

综述与评论

# 基因调控过程中的典型控制环节<sup>1)</sup>

卢欣 李衍达

(清华大学生物信息研究所 北京 100084)

(E-mail: luxin@jerry. au. tsinghua. edu. cn)

**摘要** 生物系统的控制和机械系统的控制一样,都遵循控制论的基本原理.生物学家对于基因调控和细胞信号传导过程已经进行了许多实验研究,急需对实验结果进行理论总结.综述了这些研究成果,归纳出7种典型的控制环节.基因调控和信号传导路径可以由这些控制环节组合而成,可以用控制论进行分析研究,寻找其中的共同规律.

**关键词** 基因调控,信号传导,控制论.

## TYPICAL CONTROL ELEMENTS IN GENE REGULATION PATHWAY

LU Xin LI Yanda

(*Bioinformatics Institute, Tsinghua University, Beijing 100084*)

**Abstract** The control of biology system is like the control of machine system, they all follow the principle of control theory. The genomic regulation and signal transduction have been studied in great detail by biologists, but theoretical integration is far from needed. This paper summarizes these research results, and induces seven typical control elements. The gene regulation network and signal transduction pathway are the assembly of these control elements, and can be analyzed by cybernetics so as to find the common disciplinary.

**Key words** Genomic regulation, signal transduction, cybernetics.

### 1 引言

人类基因组计划是继曼哈顿计划、阿波罗计划之后的又一项宏伟的科学研究计划,预计将在2001年提前完成<sup>[1]</sup>.届时,将获得人类的全部遗传图谱、物理图谱,由23对染色体、 $3 \times 10^9$ 个氨基酸组成的全基因组序列,以及包括细菌、酵母、线虫、小鼠在内的一些模型生

1)国家自然科学基金(69872018)、北京市科技合同及清华大学基础创新研究基金资助项目.

物的全基因组序列. 这些数据像一本厚厚的“天书”摆在我们面前. 但生命并不仅仅是这些静态的数据, 它是一个复杂的对象, 是由蛋白质、核酸以及其他物质组成的生命体, 及其与环境交互作用的复杂过程. 如何真正读懂这本“天书”, 理解正常与异常的生命, 理解生命的发生、进化、生长、疾病、衰老和死亡, 理解生物的新陈代谢、应激与调控、理解生命的本质, 是生命科学面临的更加艰巨和深远的任务. 这将是后基因组时代的核心任务, 也将是二十一世纪自然科学的核心任务之一.

传统的生物学是以观察和实验为主的科学, 但近年来随着以基因自动测序为代表的大规模、自动化实验手段的广泛应用, 实验结果堆积如山, 已不能用传统的办法来分析理解, 急需结合数学、信息科学、计算机科学等理论和技术方法, 来更好地分析理解这些数据, 更深入地理解生命现象. 因此, 80年代末期以来, 生物信息学应运而生, 以基因组信息学为核心, 在分析基因数据、寻找新基因、分析和预测蛋白质结构功能、分子进化、药物设计等方面发挥了实验生物学所不能取代的巨大作用. 但生命现象就其本质来说, 是在基因信息指导下的复杂动态过程<sup>[2]</sup>. 得到基因组数据只是万里长征的第一步, 在基因测序工作接近完成的时候, 对生命基本过程的分析研究, 对基因调控、信号传导的研究, 将是下一步工作的重点.

早在本世纪初, 控制论的创始人维纳就指出, 机械系统与生物系统的控制都遵循相同的规律<sup>[3]</sup>. 维纳创立的控制论就受到了许多生物学研究成果的启发. 著名的量子物理学家薛定谔的著作《what is life》<sup>[4]</sup>被公认为是分子生物学的经典之作, 他在1944年出版的这本书中指出, 遗传物质是一种“非周期晶体”(就是今天认识的 DNA), 而生命也符合基本的物理规律. 今天, 生命科学发展到基因和分子水平, 仍然验证着这些深刻的洞见.

为了在基因和分子水平上研究基因调控和信号传导的过程, 无数生物学家付出了艰辛的劳动. 他们的研究成果又一次证明了, 生命活动在最基本的层次上依然符合控制论的原理, 可以用控制理论来分析解释.

基因是生物信息的携带者, 基因及其产物蛋白质是生命活动的基本成员, 而细胞则是生命的最小单位. 信号传导过程必须在基因和分子水平上进行分析研究, 而对基因调控和基因功能的研究也必须结合具体的细胞环境和作用过程才能真正得到有意义的答案.

本文中, 我们将综述基因调控和信号传导的研究成果, 归纳出七种典型的控制环节. 这几种基本控制环节的组合涵盖了大部分已知的基因调控和信号传导过程. 从这些控制环节可以看出, 在基因和分子水平上, 生物的控制方式也是符合控制论的基本原理的. 这也告诉我们, 对基因调控和信号传导的研究, 不仅需要生物学家的实验观察, 也需要控制论专家的积极参与, 对控制过程进行分析、研究、仿真实验, 寻找其中共同的规律. 生物信息学将是又一个控制论可以大有作为的科学领域.

## 2 基因调控与信号传导中的典型控制环节

### 2.1 串联控制

在信号传导和调控过程中, 从细胞膜受体接受信号, 到基因转录表达水平发生变化, 对外界信号产生响应, 中间经历了一系列的生化反应, 涉及到一系列的分子作用过程. 信息从一个分子传递到另一个分子, 可以看作是一个串联控制的过程. 典型的例子包括 Ras

蛋白作用途径. Ras 途径是包括生长因子受体等众多细胞受体发挥作用,调控基因表达的重要作用途径,是已知的最保守的一族癌基因,对细胞生长、增殖、发育分化及癌变产生重要作用.

Ras 途径可以用图1简单表示:生长因子等配体与酪氨酸蛋白激酶受体(RPTK)结合后,受体膜内部分的酪氨酸激酶活性部位将生长因子结合蛋白(Grb2)磷酸化而激活,进而依次激活鸟苷酸释放因子(GRF)、Ras 蛋白、Raf-1蛋白激酶、MAPK 激酶(MEK)、有丝分裂原蛋白激酶(MAPK)、转录因子(TF),最终开始基因转录<sup>[5,6]</sup>.

实际上,串联控制是所有基因调控通路中的基本组成部分. 基因调控需要针对复杂多变的外界环境,及时准确地作出反应. 信号从细胞外经过细胞膜、细胞质、核膜到达染色体的过程中,经历了不同的环境,不可能要求一种信号分子适应所有的环境,必须多种分子,包括跨膜受体、胞内第二信使、以及各种蛋白激酶各司其职,协同工作. 而且,在串联调控通路中,一个信号分子可以与众多的底物分子发生作用,实现信号的级联放大. 少量、短时间出现的信号就可以产生重大、持久的细胞响应. 最后,基因调控必须要对各种情况作出反映,众多的控制环节也为引入更多的控制机制提供了“接口”.

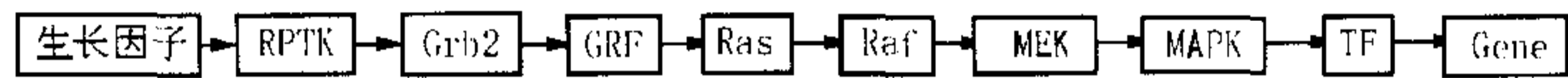


图1 以 Ras 蛋白作用途径为代表的串联控制

### 2.2 分支控制

生命活动是一个复杂的过程,因此调控过程中的信号传递多数也并非简单的“单线联系”. 生命活动中存在大量的正负控制环节,保证系统状态的稳定,所以要改变某种生命活动的状态,也往往要将信号从多个通路传递出去,协同工作以达到目的.

典型的例子包括糖原代谢的过程. 生物体内糖原的合成和分解是一对动态平衡的过程. 当饥饿、寒冷或受到惊吓刺激时,需要增加体内的葡萄糖浓度以提供更多的能量,这需要通过加速糖原分解和抑制糖原合成两个途径来实现.

糖原合成和分解通过环磷酸腺苷(cAMP)来调控,如图2所示,当 cAMP 水平升高时,依赖 cAMP 的蛋白激酶(PKA)被磷酸化激活,并进而磷酸化三个与糖原代谢有关的蛋白:首先是糖原合成酶,被磷酸化而失活,从而抑制糖原合成;其次是糖原磷酸化酶激酶被激活,进而磷酸化并激活糖原磷酸化酶,促进糖原分解;此外,还有蛋白磷酸酶抑制蛋白(I-1)被激活,从而与蛋白磷酸酶(PP-1)结合而抑制其脱磷酸活性,保持其它被 PKA 磷酸化而改变活性状态蛋白在一定时间内发挥功能<sup>[7]</sup>.

### 2.3 负反馈控制

负反馈是控制工程中常用的一种手段,比起开环控制,闭环负反馈可以大大降低控制精度的要求,而且更具有鲁棒性. 自然界总是以最节省的方式达到其目的,因此生物体很自然地选择了闭环反馈控制来实现输出的稳定.

基因调控中负反馈控制最简单的例子是组蛋白的表达调控(图3). 组蛋白是细胞

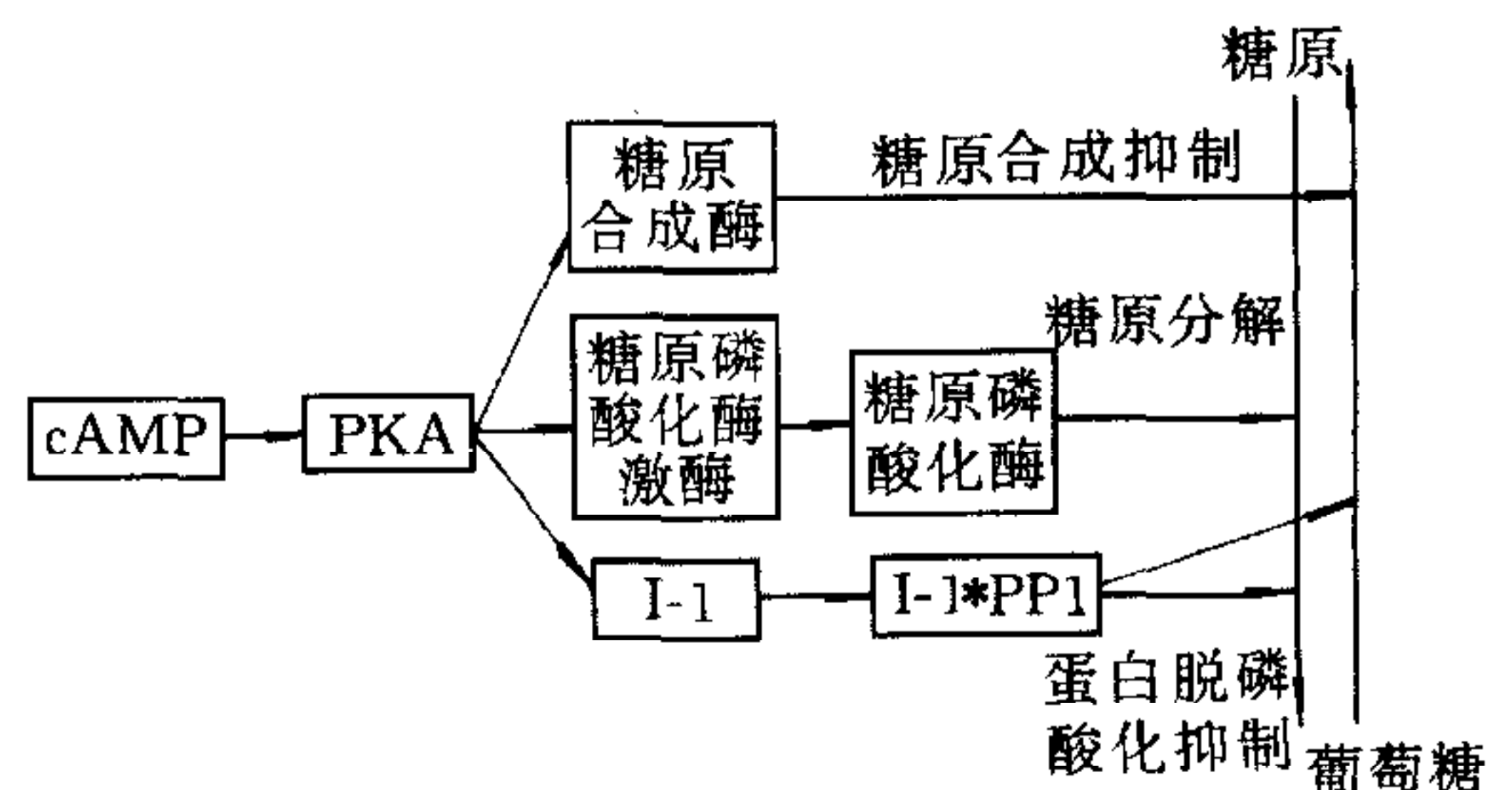


图2 以糖原代谢的控制为代表的分支控制

内广泛存在的一类重要的蛋白质,起到了形成染色体、构建细胞骨架等重要作用.当启动信号作用于组蛋白的启动子时,开始组蛋白的转录表达,但表达的组蛋白同时又对自身的启动子有抑制作用,防止组蛋白表达水平的过分波动,保证其功能的稳定<sup>[8]</sup>.

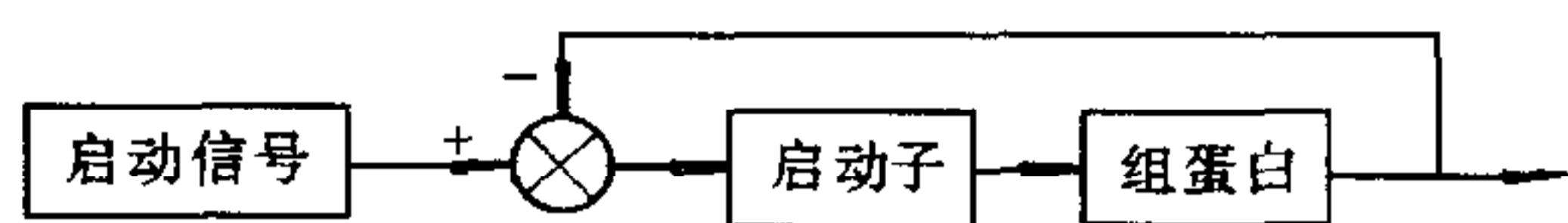


图3 以组蛋白表达控制为代表的负反馈控制

的稳定进行.

## 2.4 简单正反馈

控制的实质是信息传递和作用的过程<sup>[3]</sup>.在生物体内,信号分子的含量是十分稀少的,但在生物体内却起到了巨大的作用.这种“信号放大”作用很多是通过正反馈环节来实现的.简单正反馈是指信号分子作用的靶点在被信号分子激活后,可以再自我激励,从而在信号分子短时间出现之后发挥长时间的作用.这类似于电子技术中常见的触发器,只要一个短暂的脉冲就能改变输出的状态,并在输入脉冲消失之后仍保持输出状态.电子触发器实际上也是通过正反馈来实现的.

几乎所有的蛋白激酶都可以进行自身磷酸化作用,即蛋白激酶在被磷酸化激活之后,可以利用其自身作为底物,磷酸化并激活另一个同样的酶分子或自身的其他部分.以一个典型的蛋白激酶——胰岛素受体为例,在接受胰岛素信号之后,胰岛素受体可以通过自身磷酸化作用而维持活性,其活性不再依赖于胰岛素<sup>[7]</sup>.

## 2.5 复杂正反馈

触发器型的简单正反馈控制可以在受到一个短暂但足够强度的信号激励时,迅速改变并保持系统的状态,这有利于对外界信息作出快速反应.而在其它一些情况下,系统状态的改变需要综合多种因素和条件,需要对是否改变、何时改变系统状态进行精细的控制.为了满足这种需要,生物体发展了复杂正反馈控制方式.在这种方式下,通过正反馈逐渐积累某种物质,并通过复杂的控制环节对累积的速度进行调节,通过累积速度的快慢、停止甚至逆转,就可以精细地控制系统状态改变的时间,或阻止状态改变.这就像“计数器”一样,计数到一定时候引发某种动作,而计数的速度就控制了动作的时间.

典型的例子是细胞分裂周期的控制.处于DNA合成前期(G1期)的细胞,在不同的外界信号和环境条件下,可能进入DNA合成期(S期),并进而进行有丝分裂;可能停止在G1期,进而进入静止期(G0期);也可能分化或进入程序化死亡.控制细胞从G1期进入S期的核心进程可以用图5简要表示.

细胞增殖信号激活合成期促进因子(SBF),从而提高了细胞周期素(CLN)的表达.CLN与一种周期素依赖性蛋白激酶P34结合,这一复合物又可以进一步激活SBF的活性,形成正反馈.当CLN-P34复合物浓度上升到一定水平时,就促使细胞从G1期过渡到S期,进而引发有丝分裂<sup>[8]</sup>.

这里,CLN-P34复合物充当了计数器的作用,其浓度通过正反馈而累积,又受到

组蛋白表达水平的负反馈控制是基因调控中最简单的例子,在生命科学的研究中还发现了许多例子,通过复杂的过程形成负反馈,来精细地控制和调节生命活动的

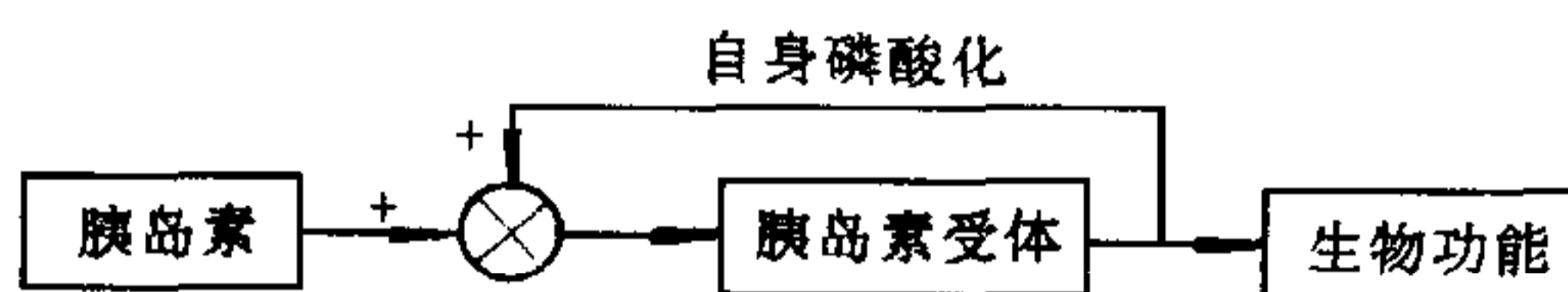


图4 以胰岛素受体自身磷酸化作用为代表的简单正反馈控制

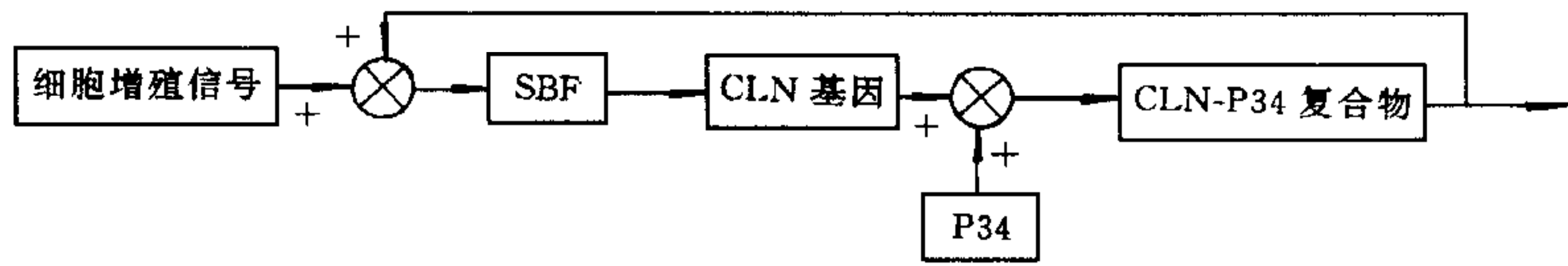


图5 以细胞周期素控制为代表的复杂正反馈控制

SBF、CLN 基因表达、P34等多种因素的控制. 通过引入众多的影响因素,细胞可以很好地适应外界环境和体内信号的要求,进行、不进行或暂缓进行分裂. 如果这些控制环节失灵,细胞就有可能不受限制地分裂,而导致肿瘤的发生.

### 2.6 条件控制

细胞的活动并不仅仅是被动地随外界信号和环境的变化而产生动作,也要受细胞内部条件和状态的控制. 仍以细胞分裂周期的控制为例,在细胞生长过程中,化学物质、辐射、病毒感染等都可能导致 DNA 损伤. 当 DNA 损伤时,即使存在细胞增殖信号和适当的环境条件,细胞也不能马上进入 S 期,必须等待 DNA 损伤修复之后再行 DNA 的复制. 否则,就会把错误的遗传信息带入到下一代细胞中.

这可以看作是一个以 DNA 是否正常为条件的控制环节. 当 DNA 正常时,可以在增殖信号的作用下,通过细胞周期素的作用而进入 S 期;而当 DNA 不正常时,在损伤修复之前,会激活 P53 基因的表达, P53 导致 P21 的表达,而 P21 可以和 CLN-P34 复合物结合抑制其活性,从而阻止细胞进入 S 期.

如果 P53 或 P21 发生突变,就不能阻止损伤的 DNA 上错误的遗传信息扩散到下一代细胞中,其结果往往是肿瘤的发生. 在约 50% 的癌细胞中都能检测到 P53 的突变,因此这个基因最早被称为“癌基因”,后来随着对其功能的深入了解, P53 被正名为“抑癌基因”<sup>[9]</sup>.

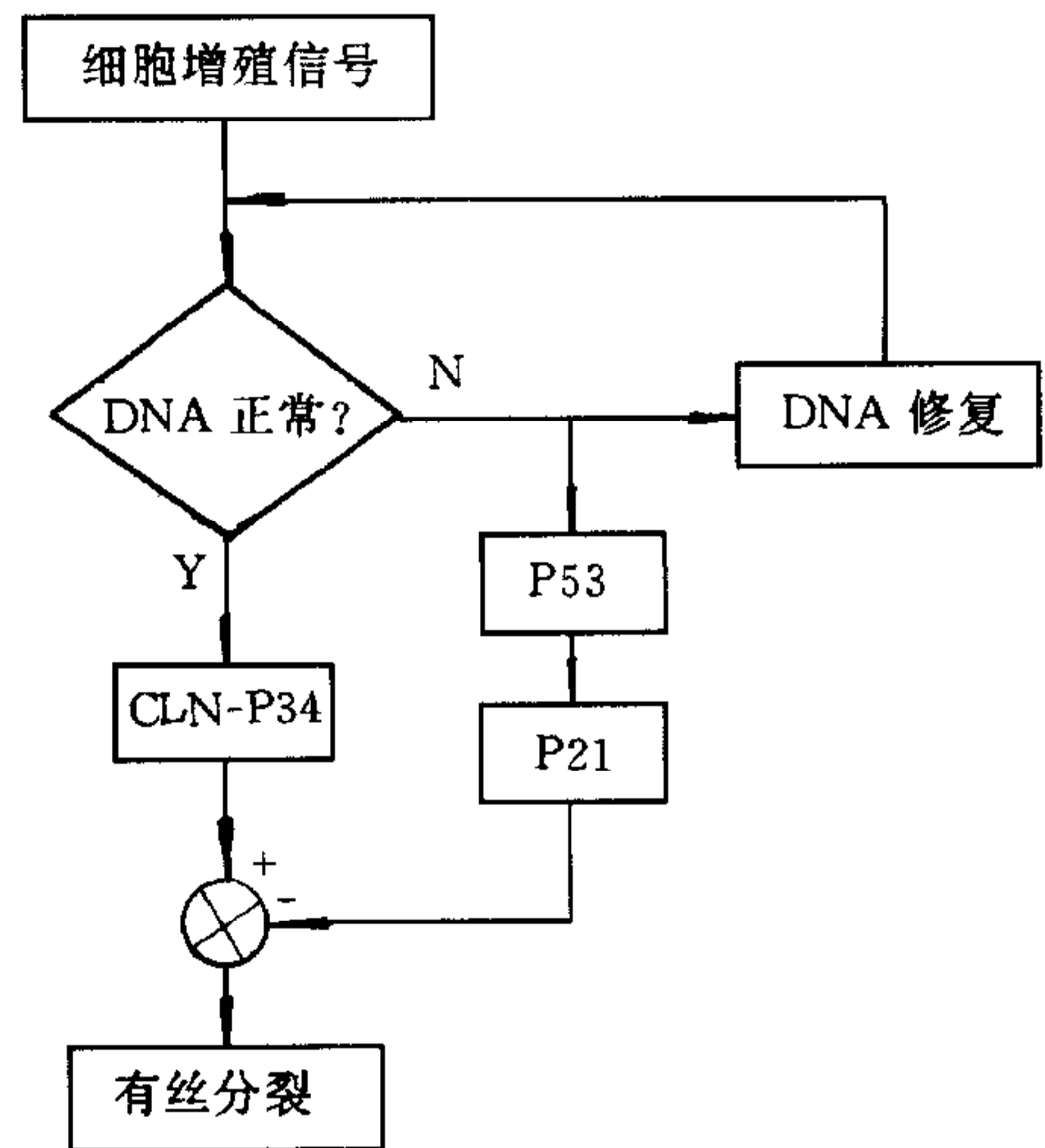


图6 以细胞分裂周期控制为代表的条件控制

### 2.7 延时控制

在正常的生命活动中,及时、准确地对外界信号刺激作出反应固然是十分必要的,而在适

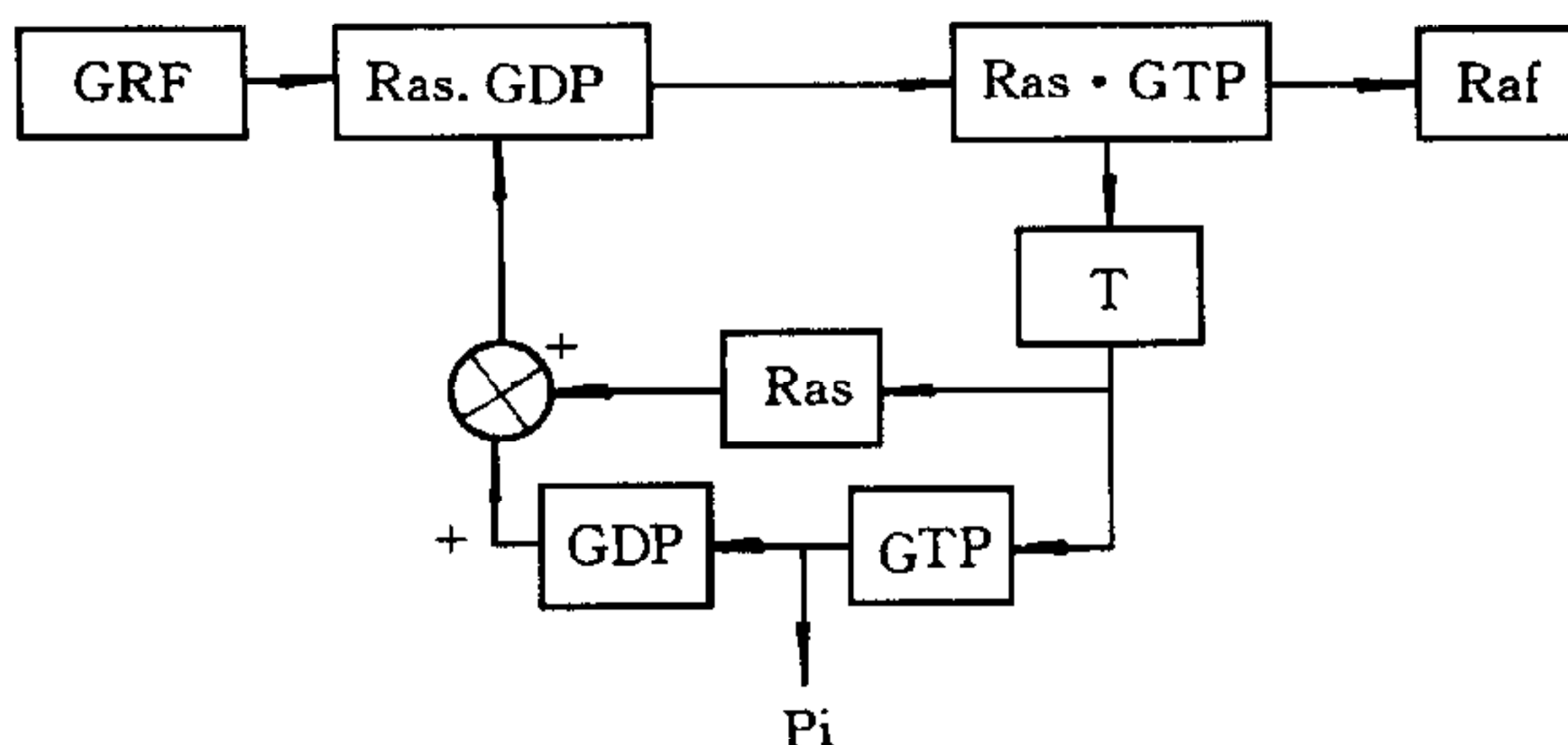


图7 以 Ras 蛋白活性控制为代表的延时控制

当时间停止反应也是同样必要的. 短时出现的信号可以通过正反馈作用而长时间地改变系统的状态,但在适当的响应之后,还需要某种机制恢复系统的初始状态,等待下一次信号的到来. 这实际上也是一种反馈控制,只是其反馈作用不是在激励信号作用时就产生的,而要像“定时器”一样等待一段时间后再作用.

延时控制可以以 Ras 蛋白的活性控制为代表(图7). Ras 蛋白平时与二磷酸鸟苷(GDP)结合,没有活性.当 GRF 作用时,Ras 释放 GDP 而结合 GTP(三磷酸鸟苷),从而产生活性.在发挥作用传递信息的同时,Ras 还有 GTP 酶活性,可以催化 GTP 水解成 GDP.这样,在一定时间之后,Ras 又将自身结合的 GTP 催化水解为 GDP,回到初始状态.Ras 的作用发挥也就限制在了 GTP 水解之前的时间之内<sup>[6]</sup>.

### 3 基因调控网络

基因调控与信号传导是一个问题的两个方面.信号传导所传递的信息最终大多落实到基因表达的改变,所以基因调控是信号传导的结果,而基因调控也是对信号传导作出正确应答的必要条件.信号传导是90年代以来生命科学研究领域的热点和前沿.目前,生物学家已经通过种种实验研究,掌握了大量细胞信号传导、基因调控的知识,但这些知识大多是零散的,是整个生物调控网络中间的一些片段.所以,在今后的工作中,不仅需要继续对细胞信号传导、基因调控过程进行实验研究,积累更多的第一手材料,更重要的是,需要对这些实验材料在分析的基础上进行综合,总结其中的共同规律.

生命活动是复杂多样的,基因调控不仅受到体外、体内环境条件的制约,而且近年来通过对兼具胞内、胞外功能的信号分子的研究,发现信号传递存在着细胞内外的双向通路,以及细胞间复杂的物理、化学联系<sup>[10]</sup>.对每一个基因在每种环境条件下的调控进行实验研究至少在目前是不现实的.但采用复杂系统的分析方法,对系统进行解耦,分解为一些相对独立的子系统来考虑,然后在分析的基础上综合,逐步扩充对基因功能的了解,可能是一种化繁为简的捷径.在生物体内,每一种基因调控活动都涉及到一个或多个典型控制环节,它们由众多的体内外因素、信号分子和基因产物构成,既相对独立,又相互联系,或称为 cross talk,组成复杂的调控网络.网络的稳定运行就对应着正常的生理活动,而网络的异常就会导致疾病,网络的崩溃就是生物的死亡<sup>[11]</sup>.

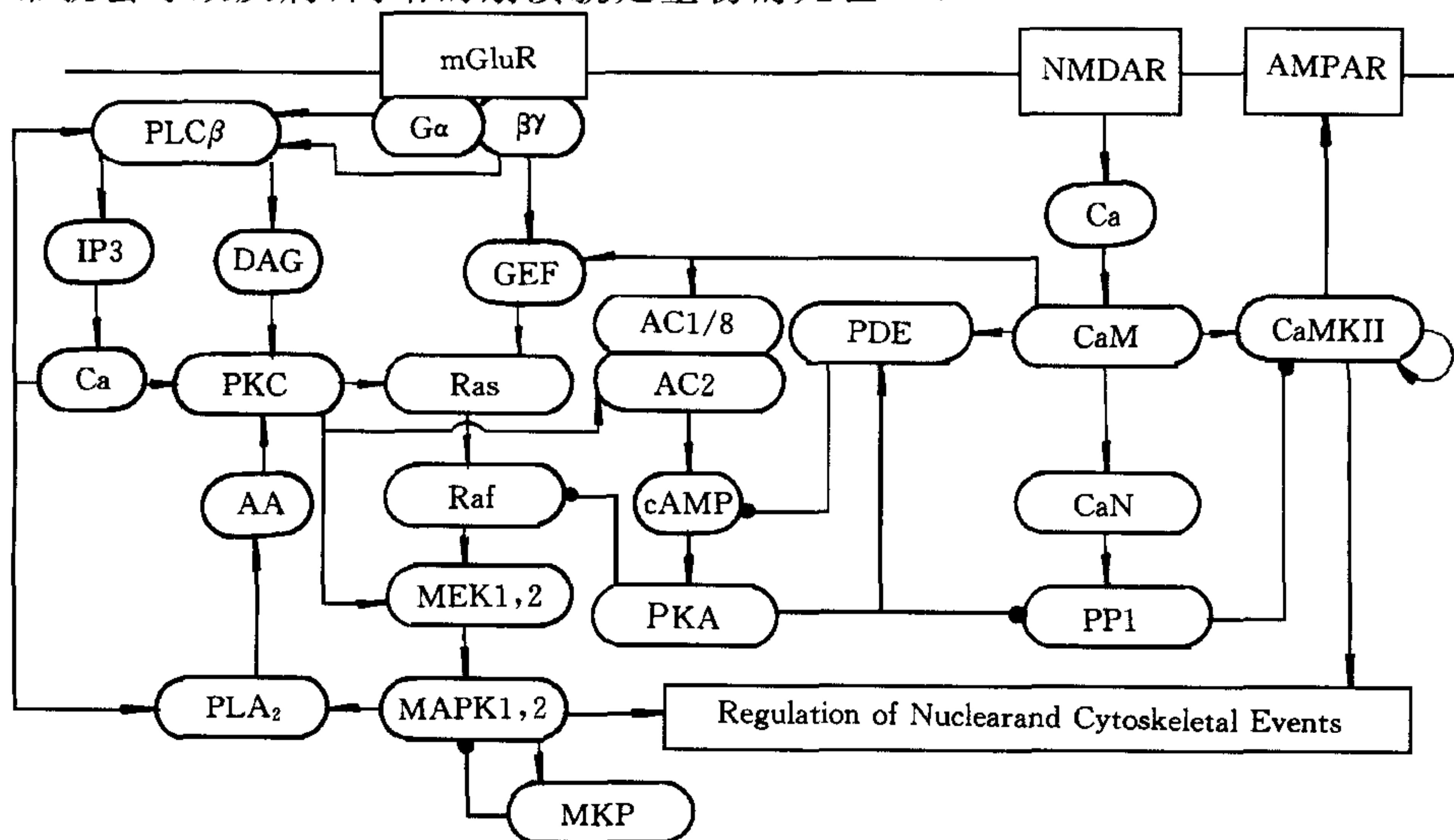


图8 一个基因调控网络的模型<sup>[13]</sup>

目前,国际上已经有一些科学家开始了这方面的工作.比较典型的是 Iyengar 实验室的工作.他们分析了生物信令系统所表现出来的复杂性<sup>[12]</sup>,并且建立了一些基因信令通路的模型,模拟在各种激励条件下信令网络所表现出来的响应特征<sup>[13]</sup>.图8是 Bhalla<sup>[13]</sup>所用的模型之一,在这个模型之中,包含了多个控制环节,互相交织成复杂的调控网络.此外, Tomita<sup>[14]</sup>等还建立了交互式的软件环境来模拟细胞基因调控、代谢及信令传输等过程,并在此基础上研究生命所必须的基因组成.生物学实验手段是无法对整个网络进行跟踪观察的,一次只能在很少的几个节点上观测,用节点的数据来代表整个网络运行的状况.而在实验的基础上,建立基因调控网络的模型,并用数字化媒体来模拟真实生物的生命过程,可以更方便地对基因调控网络的动态特性,对网络在正常、异常等各种激励和环境条件下的反应特性进行深入细致的研究.当然,这种研究的前提是需要实验生物学家提供基因调控通路的知识,和重要的反应参数,而模拟调控网络运行的结果也必须符合生物实验的结果.只有生物学家和信息科学家的通力合作,才能真正了解生命的基本过程,探索生命的本质.

## 4 结束语

分子生物学最重要的成就在于阐明生命本质的一致性<sup>[15]</sup>.基因调控是由众多分子组成的复杂网络,不同生物、不同组织、不同时间和环境条件下这一网络有着各自不同的运行特征,有很大的灵活性和鲁棒性.但生物都是从共同的祖先进化来的,不仅构成生物分子是相同的,而且这些分子发挥作用的方式也是类似的.功能相近的子系统在结构和发挥功能的方式上都有其保守性,而且越是重要的子系统其结构和功能的保守性越高.这些都给分析研究这些生物调控的子系统提供了对照的线索.正如那句法国名言:“变化的东西越多,就有越多的东西保持不变.”随着新的实验技术手段的发展,生命科学也正在突飞猛进地发展,必将有越来越多的生命基本过程、基本规律被认识,被了解.这些都将验证一个真理:生命是复杂的动态系统,生物的控制过程也符合控制理论的基本规律,可以用控制论来分析解释.而对这些生命基本过程的分析解释,正是信息科学、控制科学的一个大有作为的舞台.

因此本文的目的,不仅在于分析总结基因调控过程的重要控制环节,更重要的是,希望将生物信息学这个新兴学科介绍给信息和控制专家,希望引起广泛的兴趣和关注,吸引更多的研究者投身其中,共同探索生命的奥秘.

## 参 考 文 献

- 1 Marshall E. A high-stakes gamble on genome sequencing. *Science*, 1999, **284**(5422): 1906~1909
- 2 李衍达. 信息科学与生物之谜. 世界科技研究与发展, 1999, **21**(3):26~30
- 3 Wiener N. *Cybernetics, or Control and Communication in the Animal and the Machine*. The MIT Press and John Wiley & Sons, Inc. 1961
- 4 Schrodinger E. *What is Life?* London: Cambridge University Press, 1967
- 5 邢东琦. AngII 的胞内信号转导通路. 国外医学, 生理病理与临床分册, 1998, **18**(3): 233
- 6 裴钢. 真核细胞的信息传递与基因表达. 沈翊菲, 方福德主编, 真核基因表达调控. 北京: 高等教育出版社, 施普林格出版社. 1997

- 7 孙大业,郭艳林,马力耕编著. 细胞信号转导. 北京:科学出版社. 1999
- 8 吴宁华. 细胞周期的基因调控. 沈翊珩,方福德主编,真核基因表达调控. 北京:高等教育出版社,施普林格出版社. 1997
- 9 张文岚,杨绍娟,陈岩. G1细胞周期调控因子与癌. 白求恩医科大学学报,1997, **23**(3): 335~338
- 10 孙大业. 兼有胞内、胞外功能的信号分子的普遍性及生物学意义. 科学通报, 1999, **44**(15): 1576~1581
- 11 刘德麟. 中药分子药理学导论. 北京:清华大学出版社,1998
- 12 Weng G, Bhalla U S, Iyengar R. Complexity in biological signaling systems. *Science*, 1999, **284**(5411): 92~96
- 13 Bhalla U S, Iyengar R. Emergent properties of networks of biological signaling pathways. *Science*, 1999, **283**(5400): 381~387
- 14 Tomita M, Hashimoto K, Takahashi K, *et al.* E-CELL: software environment for whole-cell simulation. *Bioinformatics*, 1999, **15**(1), 72~84
- 15 杨福愉. 展望21世纪的分子生物学. 生物物理学报,1999, **15**(1): 1~4

卢 欣 1996年毕业于清华大学,获学士学位.现为清华大学自动化系博士研究生.研究方向为生物信息学.

李衍达 清华大学自动化系教授、信息学院院长、中国科学院院士.长期从事信号处理、计算智能理论与应用及生物信息学研究.