联线路对于不同 aTc 输入浓度的响应曲线。由图中可以看出，三阶级联线路比一阶级联线路的响应曲线更接近阶跃响应，且对于 aTc 输入浓度的响应范围比一阶级联线路要窄，响应曲线的超敏感特性更明显。

此外，作者还通过实验揭示了级联线路在抵抗内源噪声等方面的优势，为进一步从机理上研究级联线路拓扑结构对于系统性能的影响提供了有利佐证。

4.8 工程化的支架蛋白

前述各种合成生物系统都是在转录或者翻译水平上通过对基因的调控实现特定的功能。除此之外，构建拥有复杂连接关系的蛋白质相互作用系统是合成生物学又一个崭新的研究方向。美国加州大学旧金山分校（UCSF）的 W. A. Lim 小组利用工程化的合成支架蛋白作为空间上区域化的信号节点，将基于支架蛋白的反应整合在一起，得到完全依赖于蛋白质相互作用的、可重新配置的人工信号通路。由于蛋白质之间的相互作用不需要转录、翻译等步骤，具有更快速的信息处理能力，且耗费的细胞资源更少，因此此项研究具有广泛的应用前景。下面我们就来详细介绍 W. A. Lim 小组的研究内容。

支架蛋白（scaffolding protein）号称分子胶水，是信号转导系统中存在的一种本身不具备酶活性的蛋白质。支架蛋白在调节信号转导中起着重要作用，通常含有多个蛋白结合域，能同时结合多个蛋白质。细胞中，构成通路的信号蛋白分子通常由支架蛋白物理地组装成联合体，通过支架来定向信息流，提高匹配蛋白的信号传递，同时通过辅助蛋白激酶和磷酸酶与它们各自的底物结合而起作用，保证信号传递的特异性和高效性。支架蛋白还能起到生理性隔离的作用，阻止不匹配蛋白间的相互交流，防止与其他通路发生相互作用。同时，支架联合体作为反馈回路的集散中心承担着调节蛋白质活性和实现蛋白质重新组装的作用，对信号通路动态响应行为进行调节。因此，支架蛋白对于通路响应的动态行为至关重要。

在各种信号转导通路中，有丝分裂原激活蛋白激酶（mitogen-activated protein kinase，MAPK）信号转导通路不仅广泛存在，且是目前研究比较透彻的一种。因此

W. A. Lim 研究小组的 C. J. Bashor 等采用了酿酒酵母 α 细胞型的单倍体细胞交配的 MAPK 通路作为背景系统。

图 4-36 所示为酿酒酵母交配通路。酿酒酵母单倍体细胞能够与相反交配型的单倍体细胞发生交配。交配时酿酒酵母放弃原有出芽位点，根据信息素的浓度梯度，重新选择生长位点，向相反交配型细胞伸出突起进行极性生长。在性信息素 α 因子的诱导下，交配因子受体指导选择交配突起的位点。通过 G 蛋白激活 Ste20 蛋白将信号经由 Ste11（MAP kinase kinase kinase，MAPKKK）、Ste7（MAP kinase kinase，MAPKK）、Fus3（MAP kinase，MAPK）组成的 MAPK 模块传递到 Far1 和 Ste12 等因子，调控相关基因的转录，抑制原有的出芽位点，选择新的生长位点，并使细胞周期停止在 G_1 期。G 蛋白与 Cdc24、Cdc42 和 Bem1 等蛋白质作用，使肌协蛋白细胞骨架在交配突起处聚集，呈极性化分布，细胞发生极性生长[180]。

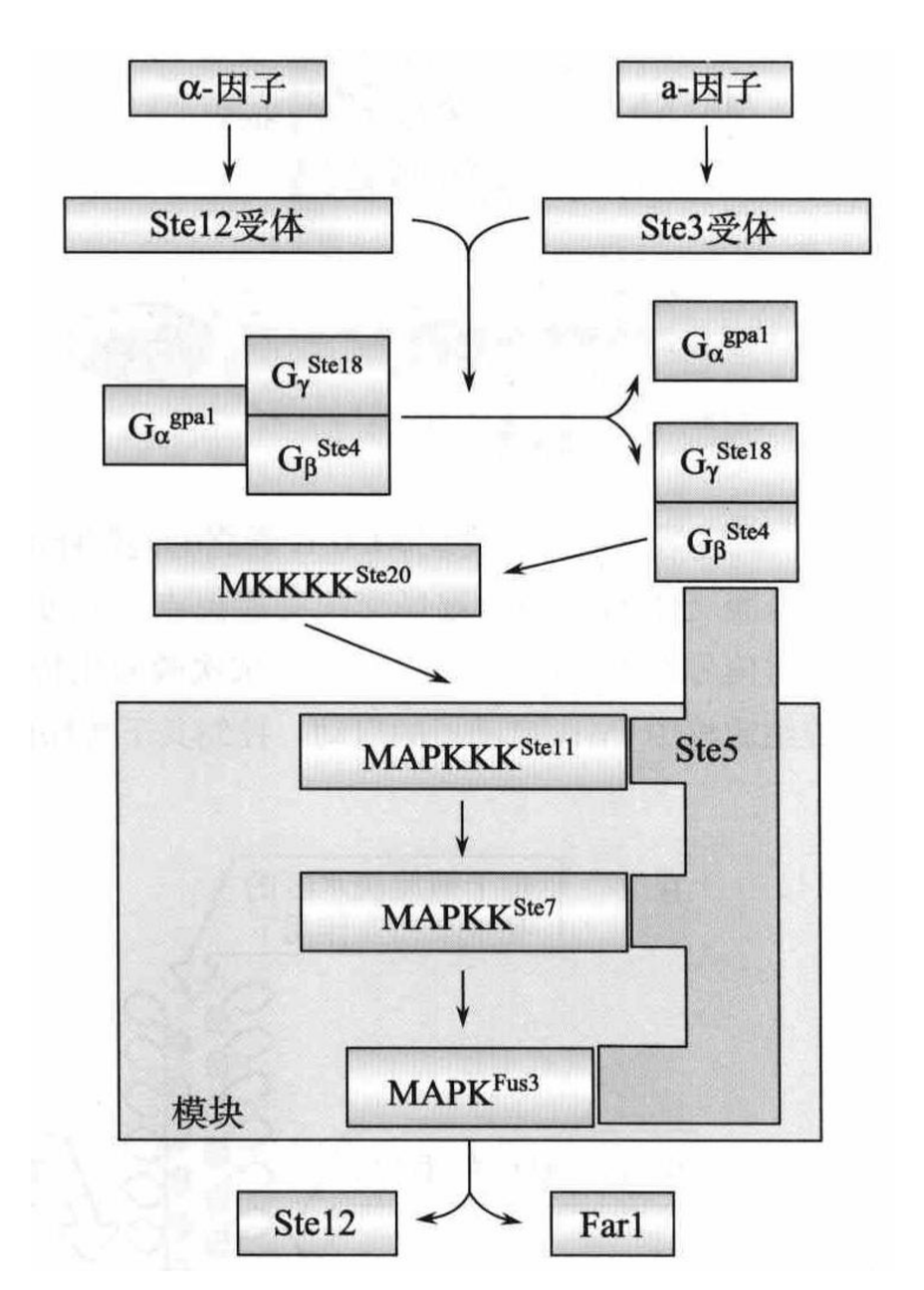


图 4-36　酿酒酵母交配通路示意图[180]

在性信息素 α 因子的诱导下，交配因子受体指导选择交配突起的位点。通过 G 蛋白激活 Ste20 蛋白将信号经由 Ste11（MAP kinase kinase kinase，MAPKKK）、Ste7（MAP kinase kinase，MAPKK）、Fus3（MAP kinase，MAPK）组成的 MAPK 模块传递到 Far1 和 Ste12 等因子，调控相关基因的转录，抑制原有的出芽位点，选择新的生长位点，并使细胞周期停止在 G_1 期。

由分析可知，这一通路的连通性主要取决于支架蛋白 Ste5，通过 Ste5 将三种核心激酶 Ste11、Ste7 和 Fus3 结合在一起，实现彼此的磷酸化和激活，对 Ste5 支架蛋白的改造可能会得到不同的通路输出。除了改造 Ste5 支架蛋白，C. J. Bashor 等还在通路中加入了能够产生荧光蛋白的基因模块以方便检测，如图 4-37 所示。为了方便读者理解，图中略去其他细节，只突出显示了与 C. J. Bashor 等研究密切相关的部分[24]。

为了实现对支架蛋白的调节，C. J. Bashor 在 Ste5 支架蛋白的 C 端融合了一个亮氨酸拉链结构的新蛋白募集位点，称为 Ste5-zipper；在调控蛋白（modulator）上融合与其互补的另一段亮氨酸拉链结构，亮氨酸拉链的疏水作用能够使调控蛋白和支架蛋白互补配对的部分形成异源二聚体，实现两者的结合，由此即可利用调控蛋白调节通路的输出，如图 4-38 所示。

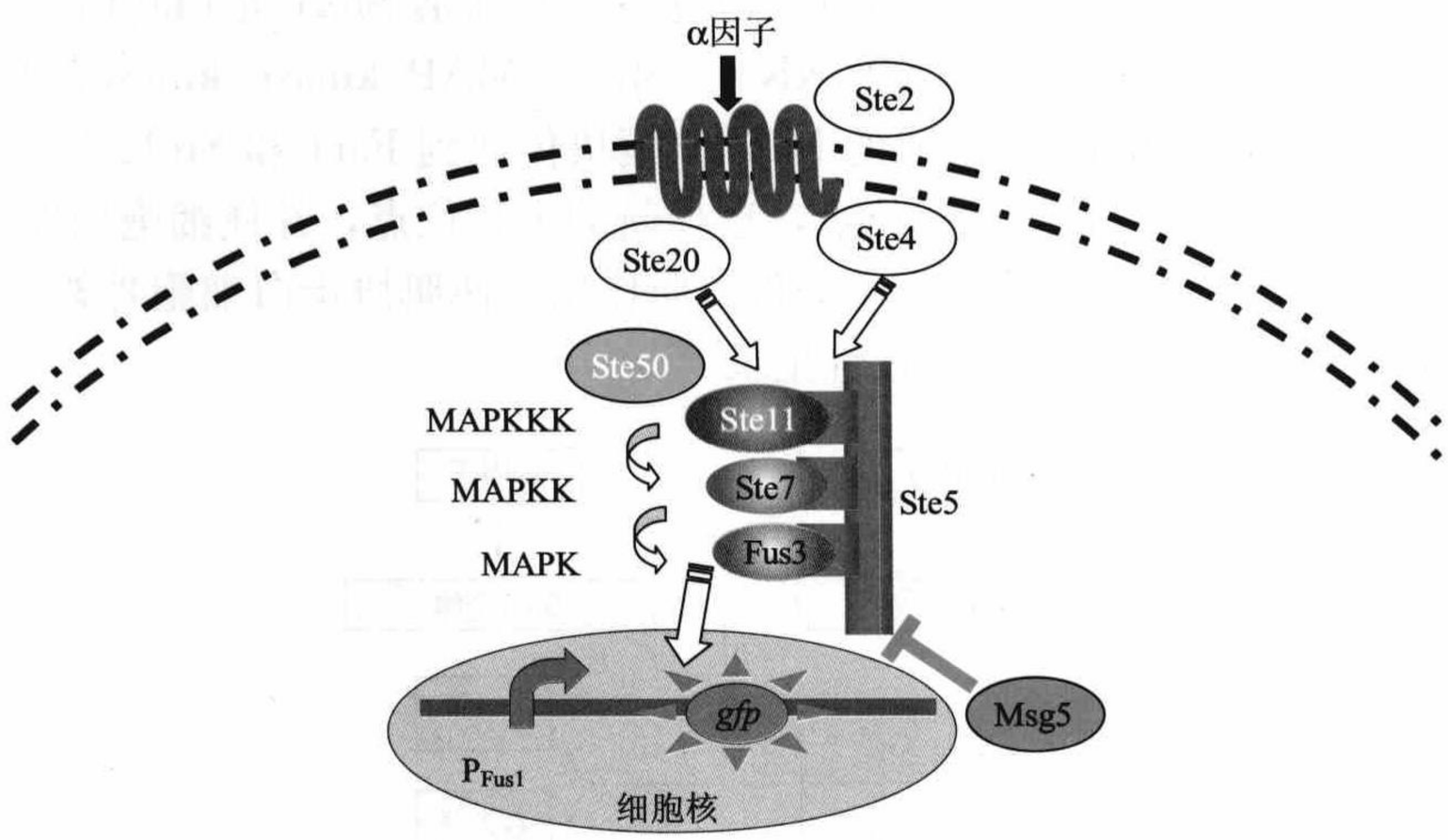


图 4-37　调控蛋白对酵母交配 MAPK 通路的调控作用

在性信息素 α 因子的诱导下，跨膜受体 Ste2 指导选择交配突起位点。通过 G 蛋白激活 Ste20 蛋白（其中 Ste4 为 G 蛋白 β 亚基）将信号经由 Ste11、Ste7、Fus3 依次磷酸化传递到目的因子，目的因子进入核区作为转录因子调控细胞核中 P_{FUS1} 启动子的启动，控制其下游绿色荧光蛋白的表达。

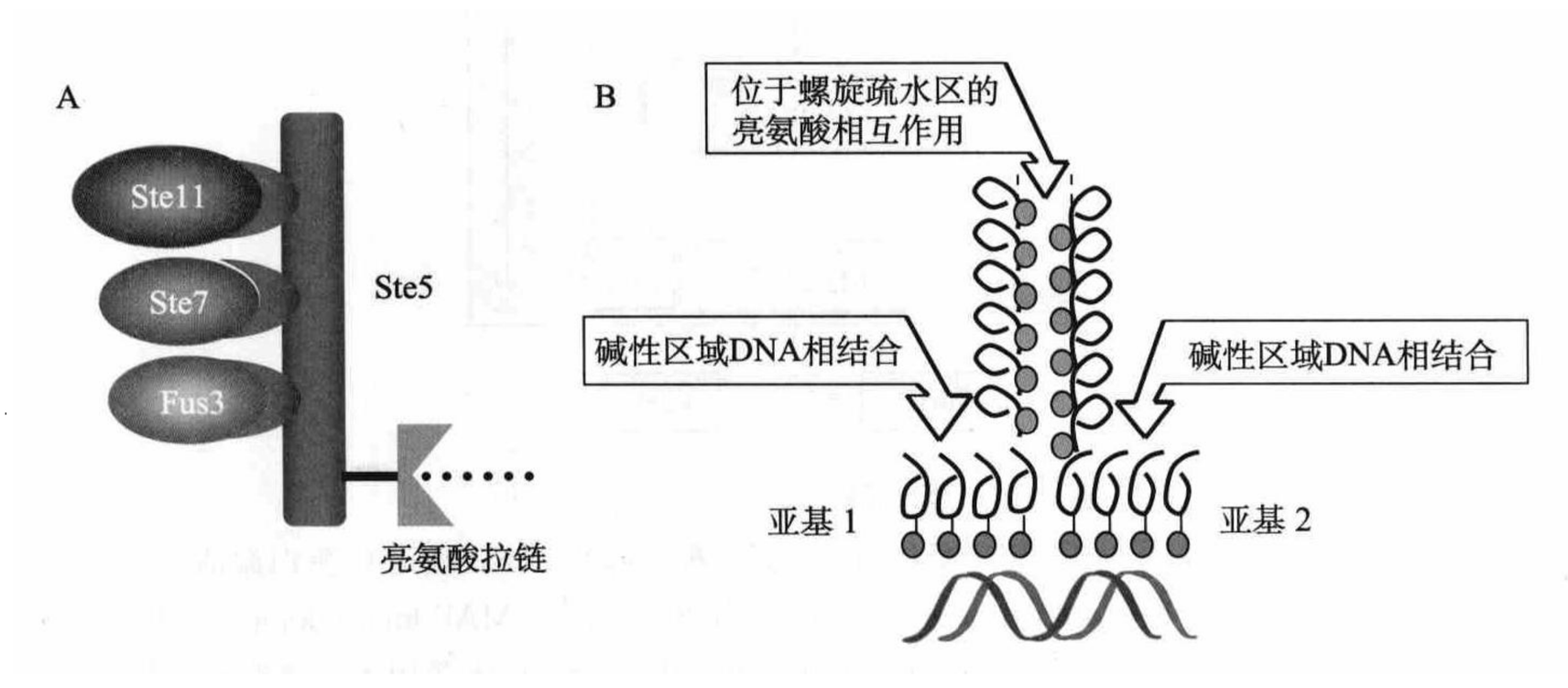


图 4-38　Ste5-Zipper 与亮氨酸拉链激活蛋白域调控区 DNA 相结合的示意图

A. 融合了新募集位点的 Ste5-Zipper；B. 亮氨酸拉链激活蛋白域调控区 DNA 相结合的示意图

亮氨酸拉链的疏水作用能够使调控蛋白和支架蛋白互补配对的部分形成异源二聚体，实现两者的结合，由此即可利用调控蛋白调节通路的输出。

C. J. Bashor 等共构建了两种调控蛋白：Ste50 和 Msg5，如图 4-39 所示。Ste50 是一个正调控蛋白，作为适配子促进 MAPKKK Ste11 与其上游激活因子 Ste20 的相互作

用，提高激活通路的稳态输出。Meg5 是一个负调控蛋白，作为 MAPK 通路中的磷脂酶使磷酸化的 Fus3 失活而降低通路输出。Meg5 的作用非常强烈，甚至可以使通路对于任何输入激励均无响应。当通过亮氨酸拉链相互作用人工重新募集到 Ste5 支架上时，Ste50 和 Msg5 即可对通路输出产生强烈的调节作用。

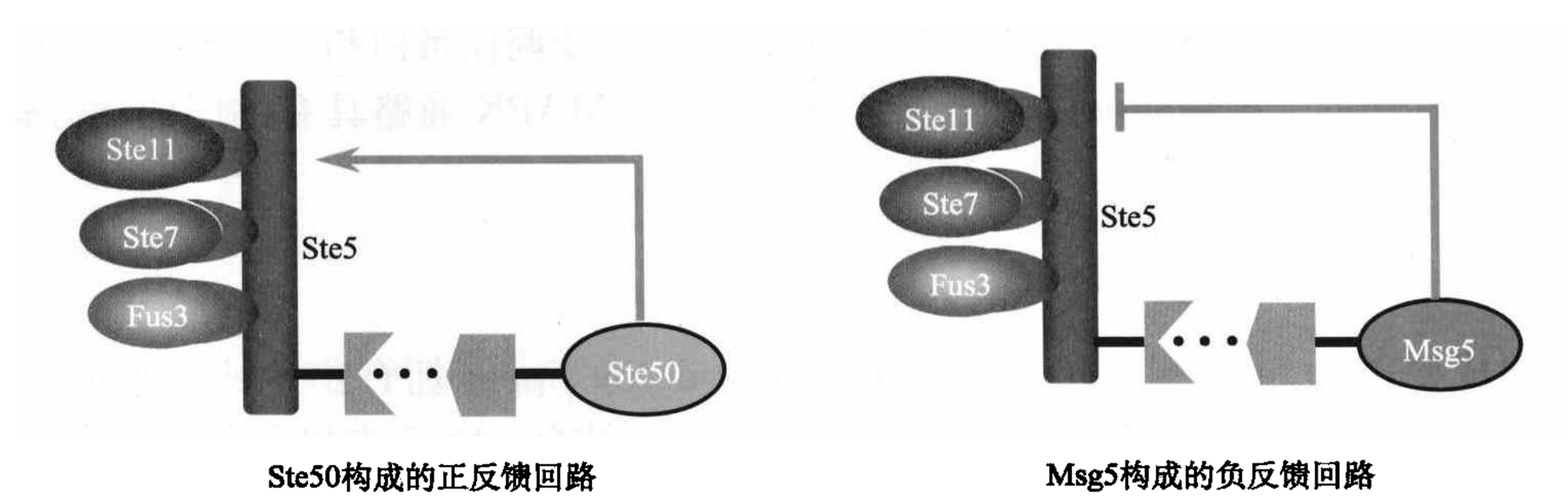


图 4-39　人工构建的调控蛋白及其逻辑控制线路

正调控蛋白 Ste50 作为适配子促进 MAPKKKSte11 与其上游激活因子 Ste20 的相互作用，提高激活通路的稳态输出。负调控蛋白 Meg5 作为 MAPK 通路中的磷脂酶使磷酸化的 Fus3 失活而降低通路输出。

为了实现对合成反馈回路的人工调控，研究者引入依赖于交配信息的启动子 P_{FIG1} 来控制调控蛋白的表达。图 4-40 所示为检验调控蛋白作用的实验结果图。从图中可以看出，结合了 Ste50 而形成的正反馈回路能够提高通路的稳态输出并使输出更加类似于阶跃响应，其希尔系数较之野生型有显著提高（希尔系数从 1.12 上升到 2.42）。而由 Msg5 构成的负反馈回路中，最初的响应与野生型类似，在 35min 时，其输出量达到峰值，其后输出量下降，动力学特性较之野生型没有明显改变，希尔系数上升不大。此外，还可以看出负反馈回路具有一定的自衡作用，35min 后，即使激励持续存在，输出仍持续降低。自衡作用对于系统的自我平衡至关重要，能够限定输出的幅值和持续时间，防止有害作用产生或者耗费太多的代谢能量。

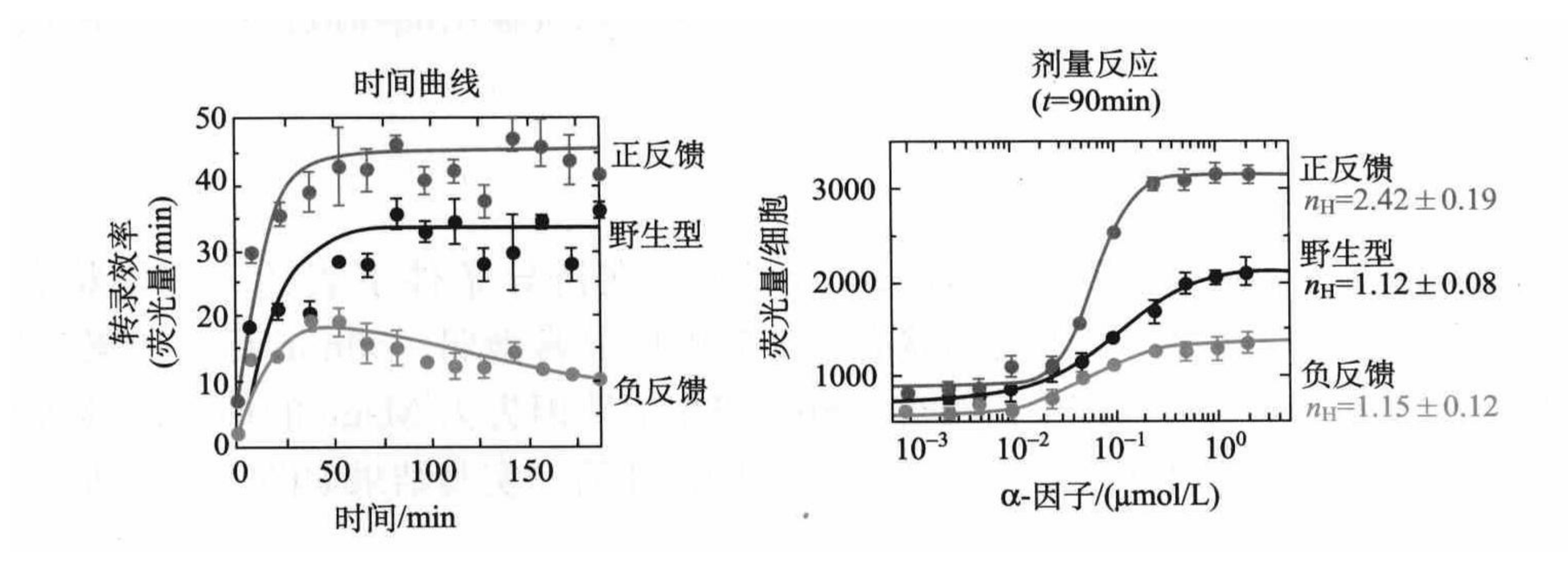


图 4-40　调控蛋白调控能力检验图[24]

为了实现对合成反馈回路的人工调控，引入了依赖于交配的启动子 P_{FIG1} 控制调控蛋白的表达。结合了 Ste50 而形成的正反馈回路能够提高通路的稳态输出，使输出响应曲线的希尔系数从 1.12 上升到 2.42，更加类似于阶跃响应。由 Msg5 构成的负反馈回路具有一定的自衡作用，细胞初始响应类似于野生型，35min 时输出达到峰值，之后即使激励持续输出仍然降低。

为了能够人工阻止调控蛋白的作用，C. J. Bashor 等还构建了亮氨酸“诱捕拉链”(decoy-zipper)，通过 decoy-zipper 竞争性相互作用阻止调控蛋白与支架蛋白的结合而达到调节回路的目的。Decoy-zipper 具有更高的亲和力，当通路激活以后，新表达的负调控蛋白 Msg5 首先与 decoy-zipper 结合，只有当 decoy-zipper 饱和以后，多余的负调控蛋白才会与支架结合抑制通路输出。由此，通过改变调控蛋白和 decoy-zipper 的表达水平、启动子的强度等能够以支架蛋白为平台，使 MAPK 通路具有不同的动态输出，从而实现一系列复杂功能。

1. 脉冲发生器功能

组成型表达 decoy-zipper。对 Msg5 而言，decoy-zipper 拥有比 Ste5-zipper 更高的亲和力，当回路启动后，Msg5 首先与 decoy-zipper 结合，Ste5 支架蛋白上无法形成负反馈回路，此时 MAPK 通路输出正常开启；一段时间后 decoy-zipper 被饱和，新产生的 Msg5 才与 Ste5 结合抑制通路输出，此时通路输出缓慢下降，由此得到一个类似脉冲发生器功能的蛋白质装置。decoy-zipper 与 Msg5 的亲和力越强、表达水平越高，饱和 decoy-zipper 所需 Msg5 的量越多，与 Ste5-zipper 的结合越滞后，对系统输出的抑制作用越滞后，通路输出有较多时间积累，得到的输出峰值越高。其实验结果如图 4-41A 所示。

2. 加速器功能

正调控蛋白 Ste50 组成型表达，负调控蛋白 Msg5 诱导表达。其中 Msg5 具有比 Ste50 更强的与支架蛋白结合的亲和力，通路激活后，首先在正调控蛋白 Ste50 的作用下迅速积累输出；随后诱导表达的 Msg5 逐渐取代 Ste50 与支架蛋白结合抑制通路输出。实验结果如图 4-41B 所示，可以看出虽然稳态输出值有所降低，但由 Msg5 组成的负反馈回路能够缩短系统进入稳态的响应时间。此外，由于初始阶段 Ste50 的正反馈加速作用，通路输出在 Msg5 发挥作用之前快速积累，系统输出的峰值接近野生型系统的输出峰值。

3. 延迟器功能

负调控蛋白 Msg5 组成型表达，decoy-zipper 在诱导条件下表达。初始状态下，Msg5 的作用使得通路几乎处于关闭状态，此时响应非常微弱。50min 后，积累到特定水平的 decoy-zipper 与负调控蛋白结合，Ste5 支架通路因失去 Msg5 的抑制而激活通路输出，由此系统输出对于输入而言具有一定时间的滞后。实验结果如图 4-41C 所示。

4. 超敏开关功能

Msg5 组成型表达，Ste50 在诱导条件下表达。通路首先在 Msg5 的抑制作用下维持较低水平的输出，当不断诱导表达 Ste50 时，Ste50 的正反馈作用迅速提高通路输出，此时通过 decoy-zipper 减少与支架蛋白结合的 Msg5 使其对回路的抑制作用降低。由此得到的系统输出响应类似于超敏开关，其希尔系数较野生型显著提高。实验结果如图

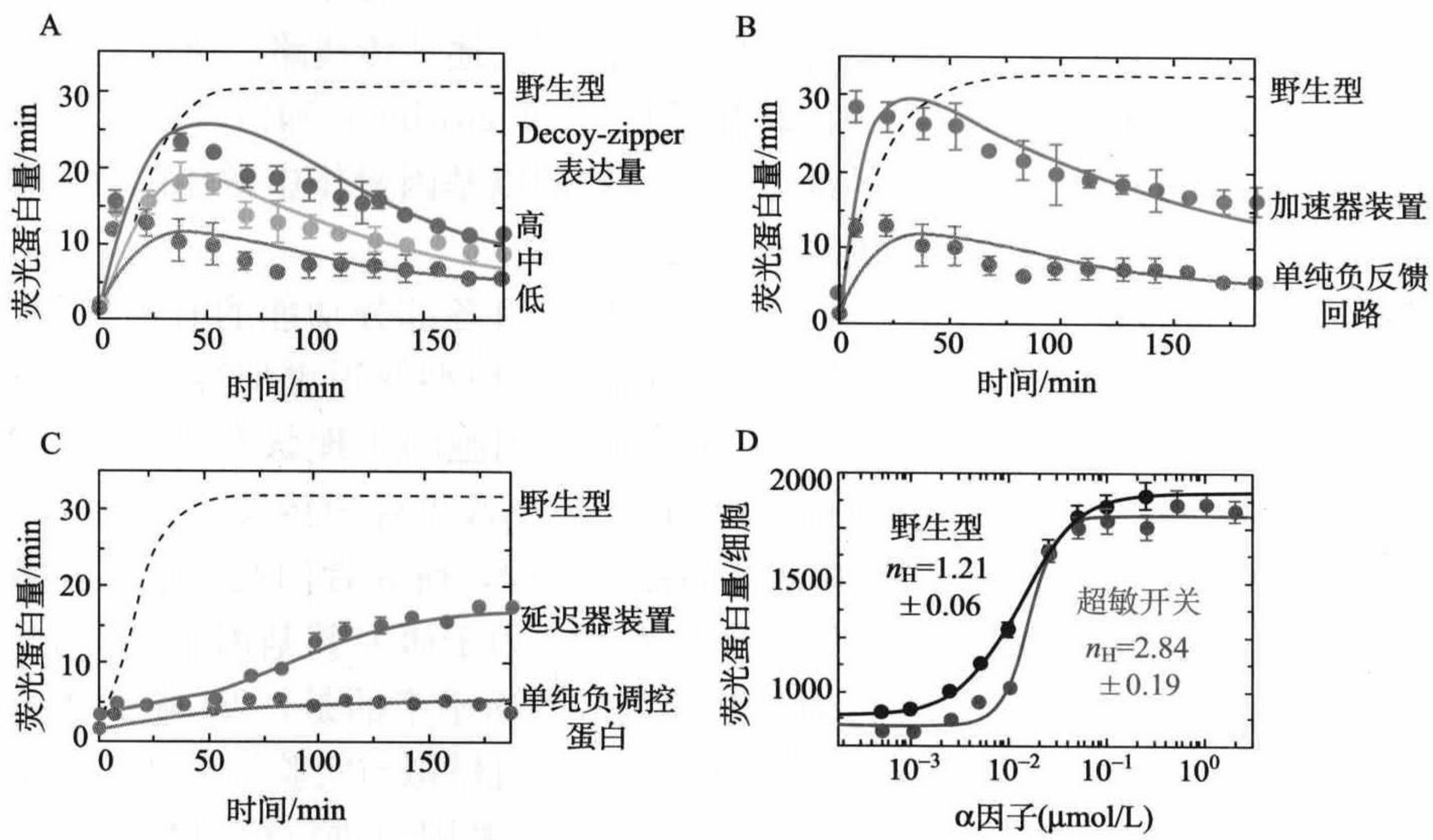


图 4-41　工程化 Ste5 支架蛋白实现的复杂动力学响应测试图[24]

图 A 为脉冲发生器实验测试图。组成型表达 decoy-zipper。具有较高亲和力的 Msg5 首先与 decoy-zipper 结合，Ste5 支架蛋白上无法形成负反馈，MAPK 通路正常输出；当 decoy-zipper 被饱和后，新产生的 Msg5 才与 Ste5 结合使通路输出缓慢下降。Msg5 与 decoy-zipper 的亲和力越强、表达水平越高，通路得到的输出峰值越高。图 B Ste50 组成型表达，Msg5 诱导条件下表达。Msg5 具有比 Ste50 更强的与支架蛋白结合的亲和力，通路激活后，首先在正调控蛋白 Ste50 的作用下迅速积累输出；随后诱导表达的 Msg5 逐渐取代 Ste50 与支架蛋白结合抑制通路输出。虽然稳态输出值有所降低，但加速器回路能够缩短系统达到稳态输出的响应时间。Ste50 的正反馈加速作用使系统输出的峰值接近野生型系统的输出峰值。图 C 为延迟器实验测试图，组成型表达负调控蛋白 Msg5，使得通路几乎处于关闭状态。50min 后，积累到特定水平的 decoy-zipper 与负调控蛋白结合，激活通路输出，由此系统输出滞后于输入。图 D 为超敏开关实验测试图，Msg5 组成型表达，Ste50 在诱导条件下表达。通路首先在 Msg5 的抑制作用下维持较低水平的输出，当不断诱导表达 Ste50 时，Ste50 的正反馈作用使 Msg5 对回路的抑制作用降低。由此得到的系统输出响应类似于超敏开关，其希尔系数较野生型显著提高。

4-41D 所示。

综上所述，这种工程化策略让支架蛋白成为一个灵活的平台，利用支架蛋白在空间上的局域化作用建立精细的微观结构，在诱导信号的调控下重新配置信号通路，实现期望的时序和动力学行为。

在哺乳动物和植物细胞内，MAPK 级联转导途径广泛存在，在调控细胞的分化、增殖、凋亡、免疫、应激反应等多方面起着重要作用，因此此工程化支架蛋白有望用来改造细胞，使其在疾病治疗等方面发挥新的作用。

4.9　重构 T7 噬菌体基因组[49]

如本书第 1 章中所述，物种必须不断进化才能在“物竞天择”的自然规律下维持自身的存在和发展。进化，对于一个物种的存在和发展来说具有至关重要的作用。但是对

于使用合成生物学方法构建生物体来说，自然的进化则往往会导致生物体过于复杂而大大增加研究的难度以致无法对其进行有效操控。除了上述遗传线路、蛋白质信号转导途径的设计和开发，对于宿主基因组的模块化重构（refactoring）和仿真模型的开发在合成生物学的研究中占有着非常重要的地位。下面我们以基因操作中常用的 T7 噬菌体为例，为读者讲解这方面工作的主要内容和意义。

T7 噬菌体可以侵染大肠杆菌，是研究细菌基因组各部分功能的有力工具。此外，T7 噬菌体独特的生物学特性使得构造 T7 噬菌体系统模型变得更加容易：①与其他噬菌体相比，T7 噬菌体具有较高的独立性，对于宿主细胞的生理状态没有特殊的要求；②T7 噬菌体基因组的后期基因是被自己编码的 T7 RNA 聚合酶转录的。

为了更加系统地研究和预测 T7 噬菌体的侵染过程，研究者们在大量实验数据的基础上建立了 T7 噬菌体侵染的计算机量化仿真模型，用于研究其基因组各部分的功能，预测当环境发生变化时的基因表达行为及其控制。然而不幸的是，T7 噬菌体基因组中基因片段的重叠和多效性等复杂情况几乎使所有计算机模拟与实验结果严重失配。

为此，D. Endy 课题组采用重构方法对 T7 噬菌体基因组进行模块化改造，在不改变基因组外部功能的基础上，改善其内部结构、质量和性能，去除基因的重叠性和多效性，为合成生物学的研究提供更易维护和扩展的基因组平台。重构后的 T7 噬菌体被命名为 T7.1 噬菌体。

重构过程中需要特别关注下面三个问题。

（1）在 T7 噬菌体的基因组中，已经找到或者推测得到了编码 60 个蛋白质的 57 个基因。在这 60 个蛋白质当中，只有 35 个蛋白质具有已知的一个或一个以上的功能。在另外 25 个非重要的蛋白质当中，只有 12 个存在于不同的类似于 T7 噬菌体的家族中。在研究 T7 噬菌体侵染的模型中是否可以忽略这些功能不明确的蛋白质？是否应该在重构 T7 基因组时去除这些蛋白质的编码基因？

（2）通过胞内转录的研究已经发现、测序和验证了 T7 噬菌体基因组中被大肠杆菌 RNA 聚合酶识别的启动子（A0、A1～3、B、C、E）。但是，通过随机片段克隆等其他方法又发现了其余十几个启动子，对于这些启动子，之前没有任何关于它们在野生型 T7 噬菌体基因组中的作用的证据。在重构 T7 基因组和对其进行仿真模拟的过程中是否应该去除这些启动子？

（3）是否还有其他的信息包含在野生型 T7 噬菌体的基因组中？在研究中是否应该包括这些信息？

重构的第一步是重新注释野生型噬菌体 T7 基因组，确定下列功能片段的边界：编码 60 个蛋白质的带有 57 个核糖体结合位点的 57 个开放读码框；调控噬菌体基因表达的 51 个操纵元件；DNA 复制等功能片段。

图 4-42 所示为野生型 T7 噬菌体和 T7.1 噬菌体基因组示意图。图 4-42A 中将野生型 T7 噬菌体的基因组分为 α～ζ 六个区域，每个区域之间使用独立于整个基因组的限制性酶切位点进行区分，确保可以独立地重组、检测和调控每个区域。图 4-42B 所示为野生型 T7 噬菌体的 α 区域的基因组成，其中包括蛋白质编码区域（蓝色）、核糖体结合位点（紫色）、RNaseIII 识别位点（粉色）、转录终止子（黄色）和其他区域（灰色）。

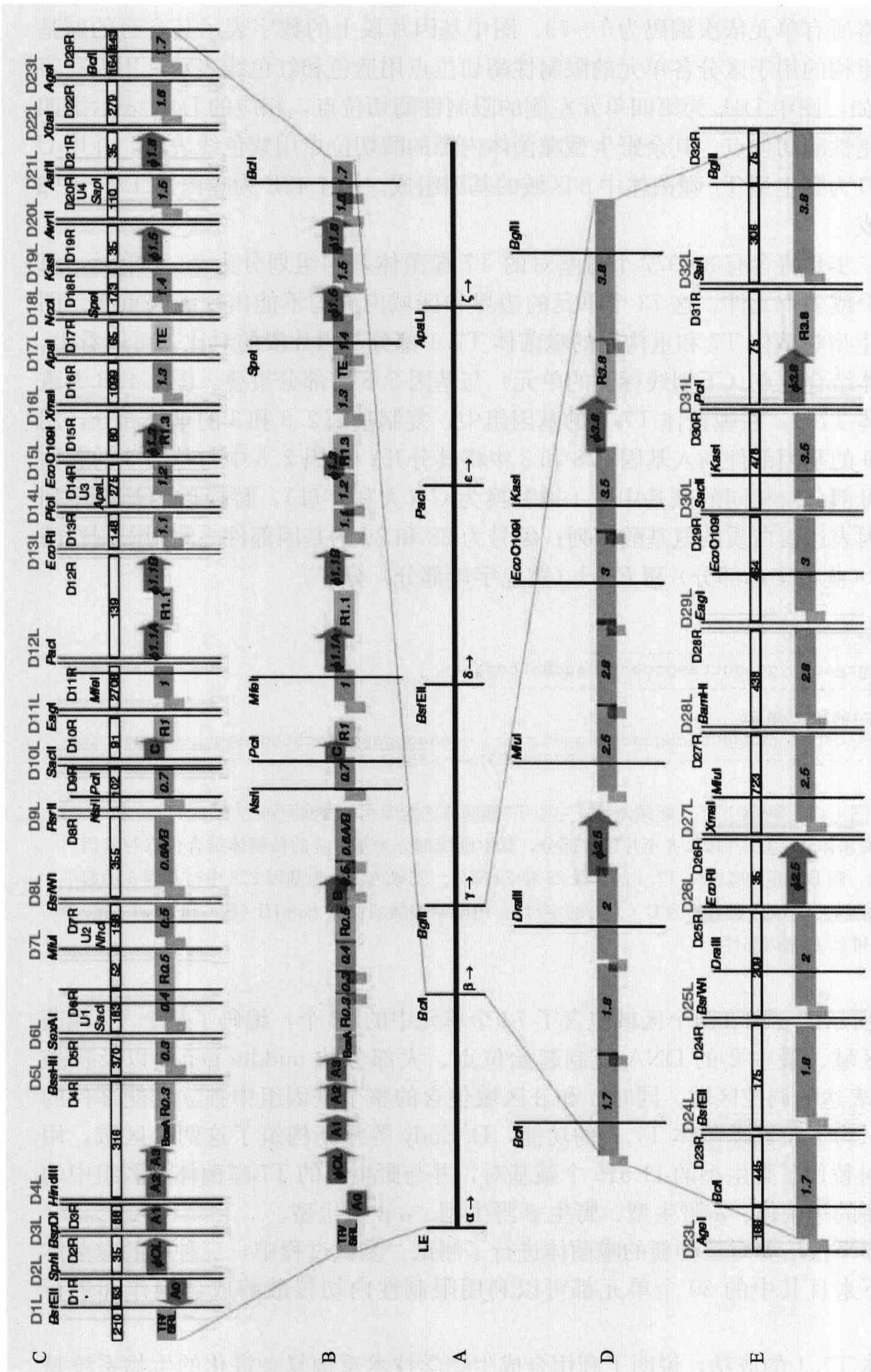


图 4-42　野生型 T7 噬菌体和 T7.1 噬菌体基因组对比示意图

图 A 中将 T7 噬菌体的基因组分为 α 到 ζ 六个区域，为确保各区域可以独立的被重组、检测和调控，区域间用独立于整个基因组的限制性酶切位点进行区分。图 B 为 T7 噬菌体 α 区域的基因组成。其中包括蛋白质编码区域（蓝色），RBS（紫色），转录终止子（黄色）和其他区域（灰色）。内源限制性酶切位点用黑色线表示。图 C 为 T7.1 噬菌体的 α 区域。从左到右将所有单元依次编码为 1－23。图中基因片段上的数字表示其含有的碱基对个数。人工重构的用于区分各单元的限制性酶切位点用蓝色和红色线表示，用大写字母 D 标注。其余野生型噬菌体内源酶切位点用紫色线表示，并用 U 标注。图 D 为 T7 噬菌体中 β 区域的基因组成。图 E 为 T7.1 中 β 区域的基因组成。

自然存在的限制性酶切位点用黑色线表示。图 4-42C 为重构后的噬菌体 T7.1 的 α 区域。从左到右将所有单元依次编码为 1～73。图中基因片段上的数字表示其含有的碱基对个数。人工重构的用于区分各单元的限制性酶切位点用蓝色和红色线表示，用大写字母 D 标注，例如，图中 D4L 为第四单元左侧的限制性酶切位点，相应的 D4R 表示第四单元右侧的限制性酶切位点。其余野生型噬菌体内源的酶切位点用紫色线表示，并用 U 标注。图 4-42D 为野生型 T7 噬菌体中 β 区域的基因组成。图 4-42E 为噬菌体 T7.1 中 β 区域的基因组成。

重构的第二步是将含有 39 977 个碱基对的 T7 噬菌体基因组划分为 73 个单元，每个单元包含一个或多个元件。这 73 个单元的边界和区域内部均不能出现编码重叠。图 4-43 所示为野生型噬菌体 T7 和重构后的噬菌体 T7.1 部分基因片段的对比。可以看出，基因 3 的核糖体结合位点（下划线标注的单元）与基因 2.8 有部分重叠。图 4-43B 为重构以后的噬菌体 T7.1。在噬菌体 T7.1 的基因组中，复制基因 2.8 和 3 的重叠部分，用编号为 28 和 29 的基因部件插入基因 2.8 和 3 中将其分开；基因 2.8 中与基因 3 的起始区域（绿色字母部分）相同的碱基中的 t 被替换为 C（大写字母），替换后的基因序列并未改变该基因表达蛋白质的氨基酸序列；编号为 28 和 29 的基因部件之间用限制性酶切位点 *Bam*HI（蓝色字母部分）和 *Eag*I（黄色字母部分）分开。

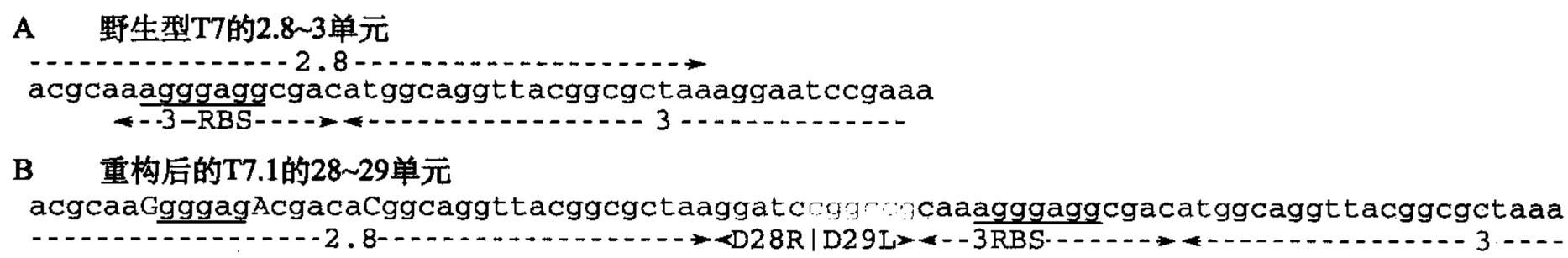


图 4-43 重构的 T7.1 噬菌体与野生型 T7 噬菌体功能单元基因编码对比[49]

图 A 为野生型噬菌体 T7 基因片段 2.8 至片段 3 部分，其中画线部分为基因 3 的核糖体结合位点与基因 2.8 的重叠部分。图 B 为重构噬菌体 T7.1 的片段 28 和 29 部分。其绿色部分为基因 2.8 中与基因 3 的起始区域相同的碱基，其中的 t 被替换为 C（大写的字母）；用限制性酶切位点 *Bam*HI（蓝）和 *Eag*I（黄）分开编号为 28 和 29 的基因部件。

图 4-42 中所示的 α 和 β 两个区域包含了 73 个单元中的 32 个，编码了整个 T7 噬菌体基因的早期区域、最主要的 DNA 复制起始位点、大部分的 middle genes 以及操控 T7 噬菌体基因表达的调控区域。同时 α 和 β 区域包含的整个基因组中独立功能部件的密度是最大的。为了验证噬菌体 T7.1 的功能，D. Endy 等重新构筑了这两个区域，用 12 179 个碱基对替代了野生型的 11 515 个碱基对，并与野生型的 T7 噬菌体基因组中的部件组成三种新的噬菌体：α-野生型、野生-β-野生型、α-β-野生型。

通过侵染和平板培养对三种新的噬菌体进行了测试。测试过程中，三种新的噬菌体都成功地存活下来且其中的 30 个单元都可以利用限制性内切核酸酶成功操作和进行 DNA 提取。

重构噬菌体 T7.1 的成功，说明了利用合成生物学技术重构复杂进化的生物系统是完全可能的。噬菌体 T7.1 的存活有力地说明了以下 3 点。

（1）α 和 β 区域中的部分可以通过外源的 DNA 进行分隔。

（2）总体来说，在实验室条件下，去除所有的基因重叠并不会影响噬菌体的功能，也不会影响其存活。

（3）噬菌体 T7.1 的侵染与其侵染宿主细胞过程的物理位置有关，而与宿主细胞的生长状态无关。

通过上述两大步骤的重构得到的噬菌体 T7.1 的基因组比野生型 T7 噬菌体的基因组增加了 1424 个碱基对，不仅去除了基因片段的重叠和多效性，而且各种限制性酶切位点的插入使得研究人员可以利用限制性内切核酸酶来独立地操作各个部件，大大增加了新噬菌体的可扩展性和可维护性。

科学成就是由一点一滴积累起来的，
唯有长期的积聚才能由点滴汇成大海。
——华罗庚

第 5 章　合成生物学的应用研究

基础提要

- 氨基酸的合成与代谢
- 大肠杆菌双组分系统
- 代谢工程基础知识

淡水资源在减少、陆地在消失、森林面积在缩小、物种在灭绝、能源在枯竭，地球上暂时能为人类利用的资源正在以残忍的速度消失；SARS、甲型 H1N1 型流感、艾滋病等新型疾病威胁着人类的健康。人类是在资源的衰竭中坐以待毙，是在各种病毒的肆虐中走向灭亡，还是积极地利用各种方法寻找解决的途径？合成生物学的出现，使人类又获得了崭新的工具。

5.1　合成生物学应用于维护人类健康

如何利用各种人工合成的生物体在疾病诊断、治疗和药物生产方面做出贡献一直是备受合成生物学家关注的问题。在工程化设计思想的指导下，通过利用标准化、可互换的生物部件，合成生物学在疾病机理和药物研究方面已经取得了一些令人瞩目的成绩[181]：

- *在疾病机理研究方面*

（1）人工重构 B 细胞抗原受体信号处理过程。V. Rolli 等在果蝇 Schneider-2 细胞（简称 S2 细胞）中人工重构了 B 细胞抗原受体（B cell antigen receptor，BCR）信号处理过程[]，使研究者对于信号网络拓扑结构有了更深入的了解，并且证实了是由于正反馈回路导致 BCR 信号的强烈放大。在 BCR 信号处理过程中，别构的酪氨酸激酶 Syk 具有酪氨酸活化模体（immunoreceptor tyrosine-based activation motif，ITAM），它能够与 BCR 免疫受体结合并被磷酸化而激活。激活的 Syk 使邻近的 ITAM 磷酸化导致更多

的 Syk 结合和激活，从而形成正反馈，并最终引起 BCR 信号的快速放大[182]。随后，这一研究小组又利用此重构的 BCR 信号处理系统证明了罕见的血液丙种球蛋白血症（agammaglobulinemia，一种以 B 细胞早期发育阻断为特征的免疫缺陷病）与 IgB 的一种无义突变有关[183]。

（2）人工改造蝙蝠冠状病毒。生物信息学的分析表明，21 世纪初 SARS（severe acquired respiratory syndrome）传染病来源于一种蝙蝠冠状病毒[184]。蝙蝠冠状病毒一般不能在细胞培养物和实验室动物之中繁殖，因此 M. M. Becker 等用合成生物学方法使蝙蝠病毒刺突蛋白的针结构域与感染人类的 SARS 病毒相应蛋白的外结构域互换，使这种工程化病毒能够在鼠科和灵长类动物细胞中复制，为 SARS 病毒的研究开发了通用平台，从而帮助人们了解动物源性病毒怎样获得感染人类的向性（亲和性），促进对 SARS 病毒变异的研究和预防[184,185]。

• *在发现新药方面*

（1）利用合成基因线路筛选提高抗结核药物活性的小分子。W. Weber 等在人类的胚肾细胞中植入了一个基于 EthR 的合成基因线路，通过此基因线路筛选了能够关闭结核分枝杆菌（*Mycobacterium tuberculosis*）对抗生素乙硫异烟胺（ethionamide，ETH）抗性的小分子[186]。ETH 是一种重要的抗结核药物，常用于感染具有多种耐药抗性结核分枝杆菌患者的治疗，但 ETH 需要被分枝杆菌酶激活后才能发挥抗生素活性。结核分枝杆菌中的 *ethA* 基因编码的单加氧酶 EthA 能够催化 ETH 的活性。然而，由于 EthA 在转录水平被 TetR/CamR 类阻遏蛋白 EthR 所抑制，因此这种激活常常效率很低[187]。为了筛选能够穿过细胞膜到达细胞内的病原体使 EthR 失活且对哺乳动物细胞没有毒性的复合物，W. Weber 等将 EthR 与 VP16 转录反激活因子（transactivator）融合成为一个新的转录因子，使其能够与 EthR 专一的操纵子 O_{ethR} 结合激活下游哺乳动物启动子并表达报告基因。W. Weber 等利用此种方法对一个合成化合物库进行筛选，发现几种复合物能抑制 EthR 活性，并最终发现小分子酯类 2-丁酸苯乙酯（2-phenylethyl-butyrate）满足上面的所有要求且能激活结核分枝杆菌中 EthA 的表达，使病原体对抗生素 ETH 更加敏感。这一发现为抗结核病的新药发现提供了崭新思路[186]。

（2）抗感染分子的筛选。PIP 是一种源于始旋链霉菌（*Streptomyces pristinaespiralis*）的链阳菌素响应的阻遏蛋白。D. Aubel 等把 PIP 工程化入哺乳动物细胞中，利用 PIP 对相应基因的阻遏作用以及链阳菌素类抗生素——普纳霉素 I（streptogramin antibiotic prinstinamycin I）对 PIP 的解阻遏作用来实现对报告基因的调控，并利用此系统筛选链霉菌代谢物库。识别其代谢化合物能够抵抗临床上病原体、渗透入哺乳动物细胞且没有毒性的链霉菌菌株[188]。

（3）抗癌药物筛选的人工基因网络。抗癌药物的关键特性之一是必须能够选择性地杀死分裂细胞而不影响无有丝分裂活性的细胞（mitotically inactive cell）。为了促进抗癌药物的研究，V. Gonzalez-Nicolini 等利用一个合成基因网络模拟 G_1 期停滞细胞群体出现癌细胞增长的现象，作为开发新药的癌细胞模型。V. Gonzalez-Nicolini 等研究

者在中国仓鼠卵巢细胞（Chinese hamster ovary cell）内构建了一个由四环素激活的工程化基因网络来控制细胞周期蛋白依赖性激酶抑制剂 $p27^{Kip1}$ 的表达。当细胞中存在四环素时，转录激活因子 tTA 失活，$p27^{Kip1}$ 的表达被沉默；当缺乏四环素时，tTA 结合到其适配启动子 P_{TET} 上并启动 $p27^{Kip1}$ 的表达，抑制 G_1 期生长。在持续的生长控制下，细胞以一定的概率脱离 $p27^{Kip1}$ 导致的 G_1 期停滞而进入增殖状态，这意味着癌症的发育[189]。由此，可利用这一人工基因网络进行抗癌药物的筛选。如果某些化合物能够选择性杀死诱导增殖的工程化细胞而不影响生长阻滞在 G_1 期的细胞群体，则此化合物可能是有效的抗癌药物。

（4）合理设计抗菌肽。G. Stephanopoulos 课题组分析了天然抗菌肽的序列并将其氨基酸序列作为具有规律性语法的正式语言，用以设计新的非天然抗菌肽。这些人工设计的新型抗菌肽遵循天然抗菌肽的语法准则但位于不同的序列区域。与常规抗生素相比，这些抗菌肽具有受细菌抗性影响小的优势，能够抑制包括金黄色葡萄球菌（*Staphylococcus aureus*）和炭疽杆菌（*Bacillus anthracis*）在内的多种致病菌[190]。

• *在疾病治疗与防御方面*

J. J. Collins的研究小组利用合成生物学方法对噬菌体进行改造提高其抗菌活性，旨在利用工程化噬菌体杀灭人类的致病细菌[191,192]。病菌的生物膜（biofilm）能够抵抗抗菌治疗或者避开宿主免疫系统的攻击，为此摧毁病菌生物膜对于杀灭致病细菌至关重要。为了溶解生物膜，研究者对 T7 噬菌体进行工程化改造使其能够表达伴放线放线杆菌（*Actinobacillus actinomycetemcomitans*）的 dispersin B（DspB）。DspB 是一种能够水解细菌生物膜形成所需的关键胞外聚合物（EPS）的 β-1，6-乙酰-D-葡萄糖胺糖苷键的酶。DspB 能够通过降解 EPS 松弛病菌生物膜结构，释放出的噬菌体又能够攻击邻近的病菌细胞。对这些噬菌体的工程化改造可使其灭菌能力较野生型噬菌体提高 4～5 个数量级。

除了生物膜结构，细菌对抗生素的抗药性是另外一个对抗细菌感染的主要课题。在此之前人们主要是利用小分子来对抗细菌的抗药性，J. J. Collins 课题组率先针对大肠杆菌的抗生素抗性对 M13 噬菌体进行改造使其产生 LexA3（一种 SOS 反应系统的阻遏蛋白）[192]。通过 LexA3 介导的 SOS 抑制，使细菌更易被抗生素攻击而不易产生抗药性[193]。进一步的实验表明，LexA3 的表达与 β-内酰胺（β-lactam）和氨基糖苷类（aminoglycoside）抗生素协同作用能够杀死细菌，提高被大肠杆菌感染的鼠的存活率[192]。

5.1.1　环境控制的大肠杆菌对癌细胞的侵入系统

癌也叫恶性肿瘤，癌组织的基本单位是癌细胞。人体的绝大部分正常细胞可以进行有限度的增殖，老化死亡后会有新生细胞取代它，以维持机体功能。而癌细胞的增殖则是无止境的，这使患者体内的营养物质被大量消耗。同时，癌细胞还能释放出多种毒素甚至转移到全身各处生长繁殖，破坏正常组织，使人体产生一系列症状，最后导致人体消瘦、无力、贫血、食欲不振、发热及脏器功能受损直至死亡等。目前人类对于癌症的致病机制还不甚了解，世界卫生组织预测，21 世纪癌症将取代心脑血管疾病成为人类

的“第一杀手”。

静脉注射法是自静脉注入药液的方法，其特点是药液从静脉直接进入血液循环而迅速发生药效。如果具有抗癌功能的工程菌能够通过静脉注射直接进入血液循环，并在随血液流经全身的过程中探测及趋向于癌细胞，进而侵入癌细胞，释放所携带的抗癌成分杀死癌细胞而不损害任何其他正常细胞，则人类完全有可能拥有更加强大的对抗癌症这一“世纪杀手”的能力。

通过利用合成生物学行之有效的工程化理念和部件，人类已经朝着这个目标迈出了可喜的一大步。2007 年，J. C. Anderson 等开发出了能够定向并侵入癌细胞的大肠杆菌，下面我们来仔细讲解一下 J. C. Anderson 等的工作。

大肠杆菌（*E. coli*）是一种在人和温血动物肠道内常见的细菌。除了肠出血性大肠杆菌（EHEC）等极少数菌种外，大多数大肠杆菌菌株在肠道内并不是致病菌，而且由于其遗传背景研究得比较透彻，大肠杆菌已经是实验室应用最广泛的工程菌。因此，J. C. Anderson等的研究仍然是从大肠杆菌开始，并希望在此基础上将研究成果扩展到其他细菌。

由前面的描述可知，如果能够实现通过静脉注射大肠杆菌治疗癌症，那么必须解决三个难题：①由静脉注射进入人体血液循环系统的大肠杆菌必须能够选择性地定位于实体瘤组织；②定位于实体瘤组织的大肠杆菌必须能够选择性地侵入癌细胞内部而不伤害周围的正常细胞；③顺利侵入癌细胞中的大肠杆菌必须能够释放或合成某种能杀死癌细胞的物质。为此 J. C. Anderson 等主要针对前两步进行研究，开发能够选择性定位并侵入癌细胞的大肠杆菌。他们的工作主要包括以下几部分[194]。

1. 大肠杆菌对哺乳动物细胞侵入机制的构建与模块化

如同所有细菌一样，大肠杆菌同样具有许多能够与真核细胞发生相互作用的系统。然而这些天然系统的冗余性和复杂的控制机制使得人们很难对其进行工程化和利用。为此，J. C. Anderson 等将取自假结核耶尔森氏菌（*Yersinia pseudotuberculosis*）的编码侵袭素的 *inv* 基因植入大肠杆菌体内，将含有此基因的大肠杆菌命名为 inv^+ *E. coli*。

侵袭素是一种长的、不易弯曲的蛋白质，能够锚定在细胞膜上。在肠道中，*Yersinia* 菌属能够利用侵袭素识别小鼠 M 细胞表面的 β1 整联蛋白并与其紧密结合，通过刺激 Rac-1 诱导细菌侵入。哺乳动物细胞能够接受表面涂有侵袭素的乳胶珠，说明侵袭素与 β1 整联蛋白的结合不需要其他附加的细菌蛋白。由此可见，将 *inv* 基因转入大肠杆菌中不仅能使大肠杆菌具有侵入细胞的能力，而且能够避开大肠杆菌天然系统各种复杂的干扰，实现侵入机制的模块化。

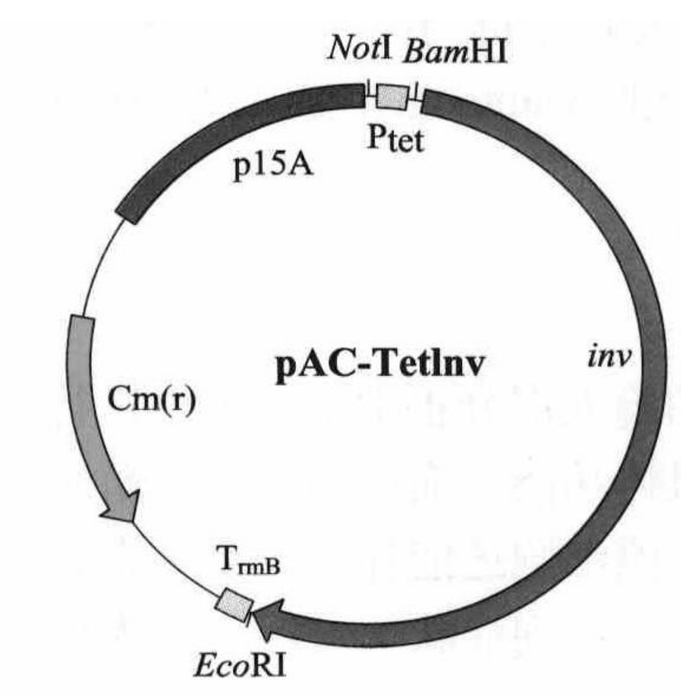


图 5-1 pAC-TetInv 质粒构建图[194]

将 *inv* 基因构筑于中等拷贝数的质粒上，利用 P_{tet} 启动子对其进行控制，并命名为 pAC-TetInv 质粒，如图 5-1 所示。

为了检验含有 pAC-TetInv 质粒的大肠杆菌对哺乳动

物细胞的侵入能力，同时检验侵入过程是否需要其他黏附素或菌毛的帮助，研究者将 pAC-TetInv 质粒转入大肠杆菌 CAMC60 菌株中（该菌株经过处理后不分泌任何黏附素，也无菌毛）并通过实验检验其对 HeLa 细胞（宫颈癌传代细胞）的侵袭能力。随后 Anderson 等还检验了 inv^+ *E. coli* 对于 U2OS（骨肉瘤，osteosarcoma）和 HepG2（肝癌，hepatocarcinoma）细胞的侵入能力。

实验结果表明 inv^+ *E. coli* 能够侵入不同来源的癌细胞，但由于 β1 整联蛋白的表达水平或可接近程度不同，inv^+ *E. coli* 对三种不同细胞侵袭的效率有所不同。需要注意的是，不是所有的哺乳动物细胞都易被 inv^+ *E. coli* 侵入。从前面的设计可知，在感染鼠上皮细胞的过程中，*Y. pseudotuburculosis* 通过肠吸收细胞（enterocyte）专一地侵袭 M 细胞。因此，类似地，inv^+ *E. coli* 也只能侵袭表达 β1 整联蛋白的细胞，如位于上皮层前沿的细胞。

2. 侵入机制的诱导控制设计

抗癌治疗的一个基本要求是尽量降低对正常细胞的侵害。为此，必须对大肠杆菌的侵入功能进行调控，使其只有在需要的时候才开启 *inv* 基因。研究者将 *inv* 基因置于 P_{araBAD}启动子的控制下，如图 5-2 所示。

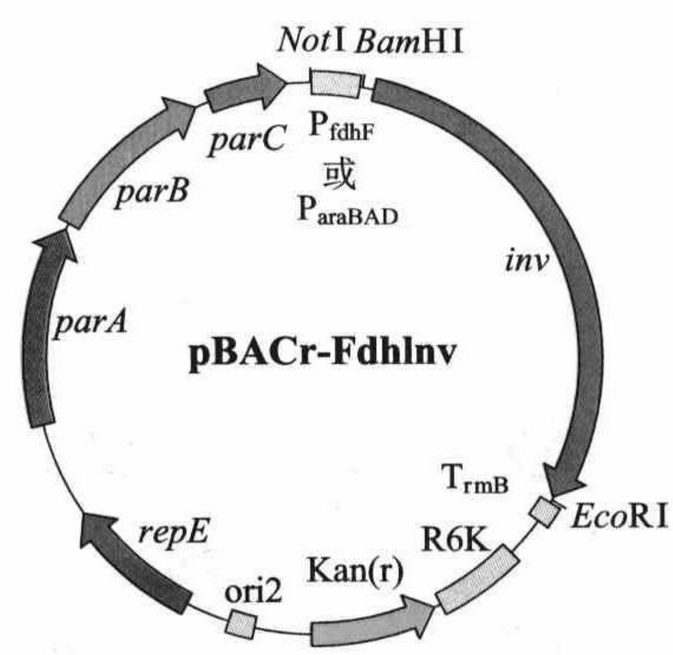


图 5-2　pBACr-FdhInv 质粒构建图[194]

因 P_{araBAD}启动子的调控需要阿拉伯糖分子激活和 AraC 转录因子的作用，研究者在原本不含有 *araBAD* 操纵子的大肠杆菌 MC1061 菌株中，用*araBAD*和*araC* 基因替换 P_{tet} 启动子。P_{araBAD}启动子被广泛用于阿拉伯糖诱导的异源蛋白的过量表达，具有较高的基础转录水平。为了降低侵袭素的基础表达，研究者将上述基因构建在单拷贝 BAC 质粒（人工细菌染色体）上。为进一步降低侵袭素的基础表达，研究者还将核糖体结合位点侧翼的 5′非翻译区（UTR）用随机序列替换，并随机排列第一、第二密码子，由此构建了容量达 10^6 个样本的 RBS 库，并采用各种方法对其进行筛选。实验结果证明了阿拉伯糖分子对于大肠杆菌的侵入能力具有调控作用。

3. 大肠杆菌对癌细胞的趋向性

具有侵入能力的 inv^+ *E. coli* 的构建目的是为了治疗癌症，因此必须使大肠杆菌具有定位于肿瘤细胞的趋向性。

1）对于缺氧微环境的趋向性

已有的研究表明，由于肿瘤细胞的快速失控生长的特性，导致血液供应不足而产生实体瘤内的缺氧（Hypoxia）微环境。因此，缺氧环境可以作为识别肿瘤并诱导 inv^+ *E. coli* 侵入的因素[195]。前期微阵列分析的研究结果表明，当生长条件从有氧环境转为缺氧环境时，大肠杆菌中的许多基因会被强烈诱导表达[196]，其中甲酸脱氢酶（formate dehydrogenase，*fdhF*）就是最强烈的被诱导基因之一[197]。因此，研究者在原有研究

的基础上，构建了含有甲酸脱氢酶基因启动子的 pBACr-FdhInv 质粒，如图 5-2 所示，并导入大肠杆菌细胞，利用此启动子将 inv^+ *E. coli* 的侵入与缺氧环境相关联。此外，采用类似于上一小节描述的方法构建容量为 10^6 个样本的 RBS 库，用厌氧环境对其进行筛选，获取厌氧诱导生长的阳性克隆。

2）依赖于细胞密度的侵入控制

由于实体瘤所在部分的缺氧微环境、缺乏免疫监视和养分充足等原因，静脉注射以后，包括非病原体在内的很多种细菌（如 *E. coli*、*Vibrio cholerae*、*Clostridium*、*Bifidobacterium*、*Salmonella* 和 *Listeria monocytogenes*）都会集中于膀胱、脑、乳腺癌等实体瘤处。例如，对小鼠进行沙门氏菌静脉注射以后，肿瘤组织内的细菌浓度大概为 10^9 cfu/g，而肝脏中是 10^6 cfu/g，肌肉中是 10^3 cfu/g。这为工程菌对肿瘤的定位提供了另一个线索。据此，研究者采用取自费氏弧菌的遗传线路（*lux*）群体感应基因作为控制 inv^+ *E. coli* 侵入的条件之一。

本书在前面的若干个例子中已经对 *lux* 群体感应遗传线路进行了较为详尽的介绍，这里就不再赘述了。值得注意的一点是，有实验表明，在 *E. coli* 中 *lux* 群体感应系统能够被激活并且与其他同源肠杆菌系统无交叉反应，由此可知利用此系统控制 inv^+ *E. coli*的侵入不会与其他内源系统发生意外的交互作用。

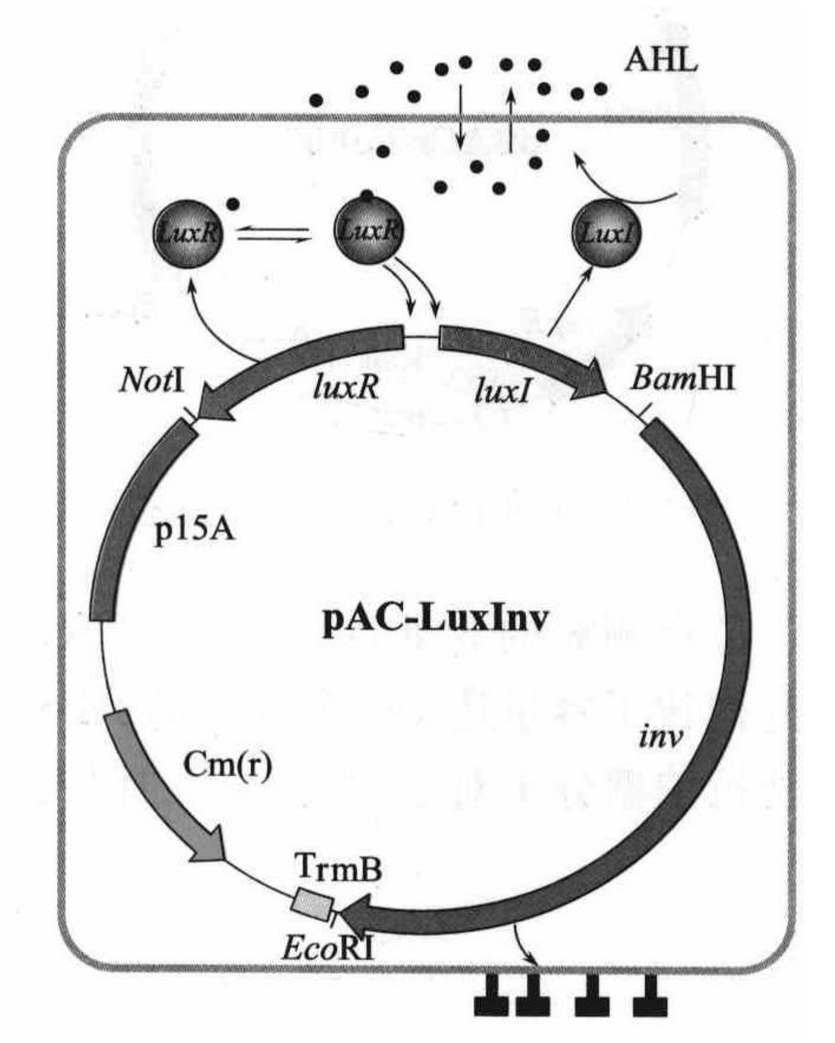


图 5-3 pAC-LuxInv 质粒构建图[194]

研究者构筑了 pAC-LuxInv 质粒，在 *luxI* 下游放置 *inv* 和 Pr_{lux}启动子的融合基因，如图 5-3 所示。含有此质粒的大肠杆菌，只有当其在肿瘤微环境中的菌体聚集密度达到并超过某一阈值时才会迅速开启 *inv* 基因而侵入癌细胞。同样的，研究者也对 RBS 进行了筛选和优化。

由上面的描述可知，J. C. Anderson 等通过分别将取自 *Y. pseudotuberculosis* 的 *inv* 基因和 *V. fisheri* 的 *lux* 操纵系统（operator）、*fdhF* 启动子和阿拉伯糖操纵子融合，使大肠杆菌细胞能够选择性地在高细胞密度、低氧微环境和化学诱导后具有侵入包括宫颈癌、肝癌、骨肉瘤在内的多种癌细胞的能力。如果在此基础上使大肠杆菌具有释放或者催化合成某种能够杀死癌细胞的物质，那么癌症的彻底治愈将指日可待。

虽然此方法离最终的临床应用还有相当一段距离，但由于该技术是通过合成生物学的“模块”技术来实现的，这些合成“零件”或“模块”具有很好的通用性。如果将此菌种进行改造还可使其结合到其他细胞的蛋白质或者碳水化合物表面从而具有侵入其他细胞的能力，用途将更加广泛。因此，该技术受到学术界及产业界的高度评价。正是由于这方面的出色研究成绩，31 岁的 J. C. Anderson 被世界著名科技杂志 *Technology Review* 评为 2007 年 TR35 奖获得者[10]。

除了上面的例子，S. Topp 等利用人工合成的 Riboswitch 使大肠杆菌具有发现并跟

踪化学信号的能力。其构建的系统机理如图 5-4 所示[198]。

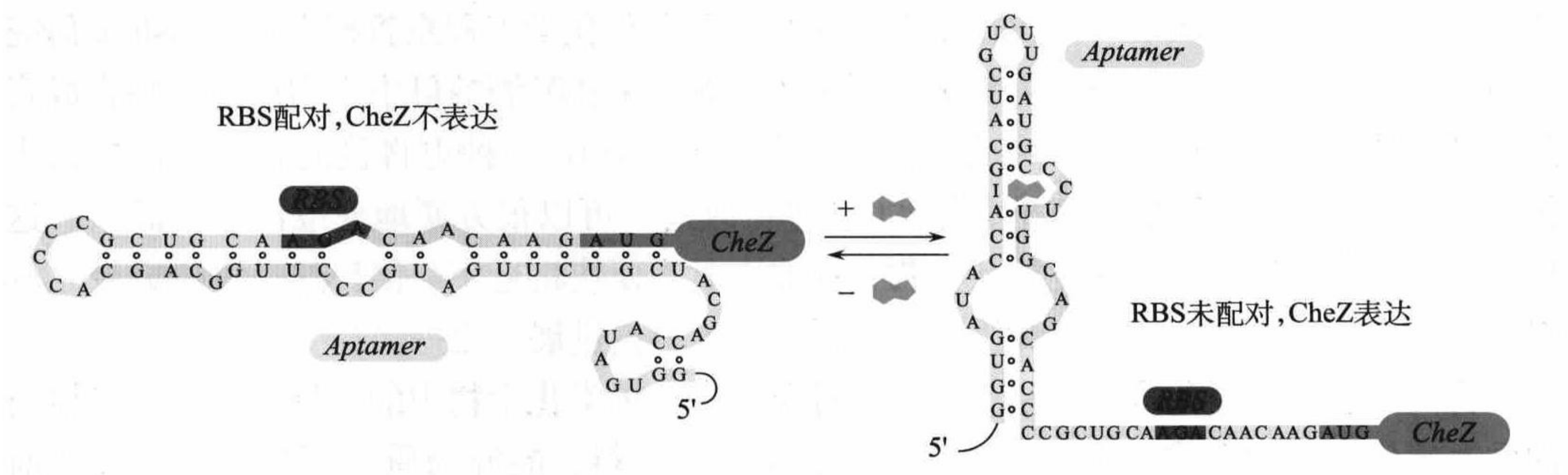


图 5-4　对茶碱敏感的合成 Riboswitch 控制 CheZ 蛋白翻译示意图[198]

当细胞中无茶碱存在时，mRNA 的构象中核糖体结合位点配对出现，其后的部分无法翻译成 CheZ 蛋白。当有茶碱存在时，mRNA 的构象中核糖体结合位点暴露在外，其后的序列顺利翻译出 CheZ 蛋白。

在大肠杆菌中，磷酸酶 CheZ 能够触发 CheY 的脱磷酸化，使大肠杆菌移动和翻滚。为了控制趋向作用，研究者将 *cheZ* 基因与对茶碱（theophylline，一种生物碱，可作为药物）敏感的适配体融合。当细胞中无茶碱存在时，mRNA 的构象中核糖体结合位点碱基配对出现，其后的部分无法翻译成 CheZ 蛋白，CheY-P 蛋白仍然保持磷酸化状态，细胞不移动；当有茶碱存在时，mRNA 的构象中核糖体结合位点暴露在外，其后的序列顺利翻译出 CheZ 蛋白，触发 CheY 的脱磷酸化，使大肠杆菌移动和翻滚，具有对于茶碱梯度分布的趋向性[198]。这一研究成果同样具有广泛的应用价值，为实现工程菌跟踪病理信使（如肿瘤标志物）侵入并杀死恶性肿瘤细胞奠定了基础。

5.1.2　合成青蒿素的微生物工厂

疟疾俗称打摆子，是由疟原虫经按蚊叮咬传播的传染病，临床上以周期性、定时性发作的寒战、高热、出汗退热及贫血和脾大为特点。疟疾是发生最频繁的人类寄生虫病，对于全球公共卫生事业的冲击仅次于结核病，每年在全球约有 5 亿宗病例，导致超过 100 万人死亡。世界卫生组织曾指出，疟疾平均每 30s 杀死一个 5 岁以下的儿童。

由于疟原虫能快速产生抗药性，许多广为人知的高效抗疟药物氯喹（chloroquine）、奎宁（quinine，又称为金鸡纳霜）和甲氟喹（mefloquine）等都已经基本失效，在某些地方，甚至发现有的疟原虫对目前所有价格低廉的一线药物都有交叉抗药性，为此全世界范围内掀起了研制新型抗疟疾药物的科研热潮。20 世纪 70 年代，我国科技工作者首次从青蒿草中提取得到青蒿素（artemisinin）。青蒿素以其高效、快速、低毒以及与氯喹无交叉影响的特点，被世界卫生组织认定为 21 世纪替代奎宁的最有效的抗疟疾药，并被作为一线抗疟疾药在全球推广。不幸的是，虽然世界大部分地区均出产野生青蒿草，但青蒿素含量普遍低于 1‰，导致植物提取青蒿素的成本很高，产量一直供不应求，费用也就居高不下。目前使用青蒿素治疗疟疾，每个疗程的费用大约是 8～15 美元，这对亚洲、非洲等贫困地区的患者来说根本无法承受。人类对抗疟疾的进程似乎又陷入了“沼泽”。

2003 年 *Nature Biotechnology*[199] 和 2006 年 *Nature*[9] 杂志上的两篇文章振奋了全球医药及生物领域的科研人员——美国加州大学伯克利分校化学工程系教授 J. D. Keasling 研究小组利用合成生物学方法分别将相关基因植入大肠杆菌和酿酒酵母中，利用微生物合成青蒿素的前体物质 amorphadiene 和青蒿酸（artemisinic acid，一种更直接的青蒿素前体衍生物）。Amorphadiene 和青蒿酸通过化学方法进行改造，可以很方便地变成活性青蒿素。这种微生物“工厂”的生产速度比从植物中提取青蒿素要快将近 100 倍！这一科研成果立刻引起了世人的广泛关注，被评为当年的“世界十大科技进展”之一。

青蒿素属于萜类化合物，是大约 5 万种类异戊二烯类化学物中的一种。类异戊二烯类物质大部分由植物合成，是许多珍贵药物、染料、香料、能源物质甚至塑料和橡胶的前体，其他著名的类异戊二烯物还包括调味薄荷醇、类胡萝卜素（可对抗紫外线损伤）和紫杉醇（一种高效抗癌药）等。但是，类异戊二烯类物质的结构非常复杂，化学合成的产量极其有限，目前大部分类异戊二烯物质的化学合成还没有工业化，其产量还远远无法满足人类的需求。为此，许多研究者尝试利用微生物合成类异戊二烯或其前体物质（只需几步简单反应即可转化成目标产物，转化的步骤被大大简化），试图利用微生物的人工培养和快速繁殖来提高产量。

J. D. Keasling 和他的研究小组首先在大肠杆菌中设计了一条合成 amorphadiene 的生物合成途径，如图 5-5 所示。整条生物合成途径共分为以下几步：①将大肠杆菌中常见的化学物质乙酰辅酶 A 转变成甲羟戊酸；②将甲羟戊酸转化成焦磷酸异戊酯（isopentanyl pyrophosphate，IPP）或者二甲（基）丙烯焦磷酸酯（dimethylallyl pyrophosphate，DMAPP）；③将 IPP 或者DMAPP转化成法尼基焦磷酸酯（farnesyl pyrophosphate，FPP）；④利用*ADS*基因编码的 amorphadiene 合成酶，将 FPP 转化成 amorphadiene。IPP 和 DMAPP 是所有类异戊二烯的通用前体物，只要再引入恰当的催化酶基因，J. D. Keasling 等开发的新菌株就可作为其他类异戊二烯化合物的生物合成平台。由此可见，这项成果具有极其巨大的应用潜力。

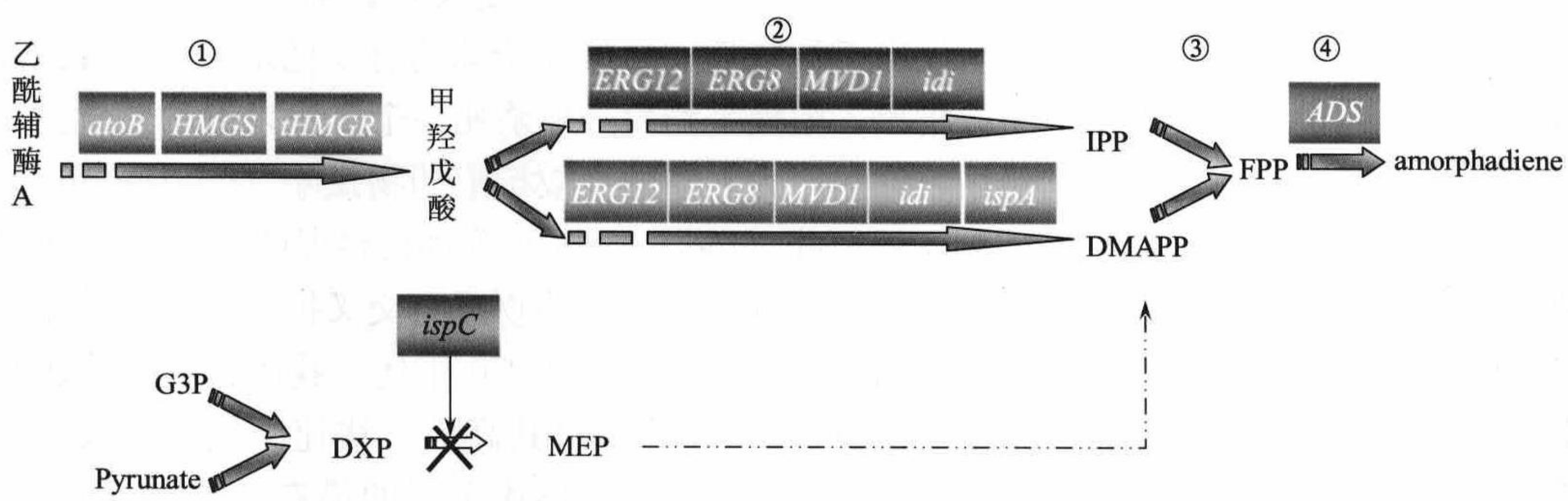


图 5-5 人工构建大肠杆菌 amorphadiene 合成途径

整条生物合成途径共分为四个大步骤：①将大肠杆菌中常见的化学物质乙酰辅酶 A 转变成甲羟戊酸；②将甲羟戊酸转化成焦磷酸异戊酯（IPP）或者二甲（基）丙烯焦磷酸酯（DMAPP）；③将 IPP 或者DMAPP转化成法尼基焦磷酸酯（FPP）；④利用*ADS*基因编码的 amorphadiene 合成酶，将 FPP 转化成 amorphadiene。为避免大肠杆菌内源 DXP 途径的干扰，引入 *ispC* 基因，其编码产物可将大肠杆菌内源的 DXP 途径阻断。

他们的工作大致可以分为以下几部分。

1. 净化大肠杆菌的代谢内环境

大肠杆菌中原有一条内源的合成 IPP 或 DMAPP 的脱氧-5-磷酸木酮糖（deoxyxylulose 5-phosphate，DXP）途径*。许多研究者都曾试图采用各种方法优化 DXP 途径来提高类异戊二烯前体的量，从而提高类胡萝卜素的产量。例如，通过平衡 glyceraldehyde-3-phosphate 和丙酮酸盐，或者提高 1-deoxy-D-xylulose 5-phosphate 合成酶（DXS）和 IPP 异构酶（*idi* 基因编码）的表达量等，但效果很不明显。另外，DXP 途径与大肠杆菌其他生理调节单元间存在着许多密切联系，是限制类异戊二烯产量进一步提高的重要因素，但是目前人们对这些联系所知不多，还无法对 DXP 途径进行有效调控。因此，J. D. Keasling 等引入了 *ispC* 基因，利用其产物将大肠杆菌原有的 DXP 途径切断，斩断所有未知因素经由 DXP 途径对生物合成 IPP 和 DMAPP 的影响，净化大肠杆菌的代谢内环境。

2. 合成和优化编码 amorphadiene 合成酶的基因 *ADS*

众所周知，在原核生物中实现植物基因的表达是件非常困难的事情，J. D. Keasling 所面临的又一困难就是如何在大肠杆菌中高效表达植物基因 *ADS*（紫穗槐二烯合成酶基因）。为了克服这一困难，提高萜类合成酶和 amorphadiene 的表达量，研究人员人工合成和优化了 *ADS* 基因，使其在大肠杆菌中大量表达，克服微生物萜类合成的瓶颈。实验表明，人工合成和优化后的 *ADS* 基因能够在大肠杆菌中顺利表达并将 Amorphadiene 的产量大幅提高。

3. 将酿酒酵母中的甲羟戊酸途径转入大肠杆菌

为了提高细胞内 FPP 的浓度，将酿酒酵母中编码甲羟戊酸途径的 8 个基因做成操纵子转入大肠杆菌中并使其表达。同时为简化操作的复杂性，将这 8 个基因分成两个操纵子，其中一个操纵子编码的酶类能够通过 3 步酶促反应将普遍存在的前体物乙酰辅酶 A 转化成为甲羟戊酸；另外一个操纵子编码的酶类能够将甲羟戊酸转化成为 IPP 和（或）DMAPP。

4. 平衡和优化合成途径

上述人工构建的合成 amorphadiene 多步反应的许多中间产物（包括 IPP），在高浓度时可导致菌体中毒。因此，必须协调 IPP 有关的基因以平衡其合成与消耗，确保在其能够杀伤大肠杆菌以前及时地转化为 amorphadiene。为了平衡和优化这条途径，

---

* IPP 及其异构体 DMAPP 的生物合成通常有甲羟戊酸类异戊二烯（mevalonate-dependent isoprenoid，MEV）途径和 DXP 途径。植物细胞会利用 MEV 和非依赖甲羟戊酸途径或者 DXP 途径合成类异戊二烯；其他真核细胞则无一例外地通过 MEV 途径将乙酰辅酶 A 转化成 IPP，IPP 随后被异构化成 DMAPP；除了某些特例外，原核生物则利用 DXP 途径分别合成 IPP 和 DMAPP。IPP 和 DMAPP 是 FPP 合成中至关重要的前体物质，而 FPP 则是 tRNA 异戊二烯化及苯醌和细胞壁生物合成中的重要物质。

J. D. Keasling和他的研究小组不仅对每个基因进行突变筛选，同时还对不同基因的突变型进行组合，以挑选出最优搭配。——读者可以尝试估量一下这一步的工作量：即使仅为每个基因制造两种突变型，所需合成的途径就多达 512 条，所需合成的核苷酸更是高达 600 万个！对于传统 DNA 合成技术而言，这项任务无疑难于登天。然而利用合成生物学标准化的操作手段，工作量可大大降低。

2006 年，在此项成果的基础上，研究人员修改了他们的最初方案，将酿酒酵母菌引入到这项研究中来。利用工业酿酒酵母中工程化的甲羟戊酸途径、取自青蒿（*Artemisia annua*）的 amorphadiene 合成酶和细胞色素 P450 单加氧酶（cytochrome P450 monooxygenase，CYP71AV1）共同作用来生产高浓度、更直接的青蒿素前体衍生物——青蒿酸（artemisinic acid）。

利用工业酵母生产青蒿酸共分三个步骤，如图 5-6 所示：

（1）优化 FPP 生物合成途径以提高 FPP 的产量。

（2）从青蒿中引入紫穗槐二烯合成酶（amorphadiene synthase，ADS）基因，其表达的产物将 FPP 转化为 amorphadiene。

（3）通过比较基因组学分析得到来自于青蒿的细胞色素 P450 单加氧酶——CYP71AV1/CPR，克隆表达后实现 amorphadiene 到青蒿酸的三步氧化还原反应。

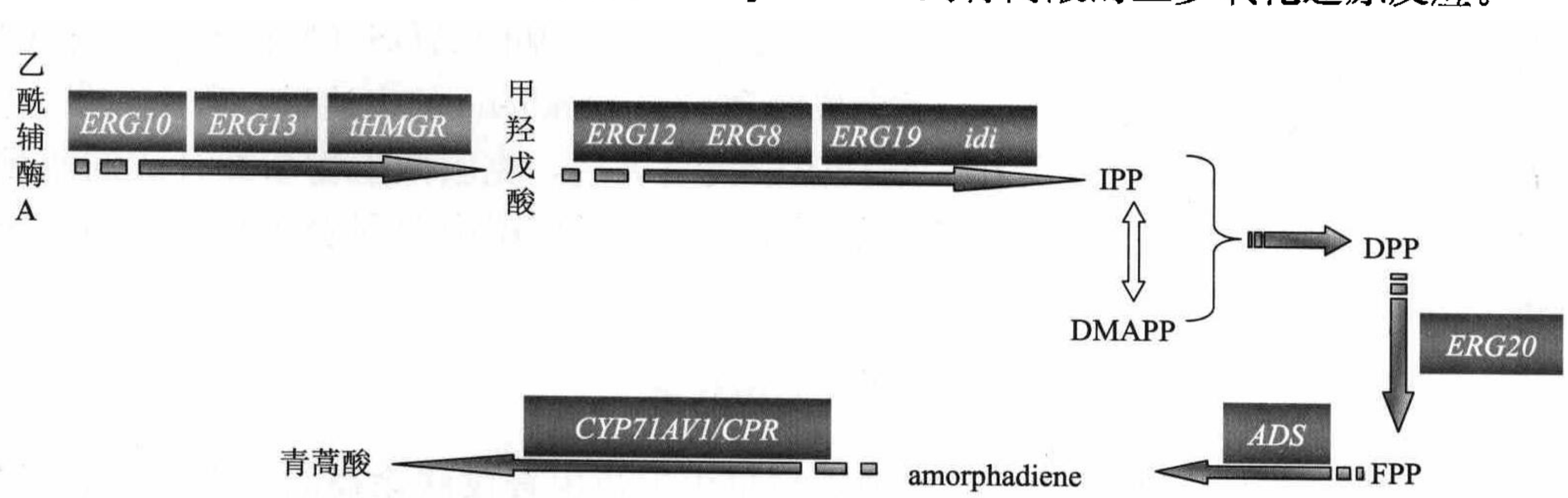


图 5-6 人工构建酿酒酵母的青蒿酸合成途径

在酿酒酵母中通过甲羟戊酸途径将乙酰辅酶 A 转化成 FPP；从青蒿中引入紫穗槐二烯合成酶基因，将 FPP 转化为 amorphadiene；通过比较基因组学分析得到来自于青蒿的细胞色素 P450 单加氧酶-CYP71AV1/CPR，克隆表达后实现 amorphadiene 到青蒿酸的三步氧化还原反应而得到青蒿酸。

在这三步中，每一步的优化都可以提高最终的青蒿酸产量，而 FPP 代谢途径的优化则效果更为显著，因此研究者主要从以下几个方面着手。

（1）过量表达 tHMGR（3-hydroxy-3-methylglutaryl-coenzyme A reductase）。tHMGR 是胆固醇合成的限速酶，可有效限制 FPP 向固醇的转化，降低 FPP 在其他方面的不必要消耗。

（2）通过甲硫氨酸可抑制启动子（P_{MET3}）下调 ERG9 编码角鲨烯（squalene）合成酶，阻断 FPP 向下合成固醇的支路，进一步避免 FPP 的不必要消耗。

（3）过表达 *upc2-1* 并结合 ERG9 的下调。半显性变异等位基因 *upc2-1* 能够增强 UPC2 的活性。UPC2 是酿酒酵母中调节合成固醇的全局性转录因子，虽然其对于 amorphadiene 产量的影响似乎并不明显，但实验表明，当过表达 *upc2-1* 并结合 ERG9

的下调时，amorphadiene 的产量可提高至 105mg/L。

（4）将所有被修饰基因都整合到染色体上，确保宿主菌中基因表达的稳定性。

（5）降低细胞密度，确保养分充足并避免不必要的互相干扰。

综合上述，所有修饰和调整的新菌株 EPY224 中 amorphadiene 的产量可达 153mg/L，比之前的相关报道提高了近 500 倍。

除了上述工作外，青蒿酸的提取和纯化也是一个至关重要的步骤。通常在酵母菌的酸性培养基条件下，青蒿酸被质子化而很难排出细胞外，大大影响了实际收益。J. D. Keasling 等开发了新的纯化方法，即用醚将青蒿酸从洗脱液中提取出来，再利用硅酸凝胶进行柱状层析分离，得到的青蒿酸纯度可大于 95%。在 1L CO_2 生物反应器中，能够生产 115mg 青蒿酸，利用该方法可以提取出其中的 76mg。

相比于常规的青蒿植物提取，4.5%干重的酵母能够生产 1.9%的青蒿酸和 0.16%的青蒿素，耗时仅仅 4～5 天，而植物提取却要几个月！而且，相比于植物提取青蒿素，微生物合成青蒿酸的方法不会受到诸如气候等因素的影响，更加可靠、廉价，纯度更高。此外，这种新改良的大肠杆菌和酵母菌还可被改造用来生产其他类异戊二烯类化学物质，任何一家公司均可以对这些工程菌做相应的改造，加入一定数量与目标产物合成有关的基因，即可得到任何一种类异戊二烯物质，用于其他疾病的治疗和其他物质的生产。由此可见，J. D. Keasling 等的研究成果无疑是合成生物学在实际应用方面的一个里程碑。

5.2　合成生物学应用于生产生物能源

世界经济的现代化是建立在化石能源基础之上的，得益于石油、天然气、煤炭等化石能源的广泛投入和应用。虽然市场的调控力量和石油生产技术的改进可使石油供应继续保持到 21 世纪，然而，不可避免的，这一经济的资源载体将以比人类设想还要快的速度迅速地接近枯竭。化石能源与原料链条的中断，必将导致世界经济危机和冲突的加剧，最终葬送现代市场经济！为了避免由于能源危机而引起的战争和冲突，维系经济发展，挽救人类社会，除了开发各种节能技术之外，太阳能、地热能、风能、海洋能、核能及生物质能等存在于自然界中的能源的开发利用是解决能源危机的重要途径。下面我们就来介绍几种常规的液体燃料及其可能的替代品。

汽油是与人类生产生活密切相关的液体燃料之一，是一种复杂的碳氢化合物混合物，主要由 $C_{4\sim12}$ 的各种直链、支链和环状烷类、芳香族化合物等组成，辛烷值为 91～99。当前在各种生物燃料中，乙醇（辛烷值为 129）燃料无疑已经成为替代汽油等液体燃料的焦点，甚至连各大汽车生产商也纷纷顺应时代的潮流推出新型的乙醇驱动力车型。然而，由于乙醇的腐蚀性和高吸湿性使其并不能成为现有燃料基础结构的理想燃料[200,201]，而且很难用现有的输油管线输送，运输不方便；乙醇和水的高混合性使其很难从发酵液中提纯；乙醇的能源含量只有汽油的 70%，能量密度太低；乙醇生产时消耗人类大量的“口粮”等，上述原因大大限制了乙醇燃料的发展。由此接踵而上的则是丁醇等高级醇生物燃料的开发利用。

丁醇的分子中比乙醇多两个碳原子，其能源含量比乙醇高大约 40%；其辛烷值为 96，比乙醇低，但和汽油差不多。

（1）与乙醇相比，丁醇不易溶于水，不仅能做汽油的添加剂，还能单独作为常规引擎的燃料。

（2）丁醇的腐蚀性小。一般生物燃料会腐蚀发动机的输油管和垫圈，因此目前大部分的汽车发动机只允许使用含 5%生物燃料和 95%普通汽油或柴油的混合燃料，而丁醇在混合燃料中的混入量可达 20%。

（3）丁醇不像乙醇对于汽油混合物中的水分那么敏感，使用丁醇作为驱动能源无须改造输油管和发动机。

（4）丁醇是一种高能量生物燃料，与传统的汽油燃料相比，相同体积的丁醇能使汽车多走 10%的路程；与乙醇相比，公里数更是增加到 30%。

通常当分子构成中含有甲基链或者双键时，其辛烷值会增加。因此，支链的 C_4 和 C_5 醇类也可以作为汽油的替代物。例如，异丁醇（2-甲基-1-丙醇）的性质与丁醇相似，但比丁醇的辛烷值高；异戊醇、异戊烯醇等的辛烷值均略高于 90，且比丁醇的能源含量高；异辛烷等支链、短链烷类也是很好的汽油替代品。这些醇类理论上可以利用类异戊二烯代谢途径进行生物合成[202]或通过转化氨基酸来得到[201]，但前提是要对生物现有代谢途径进行大规模改造，因此实现起来难度较大。

除了汽油，柴油是另一种与人类生产生活密切相关的燃料。石油柴油的碳链长度为 9～23；生物柴油通常由脂肪酸甲酯（fatty acid methyl ester）组成，碳链长度为 12～22，含有 2 个以内的双键。虽然相比于石油柴油，生物柴油浊点（cloud point）和倾点（pour point）较高，不适宜管道运输，并且其能源含量比石油柴油低 11%，但生物柴油具有更优越的辛烷值、高润滑性和低污染物排放等优势，因此仍然是备受瞩目的替代能源。生物柴油也可以通过植物中的类异戊二烯途径来生产。线性或环状单萜和倍半萜烯也是生物柴油的来源[203]。

目前，应用最广泛的生物能源是从淀粉（玉米）或甘蔗中获得的乙醇以及从植物油或动物油中获取的柴油[204,205]。这不仅使粮食和饲料作物本就十分紧缺的现状更加雪上加霜，并且这种生产过程本身就是一个大量消耗能量的过程，毫无经济性可言。与药物和酶等产品不同，只有当价格低于或等于石油燃料时，生物燃料才有可能具有竞争力，因此生物燃料的前景完全取决于成本降低的前景。

与农作物生物燃料相比，来自可再生木质纤维素的微生物燃料有许多优势[203]。

（1）不会消耗本已十分紧缺的常规粮食作物，如玉米、甘蔗、大豆和棕榈油。

（2）木质纤维素是地球上最丰富的生物高聚物分子，不用担心原料的来源。

（3）能够通过工程化手段构建新的生物合成途径生产化石燃料的替代品，如短链、支链和环状醇类，以及烷类、烯烃、酯、芳香族化合物等。

因此，尽管缺乏相应的模式生物，难以优化代谢途径，难以平衡工程微生物的氧化还原态，且酶的低活性、燃料的抑制效应和上游生物质处理步骤的副产品的存在等因素限制产量的提高，但开发低成本、高能效的将木质纤维素转化成燃料的方法仍然被提上了日程，成为目前的研究焦点[206]。

2007 年的美国能源独立与安全法案（US Energy Independence and Security Act）要求到 2022 年必须有 1620 亿升的可再生能源。据估计，仅有 675 亿升能够从玉米乙醇（corn ethanol）中获得，其余的则是利用先进技术从替代的木质纤维中获得的生物燃料[207]。

充分利用合成生物学的方法，结合代谢工程和遗传工程等手段，最近，此方面的研究已经取得了一些可喜的进展，一些潜在的生物能源的合成途径在模式生物中得以表达。例如，在大肠杆菌中表达了取自丁基梭菌 *Clostriduim* 的合成异丙醇[208,209]和丁醇的相关基因[210]；除了在发酵阶段生产乙醇，工程化的酿酒酵母能够利用氨基酸高效生产乙醇和酯[211~213]；在大肠杆菌中表达 3-甲基-1-丁醇[214]；利用蓝细菌和水藻等生物合成生物柴油等[215]。

然而，合成代谢途径的加入会破坏细胞原有的代谢平衡，导致不期望的副产品的产生并降低期望代谢物的产量；ATP/ADP、NAD^+/NADH、$NADP^+$/NADPH 和 acyl-CoAs 等与细胞活性和最终生理特性密切相关的全局水平的代谢物影响和制约着合成代谢途径的表达量；新途径也需要氨基酸、氧化还原辅助因子和能量，给细胞增加了新的负担[200,203,204,216]，所有这些问题都需要妥善解决。为了预测新代谢途径对细胞生长和产量的影响，研究者开发了许多代谢模型帮助代谢途径的优化和筛选，如大肠杆菌的生化模型[217]和酿酒酵母的代谢模型等[218]，以及一些基因组水平的模型[219]、随机模型等。

虽然由于模型参数的不确定性，很多模型的解空间都比较大，然而这些模型仍然取得了一些重要的成果。

（1）大肠杆菌在不同碳源下进化时，预测其表型空间[220]。

（2）预测高通量的网络拓扑结构[221]。

（3）在利用取自 *Pichia stipitis* 的编码木糖还原酶（xylose reductase，Xyl1p）和木糖醇脱氢酶（xylitol dehydrogenase，Xyl2p）的基因对酿酒酵母进行工程化使其能够利用木糖合成乙醇时，在代谢模型分析的建议下，移除依赖于 $NADP^+$ 谷氨酸脱氢酶 GDH1 和过表达依赖于 NAD^+ 的 GDH2，增加了乙醇产量并降低了副产品的合成[222,223]。

（4）利用基因组模型指导在酿酒酵母中插入甘油三磷酸脱氢酶（glyceraldehyde-3-phosphate）基因使甘油副产品的量降低了 58%、木糖醇副产品的量降低了 33%，而乙醇的量却提高了 24%[224]。

利用工程化的大肠杆菌非发酵代谢途径生物合成高级醇[225]

如前所述，由于乙醇的低能量密度和高吸湿性，使其并不能成为理想的燃料。与此不同，高级醇却具有接近于汽油的能量密度、低吸湿性和不易挥发的性质。目前许多研究者正在致力于研究这些高级醇的生物合成。下面我们就来讲一下 J. C. Liao 研究组利用大肠杆菌生产高级醇的方法。

为了避免异源代谢途径的引入所引起的代谢失衡，防止具有细胞毒性的异源代谢物的过量累积，同时为了达到高产异源目的产物的目的，必须针对特定宿主寻找恰当的代谢途径。为此，S. Atsumi 等研究者充分利用大肠杆菌现有的代谢能力结合取自其他生

物的 2-酮酸（2-keto acid）降解 Ehrlich 途径的最后两步代谢相关基因。2-酮酸是氨基酸生物合成途径的中间物质，能够被 2-酮酸脱羧酶（2-keto-acid decarboxylase，KDC）转化成醛，继而被醇脱氢酶（alcohol dehydrogenase，ADH）转化成醇。采用这种策略，仅需两步异源代谢步骤即可将氨基酸生物合成的中间代谢物分流出来生产醇类。

氨基酸的生物合成途径能够产生多种 2-酮酸。S. Atsumi 等利用了其中的 6 种，如图 5-7 所示：

（1）异亮氨酸生物合成途径产生的 2-酮丁酸（2-ketobutyrate）和 2-酮-3-甲基戊酸（2-keto-3-methyl-valerate）可以被分别转化成 1-丙醇和 2-甲基-1-丁醇。

（2）缬氨酸生物合成途径产生的 2-酮-异戊酸可以作为异丁醇的前体物质。

（3）亮氨酸生物合成途径产生的 2-酮-4-甲基-戊酸是 3-甲基-1-丁醇的底物。

（4）苯丙氨酸生物合成途径产生的苯丙酮酸（phenylpyruvate）可用于合成 2-苯乙醇。

（5）正缬氨酸的生物合成途径是亮氨酸生物合成的副反应，产生 1-丁醇和 2-酮戊酸的底物。

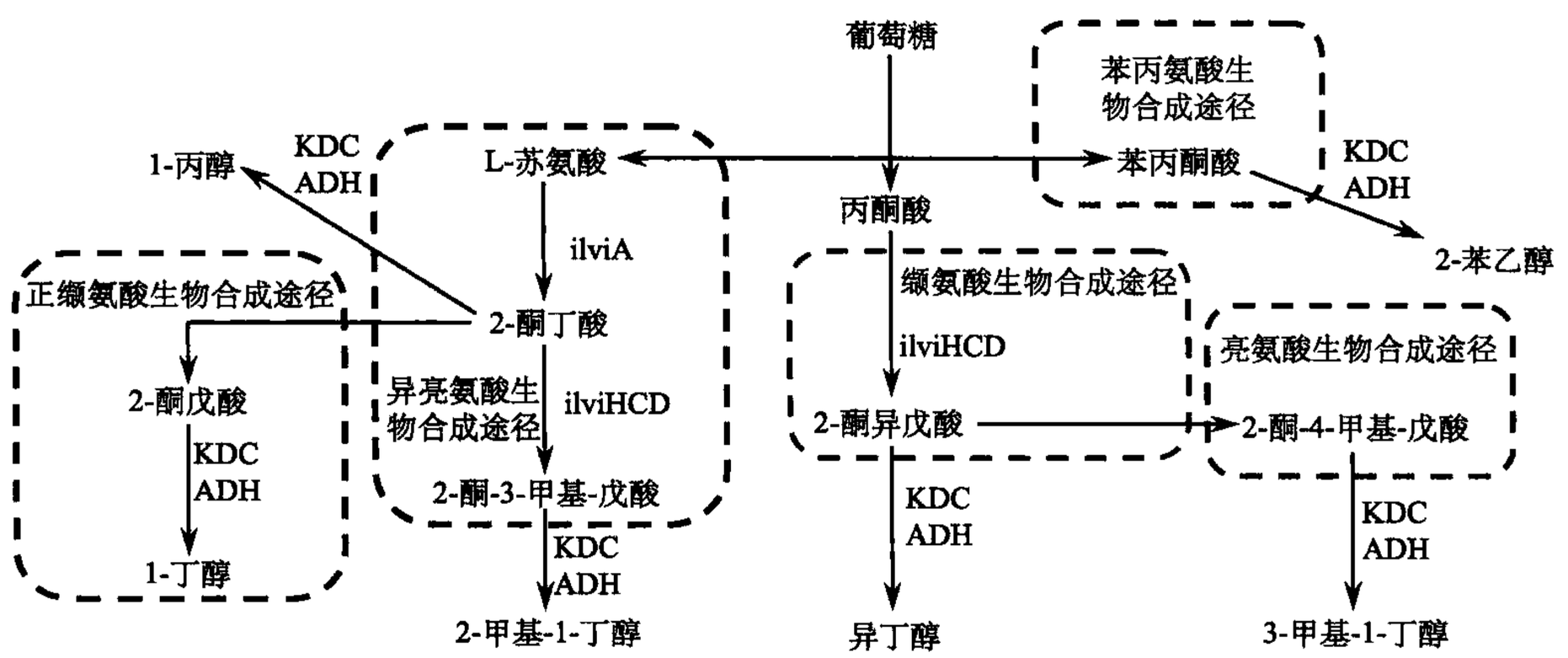


图 5-7　大肠杆菌醇类非发酵生物合成网络示意图

S. Atsumi 等的工作包括以下几部分。

1. 检验酶的活性和转化能力

在醇类合成的过程中，KDC 是关键酶，这种酶在植物和酵母中很常见，但在细菌中却比较少见。某些 KDC 的底物范围很广泛，而某些却具有特异性。为了检验大肠杆菌中不同 KDC 的活性和对内源 2-酮酸底物的转化能力，除了过表达酵母菌的醇脱氢酶 2（alcohol dehydrogenase 2，Adh2），确保将醛转化成醇外，还要过表达 5 种 KDC，分别是：①酿酒酵母的 KDC（Pdc6、Aro10、Thi3）；②乳酸乳球菌（*Lactococcus lactis*）的 Kivd；③丙酮丁醇梭菌（*Clostridium acetobutylicum*）的 Pdc。

所有实验中，只有微量的醛，说明 Adh2 具有足够的活性，能够将大部分被 KDC 转化的醛再转化成醇。气相色谱-质谱（GC-MS）分析表明：表达 Kivd 和 Aro10 的大肠杆菌

能够产生所有期望的醇类，Thi3 无任何活性，Pdc6 和 Pdc 介于中间。由此可见，Kivd 是活性最强、功能最全的脱羧酶。为此在后续实验中，选择 Kivd 作为 KDC 酶。

2. 提高中间代谢物 2-酮酸的表达量

如前所述，为了获得最高的代谢物表达量，需要对异源代谢途径进行修饰，协调相关基因的表达。本例中 2-酮酸是关键的中间代谢物，提高 2-酮酸的表达量有利于提高目的醇的产量。为此：

（1）利用 Pl_{lacO1} 启动子控制 *ilvIHCD* 基因并使其过表达以加强 2-酮异戊酸的生物合成。将 *ilv* 途径与 Kivd 和 Adh2 的醇类合成途径相结合生产异丁醇。结果表明，*ilvIHCD*过表达的菌株生产的异丁醇的产量为 23mmol/L，比普通菌株的产量提高了将近 5 倍。

（2）为进一步提高异丁醇的产量，将 *adhE*、*ldhA*、*frdAB*、*fnr* 和 *pta* 等对合成副产品有关的基因删除，降低丙酮酸的不必要消耗，提高 *ilvIHCD* 代谢通路所需的丙酮酸的水平，异丁醇的产量提高到了 30mmol/L。

（3）为提高异丁醇的转化效率，用取自 *Bacillus subtilis* 的 *alsS* 基因代替大肠杆菌的 *ilvIH* 基因。*alsS* 基因对丙酮酸具有高亲和性，而 *ilvIH* 却对 2-酮丁酸具有高亲和性。具有 *alsS* 代谢通路的菌株的异丁醇产量又提高约 1.7 倍，为 50mmol/L。

（4）删除 *pflB* 基因，进一步降低对丙酮酸的竞争。

所有这些措施使得在微好氧条件下，异丁醇的最终产量可达到 300mmol/L（22 g/L），在 40～120h，异丁醇的产量为 0.35g/g 葡萄糖，达到理论最大值的 86%。作为对照试验，在无葡萄糖培养的情况下，此菌株产生的异丁醇却几乎可以忽略不计。

3. 生物合成 1-丁醇

为了证明此方法的通用性，研究者还用其生产 1-丁醇。某些梭菌在发酵生长阶段能够产生 1-丁醇，其合成途径中许多酶对氧非常敏感且依赖于辅酶 A。在无 1-丁醇发酵代谢途径的大肠杆菌中过表达 Kivd 或 Aro10，则在非发酵生长阶段，大肠杆菌能够表达少量的 1-丁醇，说明在细胞内有 2-酮酸的前体（2-酮戊酸）存在。对于大肠杆菌而言，2-酮戊酸并不是常见的代谢物。为了提高合成 2-酮戊酸的量而提高 1-丁醇的产量，研究者引入了具有广泛底物特异性的 leuABCD 途径，其中的底物之一即是 2-酮异戊酸。研究者尝试利用 2-酮丁酸（2-ketobutyrate）通过类似于亮氨酸生物合成的步骤来合成 2-酮异戊酸。*ilvA* 基因编码的苏氨酸脱水酶（threonine dehydratase）能够将 L-苏氨酸转化成 2-酮丁酸，或者采用 *Leptospira interrogans* 和 *Methanocaldococcus jannaschii* 中的代谢途径也可获得 2-酮丁酸。在后者体内，可通过异丙基苹果酸异构酶（isopropylmalate isomerase，LeuCD）和 β-异丙基苹果酸脱氢酶（β-isopropylmalate dehydrogenase，LeuB）将甲基苹果酸（citramalate）转化成 2-酮丁酸。

因此，为了产生 1-丁醇，研究者构建了由 Pl_{lacO1} 启动子控制的编码 *ilvA-leuABCD* 代谢途径的操纵子。实验结果表明，具有 *ilvA-leuABCD* 途径的菌株能够产生

0.6mmol/L 的 1-丁醇，是非过表达此途径的菌株产量的 3 倍。

可以通过以下途径进一步提高 1-丁醇产量。

（1）采用高表达 L-苏氨酸的大肠杆菌菌株。

（2）将 *ilvD* 基因删除。该基因编码的二羟酸脱水酶能够同时帮助制造 2-酮异戊酸（亮氨酸和缬氨酸的前体）和 2-酮-3-甲基戊酸（异亮氨酸的前体）。将其敲除能够移除 *leuABCD* 途径的 2-酮异戊酸底物，减少竞争性底物的抑制作用，同时也移除了 Kivd 的竞争底物——2-酮-3-甲基-戊酸和 2-酮-4-甲基-戊酸。

（3）提高 *leuABCD* 向 2-酮丁酸的代谢途径的活性、提高 Kivd 对 2-酮戊酸的特异性。因为 2-酮丁酸也是 1-丙醇的底物，增加 2-酮丁酸也同时增加了 1-丙醇的产量。

4. 提高大肠杆菌对异源高级醇类的耐受性

相对于 1-丁醇而言，异丁醇对大肠杆菌的毒性要小些。为了使大肠杆菌能够更好地生产高级醇，必须提高其对高级醇毒性的抵抗能力。为此，研究者更换不同的培养基，发现在更换了 5 次异丁醇浓度不断增加的培养基后，野生型大肠杆菌 JCL16 发生了突变，异丁醇耐受性从 1.5%（*w*/*V*）提高到了 2%（*w*/*V*）。这一耐受性已经能够与 1-丁醇内源宿主的耐受性相匹敌。

由上述的介绍可以看出，J. C. Liao 课题组的研究成果充分利用了氨基酸合成途径的中间代谢物，避开了通常的利用辅酶 A 介导的化学合成，并且可以合成其他高级、复杂的醇类。这一方法还可用于酵母等工程微生物中，具有非常广泛的开发和应用前景。

5.3 合成生物学应用于环境治理

砷离子的检测

砷（As），又名砒，是一种广泛分布于自然界的非金属。在土壤、水、矿物、植物中都能检测出微量的砷。正常人体组织中也含有微量的砷。砷及砷化物在工农业中有着广泛的用途：农业上常用它们杀虫、毒鼠和灭钉螺；工业生产中砷及其化合物也常用于毛皮生产的消毒、防腐、脱毛，以及用作玻璃工业中的脱色剂等。然而在造福人类的同时，砷也给人类的健康带来了很大威胁。砷不溶解于水，是没有毒性的，但是砷化物，特别是三氧化二砷（As_2O_3，俗称砒霜），却是剧毒的。通常说的砷中毒，实际上是砷化物（主要是 As_2O_3）中毒。砷对黏膜具有刺激作用，可直接损害毛细血管。经黏膜或皮肤吸收的砷化物，主要沉积在毛发、指甲、骨、肝和肾等器官，破坏细胞的氧化还原能力，影响细胞正常代谢，引起组织损害和机体障碍，甚至可直接引起中毒死亡。此外，砷化物对胚胎发育也有一定的影响，可致畸胎。摄取砷化物还有可能增加患膀胱癌、皮肤癌、肺癌和其他肺部病变的机率，且部分病变的潜伏期可能长达数十年。

由于砷的广泛存在和使用，在环境化学污染物中，砷成为了最常见、危害居民健康

最严重的污染物之一。特别是随着现代工农业生产的发展，砷对环境的污染日趋严重，消除砷的污染已经成为当前环境保护的重要任务之一。剑桥大学的研究人员称，目前有 70 多个国家出现了砷污染问题。如果按照欧洲与北美洲的饮水标准来衡量，全球至少有 1.4 亿人使用的水源含砷量过高，很有可能造成严重的健康问题。

目前最常用的 Gutzeit 氏砷检测法为半定量检测法，即用无砷锌粒作还原剂，在一定浓度酸性溶液（盐酸）介质中进行还原，生成的气体从溶液中逸出。该法分析时间长、误差大（15%～35%）、费用昂贵、不易维护。最为重要的是，该方法检测的灵敏度低，最低检出限值约为 50ppb*，而世界卫生组织规定的饮用水中砷离子浓度不能超过 10ppb，当水中的砷离子浓度达到 30ppb 的时候就会对人体有显著伤害，引发一系列砷中毒现象。

为此，英国爱丁堡大学的研究者利用具有感知水中砷离子功能的启动子，设计了相应的生物感应器。该启动子在有砷离子存在的情况下能改变细胞的代谢反应，并最终改变溶液的 pH。理论上，此种砷离子生物检测器的最低砷离子检测限为 5ppb。

利用微生物作为检测器，不仅可以大大降低成本，且较之其他砷离子检测器而言，利用 pH 的变化作为砷离子存在与否的信号很容易被 pH 检测仪甚至更方便的 pH 试纸检测到，使仪器的实用性大大提高。

虽然此设计目前还处于实验室研究阶段，并未完成真正构建，在实践过程中还有一些具体困难需要克服，但其巧妙而新颖的设计为解决环境中砷污染的检测问题提供了新的思路，也为生物传感器（biosensor）的发展提供了新方法，因此本书接下来会详细讲解其设计思路供读者借鉴和参考。

爱丁堡大学设计的砷离子生物检测器系统如图 5-8 所示[226]。

整个系统只有在砷离子或乳糖存在时才有响应。装置 1 中的启动子 3 为组成型常开启动子，下游 *lacI* 基因表达的产物 LacI 抑制启动子 4 的启动。

（1）当环境中只有乳糖存在而无砷离子存在时，乳糖与 LacI 结合，解除 LacI 对启动子 4 的抑制，启动子 4 顺利启动，其下游的尿素酶基因表达产生尿素酶。尿素酶催化尿素转化为氨和二氧化碳，环境的 pH 升高，其值约为 9～10。

（2）当环境中的砷离子浓度较低时（如 5ppb），响应低浓度砷离子的高敏感装置 2 启动。砷离子与 ArsR 结合，解除其对启动子 2 的抑制，下游的 *λcI* 基因顺利表达，产生的 CI 蛋白抑制装置 1 的启动子 4，关闭尿素酶基因的表达，阻止其对尿素的催化作用。此时由于没有相应酸或碱的产生，系统的 pH 维持在中性状态（由于少量残余尿素的存在，pH 可能会略高于中性），即 pH 约为 7。

（3）当环境中砷离子浓度较高时（如 20ppb），响应高浓度砷离子的装置 3 启动。砷离子与 ArsD 结合，解除其对启动子 1 的抑制，下游的 *lacZ* 基因顺利表达，*lacZ* 基因编码的 β-半乳糖苷酶催化乳糖的发酵，使其分解为乙酸和乳糖酸，此时 pH 约为 4.5。

---

* parts per billion，1ppb=10^{-3}ppm=1×10^{-9}，此处指质量浓度。

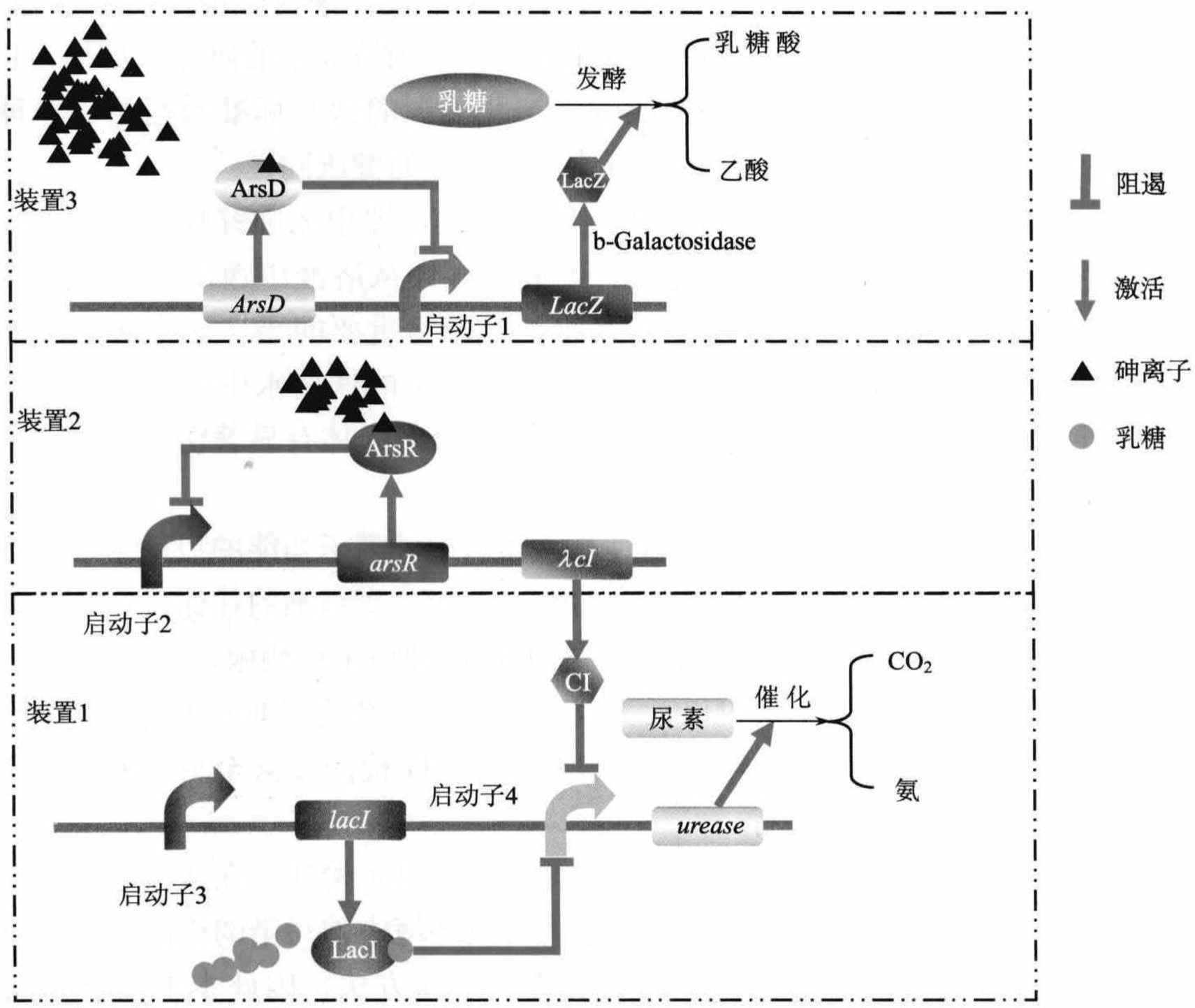


图 5-8 微生物砷离子检测器系统图

装置 1 中的启动子 3 为常开启动子，下游 *lacI* 基因表达的产物 LacI 抑制启动子 4 的启动。当环境中只有乳糖而无砷离子时，乳糖与 LacI 结合解除其对启动子 4 的抑制，启动子 4 顺利启动，下游的尿素酶基因表达产生尿素酶，催化尿素转化为氨和二氧化碳，环境的碱性增加，其 pH 约为 9～10；当环境中的砷离子浓度在 5ppb 左右时，响应低浓度砷离子的高敏感装置 2 启动。砷离子与 ArsR 结合，解除其对启动子 2 的抑制，下游的 *λcI* 基因顺利表达，产生的 CI 蛋白抑制装置 1 的启动子 4，关闭尿素酶基因的表达，尿素未被催化。因此没有相应酸或碱的产生，系统的 pH 维持在中性状态，即 pH≈7；当环境中砷离子浓度较高时（例如 20ppb），响应高浓度砷离子的装置 3 启动。砷离子解除 ArsD 对启动子 1 的抑制，下游的 *lacZ* 基因编码的 β-半乳糖苷酶催化乳糖发酵，使其分解为乙酸和乳糖酸，环境中酸性增加，pH 在 4.5 左右。

5.4 合成生物学应用于其他方面

5.4.1 大肠杆菌成像系统[227]

除了应用于上述人类最关注的健康、环境、能源三方面的课题研究外，合成生物学的研究还涉及人类生活和科学研究的其他方面。

2005 年前后，美国加利福尼亚大学旧金山分校（UCSF）的 C. A. Voigt 课题组利用合成生物学手段在生物胶片方面取得了突破性进展，获得了人类继化学成像、电子学

成像之后的第三种成像方式——微生物成像。

C. A. Voigt 课题组开发的这种微生物成像系统包含一种人工合成感应激酶，使大肠杆菌菌苔能够发挥生物胶片的功能，通过光照产生分辨率高达 100M 像素/in^2* 的二维化学图像。不仅如此，这种对于微生物基因表达的时空调控，能够“印制”出复杂的生物元素，可用于研究微生物特定空间和时间控制下磷酸化步骤的信号转导途径等多种复杂的代谢途径。为此，年仅 30 岁的助教 C. A. Voigt 被世界著名科技杂志 *Technology Review* 评为 2006 年 35 名 35 岁以下（TR35）重大技术创新奖获得者之一。下面我们就来详细讲解一下 C. A. Voigt 课题组的“大肠杆菌成像系统”。

众所周知，一个成像系统至少要包括感光和显影两大子系统。为此，研究者构建了感受光照刺激的光感应器和调控遗传线路响应的应答调节子，利用光感应器来接收光照刺激，将刺激信号传递给应答调节子，通过应答调节子对遗传线路进行调控从而产生能够被人类视觉感官观察到的色彩，进而成像。

1. 光感应器的构建

光敏色素（phytochromes）也称作“感光蛋白”，是由结合在细胞膜上的胞外感受器和胞内应答调节子组成的一个双组分系统，其单体包括一个生色团（chromophore）和一个脱辅基蛋白（apoprotein）。自然界中的植物和某些微生物通常是利用光敏色素来感受光照刺激的，以此来控制自身的趋光性、光合作用和保护色的分泌等。然而大肠杆菌中并没有感光蛋白，无法对光照刺激做出响应。因此研究者将取自 *Synechocystis* 的藻胆青素（phycocyanobilin，PCB）合成基因（*hol* 和 *pcyA*）构建在质粒上转入大肠杆菌，这两个基因表达的产物能够将血红素转化为藻胆青素（藻胆青素是藻胆蛋白光敏色素的生色团，能够接受光照刺激）。有了含有这两个基因的质粒的帮助，大肠杆菌就有了能够感受光照刺激的 PCB，也就具有了感受光照刺激的功能。

2. 应答调节子

然而，仅仅能够感受光照刺激还远远不够，还要对这种刺激做出响应才能成像。PCB 虽然能够感受光照刺激，但与其交联的脱辅基蛋白并不具有 DNA 结合域，无法调控基因的表达产生用来“显影”的化合物。因此还需构建应答调节子以便执行基因表达调控的功能。大肠杆菌中这种接受外界刺激信号做出响应的功能通常是由各种各样的“双组分系统”来完成的。其中，研究比较透彻的是广泛存在的“渗透压感应器”EnvZ-OmpR 双组分系统。这种双组分系统能够调节膜孔蛋白（porin）作为对渗透压刺激的响应。EnvZ-OmpR 系统由 EnvZ 和 OmpR 两种蛋白质组成，前者是一种组氨酸激酶，在高渗环境下能发生自身磷酸化，起感应器的作用；后者是反应调节子，含有天冬氨酸残基，能接受来自 EnvZ 的磷而被磷酸化，磷酸化的 OmpR 可作为转录因子而将来自 EnvZ 的信号输出，对 *ompC* 基因的启动子进行调控。

---

* $1in^2 \approx 6.45cm^2$

研究者通过特殊的处理和筛选手段将 EnvZ 的组氨酸激酶结构域与光敏色素脱辅基蛋白 Cph1 进行交联，组成 Cph-EnvZ 嵌合体光感应系统。由于连接感光器和应答调节子的肽链的长度和组成会影响信号的转导过程，因此研究者构建了一系列具有不同连接长度的嵌合体，并对其进行筛选和优化，挑选出其中响应最强烈的 Cph8 嵌合体作为最终实际应用的感光系统。

3. 成像器

为了使这个感光系统产生“可视化”的图像效果，研究人员选用了特殊的大肠杆菌菌种，同时引入了 *lacZ* 基因，将 *lacZ* 基因与大肠杆菌染色体上的 *ompC* 基因做成融合基因，由 OmpR 蛋白控制这个融合基因的表达。*lacZ* 基因编码的 LacZ 蛋白能够催化 S-Gal（一种稳定的、可溶于水的 DNA 染色剂）产生黑色化合物。

由此整个成像系统的所有部分构建完成，如图 5-9 所示。

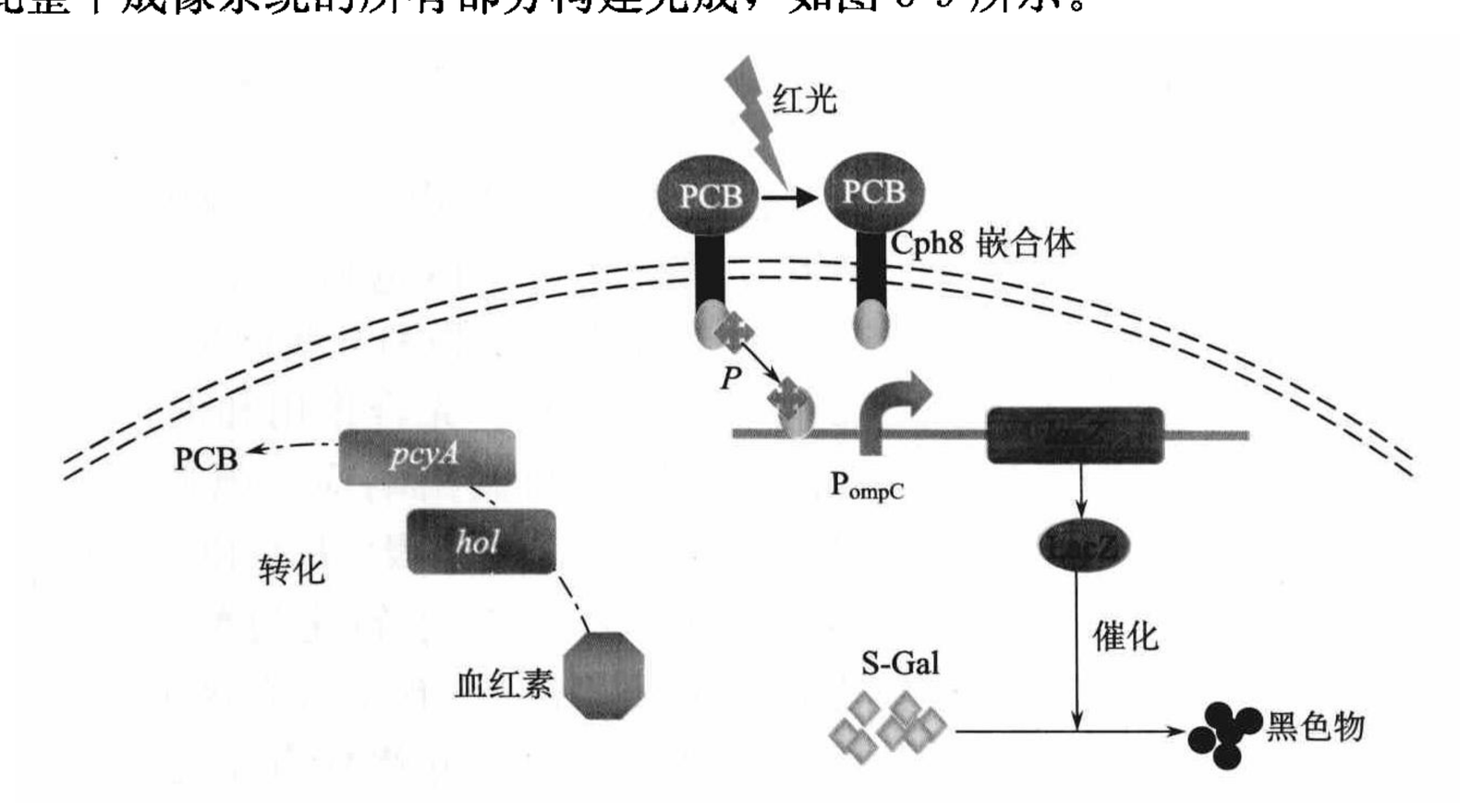
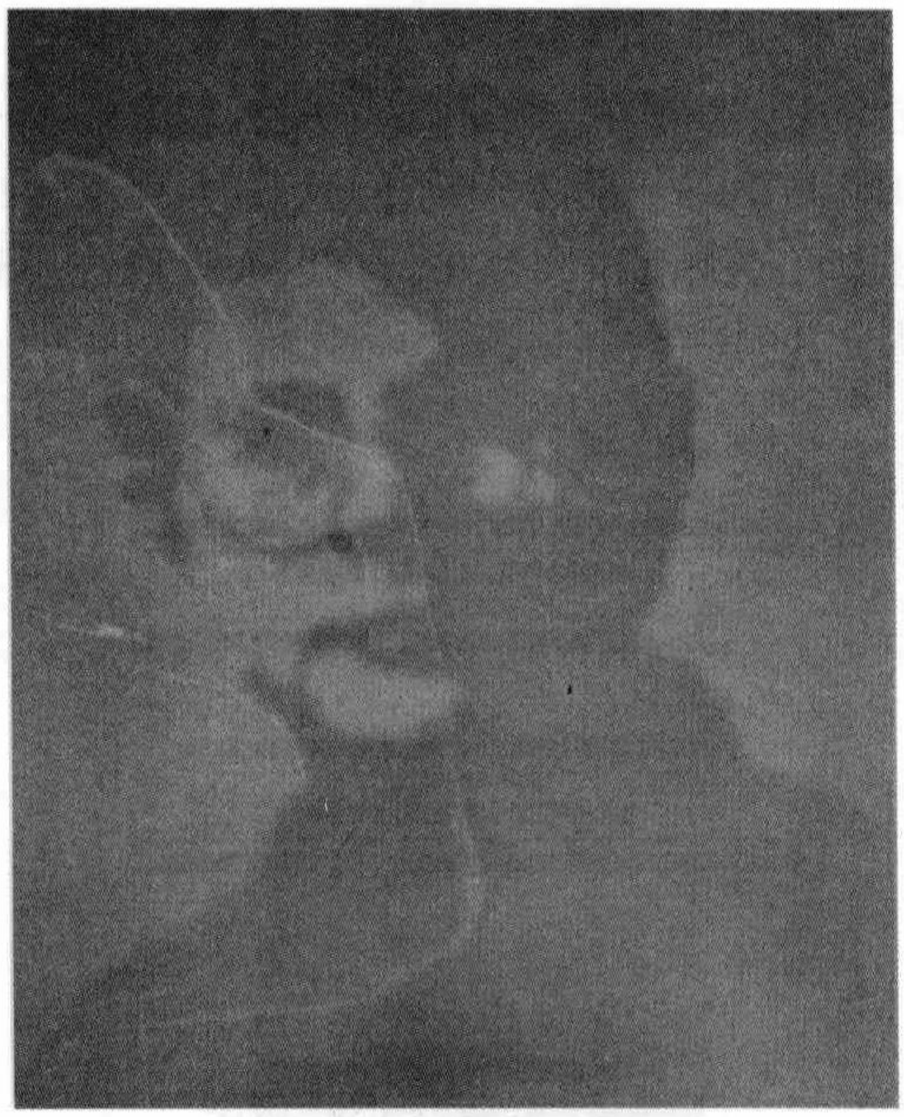


图 5-9 大肠杆菌成像系统示意图

新构建的 Cph8 嵌合体由光敏色素 Cph1（蓝色，PCB 是其组成部分）和 EnvZ-OmpR（橙色）双组分系统的组氨酸激酶结构域和响应调节域组成。红色光照能够抑制 Cph8 中 EnvZ 蛋白的自磷酸化过程（右侧嵌合体所示），EnvZ 蛋白通过使 OmpR 蛋白也无法磷酸化，从而将 P_{ompC} 关闭，作为融合基因的 *lacZ* 基因无法表达，此区域中无黑色化合物产生为菌苔原有的颜色；相反，无光照的区域，EnvZ 的自磷酸化过程顺利进行（左侧嵌合体所示），通过 OmpR 蛋白开启 *ompC* 基因的启动子，*lacZ* 基因顺利表达，编码的 LacZ 蛋白催化事先加入培养基中的 S-Gal 染色剂产生黑色化合物，此区域为黑色。

当有外界红色光照刺激时，感光蛋白接受光照刺激并抑制 EnvZ 蛋白的自磷酸化过程，EnvZ 蛋白使 OmpR 蛋白也无法磷酸化，从而将 *ompC* 基因的启动子关闭，作为融合基因的 *lacZ* 基因也就无法表达，此区域中无黑色化合物产生；相反，无光照的区域，EnvZ 的自磷酸化过程顺利进行，通过 OmpR 蛋白的磷酸化开启 *ompC* 基因的启动子，*lacZ* 基因顺利表达，编码的 LacZ 蛋白催化事先加入培养基中的 S-Gal 染色剂产生黑色化合物，此区域为黑色。

由此可见，有光照的区域为菌苔原有的颜色，无光照的区域呈现黑色，随着照射在菌苔上光照区域形状的不同，培养基会呈现不同的图像，如图 5-10 所示。



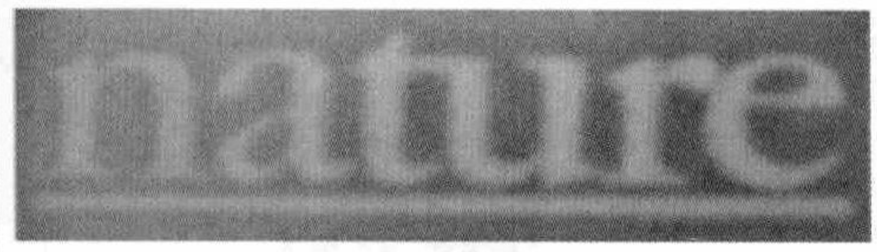


图 5-10　大肠杆菌成像系统演示图[227]

5.4.2　微生物计算系统

随着计算机软硬件技术的迅猛发展，计算机的运算能力突飞猛进，然而由于字长、串行运算等的限制，当计算机要在成千上万个可能的解中搜索可行解或者最优解时，所耗费的时间和存储空间仍然让人无法忍受。如何优化搜索过程中占用的内存和时间已成为目前数学家、软件人员和计算机研究者正在攻克的难关。与此不同的是，如果将微生物中的每个质粒都看作一个软件处理器，而将微生物看作一个计算机，则每个生物计算机体内就可以包含好几个这样的质粒，那么一瓶培养基中就可以包含数以亿计的微生物计算机。所有微生物计算机彼此并行运算，即可大大降低求解的时间，从而处理常规数字计算机不能有效处理的问题。

虽然微生物计算机在软件设计和编程（生物系统设计和构建）阶段的费用昂贵且速度很慢，但 DNA 巨大的信息存储能力和生物分子强大的自组装功能，以及硬件制造（质粒的复制）成本低廉等优势却一直激发着研究者去设计更高效、实用的生物计算机。经过一番努力，目前人类已经开发了基于 DNA 体外自组装和蛋白质-DNA 相互作用的生物计算机，但这些生物计算机无法有效利用 DNA 复制和细胞分裂所提供的并行处理能力，其效率根本无法与电子计算机相提并论。

1. *解决烧焦烧饼问题*[228]

为了证明活体生物系统运算的可行性，充分利用微生物繁殖、DNA 重组和信息存储的能力，Davidson 大学的 K. A. Haynes 课题组利用合成生物学对大肠杆菌进行改造，

植入了具有运算功能的质粒，解决了著名的数学问题——烧焦烧饼问题（the burnt pancake problem，BPP）。

BPP 问题可简述如下：有一叠随意叠放的烧饼，每个烧饼都有一个金黄面和一个焦黄面。每次只能翻动一个烧饼或者相邻的几个烧饼（此处不要求是最上面的一个或几个烧饼）而不能直接更换烧饼的顺序，直到将所有烧饼由大到小、自下而上顺序排列并且焦黄面向下为止。图 5-11 所示为有四块大小不同的烧饼的解题步骤。

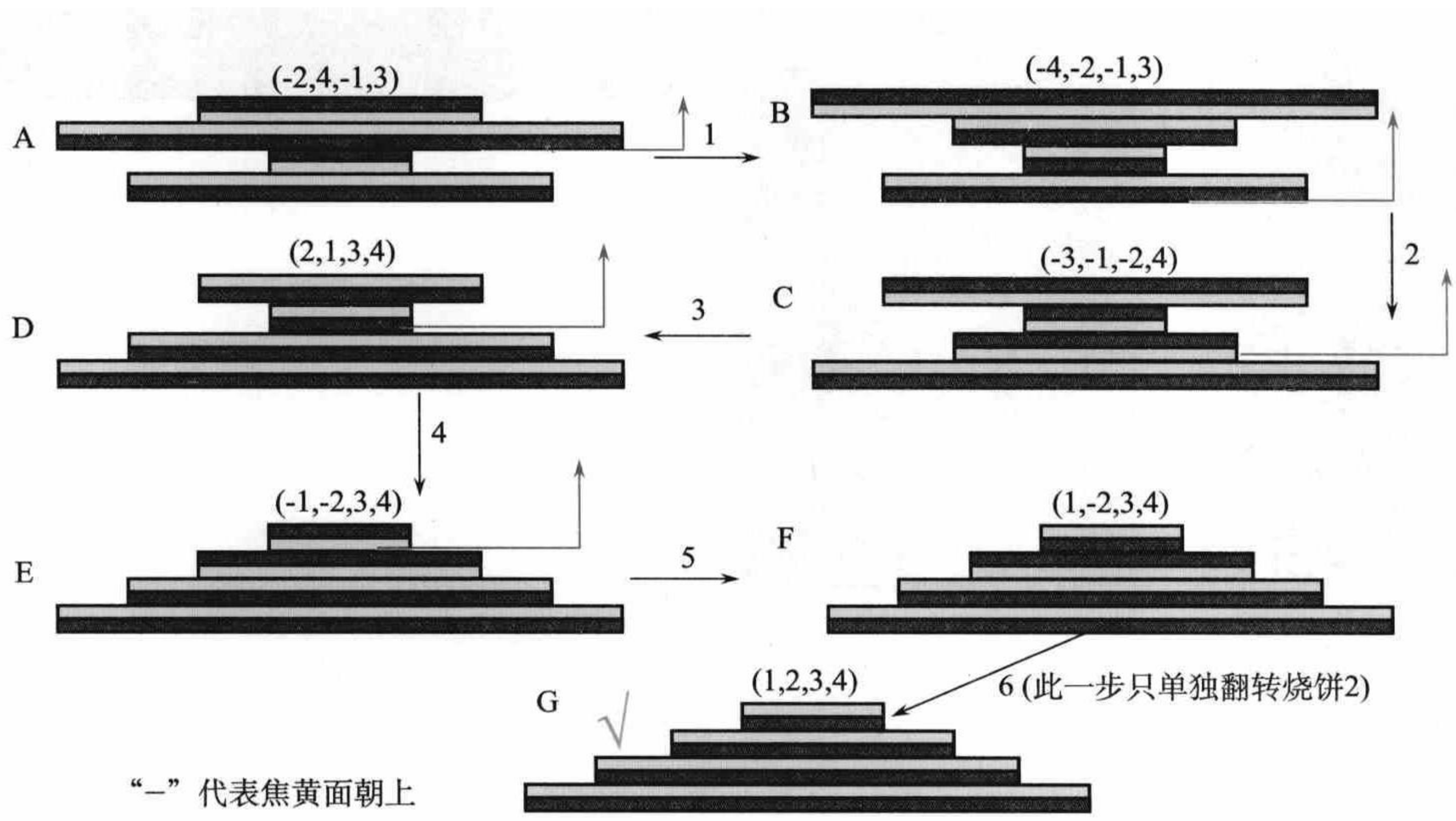


图 5-11　四块烧饼的 BPP 演示图

图中所示为有四块大小不同的烧饼的情况。深黄色代表焦黄面，浅黄色代表金黄面。第一步从最大的烧饼开始，连同其上的烧饼一起翻转得到图 B 所示结果。第二步由最底层开始将整叠烧饼翻转得到图 C 所示结果。第三步将最上面三块烧饼翻转，得到图 D 所示结果，以此类推最后得到图 G 所示最终答案。

一般情况下，两种生物的同线基因组（位于同一条染色体上）的进化距离由翻转其中一种生物基因组上的部分区域使其与另一生物基因组相应同源基因的排序和方向完全吻合所需的最小翻转次数决定。如果将基因片段的翻转看作是烧饼的翻转，那么比较生物进化关系的问题就可以看作是 BPP 问题的一种。因此，对于数学家而言，BPP 是一道典型的搜索和求解问题；而对于生物学家而言，BPP 则可以用来比较生物的进化距离。由此可见，BPP 问题具有非常巨大的应用价值。

但是，读者应该可以看出 BPP 问题是一个典型的搜索问题，当有 n 个对象（烧饼或者基因模块）时，可能的翻转数即排列组合数为 $2^n n!$；随着对象数目的增多，排列组合数呈指数增长。对于常规计算机而言，这就意味着搜索最优解所耗费的时间和存储空间也会随着对象数目的增加而呈指数增长，对象数目越多，求解越困难。当对象数目多到一定程度后，几乎无法搜索到最优解。但对于质粒 DNA 复制和微生物细胞繁殖而言，指数增长则是一件非常容易的事，不需要额外耗费更多的资源和时间，由此微生物计算机在求解 BPP 等类问题时完全有可能具有更高的效率和能力。

如果将启动子、编码区域等功能 DNA 模块看成烧饼，DNA 片段的 5′或 3′端则可以对应于烧饼的焦黄面和金黄面；基因片段功能的不同相当于烧饼大小的不同；利用 DNA 位点特异性重组实现质粒上基因片段的翻转相当于模拟烧饼的翻转；同时利用适当的筛选机制即可以实现微生物求解 BPP 问题。

为此，K. A. Haynes 小组的研究者引入鼠伤寒沙门菌（*Salmonella typhimurium*）中的 DNA 位点特异性重组机制来实现 DNA 片段的翻转。在沙门菌中，HinDNA 重组酶催化翻转反应，通过切换位于可翻转 DNA 片段上的启动子方向来实现对于鞭毛蛋白基因表达的调控。在可翻转 DNA 片段的两侧各有一段 26bp 的回文 *hixC* 序列作为翻转的识别位点。此外还有一个约 70bp 的顺式作用重组增强子（recombinational enhancer，RE）促进蛋白质-DNA 之间复杂的相互作用以提高翻转的速度。

为了实现 DNA 模块的翻转，研究者将取自鼠伤寒沙门菌的 Hin/*hix* 重组酶系统进行模块化重构，并将其植入大肠杆菌中实现质粒上 DNA 模块的翻转。如图 5-12 所示，Hin 重组酶以二聚体的形式附着在基因模块两侧的 *hixC* 序列上，使相应的 DNA 模块发生翻转（图中草绿色箭头所代表的 DNA 模块）而相邻的 DNA 模块还可以保持原方向不变。Hin 重组酶会随机地选择图中三个 *hixC* 位点中的两个进行翻转操作，两个基因模块的排列顺序则可能出现 8 种排列中的任何一种，如图 5-12C 所示，其中 1 代表模块 1 正向，2 代表模块 2 正向；−1 代表模块 1 反向，−2 代表模块 2 反向。系统会在有限次翻转后达到平衡状态。通过 DNA 模块的翻转，产生不同的模块排序，当模块的排序与预期的排序刚好吻合的时候，抗性基因启动，大肠杆菌因具有相应的抗性而能在抗生素环境中存活，其他排序不正确的大肠杆菌则会因为缺少抗性而死亡。

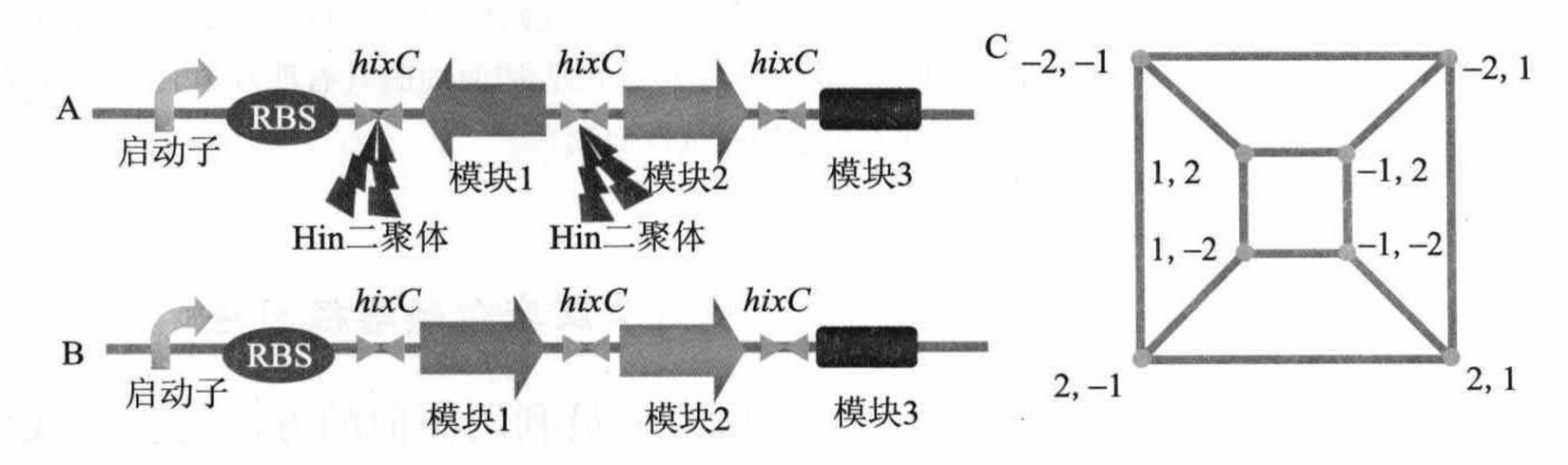


图 5-12　Hin/*hix* 重组酶系统翻转 DNA 片段示意图

在可翻转 DNA 片段的两侧各有一段回文 *hixC* 序列作为翻转的识别位点。图 A 中，Hin 重组酶以二聚体的形式附着在基因片段两侧的 *hixC* 序列上，翻转 DNA 模块 1，而相邻的 DNA 模块 2 仍保持原方向不变，翻转后的线路如图 B 所示。Hin 重组酶会随机地选择图中三个 *hixC* 位点中的两个进行翻转操作，两个基因模块的排列顺序则可能出现 8 种排列中的任何一种，如图 C 所示。图 C 中节点代表模块的状态，其中 1 代表模块 1 正向，2 代表模块 2 正向；−1 代表模块 1 反向，−2 代表模块 2 反向，节点之间的线段表示其中一个节点通过翻转可到达另一节点状态。

图 5-13 所示为利用此系统实现两块烧饼 BPP 问题的模型图和质粒构建图。在具有氨苄抗性的载体上连接 P_{lac}启动子、相应的 RBS 及 *tetA* 基因。*RBS-tetA* 序列代表烧饼 1，P_{lac}启动子代表烧饼 2，两个烧饼之间共需 3 个 *hixC* 模块用以作为翻转酶识别和切割

的位点。翻转酶系统随机地在三个位点中选择其中的两个进行切割实现翻转，代表不同烧饼的基因的初始排序随机翻转成其他所有可能的排序。将所有 BPP 翻转系统进行纯化，并利用四环素进行筛选，只有正确排序的质粒（对应于 BPP 问题的解），P_{lac}启动子才能开启 *tetA* 的表达使细胞具有四环素抗性而存活，其他细胞则会由于缺乏相应抗性而死亡。

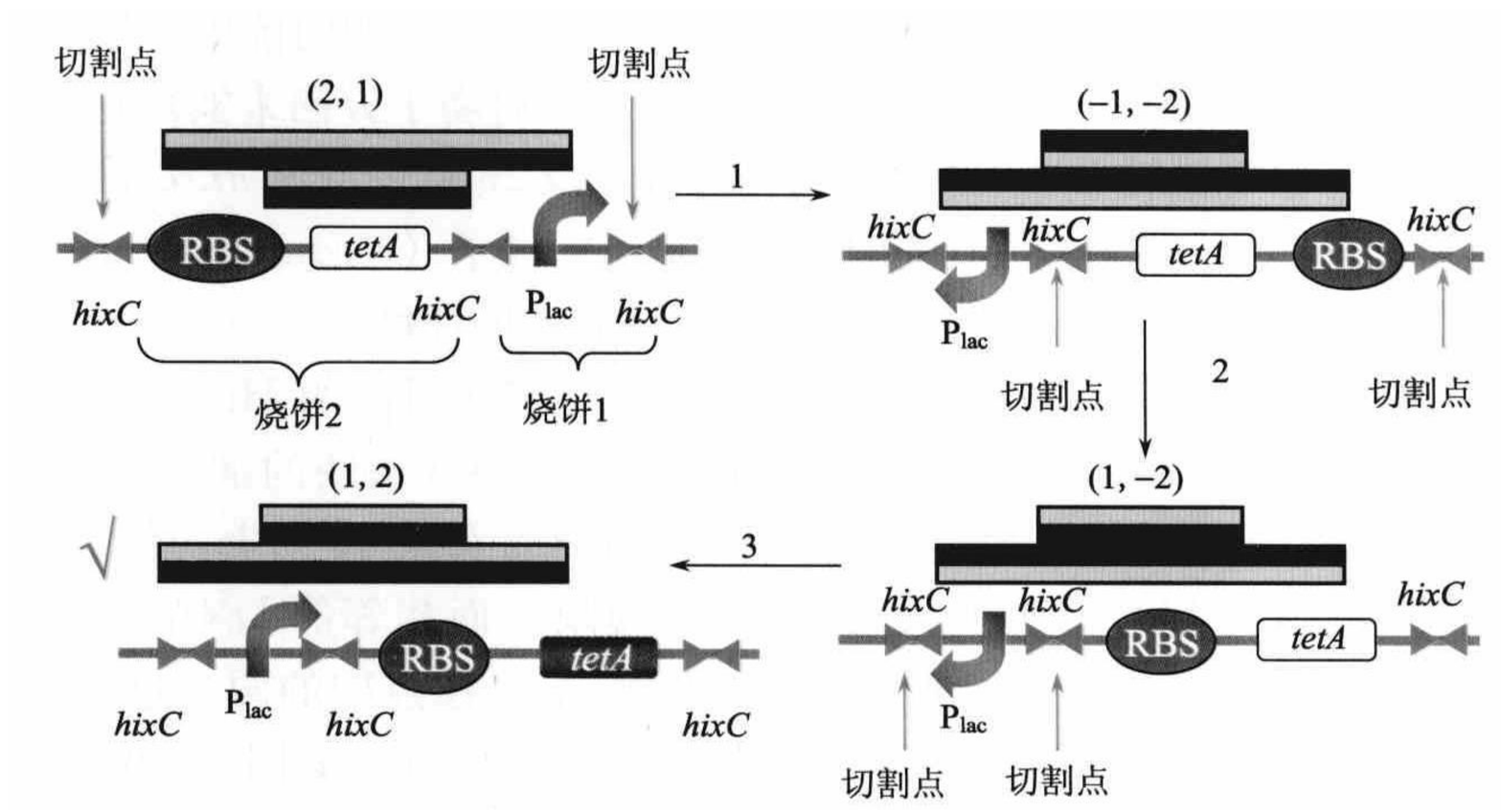


图 5-13　实现两块烧饼 BPP 问题的模型图和质粒构建图

在具有氨苄抗性的载体上连接 P_{lac}启动子、相应的 RBS 以及 *tetA* 基因。*RBS-tetA* 序列代表烧饼 1，P_{lac}启动子代表烧饼 2，两个烧饼之间共需 3 个 *hixC* 模块用以作为翻转酶识别和切割的位点。翻转酶系统随机地在三个位点中选择其中的两个进行切割和翻转反应，代表不同烧饼的基因模块的初始排序随机翻转成其他所有可能的排序。图示为其中的一种可能。第一步：翻转酶选择外侧两个位点进行切割和翻转，整个线路作为一个大模块而翻转为状态（−1,−2）；第二步：翻转酶选择烧饼 1 两侧的两个位点进行切割和翻转，线路状态为（1,−2）；第三步：翻转酶选择烧饼 2 两侧的位点进行切割和翻转，线路转换成（1,2），此时线路中各模块排序和方向正确，在利用四环素进行筛选时，P_{lac}启动子开启 *tetA* 的表达使细胞具有四环素抗性而存活，其他线路中各模块排序或方向不正确的细胞则会由于缺乏相应抗性而死亡。

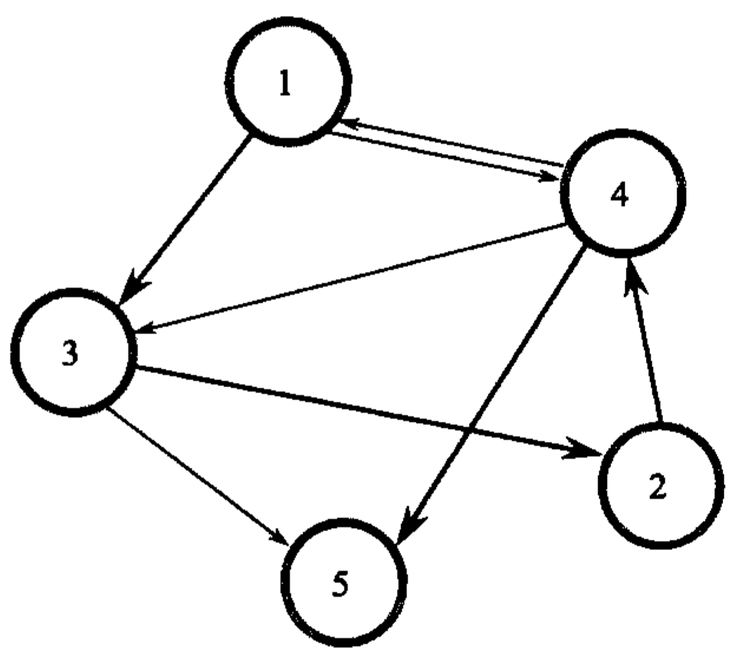


图 5-14　五个节点的汉密尔顿路径问题

图中节点之间的连线代表路径，有向边的箭头代表路径的方向，即只能沿着箭头方向行进。

2. 解决汉密尔顿路径问题[229]

研究人员利用类似的方法对生物线路进行改进，在活体细胞中解决了著名的数学问题——汉密尔顿路径问题（Hamiltonian path problem）。汉密尔顿路径问题，即如何在有向图中找到一条遍历图中所有节点一次且仅一次的最短路径（图 5-14），是计算机图论中非常著名的一道 NP 完全型问题（NP complete）。

质粒的构建由报告基因和 *hixC* 位点组成。将 *hixC* 位点插入报告基因编码序列之内，从而将编码每种荧光蛋白的报告基因分为两部分。每一个完整的报告基因代表一个节点。连

接两个分属于不同报告基因各半段序列的不包含 *hixC* 位点的序列代表有向边，如图 5-15 所示。

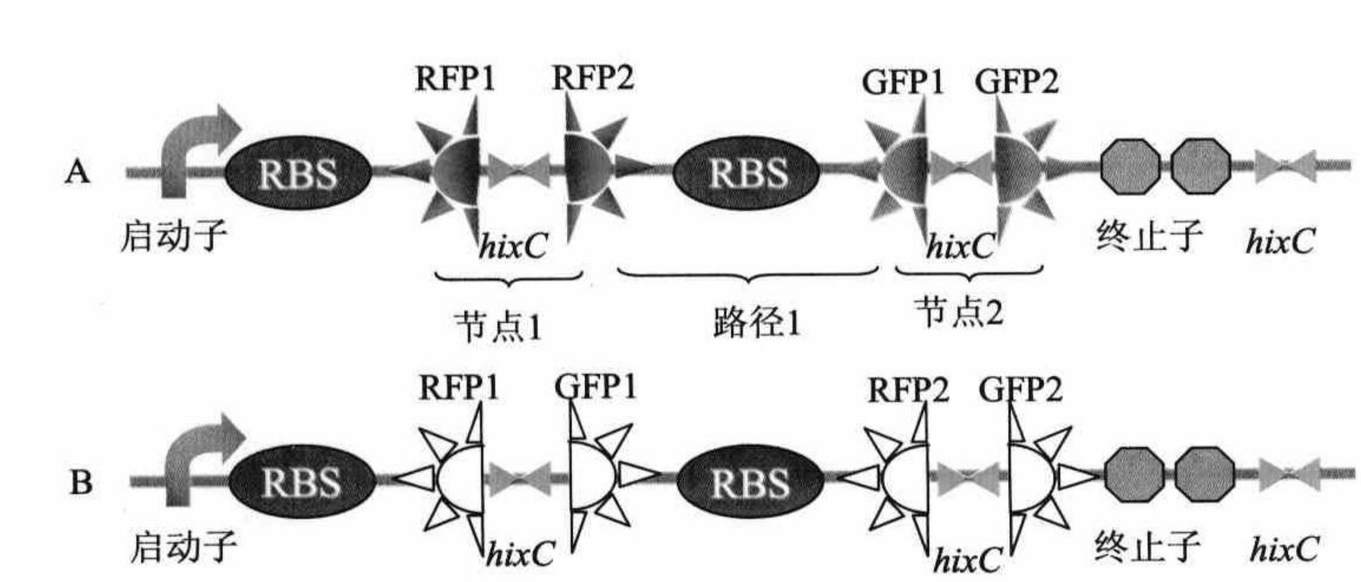


图 5-15　汉密尔顿问题生物线路图

将 *hixC* 位点插入报告基因编码序列之内。每一个完整的报告基因代表一个节点。连接两个分属于不同报告基因各半段序列的不包含 *hixC* 位点的序列代表有向边。Hin 重组酶随机选择一对 *hixC* 位点将片段翻转。图 A 所示为当同一种报告基因的两部分翻转至一起时，该报告基因顺利转录和翻译，细胞中有荧光蛋白产生。图 B 所示为报告基因两部分不匹配时细胞中无荧光蛋白表达的情况。

Hin 重组酶随机选择一对 *hixC* 位点将片段翻转，只有当同一种报告基因的两部分翻转至一起时该报告基因才会进行转录和翻译，如图 5-15A 所示。翻转的随机性可以实现不同路径的排列组合。每种排列组合代表可能的解，只有满足要求的排列组合才能使所有的荧光蛋白都得以表达。通过观察表达的荧光蛋白颜色可以判断求解的进度及每种可能解经过的节点数和相应的路径。

参考文献

1 Luria S E, Human M L. A nonhereditary, host-induced variation of bacterial viruses. *J. Bact.*, 1952, 64: 557-569.

2 Bertani G, Weigle J J. Host controlled variation in bacterial viruses. *J. Bacteriol.*, 1953, 65: 113-121.

3 Arber W, Dussoix D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.*, 1962, 5: 18-36.

4 Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, 1961, 3: 318-356.

5 Szybalski W, Skalka A. Nobel prizes and restriction enzymes. *Gene*, 1978, 4(3): 181-182.

6 Hobom B. Surgery of genes. At the doorstep of synthetic biology. Medizin. Klinik, 1980, 75: 14-21.

7 Rawls R. 'Synthetic Biology' makes its debut. *Chem. Eng. News*, 2000, 78: 49-53.

8 Kolisnychenko V, Plunkett G, Herring C D, et al. Engineering a reduced *Escherichia coli* genome. *Genome Res.*, 2002, 12: 640-647.

9 Ro D K, Paradise E M, Ouellet M, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 2006, 440(13): 940-943.

10 赵学明，王庆昭．合成生物学：学科基础、研究进展与前景展望．前沿科学，2007，3：56-66.

11 张春霆．合成生物学：我国急需发展的前沿科学．前沿科学，2007，3：55.

12 Drubin D A, Way J C, Silver P A. Designing biological systems. *Genes & Dev.*, 2007, 21: 242-254.

13 Blake W J, Isaacs F J. Synthetic biology evolves. *Trends Biotechnol.*, 2004, 22(7): 321-324.

14 Greber D, Fussenegger M. Mammalian synthetic biology: Engineering of sophisticated gene networks. *J. Biotechnol.*, 2007, 130: 329-345.

15 Sprinzak D, Elowitz M B. Reconstruction of genetic circuits. *Nature*, 2005, 438(24): 443-448.

16 Meyer A, Pellaux R, Panke S. Bioengineering novel in vitro metabolic pathways using synthetic biology, *Curr. Opin. Microbiol.*, 2007, 10: 246-253.

17 Alon Z, Mayo A E, Rosenberg R, et al. Just-in-time transcription program in metabolic pathways. *Nat. Genet.*, 2004, 36: 486-491.

18 Sera T, Uranga C. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochem.*, 2002, 41: 7074-7081.

19 Dreier B, Fuller R P, Segal D J, et al. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, 2005, 280: 35588-35597.

20 Segal D J, Dreier B, Beerli R R, et al. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl.*

Acad. Sci. USA, 1999, 96: 2758-2763.

21 Looger L L, Dwyer M A, Smith J J, et al. Computational design of receptor and sensor proteins with novel functions. *Nature*, 2003, 423: 185-190.

22 Dwyer M A, Looger L L, Hellinga H W. Computational design of a biologically active enzyme. *Science*, 2004, 304(5679): 1967-1971.

23 Dueber J E, Yeh B J, Chak K, et al. Reprogramming control of an allosteric signaling switch through modular recombination. *Science*, 2003, 301(5641): 1904-1908.

24 Bashor C J, Helman N C, Yan Shude, et al. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science*, 2008, 319(14): 1539-1548.

25 Park S H, Zarrinpar A, Lim W A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science*, 2003, 299: 1061-1064.

26 Tian J, Gong Hui, Sheng N, et al. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, 2004, 432: 1050-1054.

27 Picataggio S. Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals Stephen Picataggio. *Curr. Opin. Microbiol.*, 2009, 19: 325-329.

28 Wimmer E. The test-tube synthesis of a chemical called poliovirus: The simple synthesis of a virus has far-reaching societal implications, EMBO Rep., 2006, 7 (SI): S3-S9, doi: 10.1038/sj. embor. 7400728.

29 Young L, Dong Q H. Two-step total gene synthesis method. *Nucleic Acids Res.*, 2004, 32 (7): e59.

30 Binkowski B F, Richmond K E, Kaysen J, et al. Correcting errors in synthetic DNA through consensus shuffling. *Nucleic Acids Res.*, 2005, 33: e55.

31 Carr P A, Park J S, Lee Y J, et al. Protein mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res.*, 2004, 32: e162.

32 Ball P. Synthetic biology: starting from scratch. *Nature*, 2004, 431: 624-626.

33 Benner S A, Sismour A M. Synthetic biology. *Nat. Rev. Genet.*, 2005, 6: 533-543.

34 Lynch S R, Liu H, Gao Jianmin, et al. Toward a designed, functioning genetic system with expanded-size base pairs: solution structure of the eight-base xDNA double helix, *J. Am. Chem. Soc.*, 2006, 128(45): 14704-14711.

35 Nicole K, Simoni R D, Hill R L. Total synthesis of a tyrosine suppressor tRNA: the work of H. Gobind Khorana. *J. Biol. Chem.*, 2009, 284(22): e5.

36 Cello J, Paul A V, Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, 2002, 297(5583): 1016-1018.

37 Smith H O, Hutchinson C A 3rd., Pfannkoch C, et al. Generating a synthetic genome by whole genome assembly: Phi X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA*, 2003, 100(26): 15440-15445.

38 Tumpey T, Basler C, Aguilar P, et al. Characterization of the reconstructed 1918 Spanish Influenza Pandemic Virus. *Science*, 2005, 310: 77-80.

39 Taubenberger J K, Reid A H, Lourens R M, et al. Characterization of the 1918 influenza virus polymerase genes. *Nature*, 2005, 437: 889-893.

40 Gibson D G. Complete Chemical Synthesis, Assembly and cloning of a *Mycoplasma genitalium* Ge-

nome. *Science*, 2008, 319: 1215-1220.

41 张春霆. 合成生物学研究的进展. 中国科学基金, 2008, 2: 65-69.

42 Lartigue C. Genome transplantation in bacteria: changing one species to another. *Science*, 2007, 317: 632-638.

43 Hashimoto M, Ichimura T, Mizoguchi H, et al. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.*, 2005, 55, 137-149.

44 Glass J I, Assad-Garcia N, Alperovich N, et al. Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. USA*, 2005, 103: 425-430.

45 Forster A C, Church G M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, 2006, 2, doi: 10.1038/msb4100090.

46 Pennisi E. Synthetic biology remakes small genomes. *Science*, 2005, 310: 769-770.

47 Zhang R, Ou H Y, Zhang C T. DEG: a database of essential genes. *Nucleic Acids Res.*, 2004, 32: D271-D272.

48 ETC Group. Extreme genetic engineering: an introduction to synthetic biology. The. ETC Group, January 2007, www.etcgroup.org.

49 Chan L Y, Kosuri S, Endy D. Refactoring bacteriophage T7. *Mol. Syst. Biol.*, 2005, doi: 10.1038/msb4100025.

50 Pitera D J, Paddon C, Newman J D, et al. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.*, 2007, 9: 193-207.

51 Keasling J D. Synthetic biology for synthetic chemistry. *ACS Chem. Biol.*, 2008, 3: 64-76.

52 Monod J, Jacob F. General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harbor Symposia on Quantitative Biology, 1961, 26: 389-401.

53 Mjolsness E, Sharp D H, Reinitz J. A connectionist model of development. *J. Theor. Biol.*, 1991, 152: 429-453.

54 Reinitz J, Sharp D H. Gene circuits and their uses. In: Collado J, Magasanik B, Smith T. Integrative approaches to molecular biology. MIT Press, Cambridge, MA, 1996, Chap. 13: 253-272.

55 Sharp D H, Reinitz J. Prediction of mutant expression patterns using gene circuits. *BioSystems*, 1998, 47: 79-90.

56 McAdams H H, Shapiro L. Circuit simulation of genetic networks. *Science*, 1995, 269(5224): 650-656.

57 Weiss R. Cellular computation and communications using engineered genetic uegulatory networks. Ph. D. Thesis, Massachusetts Institute of Technology, September 2001.

58 Gardner T S, Cantor C R, Collins J J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 2000, 403: 339-342.

59 Atkinson M R, Savageau M A, Myers J T, et al. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 2003, 113: 597-607.

60 Elowitz M B, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 2000, 403: 335-338.

61 Kobayashi H, Kærn M, Araki M, et al. Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA*, 2004, 101(22): 8414-8419.

62 Balagadde F K, You L, Hansen C L, et al. Long-term monitoring of bacteria undergoing programmed population control in a Microchemostat. *Science*, 2005, 309(5731): 137-140.

63 Garcia-Ojalvo J, Elowitz M B, Strogatz S H. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. *Proc. Natl. Acad. Sci. USA*, 2004, 101: 10955-10960.

64 McMillen D, Kopell N, Hasty J, et al. Synchronizing genetic relaxation oscillators by intercell signaing. *Proc. Natl. Acad. Sci. USA*, 2002, 99: 679-684.

65 Hasty J, McMillen D, Collins J J. Engineered gene circuits. *Nature*, 2002, 420(14): 224-230.

66 Guido N J, Wang X, Adalsteinsson D, et al. A bottom-up approach to gene regulation. *Nature*, 2006, 439(16): 856-860.

67 Isaacs F J, Hasty J, Cantor C R, et al. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*, 2003, 100(13): 7714-7719.

68 Endy D, Brent R. Modelling cellular behaviour. *Nature*, 2001, 409(18): 391-395.

69 Dhar K P, Weiss R, Enabling the new biology of the 21st century. *Systems and Synthetic Biology*, 2007, 1(1) : 1-2.

70 Csete M E, Doyle J C. Reverse engineering of biological complexity. *Science*, 2002, 295: 1664-1669.

71 Brent R. A partnership between biology and engineering. *Nat. Biotechnol.*, 2004, 22(10): 1211-1214.

72 Brian Y J, Wendell L A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.*, 2007, 3(9): 521-525.

73 Baker D, George C, Collins J J, et al. Engineering life: building a fab for biology. *Sci. Am.*, 2006, 294(6): 44-51.

74 Matthias H, Sven P. Synthetic biology: putting engineering into biology. *Bioinformatics*, 2006, 22(22): 2790-2799.

75 Bhutkar A. Synthetic biology: navigating the challenges ahead. *J. Biolaw & Bus.*, 2005, 8(2): 19-29.

76 Andrianantoandro E, Basu S, Karig D K, et al. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, 2006, doi: 10.1038/msb4100073.

77 Endy D. Foundations for engineering biology. *Nature*, 2005, 438, 449-453.

78 Keasling J. The promise of synthetic biology. *Bridge Natl. Acad. Eng.*, 2005, 35: 18-21.

79 Nelson D L, Cox M M. Lehninger's Principles of Biochemistry. 4th ed. New York: W. H. Freeman and Company, 2005, 1039.

80 Church G M. From systems biology to synthetic biology. *Mol. Syst. Biol.*, 2005, doi: 10.1038/msb4100007

81 Barrett L C, Kim T Y, Kim H U, et al. Systems biology as a foundation for genome-scale synthetic biology. *Curr. Opin. Microbiol.*, 2006, 17: 488-492.

82 Peer S, Markus B, Gao X, et al. Another side of genomics: synthetic biology as a means for exploitation of whole-genome sequence information. *J. Biotechnol.*, 2006, 124: 206-212.

83 Brown J. The iGEM competition: building with biology. *IET Synth. Biol.*, 2007, 1(1-2): 3-6.

84 Balmer A, Martin P. Synthetic Biology: social and ethical challenges. An independent review commissioned by the Biotechnology and Biological Sciences Research Council(BBSRC), May 2008.

85 Bügl H, Danner J P, Molinari R J, et al. DNA synthesis and biological security. *Nat. Biotechnol.*, 2007, 25(6): 627- 629.

86 Kumar S, Rai A. Synthetic biology: the intellectual property puzzle. *Tex. Law Rev.*, 2007, 85:

1745-1768.

87 Rai A, Boyle J. Synthetic biology: caught between property rights, the public domain, and the commons. *PLoS Biol.*, 2007, 5(3): e58.

88 Church G. Let us go forth and safely multiply. *Nature*, 2005, 438(24): 423.

89 Chopra P, Kamma A. Engineering life through synthetic biology. *In Silico Biol.*, 2006, 6: 401-410.

90 Schmidt M, Kelle A, Ganguli-Mitra A, et al. Synthetic biology: the technoscience and its societal consequences, Springer Netherlands, 2010, Online.

91 Tucker J B, Zilinskas R A. The promise and perils of synthetic biology. *The New Atlantis*, 2006, 12: 25-45.

92 IRGC, Synthetic Biology: Risks and opportunities of an emerging field. International Risk Governance Council, Geneva, 2008, http: //www. irgc. org/IMG/pdf/IRGC _ ConceptNote _ SyntheticBiology _ Final _ 30April. pdf

93 Stromgaard A, Jensen A A, Stromgaard K. Site-specific incorporation of unnatural amino acids into proteins. *ChemBioChem*, 2004, 5: 909-916.

94 Wang L, Brock A, Herberich B, et al. Expanding the genetic code of *Escherichia coli*. Science 2001, 292: 498-500.

95 Anderson J C, Wu N, Santoro S W, et al. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA*, 2004, 101: 7566-7571.

96 Forster AC, Church GM. Synthetic biology projects *in vitro*. *Genome Res.*, 2007, 17: 1-6.

97 Aldhous P. The Bioweapon is in the Post. *New Scientist*, 2005, 12: 8-9.

98 Ball P. Synthetic biology for nanotechnology. *Nanotechnology*, 2005, 16: R1-R8.

99 Knight T. Idempotent vector design for standard assembly of BioBricks, MIT Artificial Intelligence Laboratory. 3 December, 2009, http: //hdl. handle. net/1721. 1/21168.

100 Noireaux V, Bar-Ziv R, Libchaber A. Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. USA*, 2003, 100(22): 12672-12677.

101 Shimizu Y, Inoue A, Tomari Y, et al. Cell-free translation reconstituted with purifiedcomponents. *Nat. Biotechnol.*, 2001, 19: 751-755.

102 Simpson M L. Cell-free synthetic biology: a bottom-up approach to discovery by design. *Mol. Syst. Biol.*, 2006, 2: 69, doi: 10. 1038/msb4100104.

103 Shetty R P, Endy D, Knight T F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.*, 2008, 2: 5.

104 Canton B, Labno A, Endy D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.*, 2008, 26: 787-793.

105 Lee T I, Rinaldi N J, Robert F, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 2002, 298(25): 799-804.

106 Huang C Y, Ferrell J E Jr. Ultrasensitivity in the mitogenactivated protein kinase cascade. *Proc. Natl. Acad. Sci. USA*, 1996, 93: 10078-10083.

107 Sayut D J, Kambam P K, Sun L. Engineering and applications of genetic circuits. *Mol. Biosyst.*, 2007, 3: 835-840.

108 Achimescu S, Lipan O. Signal propagation in nonlinear stochastic gene regulatory Networks. *Systems Biology*, *IEE Proceedings*, 2006, 153(3): 120-134.

109 Blake W J, Kærn M, Cantor C R, et al. Noise in eukaryotic gene expression. *Nature*, 2003, 422: 633-637

110 Shen-Orr S S, Milo R, Mangan S, et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, 2002, 31: 64-68.

111 Thattai M, van Oudenaarden A. Attenuation of noise in ultrasensitive signaling cascades. *Biophys. J.*, 2002, 82: 2943-2950.

112 Rosenfeld N, Alon U. Response delays and the structure of transcription networks. *J. Mol. Biol.*, 2003, 329: 645-654.

113 Bulter T, Lee Sun-Gu, Wong W W, et al. Design of artificial cell-cell communication using gene and metabolic networks. *Proc. Natl. Acad. Sci. USA*, 2004, 101(8): 2299-2304.

114 Weber W, Daoud-El Baba M, Fussenegger M. Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl. Acad. Sci. USA*, 2007, 104(25): 10435-10440.

115 Miller M B, Bassler B L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.*, 2001, 55: 165-199.

116 Collins C H, Arnold F H, Leadbetter J R. Directed evolution of *Vibrio fischeri* LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. *Mol. Microbiol.*, 2005, 55 (3): 712-723.

117 Chen M, Weiss R. Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*. *Nat. Biotechnol.*, 2005, 23(12): 1551-1555.

118 Wall M E, Hlavacek W S, Savageau M A. Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.*, 2004, 5: 34-42.

119 Alon U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 2007, 8: 450-461.

120 Savageau M A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*, 1974, 252: 546-549.

121 Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. *Nature*, 2000, 405: 590-593.

122 Becskei A, Seraphin B, Serrano L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.*, 2001, 20(10): 2528-2535.

123 Rosenfeld N, Elowitz M B, Alon U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.*, 2002, 323: 785-793.

124 Kramer B P, Viretta A U, Baba M D, et al. An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.*, 2004, 22: 867-870.

125 Mangan S, Alon U. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA*, 2003, 100: 11980-11985.

126 Dekel E, Mangan S, Alon U. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys. Biol.*, 2005, 2: 81-88.

127 Wall M E, Dunlop M J, Hlavacek W S. Multiple functions of a feed-forward-loop gene circuit. *J. Mol. Biol.*, 2005, 349: 501-514.

128 Mangan S, Zaslaver A, Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.*, 2003, 334: 197-204.

129 Elowitz M B, Levine A J, Siggia E D, et al. Stochastic gene expression in a single cell. *Science*,

2002, 297: 1183-1186.

130 McDaniel R, Weiss R. Advances in synthetic biology: on the path from prototypes to applications. *Curr. Opin. Microbiol.*, 2005, 16: 476-483.

131 Rosenfeld N, Young J W, Alon U, et al. Gene regulation at the single-cell level. *Science*, 2005, 307: 1962-1965.

132 Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*, 2001, 98: 8614-8619.

133 Ronen M, Rosenberg R, Shraiman B I, et al. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. USA*, 2002, 99(16): 10555-10560.

134 Braun D, Basu S, Weiss R. Parameter estimation for two synthetic gene networks: a case study. IEEE International Conference on Acoustics, Speech, and Signal Processing March 2005.

135 Witteman B. (supervisors: Marc Hulsman and Marcel Reinders), Synthetic biology, Delft, Delft University of Technology, 2007, http: //ict. ewi. tudelft. nl/index. php? option = com _ pub&task = view&id=2546.

136 Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, 2002, 9(1): 67-103.

137 Rao C V, Wolf D M, Arkin A P. Control, exploitation and tolerance of intracellular noise. *Nature*, 2002, 420(14): 231-237.

138 Koudrivtsev A B, Jameson R F, Linert W. The law of mass action. Berlin: Springer-Verlag, 2001.

139 Michaelis L, Menten M L. Die kinetik der invertinwirkung. *Biochem. Z.*, 1913, 49: 334-336.

140 Hill A V. The combinations of haemoglobin with oxygen and with carbon monoxide. *Biochem. J.*, 1913, 7: 472-480.

141 Segal I H. Enzyme Kinetics. New York: John Wiley & Sons, 1975.

142 Kepler T B, Elston T C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.*, 2001, 81: 3116-3136.

143 Kærn M, Elston T C, Blake W J, et al. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, 2005, 6: 451-464.

144 Gillespie D T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 1977, 25(81): 2340-2361.

145 Gillespiea D T. Approximate accelerated stochastic simulation of chemically reacting. *J. Phys. Chem.*, 2001, 4(115): 297-306.

146 Haseltine E L, Rawlings J B. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 2002, 117(1515): 6959-6969.

147 Salis H, Kaznessis Y. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.*, 2005, 122(5): 54103.

148 Griffith M, Courtney T, Peccoud J, et al. Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network. *Bioinformatics*, 2006, 22(22): 2782-2789.

149 Driscoll M E, Gardner T S. Identification and control of gene networks in living organisms via supervised and unsupervised learning. *J. Pro. Cont.*, 2006, 16: 303-311.

150 Datta S, Datta S. Evaluation of clustering algorithms for gene expression data. *BMC*

Bioinformatics, 2006, 7(Suppl4): S17.

151 Liao T W. Clustering of time series data-a survey. *Patt. Recog.*, 2005, 38: 1857-1874.

152 Kim S. Protein β-turn prediction using nearest neighbour method. *Bioinformatics*, 2004, 20: 40.

153 Tominaga Y. Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, ANNs, and *k*-NN. *Chemo Intel. Lab. Syst.*, 1999, 49: 105-115.

154 Liao S, Wen C. Artificial neural networks classification and clustering of methodologies and applications - literature analysis from 1995 to 2005. *Expert Syst. Appl.*, 2007, 32: 1-11.

155 Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach. Learn.*, 1997, 9: 131-163.

156 Batt G, Yordanov B, Weiss R, et al. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics*, 2007, 23(18): 2415-2422.

157 Feng X J, Hooshangi S, Chen D, et al. Optimizing genetic circuits by global sensitivity analysis. *Biophys. J.*, 2004, 87: 2195-2202.

158 Haseltine E L, Arnold F H. Synthetic gene circuits: design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.*, 2007, 36: 1-19.

159 Austin D W, Allen M S, McCollum J M, et al. Gene network shaping of inherent noise spectra. *Nature*, 2006, 439(2): 608-611.

160 Albert R, Jeong H, Barabasi A L. Error and attack tolerance of complex networks. *Nature*, 2000, 406(6794): 378-382.

161 Chaussade C, Rewcastle G W, Kendall J D, et al. Evidence for functional redundancy of class ia pi3k isoforms in insulin signalling. *Biochem. J.*, 2007, 404(3): 449-458.

162 Hooshangi S, Weiss R. The effect of negative feedback on noise propagation in transcriptional gene networks. *Chaos*, 2006, 16: 026108.

163 Maeda Y T, Sano M. Regulatory dynamics of synthetic gene networks with positive feedback. *J. Mol. Biol.*, 2006, 359: 1107-1124.

164 Komarova N L, Zou X, Nie Q, et al. A theoretical framework for specificity in cell signaling. *Mol. Syst. Biol.*, 2005, Report: E1-E5.

165 Rackham O, Chin J W. Cellular logic with orthogonal ribosomes. *J. Am. Chem. Soc.*, 2005, 127 (50): 17584-17585.

166 Yokobayashi Y, Weiss R, Arnold F H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA*, 2002, 99: 16587-16591.

167 Isaacs F J, Hasty J, Cantor C R, et al. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*, 2003, 100(13): 7714-7719.

168 Michalowski C B, Short M D, Little J W. Sequence tolerance of the phage λPRM promoter: implications for evolution of gene regulatory circuitry. *J. Bacteriol.*, 2004, 186: 7899-7999.

169 Mandal M, Breaker R R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, 2004, 5: 451-463.

170 Isaacs F J, Dwyer D J, Ding C, et al. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.*, 2004, 22: 841-847.

171 Bayer T S, Smolke C D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.*, 2005, 23: 337-343.

172 Pfleger B F, Pitera D J, Smolke C D, et al. Combinatorial engineering of intergenic regions in

operons tunes expression of multiple genes. *Nat. Biotechnol.*, 2006, 24(8): 1027-1032.

173 Isaacs F J, Dwyer D J, Collins J J. RNA synthetic biology. *Nat. Biotechnol.*, 2006, 24(5): 545-554.

174 Werstuck G, Green M R. Controlling gene expression in living cells through small molecule-RNA interactions. *Science*, 1998, 282: 296-298.

175 Kramer B P, Viretta A U, Baba M D, et al. An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.*, 2004, 22: 867-870.

176 Basu S, Mehreja R, Thiberge S, et al. Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl. Acad. Sci. USA*, 2004, 101(17): 6355-6360.

177 You L, Cox R S III, Weiss R, et al. Programmed population control by cell-cell communication and regulated killing. *Nature*, 2004, 428(22): 868-871.

178 Basu S, Gerchman Y, Collins C H, et al. A synthetic multicellular system for programmed pattern formation. *Nature*, 2005, 434: 1130-1134.

179 Hooshangi S, Thiberge S, Weiss R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. USA*, 2005, 102(10): 3581-3586.

180 Widmann C, Gibson S, Jarpe M B, et al. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.*, 1999, 79(1): 143-180.

181 Weber W, Fussenegger M. The impact of synthetic biology on drug discovery. *Drug Disco Today*, 2009, 14(19-20): 956-963.

182 Rolli V, Gallwitz M, Wossning T, et al. Amplification of B cell antigen receptor signaling by a Syk/ITAM positive feedback loop. *Mol. Cell*, 2002, 10: 1057-1069.

183 Ferrari S, Lougaris V, Caraffi S, et al. Mutations of the Igbeta gene cause agammaglobulinemia in man. *J. Exp. Med.* 2007, 204: 2047-2051.

184 Becker M M, Graham R L, Donaldson E F, et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl. Acad. Sci. USA*, 2008, 105: 19944-19949.

185 Rockx B, Sheahan T, Donaldson E, et al. Synthetic reconstruction of zoonotic and early human severe acute respiratory syndrome coronavirus isolates that produce fatal disease in aged mice. *J. Virol.*, 2007, 81: 7410-7423.

186 Weber W, Schoenmakers R, Keller B, et al. A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc. Natl. Acad. Sci. USA*, 2008, 105: 9994-9998.

187 Engohang Ndong J, Baillat D, Aumercier M, et al. EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. *Mol. Microbiol.*, 2004, 51(1): 175-188.

188 Aubel D, Morris R, Lennon B. Design of a novel mammalian screening system for the detection of bioavailable, non-cytotoxic streptogramin antibiotics. *J. Antibiot.* (Tokyo), 2001, 54: 44-55.

189 Gonzalez-Nicolini V, Fux C, Fussenegger M. A novel mammalian cell-based approach for the discovery of anticancer drugs with reduced cytotoxicity on non-dividing cells. *Invest. New Drugs.*, 2004, 22: 253-262.

190 Loose C, Jensen K, Rigoutsos I, et al. A linguistic model for the rational design of antimicrobial peptides. *Nature*, 2006, 443: 867-869.

191 Lu T K, Collins J J. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl.*

Acad. Sci. USA, 2007, 104: 11197-11202.

192 Lu T K, Collins J J. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc. Natl. Acad. Sci. USA*, 2009, 106: 4629-4634.

193 Cirz R T, Chin J K, Andes D R, et al. Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol.*, 2005, 3: e176.

194 Anderson J C, Clarke E J, Arkin A P. Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.*, 2006, 355: 619-627.

195 Henning T, Kraus M, Brischwein M, et al. Relevance of tumor microenvironment for progression, therapy and drug development. *Anticancer Drugs*, 2004, 15: 7-14.

196 Salmon K A, Hung S P, Steffen N R, et al. Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J. Biol. Chem.*, 2003, 278: 29837-29855.

197 Wang H, Gunsalus R P. Coordinate regulation of the *Escherichia coli* formate dehydrogenase fdnGHI and fdhF genes in response to nitrate, nitrite, and formate: roles for NarL and NarP. *J. Bacteriol.*, 2003, 185: 5076-5085.

198 Topp S, Gallivan J P. Guiding bacteria with small molecules and RNA. *J. Am. Chem. Soc.*, 2007, 129: 6807-6811.

199 Martin V J J, Pitera D J, Withers S T, et al. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.*, 2003, 21: 796-802.

200 Stephanopoulos G. Challenges in engineering microbes for biofuels production. *Science*, 2007, 315: 801-804.

201 Atsumi S, Hanai T, Liao J C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 2008, 451: 86-89.

202 Withers S T, Gottlieb S S, Lieu B, et al. Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. *Appl. Environ. Microbiol.*, 2007, 73: 6277-6283.

203 Lee S K, Choul H, Ham T S. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Microbiol.*, 2008, 19: 556-563.

204 Fortman J L, Chhabra S, Mukhopadhyay A, et al. Biofuels alternatives to ethanol: pumping the microbial well. *Trends Biotechnol.*, 2008, 26: 375-381.

205 Blanch H W, Adams P D, Andrews-Cramer K M, et al. Addressing the need for alternative transportation fuels: the joint BioEnergy institute. *ACS Chem. Biol.*, 2008, 3: 17-20.

206 Mukhopadhyay A, Redding A M, Rutherford B J, et al. Importance of systems biology in engineering microbes for biofuel production. *Curr. Opin. Biotechnol.*, 2008, 19: 228-234.

207 Picataggio S. Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals. *Curr. Opin. Microbiol.*, 2009, 19: 325-329.

208 Chen J S, Hiu S F. Acetone butanol isopropanol production by Clostridium-Beijerinckii (Synonym, Clostridium-Butylicum). *Biotechnol. Lett.*, 1986, 8: 371-376.

209 Hanai T, Atsumi S, Liao J C. Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl. Environ. Microbiol.*, 2007, 73: 7814-7818.

210 Atsumi S, Canna A F, Connora M R, et al. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.*, 2008, doi: 10.1016/j.ymben.2007.08.003.

211 Schoondermark-Stolk S A, Jansen M, Veurink J H, et al. Rapid identification of target genes for

3-methyl-1-butanol production in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, 2006, 70: 237-246.

212 Sentheshanmuganathan S. Mechanism of the formation of higher alcohols from amino acids by *Saccharomyces cerevisiae*. *Biochem. J.*, 1960, 74: 568-576.

213 Yoshimoto H, Fukushige T, Yonezawa T, et al. Genetic and physiological analysis of branched-chain alcohols and isoamyl acetate production in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, 2002, 59: 501-508.

214 Connor M R, Liao J C. Engineering *Escherichia coli* for the production of 3-methyl-1-butanol. *Appl. Environ. Microbiol.*, 2008, 74: 5769-5775.

215 Chisti Y. Biodiesel from microalgae. *Biotechnol. Adv.*, 2007, 25: 294-306.

216 Keasling J D, Chou H. Metabolic engineering delivers next-generation biofuels. *Nat. Biotechnol.*, 2008, 26: 298-299.

217 Reed J L, Vo T D, Schilling C H, et al. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/ GPR). *Genome Biol.*, 2003, 4: R54.

218 Forster J, Famili I, Fu P, et al. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.*, 2003, 13: 244-253.

219 Patil K R, Akesson M, Nielsen J. Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.*, 2004, 15: 64-69.

220 Ibarra R U, Edwards J S, Palsson B O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 2002, 420: 186-189.

221 Almaas E, Kovacs B, Vicsek T, et al. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, 2004, 427: 839-843.

222 Jeffries T W. Engineering yeasts for xylose metabolism. *Curr. Opin. Biotechnol.*, 2006, 17: 320-326.

223 Chu B C, Lee H. Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation. *Biotechnol. Adv.*, 2007, 25: 425-441.

224 Bro C, Regenberg B, Forster J, et al. *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab. Eng.*, 2006, 8: 102-111.

225 Atsumi S, Hanai T, Liao J C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 2008, 451: 86-89.

226 Aleksic J, Bizzari F, Cai Y, et al. Development of a novel biosensor for the detection of arsenic in drinking water. *IET Synth. Biol.*, 2007, 1: 87-90.

227 Anselm L, Chevalier A A, Tabor J J, et al. Engineering *Escherichia coli* to see light. *Nature*, 2005, 438: 441-442.

228 Haynes K A, Broderick M L, Brown A D, et al. Engineering bacteria to solve the Burnt Pancake Problem. *J. Biol. Eng.*, 2008, 2: 8.

229 Baumgardner J, Acker K, Adefuye O, et al. Solving a Hamiltonian Path Problem with a bacterial computer. *J. Biol. Eng.*, 2009, 3: 11, doi: 10.1186/1754-1611-3-11.

主要网站

1. http://syntheticbiology.org

主要内容：关于合成生物学的内容描述、讨论、新闻、资料下载、研究进展以及相关的会议和研究团体链接。

点　　评：合成生物学委员会的官方网站，其中对合成生物学的描述目前被全世界普遍接受。里面的资料权威而且全面。

2. http://partsregistry.org/Main _ Page

主要内容：关于已经注册的 BioBrick 的全部资料，包括标准化 BioBrick、质粒的基本概念、详细构成、操作方法、功能描述以及使用者论坛。

点　　评：每年参加 iGEM 竞赛的队伍都会在赛前在该网站上注册自己队伍新开发的 BioBrick，组委会会定期对已注册的组件进行整理和验证，并将样本免费提供给所有参赛队伍。网站的更新和扩展速度很快。该网站对于实际从事合成生物学实验的研究者非常有用。

3. http://biobricks.org/

主要内容：关于 BBF 基金会构成、性质、目的的介绍，提供与 BioBrick 和 BioBricks 两个专利相关的一些手机软件等软件下载。

点　　评：BBF 基金会（The BioBricks Foundation）官方网站，对基金会的介绍较全面。

4. http://synbio.tju.edu.cn

主要内容：介绍中国国内历届 iGEM 的参赛内容、国内较有影响的合成生物学论坛、会议，以及天津大学“合成生物学兴趣小组”的相关情况。

点　　评：国内第一家有关合成生物学网站，中英文双语，对参赛队的参赛项目介绍的很全面，是一个了解国内高校合成生物学现状的主要网站。由张春霆院士创建和维护。

5. http://kn.theiet.org/communities/biology/index.cfm

主要内容：系统生物学与合成生物学的研究进展，IET（The Institution of Engineering and Technology）委员会主办的杂志内容、论坛等。

点　　评：IET 委员会主办的有关系统生物学与合成生物学的网站。

6. http://www.syntheticgenomics.com/

主要内容：介绍 Synthetic Genomics 公司在利用合成生物学等方法进行基因组研究与合成方面的有关内容。

点　　评：Synthetic Genomics 公司的网站。便于了解 J. C. Venter 等的研究现状。

7. http://www.synbiosafe.eu/

主要内容：欧盟关于合成生物学相关的安全、道德的网站，包括这方面的议题、文章和措施等内容。

点　　评：有关生物安全和对道德的冲击一直是全世界广泛关注的课题，这个网站正是围绕这方面的内容建立的。

8. http://www.etcgroup.org

主要内容：ETC 技术监督组织（The Action Group on Erosion，Technology and Concentration）的官方网站。

点　　评：ETC 技术监督组织一直致力于维护和监督人类文明、生态多样性和人权。涉及的范围非常广泛，是国际上较有影响力的非官方组织之一。

9. http://openwetware.org

主要内容：一个非常开放的非官方的生物学和生物工程有关的网络信息平台，涉及生物研究相关的方方面面。

点　　评：用户可以申请建立自己的网页，方便而灵活。其中 BBF 基金会和 iGEM 组委会就利用这个平台建立了自己的网页来推广和普及合成生物学。

附录A 名词解释

“限制-修饰”现象 1952年S. E. Luria和M. L. Human在研究T偶数噬菌体对大肠杆菌的侵染实验中发现细菌能将外来的DNA片段在某些专一位点上切断，从而保证自身不被外来噬菌体所侵染，而其自身的DNA由于被一种特殊的酶所修饰（甲基化酶）而受到保护，即所谓的“限制-修饰”现象。

合成生物学 合成生物学是指按照一定的规律和已有的知识，①设计和建造新的生物零件、装置和系统；②重新设计已有的天然生物系统为人类的特殊目的服务。[Synthetic Biology is: A) the design and construction of new biological parts, devices, and systems, and B) the re-design of existing, natural biological systems for useful purposes.]

即插即用 即插即用（Plug and Play，PnP）是一种计算机硬件的一般术语，指在计算机上加上一个新的外部装置时，能自动检测和配置相应系统资源，而不需要重新组态或手动安装驱动程序。

开放读码框 开放读码框（open reading frame，ORF）是一段未知功能的DNA序列的称谓，它可能为一个未知功能的基因编码，或者是一个未知基因的一部分。通常情况下，在不为基因编码的随机DNA序列中，每20个左右连续密码子就可能出现一个终止密码子（3/64），因此，如果某一段读码框含有50个以上连续密码子而无终止密码子，则称这一段DNA含有一个“开放读码框”。

遗传线路 遗传线路（genetic circuit），俗称“基因线路”（gene circuit）。在合成生物学中，由各种调节元件和被调节的基因组合成的遗传装置（genetic device），可以在给定条件下定时、定量地表达基因产物。——此定义由南开大学生命科学学院生物化学和分子生物学系黄熙泰教授拟定。

遗传工程 遗传工程（genetic engineering）俗称“基因工程”，是生物工程的一个重要分支，指在分子水平上对基因进行操作，将外源基因通过体外重组后导入受体细胞内获得“克隆”，并使其能在受体细胞内复制、转录、翻译表达的操作。

分子生物学 分子生物学是指在分子水平上研究生物现象的科学，通过对生物大分子（核酸、蛋白质）的结构、功能和生物合成等方面的研究，来阐明各种生物现象的本质。其研究内容包括遗传信息及传递、生物进化、光合作用、发育的分子机制、神经活动的机理和癌症的发生等过程。

细胞生物学 细胞生物学是以细胞为研究对象，从细胞的整体水平、亚显微水平、分子水平三个层次，以动态的观点研究细胞和细胞器的结构和功能、细胞的生活史及各种生物活动规律的学科。

系统生物学 系统生物学就是在细胞、组织、器官和生物体整体水平上研究结构和功能各异的生物分子及其相互作用，并通过计算生物学来定量阐明和预测生物系统的功

能、表型和行为的学科。

TA克隆方法 TA克隆方法（original TA cloning kit）的原理是：PCR反应中所适用的*Taq*酶具有末端转移的活性，能够在PCR产物的3′端自动加上一个非模板依赖的3′-A突出端，只有用经过特殊处理的具有3′-T突出末端的DNA片段才能通过T/A配对进行连接。而T载体是一种带有3′-T突出端的载体，在连接酶作用下，可以快速地、一步到位地把PCR产物直接插入到质粒载体的多克隆位点。所以TA克隆不需使用含限制酶序列的引物，不需将PCR产物进行优化，不需把PCR产物做平端处理，不需在PCR扩增产物上加接头，即可直接进行克隆。

PoPS PoPS（[RNA] Polymerase Per Second）即RNA聚合酶每秒，用于衡量基因的被转录水平，指对于每个DNA拷贝来讲RNA聚合酶分子每秒钟通过DNA分子上某一点的数量。

RIPS RIPS（Ribosomal initiations Per Second），用于衡量mRNA的被翻译水平，指对于每一个mRNA来讲，核糖体分子每秒钟通过mRNA分子上某一点的数量。

启动子 启动子（promoter）是操纵子（operon）的一个组成部分，专一地与RNA聚合酶结合并决定转录从何处起始的部位，控制基因表达（转录）的起始时间和表达的程度。

核糖体结合位点 核糖体结合位点（Ribosome binding site，RBS）是指mRNA分子中紧靠启动子下游的、起始密码子AUG上游的一段非翻译区序列，用于结合核糖体以便开始转录。

SD序列 在原核生物中，mRNA中与核糖体16S rRNA结合的序列称为SD序列（SD sequence），它是1974年由J. Shine和L. Dalgarno发现的，故此而命名。

终止子 终止子是指结束基因转录的DNA信号序列。

ρ因子 ρ因子是一个相对分子质量为2.0×10^5、由相同的6个亚基组成的六聚体蛋白质。它能水解各种核苷三磷酸，实际上是一种NTP酶，利用NTPase水解反应提供能量。由于催化了NTP的水解，ρ因子能促使新生的RNA链从三链转录复合物中解离出来，从而终止转录。

操纵子 操纵子（operon）是细菌的基因表达调节装置，由启动子和其他DNA调节元件与串联的多个相关基因组成，并由同一套调节蛋白调节。

报告基因 报告基因是指其产物易于被检出的基因，在分子生物学实验中用于替换天然基因的位置，以检验其启动子及调节因子的结构组成和效率。

转换器 转换器（inverter）是一种遗传装置，它在接收某种信号时停止下游基因的转录，未接收到信号时开启下游基因的转录。

信号转导装置 信号转导装置（signaling）是指环境与细胞之间或者邻近的细胞与细胞之间接收信号和传递信号的装置。

蛋白生成装置 蛋白生成装置（protein generator）是产生具有一定功能蛋白质的装置。

超敏感遗传线路 超敏感遗传线路对诱导物浓度在一个很窄范围内的变化具有快速响应能力，即使微弱的输入信号，一旦达到阈值即可快速激活遗传线路，其响应曲线非

常类似于典型的阶跃响应。

群体感应 群体感应（quorum sensing，QS）现象最早在海洋中的一种发光弧菌（*Vibrio fischeri*）中被发现。近年来的研究证明微生物之间存在着信息交流，许多微生物都能合成并释放一种被称为自体诱导物质（autoinducer，AI）的信号分子，胞外的AI浓度能随微生物密度的增加而增加，当达到一个临界浓度时，AI能启动菌体中相关基因的表达，调控微生物的生物行为，如产生毒素、形成生物膜、产生抗生素、生成孢子、产生荧光等，以适应环境的变化，这一现象被称为群体感应现象。因这一现象只有在微生物密度达到一定阈值后才会发生，所以也称为依赖细胞密度控制的基因表达（cell density dependent control of gene expression）。

AHL AHL即*N*-乙酰基高丝氨酸内酯（*N*-acyl-homoserine lactone，AHL），是一种化学小分子，能够透过细胞膜扩散到细胞周围环境中。

调控网络基元 调控网络基元（motif）是转录因子和靶基因之间相互调控关系的特定小规模组合，通常由一组相关基因和调节元件按照一定的拓扑结构构成。

基础基因线路 基础基因线路（elementary gene circuit）是比较简单的基因线路形式，即一个基因的表达受单一转录因子的调节，并在给定条件下对一种信号分子作出反应。

反馈 反馈是指系统的输出会反过来影响系统的输入，进而影响自身的一种控制机制。

正反馈 基因表达的产物会反过来促进其上游基因表达。

负反馈 基因表达的产物会反过来抑制其上游基因表达。

可诱导基因线路 如果基因的表达量随着信号分子的加入而增加，这种遗传线路就被定义为可诱导的。

可阻遏基因线路 如果基因的表达量随着信号分子的加入而减少，这种遗传线路就被定义为可阻遏的。

前馈 前馈控制也称预先控制或提前控制，其基本原理就是测取进入过程的扰动量（包括外界扰动和设定值变化），并按照其信号产生合适的控制作用去改变控制量，使被控制变量维持在设定值上。前馈控制与反馈控制的最大区别就在于它是在偏差出现之前就采取控制措施，而后者则是在偏差出现之后才采取纠正措施。

一致前馈 一致前馈（coherent feedforward）是指前馈控制中直接和间接调控途径对输出模块的作用相同的一种前馈形式。

非一致前馈 非一致前馈（incoherent feedforward）是指前馈控制中直接和间接调控对输出模块的作用相反的一种前馈形式。

无标度网络 1999年，印第安纳州圣母大学物理学教授A. L. Barabási及其同事在对万维网拓扑结构进行研究时发现考察的情况比随机网络所描述的要复杂，即钟形曲线的连接平均数或标度不见了，它所产生的是一条以负幂函数形式不断递减的曲线。Barabási把具有这种性质的网络称之为无标度网络（scale-free networks）。无标度网络中大多数节点的连接数很低，只有少数几个节点的连通性好，起着网络集散中心的作用。

常微分方程 常微分方程（ordinary differential equation ODE）指如果一个微分方

程中出现的未知函数只含一个自变量，这个方程就叫做常微分方程，也可以简单地叫做微分方程。

酶促反应动力学 酶促反应动力学简称酶动力学，主要研究各种酶在酶促反应中的特性（如速度常数），以及一些影响因子（如温度、底物浓度、抑制剂等）对反应速度的影响。

模式 所谓模式，广义上讲，是指一些供模拟用的、完美无缺的样本。存在于时间和空间中可观察的事物，如果我们可以区别它们是否相同或相似，都可以称之为模式。但模式不是事物本身，而是我们从事物中获得的信息。

模式识别 模式识别就是识别出特定客体所模拟的样本，其作用和目的在于面对某一具体事物时将其正确地归入某一类别。

监督学习方法 在进行分类器设计时，假定每个样本的类别都是已知的，即每个样本均被“标记”了类别属性，利用已知标记类别的样本集进行分类器设计的方法称为监督学习方法或有导师学习方法，也称为分类分析。

非监督学习方法 在大多数实际应用中，无法预先知道样本的类别，只能根据无样本类别标记的样本集进行分类器设计，这就是通常说的非监督学习方法或无导师学习方法，也称为聚类分析。

稳定性 稳定性（stability）反映在暂态扰动消失后的时间响应性质上，可以这样来定义：系统在任何足够小的初始偏差的作用下，对于暂态扰动的时间响应随着时间的推移而逐渐衰减并趋向于零，则该系统是稳定的；否则，该系统是不稳定的。

鲁棒性 鲁棒性（robustness）就是系统的健壮性，它是在异常和危险情况下系统生存的关键。所谓“鲁棒性”，是指系统在一定（结构、大小）的参数摄动下，维持某些性能的特性。例如，计算机软件在输入错误、磁盘故障、网络过载或有意攻击情况下，能否不死机、不崩溃，就是该软件的鲁棒性。

响应快速性 所谓的响应快速性就是处于稳定状态的系统对于外部环境改变的快速反应能力，也是系统在受到扰动后迅速进入稳态的能力。对于生物模块来讲，任何输入信号，无论是信号分子、蛋白质，还是代谢物，都会降解、排出或者转化成其他物质。生物模块必须要在这些输入信号消失之前做出响应，才能确保自身功能的发挥。

基因开关 基因开关是指某种化学诱导物存在或缺乏时，或者在两个独立的外源刺激作用下，基因处于两种可能状态中的一种状态的系统。

Riboswitch 生物体能够通过各种精巧的机制控制基因表达的时间和数量，几年前发现的 Riboswitch 就是这些精巧机制中的其中一种。Riboswitch 主要是通过核糖核酸（RNA）构象的改变来实现“开关”的功能，阻止或开启目的蛋白的生成。大部分 Riboswitch 只有一个识别靶向配体的结合位点或适配体（aptamer），适配体一般位于基因表达区域附近，当与代谢产物结合时，会改变自身结构，在转录或翻译水平上行使基因调控的功能。Riboswitch 可被一种调节 sRNA（此处指小片段的调节 RNA，非编码 RNA 的一种，对于真核和原核生物的调节作用非常重要）“打开”，重新激活转录或翻译。Riboswitch 在进化上可能是 RNA 世界遗留的分子化石，广泛存在于革兰氏阴性菌和革兰氏阳性菌的代谢相关基因中，在真菌和植物中也有发现。Riboswitch 调节基因

表达不需要任何蛋白因子作为中介，主要参与调节维生素、氨基酸和核苷酸等基础物质的代谢过程，因此可用于研究基因功能，开发新型药物及进行基因治疗。

双稳态开关 双稳态开关也称“拨动开关”，可通过人为调控实现基因线路在两种不同稳定状态间的切换。经典的转录水平双稳态开关由两个启动子组成，每个启动子的“开”和“关”状态之间具有明显的界限，只有通过瞬态的诱导因素变化才能切换拨动开关至其中的一个稳态，即使在移除其输入激励后仍会维持原来的状态。

细胞凋亡 细胞凋亡又称程序性细胞死亡（programmed cell death，PCD），是一种高度调节的生理性细胞死亡过程。与坏死不同，细胞凋亡是指细胞接受某种信号或受到某些因素刺激后，会发生主动的、由基因调控的消亡过程。

支架蛋白 支架蛋白（scaffolding protein）号称分子胶水，是信号转导系统中存在的一种本身不具备酶活性的蛋白质。支架蛋白在调节信号转导中起着重要作用，通常含有多个蛋白结合域，能同时结合多个蛋白质。细胞中，构成通路的信号蛋白分子通常由支架蛋白物理地组装成联合体，通过支架来定向信息流、提高匹配蛋白的信号传递，同时通过辅助蛋白激酶和磷酸酶与它们各自的底物结合而起作用，保证信号传递的特异性和高效性。支架蛋白还能起到生理性隔离的作用，阻止不匹配蛋白间的相互交流和防止与其他通路发生互相作用。同时，支架联合体作为反馈回路的集散中心承担调节蛋白活性和实现蛋白重新组装的作用，从而实现对信号通路动态响应行为的调节。

辛烷值 辛烷值就是衡量汽油在气缸内抗爆震燃烧能力的一种数字指标，其值高表示抗爆性好。

高级醇 分子中碳原子个数大于 2 的醇类。

浊点 油类、清漆等液体样品在标准状态下冷却至开始出现混浊的温度为其浊点（cloud point）。

倾点 倾点（pour point）是指油品在规定的试验条件下，被冷却了的油样开始连续流动时的最低温度。

光敏色素 光敏色素（phytochromes）也称作“感光蛋白”，是由结合在细胞膜上的胞外感受器和胞内应答调节子组成的一个双组分系统，其单体包括一个生色团（chromophore）和一个脱辅基蛋白（apoprotein）。

膜孔蛋白 孔蛋白又称通道形成蛋白，它的跨膜结构域形成 β 折叠片结构，反向平行的 β 折叠片相互作用形成非特异的跨膜通道，允许相对分子质量较大的小分子自由通过。膜孔蛋白（porin）就是细胞膜上的孔蛋白。

汉密尔顿路径问题 汉密尔顿路径问题，即如何在有向图中找到一条遍历图中所有节点一次且仅一次的最短路径，是计算机图论中非常著名的一道 NP 完全型问题。

NP 完全型问题 NP 完全型问题（NP complete）是指无法用多项式求解，只能将所有可能解遍历一遍之后才能得到最终解的问题。

疫苗 疫苗是灭活的抗原（antigen），是消除感染性而保留抗原性的物质。疫苗不能使人体生病，但可通过抗原使机体获得抵抗感染的免疫能力，这种免疫通常称为特异性免疫、获得性免疫或适应性免疫。当真正的病毒或微生物入侵时，获得性免疫反应就会更快、更强地产生抗体，消灭病原体。

表位 免疫细胞通常难以识别整个抗原分子，而仅识别抗原大分子上特定的部分，这种特定的部位称为表位（epitope）或抗原决定簇（antigenic determinant），是抗原引起免疫应答的关键部分。严格说来，抗体的特异性是针对表位而不是针对完整的抗原分子。表位代表了抗原分子上的一个免疫活性区，负责与免疫细胞表面的抗原受体或游离的抗体分子相结合。表位通常可以根据引起体液免疫或细胞免疫的不同分为B表位和T表位。一般情况下，当抗原分子质量大、化学组成复杂、抗原表位表露时，其有效性才会高。

空泡细胞毒素 空泡细胞毒素（vacuolating cytotoxin，VacA）是一种气泡形毒素，能够导致哺乳动物细胞质形成过量的空泡，并且能够造成膜电位的去极性、改变线粒体的膜渗透性、导致凋亡、细胞从基底膜脱离、有丝分裂活性蛋白激酶的激活、抑制抗原表达、抑制T细胞活性和增殖等多种不良反应。

受体激动剂 受体激动剂是指对受体有较强亲和力和内在活性，能通过受体兴奋发挥最大效应的药物。

干扰素 干扰素（interferon，IF）是细胞在一定诱生剂的作用下所产生的一类蛋白质。它具有广谱抗病毒、抗肿瘤以及免疫调节的活性，是细胞功能的重要调节物质。I型干扰素能够避免过度发炎，非常适合用做疫苗。

中和抗体 中和抗体是由适应性免疫应答细胞分泌的一种可溶性蛋白。病毒侵入人体之后，免疫细胞把中和蛋白分泌到血液里，后者与血液里的病毒颗粒结合，阻止病毒感染细胞，破坏病毒颗粒，相当于把病毒“中和”掉了。

艾滋病 医学全名为“获得性免疫缺陷综合征”（acquired immune deficiency syndrome，AIDS），是人体感染了人类免疫缺陷病毒（HIV）又称艾滋病病毒所导致的传染病。HIV病毒终生传染，破坏免疫系统，但其本身并不会引发任何疾病，而是在免疫系统被HIV破坏后，使人体对威胁生命的各种病原体丧失了抵抗能力，从而发生多种感染或肿瘤，最后导致死亡的一种严重传染病。

高效抗逆转录病毒治疗 “高效抗逆转录病毒治疗”（highly active antiretroviral therapy，HAART）是目前普遍采用的艾滋病治疗方法，利于缓解艾滋病患者的病痛和降低艾滋病的死亡率。HIV病毒的侵染周期主要可以分为四步：①HIV病毒侵染宿主细胞，脱掉外面的核衣壳和包膜，病毒的sRNA进入宿主细胞；②HIV病毒的sRNA逆转录为DNA；③逆转录的DNA整合到宿主细胞的染色体上；④HIV的DNA随宿主细胞的染色体DNA一起转录、表达，生成自己的蛋白质后装配成新的HIV病毒继续侵染其他宿主细胞。HAART即是针对艾滋病侵染的第二步进行治疗的，即：只有成功阻止HIV病毒RNA逆转录成DNA才能使病毒的DNA拷贝不能整合到宿主细胞的染色体DNA上，因此干扰HIV逆转录为DNA的阶段，即可以达到治疗的目的。但是该疗法也存在缺点：①对于已经和宿主细胞DNA整合的病毒来说HAART不起作用，治疗效果不够彻底；②这种疗法对人体有副作用；③其价格昂贵，难以普遍使用。

脓毒症 脓毒症（Sepsis）是败血症的一种，指宿主对微生物感染的全身性炎症反应，通常指由病原微生物引发的败血症。Sepsis的病原多数为革兰氏阳性或革兰氏阴性菌，但病毒、立克次体、真菌，以及微生物分子信号或毒素的全身播散也可引发Sepsis。

Toll样受体 Toll样受体（Toll-like receptor，TLR）均为I型跨膜蛋白，由胞外

区、跨膜区和胞内区三部分组成。① 胞外区：由富含亮氨酸的重复序列区（leucine-rich repeat，LRR）组成。② 胞内区：与 IL-1 受体（IL-1R）胞内段有很高的同源性，称为 TIR 结构域（Toll/IL-1 receptor homologous region，TIR），是一保守结构，其中包括 3 个保守盒（conserved box）。

病原相关分子模式 1989 年，美国免疫学家 C. A. Janeway Jr 和 R. Medzhitov 提出模式识别理论，认为某些病原体或其产物共有在进化上高度保守的特定分子结构，称为病原相关分子模式（pathogen-associated molecular pattern，PAMP）。PAMP 能被固有免疫细胞表面的相应受体，即模式识别受体（pattern recognition receptor，PRR）所识别，以保证免疫识别作用的相对特异性。

附录 B　iGEM 参赛项目选介

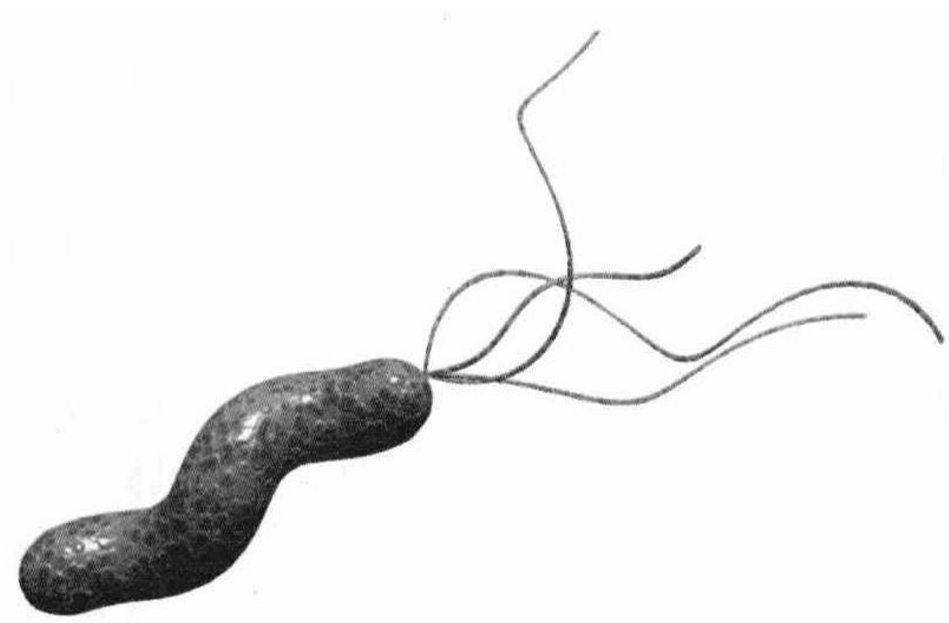
此部分内容在作者发稿前并未见有正式发表，只是为了方便老师和学生对 iGEM 有所了解，而根据参赛队的参赛网页、PPT 和海报的内容编写成的，对其可实现性和可靠性不承担任何责任，读者也应该抱着理性的态度去阅读。有兴趣的读者可以参阅相关参赛网页获取更详细的信息。

1. 幽门螺杆菌疫苗

iGEM2008，唯一大奖，斯洛文尼亚的 Ljubljana 大学参赛项目。

参赛网页：http://2008.igem.org/Team:Slovenia

幽门螺杆菌（*Helicobacter pylori*，*H. pylori*）是一种螺旋状微好氧细菌，具有极性鞭毛，能够在人体的胃黏膜表面附近存活（附图 1）。数万年来，幽门螺杆菌一直与人类共同进化、共同迁移。全世界几乎半数以上人的胃肠道中都有幽门螺杆菌，第三世界国家的人口更是如此。



附图 1　幽门螺杆菌示意图

幽门螺杆菌主要是通过飞沫传播，人类通常会在童年时就感染上幽门螺杆菌，从此携带一生。虽然 80%～90%的感染人群没有任何症状，但感染幽门螺杆菌的人患上胃炎和胃溃疡的可能性极大。近二十多年的研究发现，幽门螺杆菌感染是慢性活动性胃炎、消化性溃疡、胃黏膜相关淋巴组织（MALT）淋巴瘤和胃癌的主要致病因素。1994 年世界卫生组织/国际癌症研究机构（WHO/IARC）将幽门螺杆菌定为 I 类致癌原。因此，从医学角度讲，幽门螺杆菌是一种危险的病原微生物。

正常情况下，人类的胃壁有一系列完善的自我保护机制（分泌胃酸和蛋白酶的功能、不溶性与可溶性黏液层的保护作用、有规律的运动等），能抵御经口而入的大部分微生物的侵袭。但幽门螺杆菌却是唯一能突破这一天然屏障而在人类胃部的恶劣环境下仍然能够繁殖的生物：幽门螺杆菌可以利用它的极性鞭毛对 pH 梯度做出趋化反应；通过黏附素，牢牢地与上皮细胞连接在一起，避免随食物一起被胃排空；分泌过氧化物歧化酶（SOD）和过氧化氢酶，以保护自身不受中性粒细胞的杀伤；富含尿素酶，通过尿素酶水解尿素产生氨，在菌体周围形成“氨云”保护层，抵抗胃酸的杀灭作用。除了破坏细胞和物质障碍，幽门螺杆菌还具有逃避宿主先天和后天免疫系统的能力。其鞭毛蛋

白已经进化到能够掩护它们而不被 Toll 样受体 5（一种先天免疫受体，能够识别许多其他细菌的鞭毛）所发现。

目前治疗胃溃疡和幽门螺杆菌感染的药物主要包括具有质子泵抑制剂（proton pump inhibitor）作用的抗生素来降低胃酸；保护细胞免受化学毒素等刺激的细胞保护剂（cytoprotective agents）。虽然这些方法的治愈率高达 90%，但仍然会有严重抗药性或二次感染的情况发生。如果能够通过 DNA 或者蛋白质疫苗使人类具有抵抗幽门螺杆菌的能力，则有望将人类从与其纠缠了数万年的幽门螺杆菌的侵扰中解脱出来。为此，斯洛文尼亚的 Ljubljana 大学 iGEM2008 参赛队力图通过合成生物学设计能够编码幽门螺杆菌疫苗的融合基因，通过融合基因的转化或者融合蛋白的注入，使人体产生对幽门螺杆菌的抗体，具有相应的免疫能力。

疫苗是灭活的抗原（antigen），是消除感染性而保留抗原性的物质。疫苗不能使人体生病，但可通过抗原使机体获得抵抗感染的免疫能力，这种免疫通常称为特异性免疫、获得性免疫或适应性免疫。当真正的病毒或微生物入侵时，获得性免疫反应就会更快、更强地产生抗体，消灭病原体。

参与特异性免疫应答的免疫细胞有很多种，其中最重要的是 T 淋巴细胞和 B 淋巴细胞。B/T 淋巴细胞受病原体刺激后，引起一系列变化，所产生的抗体通过各种方式来消灭病原体。此外，具有吞食异物作用的巨噬细胞也是一种重要的免疫细胞，它对异物进行加工处理，处理后的异物（抗原）与 T 淋巴细胞和 B 淋巴细胞发生免疫反应。巨噬细胞本身也能直接杀灭异物或者产生细胞因子参与免疫反应。

免疫细胞通常难以识别整个抗原分子，而仅识别抗原大分子上特定的部分，这种特定的部位称为表位（epitope）或抗原决定簇（antigenic determinant），是抗原引起免疫应答的关键部位。严格说来，抗体的特异性是针对表位而不是针对完整的抗原分子。表位代表了抗原分子上的一个免疫活性区，负责与免疫细胞表面的抗原受体或游离的抗体分子相结合。表位通常可以根据引起体液免疫或细胞免疫的不同分为 B 表位和 T 表位。一般情况下，抗原的分子质量大、化学组成复杂、抗原表位表露，其有效性才会高。

由此可知，幽门螺杆菌的疫苗抗原中必须包含幽门螺杆菌的特异性部分才能引起特异性免疫反应。通常这些特异性部分主要由病原微生物的鞭毛蛋白组成，而幽门螺杆菌的鞭毛蛋白却无法被人体的免疫系统识别。因此，Ljubljana 大学参赛组成员利用合成生物学方法通过设计融合基因，对鞭毛蛋白进行改造，生成融合蛋白作为疫苗。融合蛋白中包括幽门螺杆菌鞭毛蛋白的核心部分，以及能够被人体免疫系统识别的大肠杆菌鞭毛蛋白的氨基酸和羧基序列。同时融合基因的设计还考虑了空泡细胞毒素（vacuolating cytotoxin，VacA）和黏附素 A 的抗原部分，并通过巧妙的设计利用强制二聚体结构避开了对疫苗辅药的需求，使得所设计的新型疫苗在不需要任何 Toll 样受体激动剂（TLR agonists）的帮助下即可发挥作用。该项目共分两大块。

1）工程化基于幽门螺杆菌鞭毛蛋白的疫苗

（1）对疫苗抗原的改造，使其能够成功激活 TLR5，从而激活免疫响应。

新疫苗的抗原核心部分是幽门螺杆菌的鞭毛蛋白，但自然状况下幽门螺杆菌的鞭毛

蛋白不会激活 TLR5（TLR5 在抵御大多数鞭毛细菌的过程中具有关键作用，主要分布在免疫系统细胞和上皮细胞表面）。而 TLR5 的激活能够触发上皮细胞产生 CCL20、Th1 响应和黏膜 IgA 产物的产生从而激活一系列强有力的免疫响应，因此需要对幽门螺杆菌的鞭毛蛋白进行改造。

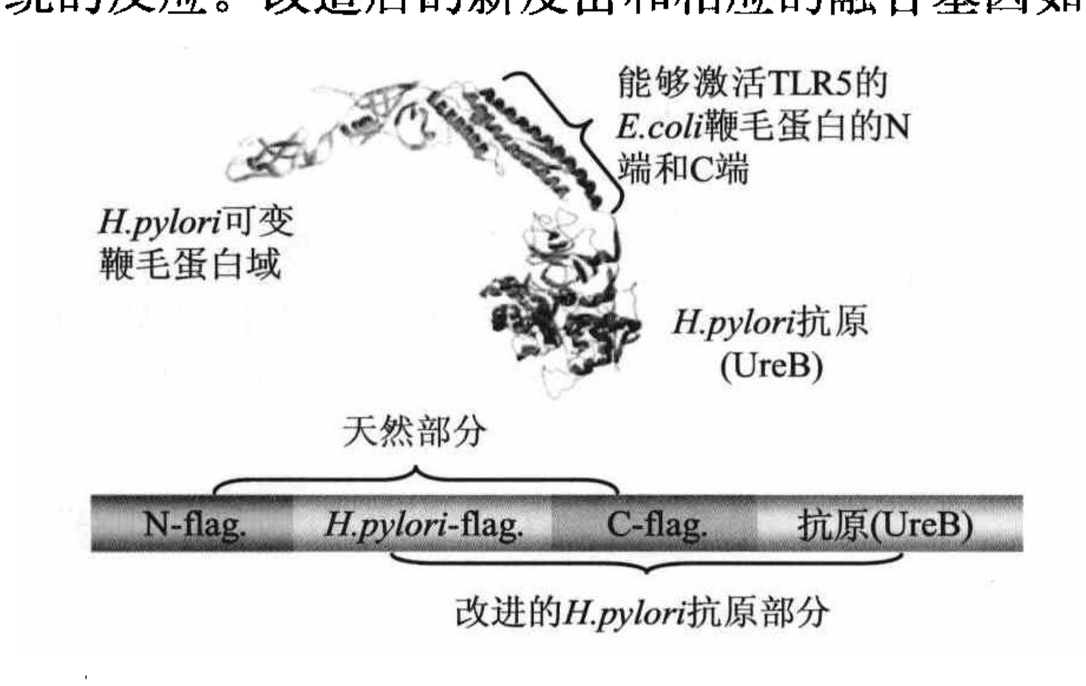
鼠伤寒沙门菌（*S. typhimurium*）、大肠杆菌等细菌鞭毛蛋白的氨基酸和羧基序列均能够激活 TLR5。如果将幽门螺杆菌鞭毛蛋白的 N 端和 C 端片段替换成大肠杆菌鞭毛蛋白的 TLR5 激活片段，则幽门螺杆菌的鞭毛蛋白就可激活 TLR5 受体。因此，新疫苗中的鞭毛蛋白需要包含幽门螺杆菌抗原片段的主要核心部分——鞭毛蛋白 FlaA，同时用大肠杆菌的 FliC 鞭毛蛋白替换其 N 端和 C 端片段。另外，将鞭毛蛋白的 C 端与对幽门螺杆菌具有毒性的抗原基因或多抗原决定簇（epitopes）片段融合，从而可以通过来自大肠杆菌的 FliC 的 N 端和 C 端激活 TLR5，还可通过幽门螺杆菌的抗原和 FlaA 核心片段激活免疫响应。

（2）设计多抗原决定簇。

为了增强免疫细胞 $CD4^+$ T、$CD8^+$ T、B 细胞之间的交互作用，研究者还设计了合成的多抗原决定簇，包含 *H. pylori* 的尿素酶 B（UreB）、空泡细胞毒素（vacuolating cytotoxin，VacA）和黏附素 A（HpaA）的抗原决定簇，并将这个多抗原决定簇与新鞭毛蛋白的 C 端融合。尿素酶 B 抗原决定簇包括一个尿素酶 β 亚基激活位点，能够引起抑制尿素酶的酶活性物质的产生，帮助根除致病细菌；VacA 是一种气泡形毒素，能够导致哺乳动物细胞质过量的空泡形成，并且能够造成膜电位的去极性、改变线粒体的膜渗透性、导致凋亡、细胞从基底膜脱离、有丝分裂活性蛋白激酶的激活、抑制抗原表达、抑制 T 细胞活性和增殖等多种不良反应。VacA 细胞毒素的激活需要 VacA 单体组合成低聚体结构，而中和抗体能够阻止低聚物的形成，从而抑制 VacA 细胞毒素。

（3）引入 MHC-II 分子。

巨噬作用早期受体-配体的交互作用是免疫过程的关键。最近的研究发现，主要组织相容性复合体 MHC-II 分子（major histocompatibility complex class 2 molecules）的存在能够促进配体与 TLR 的咬合，增强抗原递呈细胞 APC 和 T 辅助细胞之间的交互作用，促进免疫系统的反应。改造后的新疫苗和相应的融合基因如附图 2 所示。



附图 2　融合到 UreB 上的新鞭毛蛋白疫苗分子模型

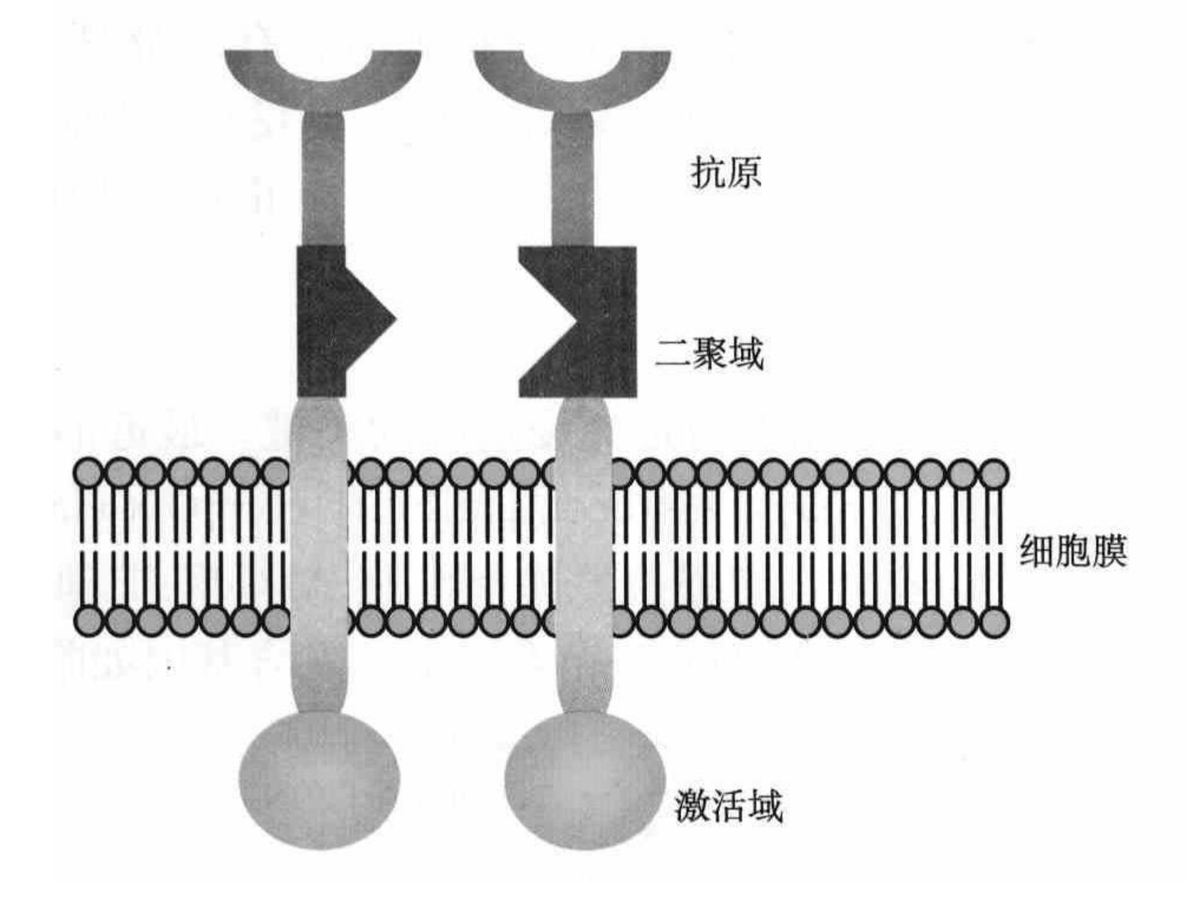
2）工程化新疫苗使其不需任何辅助 agonists

工程化新疫苗，使其不需要任何 TLR agonists 的辅助作用即可产生足够的免疫作用，同时具有触发不同信号通路或其组合通路产生抗原处理过程和一组协同刺激的分子产物的功效。这方面的工作需要克服两大难题。

（1）Agonists 是能激活受体的配体，能够帮助激活免疫响应。作为疫苗辅药的 TLR agonists 中除了鞭毛蛋白之外，最有效的部分就是能够激活 TLR9 的 CpG、激活 TLR3 的 poly（I：C）以及激活 TLR4 的 MPLA。这些配体的刺激能够触发 I 型干扰素的产生。如果将 agonists 与疫苗做成融合蛋白，则疫苗自身就会具有 agonists 的功效而不需任何外加 agonists 的辅助。然而这些 TLR agonists 的有效部分不是蛋白质，不能像前述鞭毛蛋白一样与疫苗融合编码。因此，必须选择其他不需与 TLR 配体融合编码的方法来激活这些 Toll 样受体。

（2）激活 Toll 样受体的另外一个重要因素是 TLR 与抗原在同一细胞区室中协同激活。因为即使被可溶解的 TLR agonists 激活的细胞也不能触发被吞噬蛋白的水解过程，只有在含有 TLR agonists 或者 TLR 激活时才能触发此过程。这是保证只有病原微生物（而不是内源细胞）才能特异性激活 TLR 触发免疫响应的分子机制。由此可知，为了使疫苗发挥功效，必须使 TLR 与抗原在同一细胞区室中协同激活。为此研究者设计了一个能够融合抗原和激活 TLR 的构件，确保免疫的激活和抗原过程在细胞内同一处所发生，从而避开对 TLR agonists 的需求。

为了实现上述目的，需要将下列元件整合进一个融合蛋白中，如附图 3 所示。



附图 3　抗体-TLR 融合构件示意图

通过强制胞质信号域与附加二聚域组成二聚体来缩短 TLR 信号路径和激活特定活性的 TLR。在 N 端片段上附加一个特定抗原可使其具有更好的优势。

• 特定 TLR 的胞质 TIR 域，负责胞内信号级联的激活，导致 I 型干扰素的产生，适用于 TLR3、TLR9 和 TLR4。

• 跨膜域，负责构件的准确定位。

• 二聚域，产生构件的二聚体结构，作为下游信号的适配子。

• 抗原，考虑到细胞表面 MHC-II 分子的存在而具有特定抗原决定簇。

• 多个缩氨酸标志域，用于细胞定位和跟踪细胞活性。

设计这一构件的目的是通过强制胞质信号域与附加二聚域组成二聚体来缩短 TLR 信号路径和激活特定活性的 TLR。

基于该参赛队 2007 年的研究可知：TLR3 能够利用不同的信号通路与适配分子 TRIF（不是其他 TLR 常用的 MyD88）具有良好的交互作用。与 TLR4 主要影响 NFκB 活性的作用不同，TLR3 信号通路的最终作用是增加 β 干扰素的产量。TLR3 通常参与核酸识别，不会在细胞表面出现，却可被重定向到内涵体，并在那里与从裂解的病毒中脱离的双链 RNA 结合；TLR3 和 TLR9 能够导致 I 型干扰素的产生，最大限度地支持有效疫苗接种；TLR4 能够利用多种不同的适配分子产生很强的细胞活性。因此，项目中选择 TLR4、TLR3 和 TLR9 为目标，激活 Toll 样受体作为启动免疫响应的受体。

TLR4 很容易被 TM 和 TIR 与几乎任何胞外域的融合而激活，因此设计的构件中相应于 TLR4 的部分仅用 TM 和 TIR 与特定抗原融合而没有任何附加二聚域。TLR3 的激活则没这么简单，CD4 的胞外域与任何 TLR 的 TM 和 TIR 域的融合能够极大地增强 TLR3 的活性。因此在疫苗中融合了 TLR3 和 TLR9 的 agonists：poly（I：C）和 CpG。

通过上述两大部分内容的设计，Ljubljana 大学研制的新型蛋白疫苗能够产生识别幽门螺杆菌的抗体，同时由于自身具有特定的融合蛋白，在不需任何辅助 agonists 的情况下即可发挥很好的功效，具有同时抵抗空泡细胞毒素和黏附素 A 的作用。

目前该项目已经进入了动物试验阶段，注射后小鼠体内血清抗体的实验数据证明了该疫苗的有效性。

注：为方便理解，此部分加入了一些疫苗相关的背景知识，其余以网页内容为准。

2. 艾滋病的治疗

iGEM2007，Finallist，斯洛文尼亚的 Ljubljana 大学参赛项目。

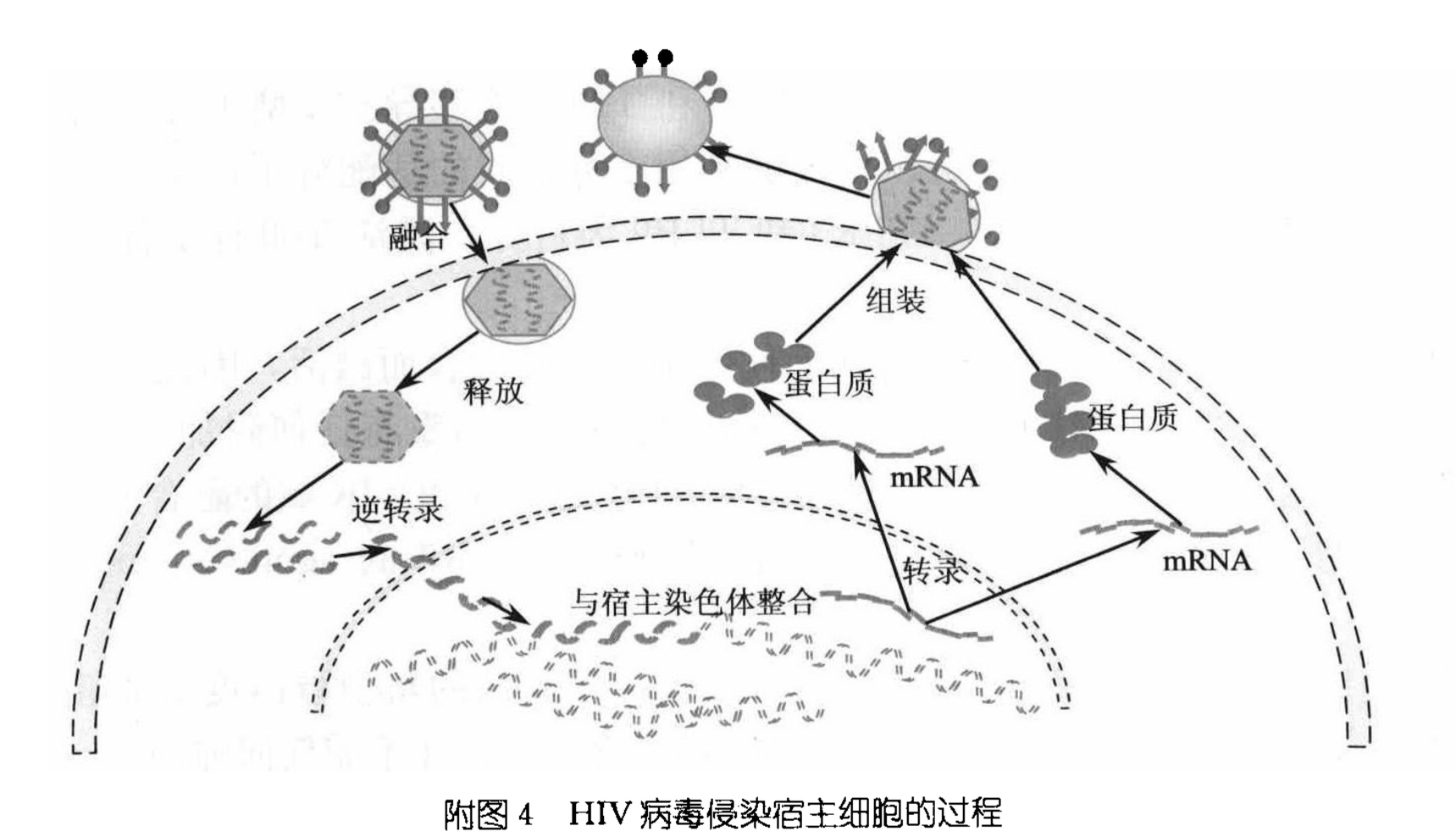
参赛网页：http://parts.mit.edu/igem07/index.php/Lethbridge

2007 年 Ljubljana 大学参赛队利用合成生物学理念对以免疫缺陷为特征的艾滋病即获得性免疫缺陷综合征进行了研究。

艾滋病，医学全名为“获得性免疫缺陷综合征”（acquired immune deficiency syndrome，AIDS），是人体感染了人类免疫缺陷病毒（HIV）又称艾滋病病毒所导致的传染病。HIV 病毒终生传染，破坏免疫系统，但其本身并不会引发任何疾病。当免疫系统被 HIV 破坏后，人体对威胁生命的各种病原体丧失了抵抗能力，从而发生多种感染或肿瘤，最后导致死亡。艾滋病主要通过血液、性行为、母婴遗传等途径传播。国际医学界至今尚无防治艾滋病的有效药物和疗法，因此艾滋病也被称为“超级癌症”和“世纪杀手”。目前治疗艾滋病的主要难题是：HIV 病毒不仅攻击正常的

免疫细胞，而且突变率相当高，一个复制周期突变率高达 3×10^{-5}，如果每天产生 $10^{9}\sim10^{10}$个艾滋病病毒的话，就意味着有几万到几百万个病毒产生了变异，这给治疗带来了相当大的困难。

HIV 病毒的侵染周期主要分为四个：①HIV 病毒侵染宿主细胞，脱掉外面的核衣壳和包膜，病毒的 sRNA 进入宿主细胞；②HIV 病毒的 sRNA 逆转录为 DNA；③逆转录的 DNA 整合到宿主细胞的染色体上；④HIV 的 DNA 随宿主细胞的染色体 DNA 一起转录、表达，生成自己的蛋白质后装配成新的 HIV 病毒继续侵染其他宿主细胞。其具体过程可见附图 4。



附图 4　HIV 病毒侵染宿主细胞的过程

HIV 病毒的侵染期主要可以分为四步：①HIV 病毒侵染宿主细胞，脱掉外面的核衣壳和包膜，病毒的 sRNA 进入宿主细胞；②HIV 病毒的 sRNA 逆转录为 DNA；③逆转录的 DNA 整合到宿主细胞的染色体上；④HIV 的 DNA 随宿主细胞的染色体 DNA 一起转录、表达，生成自己的蛋白后装配成新的 HIV 病毒继续侵染其他宿主细胞。

由此可知，要对艾滋病进行治疗可针对性地从以下四个方面加以控制：①在 HIV 病毒侵染宿主细胞阶段将其杀死；②阻止 HIV 病毒的 sRNA 逆转录为 DNA；③抑制 HIV 病毒的 DNA 整合到宿主细胞的染色体上；④抑制 HIV 蛋白合成和加工。

目前普遍采用的艾滋病治疗方法是针对艾滋病侵染的第二步进行治疗的，即“高效抗逆转录病毒治疗”(highly active antiretroviral therapy，HAART)，它对于缓解艾滋病患者的病痛和降低艾滋病的死亡率是很有用的。HAART 的作用机理是：只有成功阻止 HIV 病毒 RNA 逆转录成 DNA 才能使病毒的 DNA 拷贝不能整合到宿主细胞的染色体 DNA 上，因此干扰 HIV 逆转录为 DNA 的阶段，即可以达到治疗的目的。但是该疗法也存在缺点：①对于已经和宿主细胞 DNA 整合的病毒来说 HAART 不起作用，治疗效果不够彻底；②这种疗法对人体有副作用；③其价格昂贵，难以普遍使用。因此，更有效、经济的治疗艾滋病的方法亟待发现。

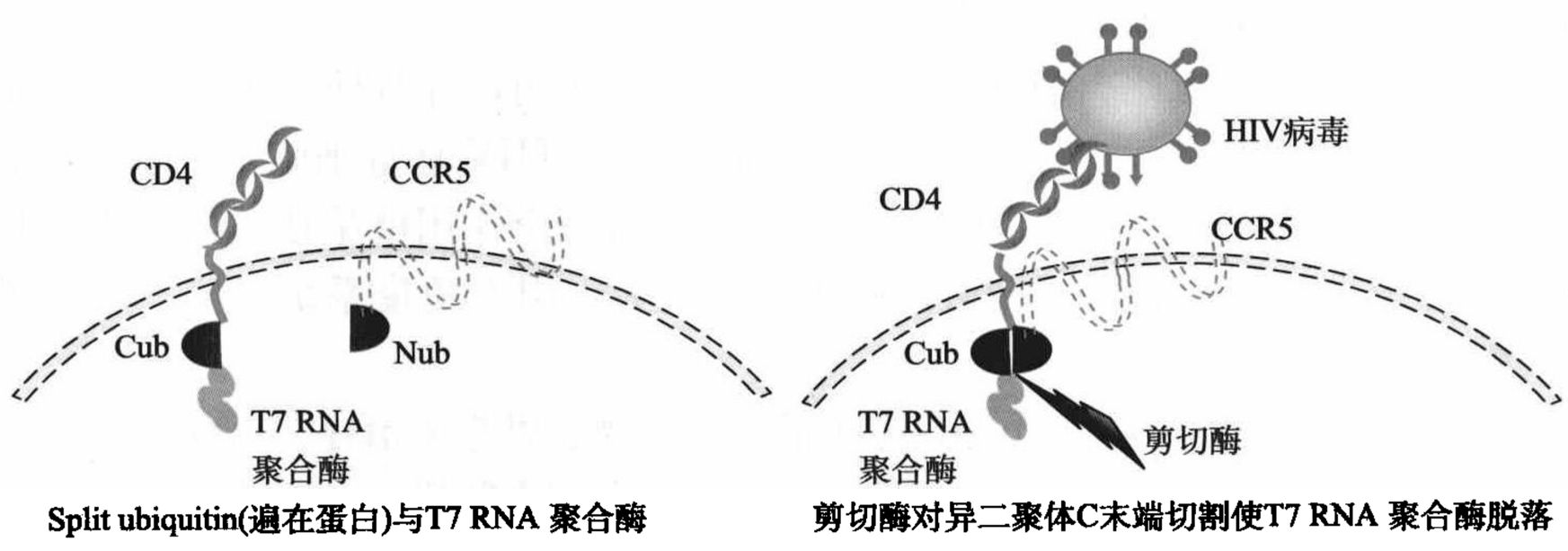
斯洛文尼亚的 Ljubljana 大学 iGEM07 小组成员想到了可以通过对病毒生命周期关键阶段进行控制来达到治疗艾滋病的目的，这样可以减少由于突变率过高而造成的治疗

上的难题。从可行性上考虑，他们最终挑选了两种方案：①在 HIV 病毒侵染宿主细胞阶段对其进行免疫；②抑制 HIV 表达成自己的蛋白质。其基本设计思路是：通过设计一个遗传学装置侦察 HIV 病毒，在 HIV 病毒感染的初期和末期分别利用特殊的脱落蛋白和 HIV 蛋白识别信号使 T7 RNA 聚合酶从细胞膜上脱落，进而利用脱落的 T7 RNA 聚合酶与 T7 启动子结合，启动下游基因的表达，产生免疫蛋白来杀死 HIV 病毒，并将信号进行转导和放大，激活机体免疫系统进行反应。

1）T7 RNA 聚合酶的脱落

（1）感染初期的抵御机制：这里面用到了特殊的脱落蛋白 Split ubiquitin（泛素）和 TEV protease（烟草噬纹病毒蛋白）。其中 Split ubiquitin 的 C 端可以被它特有的剪切酶剪切，而 TEV protease 上也有一个特殊的可被切割的位点。脱落蛋白的 C 端和 N 端可以分开也可以合在一起，当 C 端和 N 端结合以异二聚体的形式存在时，这种信号会被侦察到而引发免疫反应。

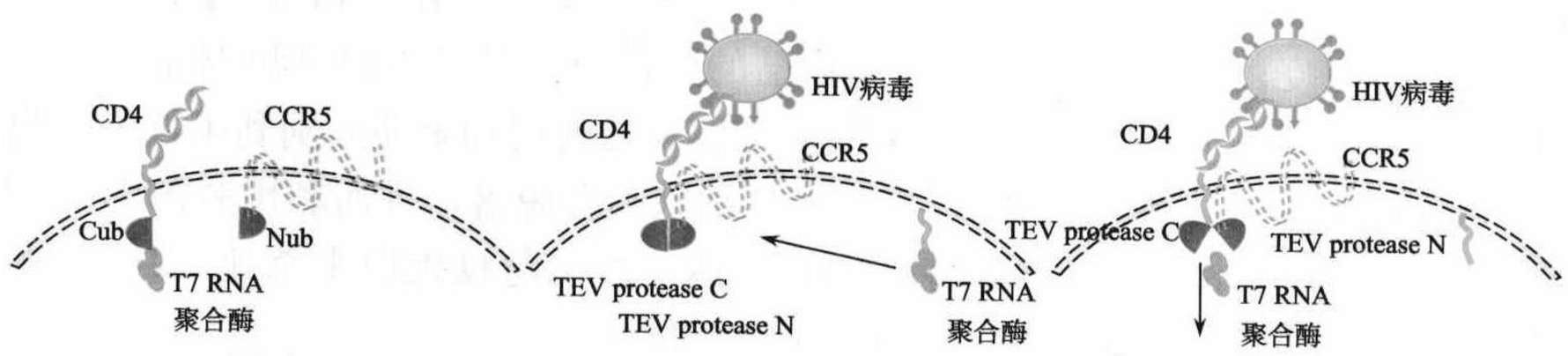
Split ubiquitin 的具体作用机理如下：初始状态时 Split ubiquitin 的 C 端和感受器 CD4 及 T7 RNA 聚合酶结合在一起，而它的 N 端和感受器 CCR5 结合在一起。当受到 HIV 病毒进攻时，Split ubiquitin 的 C 端和 N 端结合以异二聚体的形式存在，这种信号会引发 Split ubiquitin 特异性剪切酶 ubiquitin-specific protease 识别其 C 端并将连接在 C 端的 T7 RNA 聚合酶剪切下来，如附图 5 所示。



附图 5　Split ubiquitin 的存在导致 T7 RNA 聚合酶脱落

初始状态时感受器 CD4 经 Split ubiquitin 的 C 端和 T7 RNA 聚合酶结合，而感受器 CCR5 和 N 端结合。当受到 HIV 病毒进攻时，Split ubiquitin 的 C 端和 N 端结合成异二聚体，引发 Split ubiquitin 特异性剪切酶 ubiquitin-specific protease 对异二聚体 C 端进行切割，T7 RNA 聚合酶脱落。

而 TEV protease 的作用机理如下：初始状态时 TEV protease 的 C 端和感受器 CD4 结合在一起，它的 N 端和感受器 CCR5 结合在一起。当病毒进攻时会使 C 端和 N 端结合以异二聚体的形式存在。此部分机理与 Split ubiquitin 的作用机理类似，不同的是，T7 RNA 聚合酶初始时是锚定在细胞膜上的，当形成异二聚体时会使 T7 RNA 聚合酶沿着细胞膜滑动，遇到 TEV protease 上的特殊识别位点时才脱落下来，如附图 6 所示。



附图 6　TEV protease 的存在导致 T7 RNA 聚合酶脱落

初始状态时 T7 RNA 聚合酶锚定在细胞膜上，TEV protease 的 C 端和感受器 CD4 结合在一起，它的 N 端和感受器 CCR5 结合在一起。当 HIV 进攻时，C 端和 N 端结合成异二聚体，T7 RNA 聚合酶沿着细胞膜滑动，遇到 TEV protease 上的特殊识别位点时脱落下来。

（2）在感染末期的抵御机制：此时翻译出来的 HIV 蛋白识别病毒信号，使 T7 RNA 聚合酶激活并从膜上脱落。

初期和末期防御的最终目的都是使 T7 RNA 聚合酶脱落。那么 T7 RNA 聚合酶的脱落又是怎样起到抑制 HIV 病毒的作用呢？

2）脱落的 T7 RNA 聚合酶识别启动子

脱落的 T7 RNA 聚合酶可以和 T7 启动子结合，启动其下游基因的表达，转录翻译出 β 干扰素和 caspase-3 等免疫蛋白。

一旦 HIV 病毒进入此类细胞，干扰素就会发挥作用，并且使邻近细胞也产生 β 干扰素从而抑制病毒的入侵；而 caspase-3 是通过感染使 HIV 病毒细胞自身溶解从而达到杀死 HIV 病毒的目的。脱落的 T7 RNA 聚合酶还有一个作用就是使信号放大，只要有少量的 HIV 病毒侵入，信号就会被侦测到并被放大，引发免疫系统反应，从而提高治疗的灵敏度。

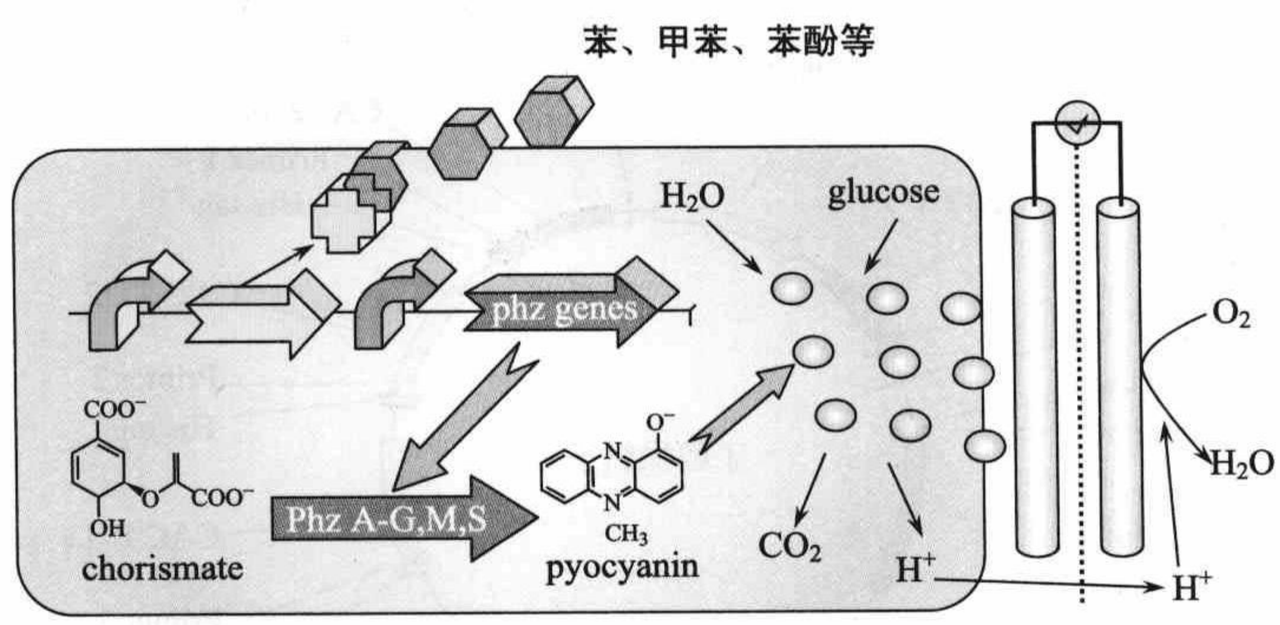
需要提醒读者注意的是，虽然上述利用合成生物学思想来治疗艾滋病的研究为病患防治提供了大胆的思路，但这些研究仍然处于实验室探索阶段，离最终的临床应用尚有相当一段距离。读者在看到其潜在应用前景的同时不要盲目乐观，要抱着科学的态度去看待这些研究成果。

3. 芳香烃类污染物的检测

iGEM2007，金牌，英国格拉斯哥（Glasgow）大学参赛项目

参赛网页：http://parts.mit.edu/igem07/index.php/Glasgow

格拉斯哥大学的研究人员构建出一种大肠杆菌，它能够感知环境中芳香烃类污染物的浓度并通过改造过的代谢途径转化为可检测的电信号，如附图 7 所示。



附图 7　芳香烃类污染物的微生物检测器

装置由三部分组成：一是感受环境中苯、甲苯等芳香烃类化学分子存在的 Biosenser，Biosenser 中表达的蛋白质能与这些分子结合，启动 *phz* 基因的表达；二是 *phz* 基因表达的酶将分支酸转化为 pyocyanin，pyocyanin 作为电子载体将大肠杆菌代谢中产生的电子迅速转移；三是生物电池，转移到正级上的电子向负极转移，氢离子透过两电极之间的选择性透过膜，与负极上的电子结合，产生电信号。

该装置由三部分组成：①Biosenser，表达蛋白能与环境中苯、甲苯等芳香烃类化学分子结合，启动 *phz* 基因的表达；②*phz* 基因表达的酶能将分支酸转化为 pyocyanin，pyocyanin 是一种电子载体，能将大肠杆菌代谢中产生的电子迅速转移；③生物电池，转移到正极上的电子向负极转移，两电极之间有选择性透过膜，氢离子可以透过膜与负极上的电子结合，产生电信号。

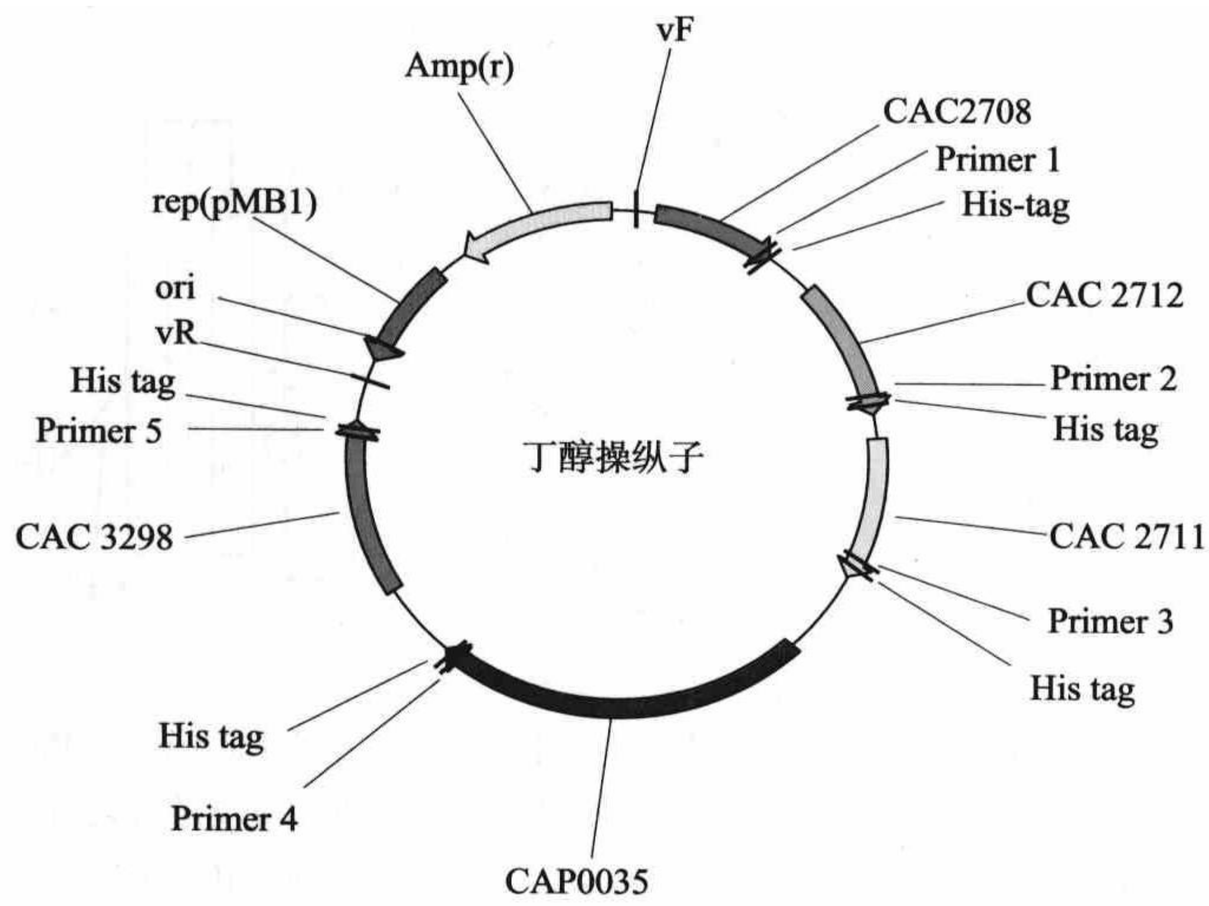
4. 丁醇的生物合成

iGEM2007，金牌，加拿大阿尔伯塔（Alberta）大学参赛项目

参赛网页：http://parts.mit.edu/igem07/index.php/Alberta

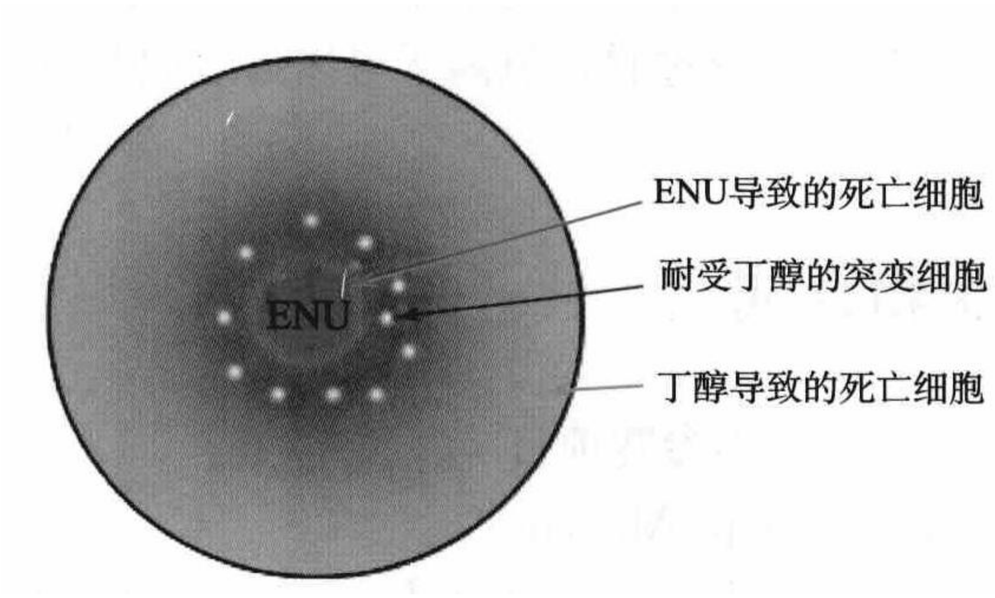
生物方法生产丁醇常用的丙酮丁醇梭杆菌（*Clostridium acetobutylicum*）耐丁醇能力不强，发酵产物成分复杂，分离工序繁多，是目前丁醇浓度和产量无法进一步提高的主要瓶颈。因此，2007 年 iGEM 参赛队之一的阿尔伯塔大学利用合成生物学理念，研究丙酮丁醇梭杆菌中的丁醇合成基因路径，并将其转入大肠杆菌中，利用大肠杆菌生物质能作为能量来源，以简化丁醇的提纯工序，并希望在此基础上将含有此功能基因的质粒转入光合自养细菌绿硫菌（*Chlorobium tepidum*）中，得到的工程菌可以利用阳光和二氧化碳直接合成丁醇，不仅大大降低了生产成本，而且有利于减少温室气体的排放。

附图 8 为阿尔伯塔大学利用合成生物学技术构建的大肠杆菌合成丁醇的操纵子。该操纵子包含除了大肠杆菌硫解酶基因之外的合成丁醇的全部必要基因，通过添加可诱导的启动子和核糖体结合位点便可对其进行必要的调控。为方便质粒的扩展利用，研究人员对操纵子进行了标准化改造，制作成标准化的功能质粒。图中每个基因羧基边界的 Histamine 标签可以帮助研究者利用 Ni-NTA 柱对标记蛋白进行纯化以分析蛋白质的表达和酶的活性。



附图 8　转入大肠杆菌中的丁醇合成操纵子

将可诱导的启动子、核糖体结合位点以及丙酮丁醇梭杆菌中合成丁醇的全部必要基因做成操纵子，为方便质粒的扩展对其进行标准化改造。图中基因羧基边界的 Histamine 标签方便研究者利用 Ni-NTA 柱对标记蛋白进行纯化以分析蛋白表达和酶活性。



附图 9　大肠杆菌细胞对于 ENU 和丁醇的耐受性检验图

图中所示为丁醇琼脂盘培养基中大肠杆菌的生长状况，在圆盘的中心区域加入一定浓度的 ENU，由于扩散作用，ENU 的浓度由中心向四周梯度下降。圆盘的中心区域是由于 ENU 的毒性过高而致死的细菌，圆盘最外围为丁醇毒性致死的细菌，两者之间则是经过 ENU 诱变而具有较高丁醇抗性的大肠杆菌细胞。

为了提高此种大肠杆菌对丁醇的耐受性，研究者还利用乙基亚硝基脲（ethylnitrosourea，ENU）对细胞进行突变诱导。附图 9 所示为丁醇琼脂盘中大肠杆菌的生长状况，其中心区域含有一定浓度的 ENU。图中中心区域圆环是由于 ENU 的毒性致死的细菌，其外围为丁醇毒性致死的细菌，两者之间则是经过 ENU 诱变而具有较高丁醇抗性的大肠杆菌细胞。

在此基础上，小组成员正在研究如何将丁醇合成操纵子质粒转入光合细菌 *Chlorobium tepidum* 中。*C. tepidum* 是一种严格厌氧绿色硫酸细菌，利用硫化物作为光合作用的电子供应者。*C. tepidum* 喜欢生长在 40℃、pH4.5～6.0、低亮度的环境中。这种细菌能够通过反三羧酸循环（reverse tricarboxylic acid）利用二氧化碳生成碳水化合物。因此转入丁醇合成操纵子质粒而得到的工程 *C. tepidum* 菌可以利用阳光和二氧化碳直接合成丁醇，这样不仅大大降低了生产成本，而且有利于减少温室气体的排放。

5. 基因反馈调节治疗脓毒症

iGEM2006，唯一大奖，斯洛文尼亚的 Ljubljana 大学参赛项目。

参赛网页：http://parts.mit.edu/wiki/index.php/Ljubljana%2C_Slovenia_2006

脓毒症（Sepsis）是败血症的一种，指宿主对微生物感染的全身性炎症反应，常指由病原微生物引发的败血症。Sepsis 的病原多数为革兰氏阳性或阴性菌，但病毒、立克次体、真菌，以及微生物分子信号或毒素的全身播散也可引发 Sepsis。

研究发现，Sepsis 的产生实际上是人体的免疫系统“过度反应”或者说“物极必反”的结果。因此，小组成员设计了一个防治 Sepsis 的基因反馈调节系统，巧妙地借助细胞自身作用的发挥来维持人体细胞环境的平衡，达到治疗疾病的目的。

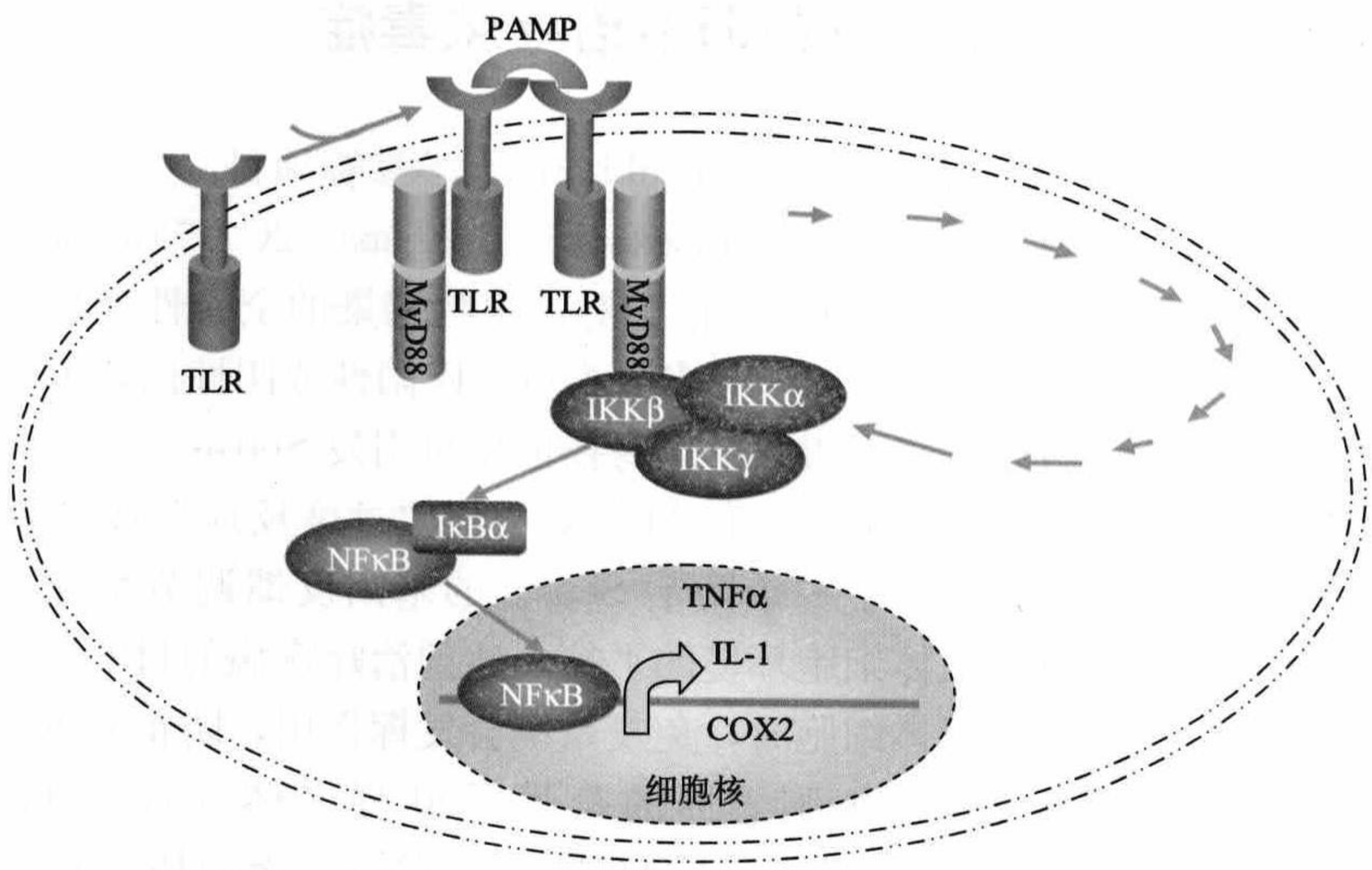
当外来有毒物质试图入侵机体细胞时，免疫系统会发挥作用，防止有毒物质的进入或者特异性地识别并杀死它们。细胞表面的跨膜 Toll 样受体（Toll-like receptor，TLR）起着警卫的作用，它们能够识别病源物质（如 RNA、蛋白质、脂多糖、脂蛋白、细菌结构蛋白等）的 PAMP（pathogen associated molecular pattern）结构捕获病源，保证免疫系统能够及时发现并采取相应防御措施，防止有毒物质对机体的进一步侵害。由此可见，只有 Toll 样受体特异性地识别并结合 PAMP 时，才能引发机体的免疫过程。当这种特异性的结合不能正常发挥作用时，Toll 样受体无法识别或者捕获 PAMP；或者报警信号的传递发生了障碍，免疫系统无法及时发挥正常的抵御作用，生物机体就会被侵袭而引发感染。相反，如果 Toll 样受体即我们的“警卫”小题大做、防卫过当，机体自身的免疫系统就会因为投入过多不必要的抵御措施而虚耗能量，当这种防卫过当情况严重时就会引发危害人类生命安全的 Sepsis。

下面我们就来分析一下 Sepsis 的产生过程及发病机理。

如附图 10 所示，TLR 与有毒物质的 PAMP 特异性的结合能够引发人体细胞的信号传递，通过一系列的电子传递和磷酸化过程，最终将信号传递至 IκBα-NFκB 处，导致 IκBα-NFκB 二者的脱离。游离的 NFκB 进入细胞核，寻找自己的靶位点，一旦找到了作用位点后，便会导致引发炎症相关物质的表达，当这种表达超过一定阈值时就会导致 Sepsis。也就是说，正常情况下 NFκB 是固定在 IκBα 分子上的，游离的 NFκB 分子是引发 Sepsis 的关键信号。只要能够阻止 NFκB 与 IκBα 的分离产生游离的 NFκB 分子，就可以阻止 NFκB 进入细胞核引发免疫系统的过激反应，阻止 Sepsis 的发生。

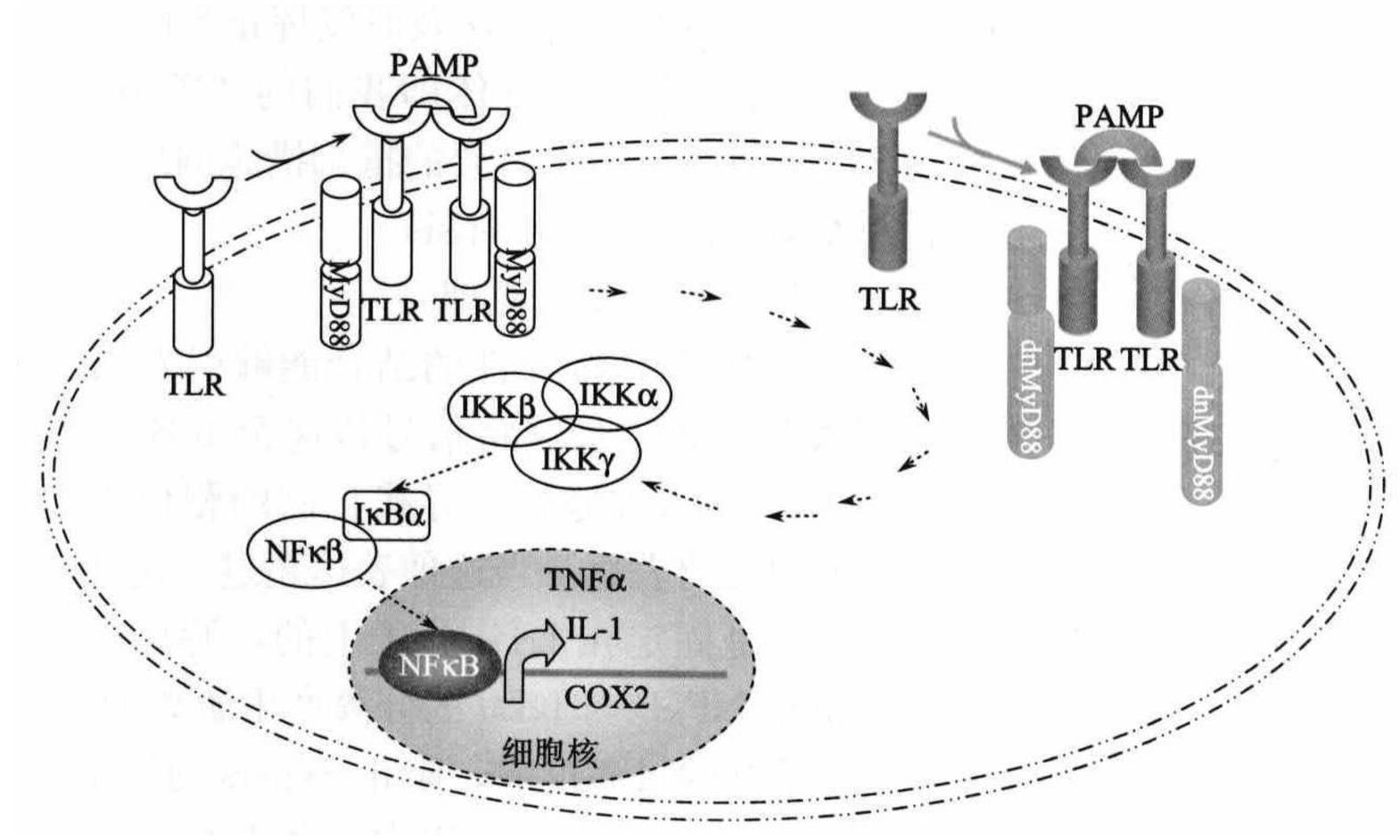
Ljubljana 大学的 iGEM 小组成员仔细分析了免疫过程中的各个相关物质分子，发现几乎所有的信号传递过程都要经过一种叫做 MyD88 的媒介物质，也就是说这种物质在免疫系统信号传递中起到了关键性的介导作用。为此，他们引入一种 dnMyD88 的反馈效应抑制 MyD88 的作用，通过切断 Sepsis 的信号链，阻止 NFκB 的游离，从而阻止 NFκB 进入细胞核引发 Sepsis。

附图 11 为小组成员设计的反馈过程模式图。研究表明，MyD88 有两种功能基团，其一作用于 TLR，即信号传递过程的上游；另一种作用于信号传递的下游。dnMyD88 是切掉第二部分功能基团的 MyD88，因此无法与下游的反应链结合。当在细胞内引入



附图 10　Sepsis 发病机理示意图

TLR 与有毒物质的 PAMP 特异性的结合能够引发一系列的电子传递和磷酸化过程，最终将信号传递至 IκBα-NFκB 处，导致 IκBα-NFκB 二者的脱离。游离的 NFκB 进入细胞核后寻找自己的靶位点并与之结合引发炎症相关物质的表达，当这种表达超过一定阈值就会导致 Sepsis。



附图 11　Sepsis 阻断系统的模式图

MyD88 中具有分别作用于信号传递过程上游和下游的两种功能基团，dnMyD88 正是切掉下游部分功能基团的 MyD88，因此无法与下游的反应链结合。引入细胞内的 dnMyD88 与 MyD88 竞争性地结合 TLR，与之结合的 TLR 无法发生信号传递，抑制了游离的 NFκB 产生，达到抑制 Sepsis 的目的。

dnMyD88 后，它与 MyD88 竞争性地结合 TLR，与之结合的 TLR 不发生信号传递，从而抑制（或减弱）了自由的 NFκB 产生，达到抑制或减弱 Sepsis 的目的。

但是如果上述免疫反应链完全被抑制，机体就会丧失防御功能而无法抵御有毒物质的侵入。因此必须在抑制免疫过当的基础上不损害正当的免疫过程，避免因免疫系统信号不能传导而引发免疫不足，也就是说既要防止 Sepsis 也要防止 dnMyD88 引起的信号中断而导致的机体感染。因此，除了能够产生 dnMyD88 这种物质的基因外，研究人员还引入了 *PEST* 基因。这段基因所表达的蛋白质可以快速降解 MyD88 和 dnMyD88 两种竞争关系的蛋白质，使生物机体恢复到未被有毒物质侵害的状态，当有毒物质再次入侵的时候，即可做出正确的反应。

要提醒读者的是，此研究仍然处于实验室探索阶段，离最终的临床应用尚有相当一段距离。读者在看到其潜在应用前景的同时不要盲目乐观，要抱着科学的态度去看待这些研究成果。

6. 大肠杆菌香味剂

iGEM2006，美国麻省理工学院参赛项目

参赛网页：http://openwetware.org/wiki/IGEM：MIT/2006

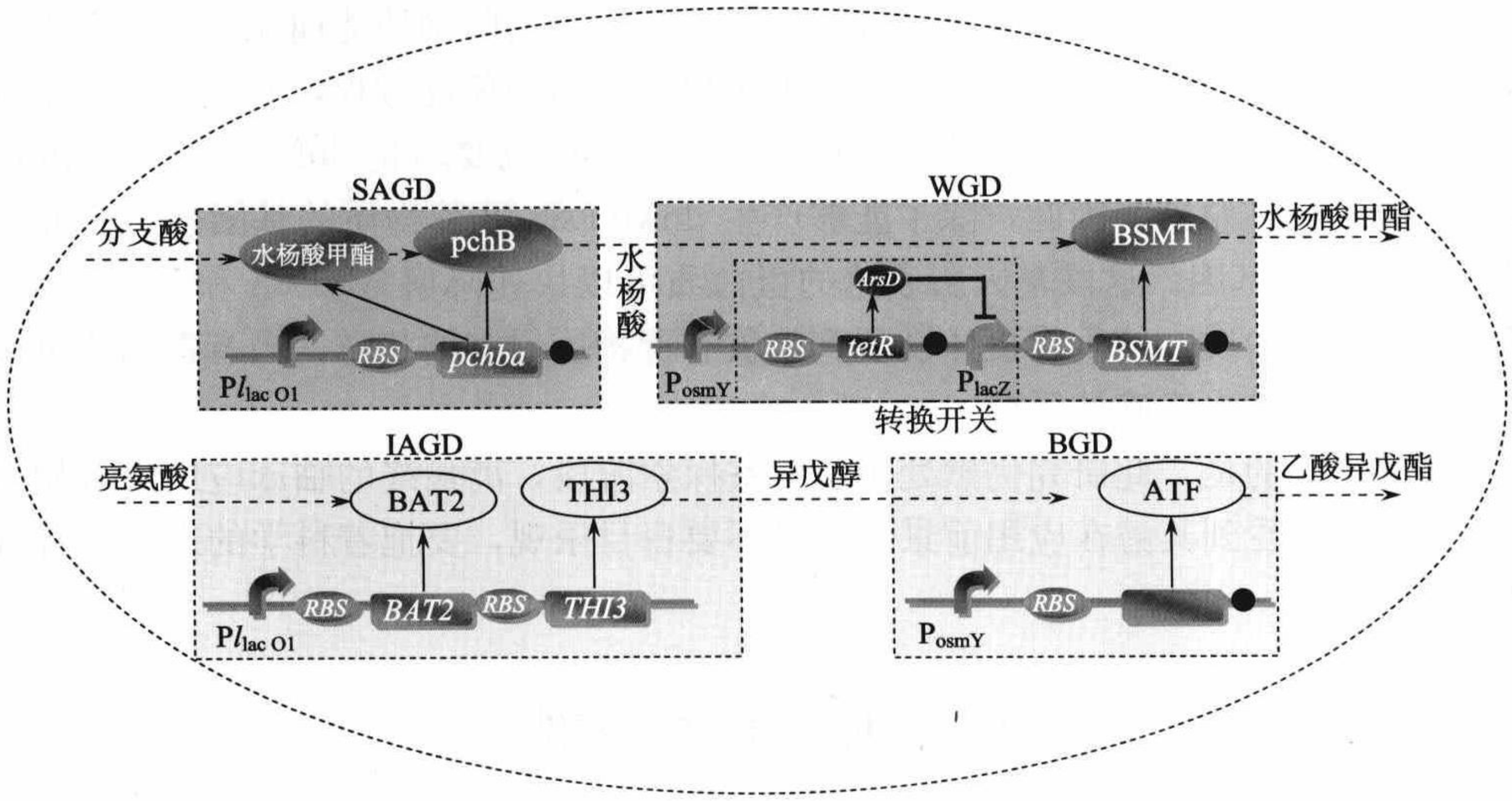
2006 年 iGEM 大赛中由麻省理工学院参赛队设计的大肠杆菌香味剂是利用合成生物学进行环境改造的一个很好的例子。下面对其进行简单介绍。

大肠杆菌（*E. coli*）是一种在人和温血动物肠道内常见的细菌，也是实验室中常用的工程菌。在肠道中，除了肠出血性大肠杆菌（EHEC）等极少数菌种外，大多数大肠杆菌菌株并无害处，但它们在代谢过程中会产生吲哚而发出的臭味却是环境污染的主要因素之一。针对这一特点，队员们希望能使大肠杆菌在对数期产生水杨酸甲酯（methyl salicylate，一种能发出冬青气味的物质），而在稳定期产生乙酸异戊酯（isoamyl acetate，一种能产生香蕉气味的物质），从而对大肠杆菌产生的环境污染进行治理，改善研究人员的工作环境。因此，队员们构建了如附图 12 所示的生物装置以达到此目的。

整个系统共有四个装置组成，分别是产生水杨酸的装置 SAGD（salicylic acid generating device）、产生冬青气味的装置 WGD（wintergreen generating device）、产生异戊醇的装置 IAGD（isoamyl alcohol generating device）及产生香蕉气味的装置 BSGD（banana smell generating device）。

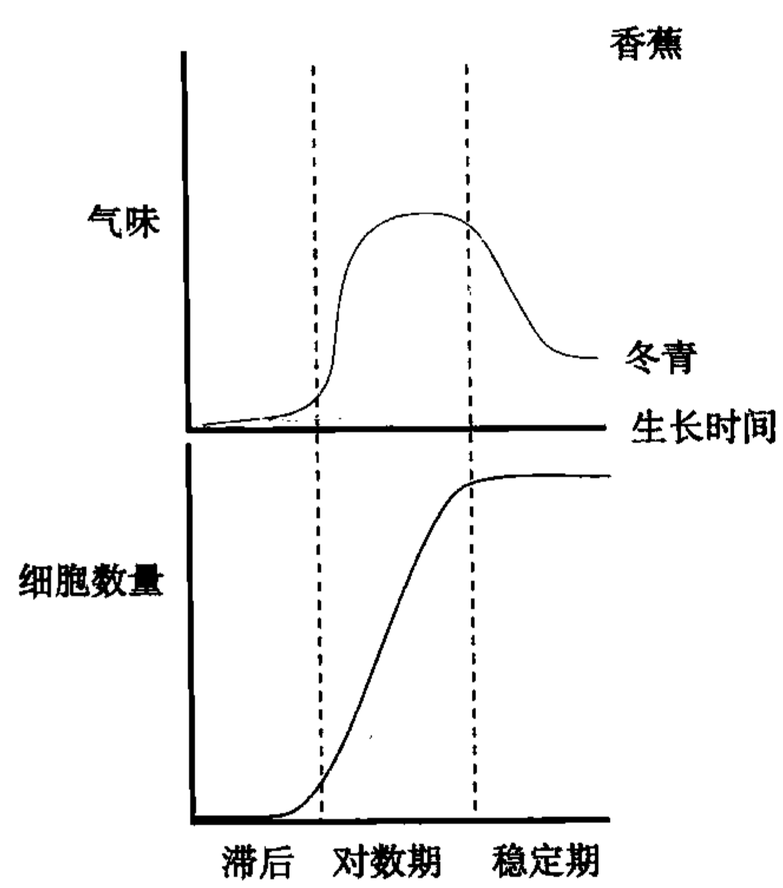
P_{osmY}是一种在稳定期活性高而在对数期活性低的启动子，转换器是一种改变 P_{osmY} 启动子作用结果的基因序列，由一个 *tetR* 基因、一个双终止子（由两个终止子连接而成，具有双重终止的功效，对 DNA 转录的终止作用更彻底）和一个 P_{lacZ}启动子组成，其中 *tetR* 基因经转录和翻译产生的 TetR 蛋白对 P_{lacZ}启动子具有抑制作用。当 P_{osmY}活性高时，TetR 蛋白的表达量高，对 P_{lacZ}启动子的抑制作用强，P_{lacZ}无法启动，此时转换器的输入为高，输出为低；当 P_{osmY}活性低时，TetR 蛋白的表达量低，对 P_{lacZ}启动子的抑制作用弱，P_{lacZ}正常启动，此时转换器的输入为低，输出却为高。

当细胞处于对数期的时候，P_{osmY}启动子活性很低，转换器中 TetR 蛋白产生的量很少，对 P_{lacZ}启动子的抑制作用很小，其后的 *BSMT* 基因能够顺利表达产生 BSMT 蛋白，



附图 12　大肠杆菌香气改造生物装置

整个系统共有四个装置组成，分别是产生水杨酸的装置 SAGD（salicylic acid generating device）、产生冬青气味的装置 WGD（wintergreen generating device）、产生异戊醇的装置 IAGD（isoamyl alcohol generating device）以及产生香蕉气味的装置 BSGD（banana smell generating device）。转换器 inverter 由 P_{lacZ}启动子、产生抑制 P_{lacZ}的 TetR 蛋白的 *tetR* 基因和双终止子（由两个终止子连接而成，具有双重终止的功效，对 DNA 转录的终止作用更彻底）组成。P_{osmY}是一种在稳定期活性高而在对数期活性低的启动子。当细胞处于对数期时，因 P_{osmY}启动子活性低，TetR 蛋白的产量很少而无法抑制 P_{lacZ}，其后的 *BSMT* 基因能够顺利表达 BSMT 蛋白，WGD 发挥功能将由分支酸经 SAGD 装置产生的水杨酸转换成具有冬青气味的水杨酸甲酯。而此时，同样是由于 P_{osmY}启动子活性很低，BGD 装置无法产生香蕉气味的乙酸异戊酯，由此大肠杆菌在对数期具有产生冬青气味的性状。而当细胞处于稳定期的时候，P_{osmY}启动子活性变高，TetR 蛋白的量也随之升高，抑制了 P_{lacZ}启动子的活性，从而导致水杨酸甲酯不能产生，而乙酸异戊酯可以顺利产生，由此进入稳定期的大肠杆菌产生香蕉气味。



附图 13　细胞数量及气味性状随生长期的变化图

由此 WGD 发挥功能，将由分支酸经 SAGD 装置产生的水杨酸转换成具有冬青气味的水杨酸甲酯。而此时，同样是由于 P_{osmY}启动子活性很低，BGD 装置无法发挥作用，也就无香蕉气味的乙酸异戊酯产生，由此大肠杆菌在对数期具有产生冬青气味的性状。当细胞处于稳定期的时候，P_{osmY}启动子活性变高，TetR 蛋白的量也随之升高，抑制了 P_{lacZ}启动子的活性，从而导致水杨酸甲酯不能产生，而乙酸异戊酯可以顺利产生，由此进入稳定期的大肠杆菌产生香蕉气味，如附图 13 所示。

名词索引

2-丁酸苯乙酯 112
伴侣蛋白 7
必需基因 7
标准化 11
拨动开关 77
茶碱 117
肠吸收细胞 115
超敏感性 36,37,68,99
程序性细胞死亡 92
蛋白募集位点 102
凋亡 92～95
调控蛋白 102～105
多效性 7,106,109
二甲(基)丙烯焦磷酸酯 118,119,120
发夹结构 35,76,77
法尼基焦磷酸酯 118～120
反向重复序列 35
非一致前馈 39,44,45,46,88
封闭性 26,27
感光蛋白 129,130
光敏色素 129,130
国际遗传机器大赛 16
合成生物学(描述性定义) 4
灰色黏质 18
基础基因线路 8,39,153
基因线路 8
基元 8,39,40,44,45
级联线路 9,28,36,37,39,46,68,99,100
甲氟喹 117
甲羟戊酸类异戊二烯途径 8,119
焦磷酸异戊酯 8,118,119
角鲨烯 120
解耦 11,12,28
金鸡纳霜 117
茎环 35
开放性 26
奎宁 117
酪氨酸活化模体 111
亮氨酸拉链 102,103,104
鲁棒性 37,59,64～68,78
氯喹 117
密集交盖调节网 41
模块 26
内聚 26,27
耦合 12,26,27
青蒿素 3,8,117,121
青蒿酸 3,8,117～121
群体感应 9,37,38,92～95,116
上升刺激 45
烧焦烧饼问题 131～134
生色团 129
生物传感器 127
生物积块 21,27～34
生物装置 11,28,34,36
适配体 75,77,117,154
树扩张型朴素贝叶斯方法 64
双稳态开关 9,36,44,77～83
双相开关 74,75
双组分系统 129,130
同线基因组 132
脱辅基蛋白 129
脱氧-5-磷酸木酮糖 119
无标度网络 46,68,153
下降刺激 45
信号敏感性延迟 45
血液丙种球蛋白血症 112

一致前馈 39,44～46
遗传线路 8
乙硫异烟胺 112
元反应 50,51
藻胆青素 129
重构 105～109,111
重组增强子 133
转换器 33,36,170
转录反激活因子 112
转录后 8,36,74,76,77
紫穗槐-4,11-二烯合成酶 8,118～120
自体诱导物 38
AHL 38,88～97
B细胞抗原受体 111
DNA从头合成 6,13
DNA重组技术 3
RNA聚合酶每秒 33
TA克隆 30,31

人名索引

A. Becskei 43
A. L. Barabási 153
A. V. Hill 52
B. Hobom 3
C. A. Janeway Jr 157
C. A. Voigt 128,129
C. D. Smolke 前言,3,77
C. J. Bashor 101,102,104
C. M. Guldberg 50
D. Aubel 112
D. Dussoix 3
D. Endy 前言,3,33,106,108
D. T. Gillespie 57
E. Drexler 18
E. L. Haseltine 57
E. T. Kool 3,6
E. Wimmer 7,18,19
F. Crick 1
F. Jacob 3,8
F. J. Isaacs 75,76
F. Kekulé 10
G. Bertani 3
G. Church 4
G. Stephanopoulos 113
H. G. Khorana 7
H. Salis 57
H. Smith 3
J. C. Anderson 114,116
J. C. Liao 123,126
J. C. Venter 7,18,20,21
J. D. Keasling 3,4,118～121
J. D. Watson 1
J. J. Collins 3,75,78,113
J. J. Weigle 3
J. Monod 3,8
J. Shine 152
J. W. Chin 74
K. A. Haynes 前言,131,133
L. Dalgarno 152
L. You 92
M. A. Savageau 43
M. Berthelot 10
M. B. Elowitz 83
M. Griffith 57
M. L. Human 2,151
M. M. Becker 112
N. Friedman 63,64
O. Rackham 74
P. Charles 18
P. F. Verhulst 53
P. Waage 50
R. Holt 4
R. Medzhitov 157
R. Rettberg 前言
R. Weiss 3,38,72,93
S. Atsumi 123,124
S. A. Benner 6
S. Basu 88,95
S. E. Luria 2,151
S. Hooshangi 99
S. Lin 3
S. Topp 116
T. Knight 前言
T. S. Bayer 77
T. S. Gardner 78
V. Gonzalez-Nicolini 112
V. Rolli 111
W. Arber 3
W. A. Lim 100,101
W. Szybalski 3,10
W. Weber 112

[General Information]
书名：合成生物学导论
作者：宋凯编著
页数：173
出版社：科学出版社
出版日期：2010.02
简介：本书分5章。第1章介绍合成生物学的基本概念和内容，使读者对合成生物学有一个全面的了解，第2章和第3章介绍合成生物学设计和分析的基本法、合成生物学的层次化结构和概念及合成生物学中的常用算法，第4章和第5章介绍合成生物学在基础和实际应用方面的研究内容、成果和进展，包括经典的双稳态开关、大肠杆菌成像系统、微生物计算机，合成肯蒿酸的微生物工厂等。
主题词：生物合成 - 概论 - 生物合成
SS号：12453992
DX号：0000006864395
http://book1.duxiu.com/bookDetail.jsp?dxNumber=0000068643 95&d=94BCD640B320B8D09FB1C677B020D3E3&fenlei=1510&sw=%BA%CF%B3%C9%C9%FA%CE%EF%D1%A7%B5%BC%C2%DB

封面
书名
版权
前言
目录
第１章　绪论
第一部分　合成生物学概述
１．１　合成生物学的诞生
１．２　合成生物学的定义
１．３　合成生物学的研究内容
１．３．１　生物大分子的合成与模块化
１．３．２　生物基因组的合成、简化与重构
１．３．３　合成代谢网络
１．３．４　遗传／基因线路的设计与构建
１．３．５　细胞群体系统及多细胞系统研究
１．３．６　数学模拟和功能预测
１．４　合成生物学的意义
１．４．１　加速合成生物系统工程化的进程
１．４．２　验证和深化对于生物现象的理解
１．５　合成生物学的工程本质
１．６　合成生物学与相关生物学科
１．６．１　与遗传工程的关系
１．６．２　与分子生物学和细胞生物学的关系
１．６．３　与系统生物学和功能基因组学的关系
１．６．４　与生物信息学的关系
１．７　合成生物学国际会议及国际遗传机器大赛
１．７．１　合成生物学国际会议（Ｓｙｎｔｈｅｔｉｃ　Ｂｉｏｌｏｇｙ　ｘ．０）
１．７．２　国际遗传机器大赛
第二部分　合成生物学引发的思考
１．８　人类造物主
１．８．１　道德规范
１．８．２　对环境无控制地排放
１．８．３　生物武器的滥用
１．８．４　对生物的专利与垄断
１．８．５　创造人工生命？
１．８．６　究竟该如何应对？
第２章　合成生物系统的设计
第一部分　合成生物系统的模块化与层次化结构
２．１　合成生物学的模块化设计
２．１．１　模块设计的三个特征因素
２．１．２　细胞本底环境对模块化设计的影响
２．１．３　无细胞合成生物系统
２．２　标准化生物模块——生物积块
２．２．１　ＢｉｏＢｒｉｃｋ命名规则
２．２．２　ＢｉｏＢｒｉｃｋ的连接
２．３　标准定量机制
２．３．１　ＰｏＰＳ
２．３．２　ＲＩＰＳ
２．４　生物系统的层次化结构
２．４．１　生物部件“Ｐａｒｔ”
２．４．２　生物装置“Ｄｅｖｉｃｅ”
２．４．３　生物系统“Ｓｙｓｔｅｍ”
２．４．４　细胞群体系统及多细胞系统
第二部分　合成生物系统的逻辑拓扑结构
２．５　串联
２．６　单输入
２．７　多输入

２．８　反馈
２．９　前馈
第３章　合成生物系统的数学模拟与性能分析
第一部分　生物系统的数学模拟
３．１　常微分方程
３．１．１　质量作用定律
３．１．２　米氏方程
３．１．３　希尔方程
３．２　Ｌｏｇｉｓｔｉｃ模型
３．２．１　单种群细胞自由生长的Ｌｏｇｉｓｔｉｃ模型
３．２．２　两种细胞自由生长的Ｌｏｇｉｓｔｉｃ模型
３．３　主方程
３．４　随机微分方程
３．５　模式识别
３．５．１　聚类分析
３．５．２　分类分析
３．６　小结
第二部分　合成生物系统性能分析
３．７　常用性能指标
３．７．１　稳定性
３．７．２　鲁棒性
３．７．３　响应快速性
３．８　稳定性分析
３．９　鲁棒性分析／敏感性分析
３．１０　提高系统稳定性和鲁棒性的机制
３．１０．１　冗余
３．１０．２　反馈控制
３．１０．３　特异性和保真度
第４章　合成生物学的基础研究
第一部分　模拟逻辑功能的基础研究
４．１　逻辑门功能遗传线路
４．２　基因调控开关
４．２．１　转换开关
４．２．２　双相开关
４．２．３　Ｒｉｂｏｓｗｉｔｃｈ
４．２．４　双稳态开关
４．３　Ｒｅｐｒｅｓｓｉｌａｔｏｒ
４．４　细胞群体系统脉冲发生器
第二部分　其他方面的合成生物学基础研究
４．５　控制群体数量的基因线路
４．６　人工细胞群体图案系统
４．７　人工转录级联线路
４．８　工程化的支架蛋白
４．９　重构Ｔ７噬菌体基因组
第５章　合成生物学的应用研究
５．１　合成生物学应用于维护人类健康
５．１．１　环境控制的大肠杆菌对癌细胞的侵入系统
５．１．２　合成青蒿素的微生物工厂
５．２　合成生物学应用于生产生物能源
利用工程化的大肠杆菌非发酵代谢途径生物合成高级醇
５．３　合成生物学应用于环境治理
砷离子的检测
５．４　合成生物学应用于其他方面
５．４．１　大肠杆菌成像系统
５．４．２　微生物计算系统
参考文献
主要网站

附录A　　名词解释
附录B　　ｉＧＥＭ参赛项目选介
　　１．幽门螺杆菌疫苗
　　２．艾滋病的治疗
　　３．芳香烃类污染物的检测
　　４．丁醇的生物合成
　　５．基因反馈调节治疗脓毒症
　　６．大肠杆菌香味剂
名词索引
人名索引