

非线性
动力学丛书

15

生物分子网络的 构建和分析

刘曾荣 王瑞琦 著
杨 凌 赵兴明



科学出版社

非线性动力学丛书 15

生物分子网络的 构建和分析

刘曾荣 王瑞琦 著
杨 凌 赵兴明

科学出版社

北京

内 容 简 介

系统生物学是一门新兴的学科. 它以系统的观点看待生物系统, 把生物系统作为一个多尺度的动态的复杂系统. 因而在研究中除了正常的生物学知识外, 还需要图论(网络)、控制论以及动力系统理论. 本书结合生物问题, 介绍了与系统生物学有关的各种所需知识.

本书可供来自高等院校生物、数学、信息和计算机科学等相关专业并从事系统生物学交叉研究的研究生和青年学者使用, 也可供相关的科研工作者和工程技术人员参考.

图书在版编目(CIP)数据

生物分子网络的构建和分析/刘曾荣等著. —北京: 科学出版社, 2012
(非线性动力学丛书; 15)

ISBN 978-7-03-034448-9

I. ①生… II. ①刘… III. ①分子生物学—研究 IV. ①Q7

中国版本图书馆 CIP 数据核字(2012) 第 107359 号

责任编辑: 钱 俊 / 责任校对: 宋玲玲

责任印制: 钱玉芬 / 封面设计: 陈 敬

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>

骏 杰 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2012 年 6 月 第 一 版 开本: B5(720 × 1000)

2012 年 6 月 第一次印刷 印张: 18 3/4

字数: 360 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

《非线性动力学丛书》序

真实的动力系统几乎都含有各种各样的非线性因素, 诸如机械系统中的间隙、干摩擦, 结构系统中的材料弹塑性、构件大变形, 控制系统中的元器件饱和特性、变结构控制策略等. 实践中, 人们经常试图用线性模型来替代实际的非线性系统, 以求方便地获得其动力学行为的某种逼近. 然而, 被忽略的非线性因素常常会在分析和计算中引起无法接受的误差, 使得线性逼近成为一场徒劳. 特别对于系统的长时间历程动力学问题, 有时即使略去很微弱的非线性因素, 也会在分析和计算中出现本质性的错误.

因此, 人们很早就开始关注非线性系统的动力学问题. 早期研究可追溯到 1673 年 Huygens 对单摆大幅摆动非等时性的观察. 从 19 世纪末起, Poincaré, Lyapunov, Birkhoff, Andronov, Arnold 和 Smale 等数学家和力学家相继对非线性动力系统的理论进行了奠基性研究, Duffing, van der Pol, Lorenz, Ueda 等物理学家和工程师则在实验和数值模拟中获得了许多启示性发现. 他们的杰出贡献相辅相成, 形成了分岔、混沌、分形的理论框架, 使非线性动力学在 20 世纪 70 年代成为一门重要的前沿学科, 并促进了非线性科学的形成和发展.

近 20 年来, 非线性动力学在理论和应用两个方面均取得了很大进展. 这促使越来越多的学者基于非线性动力学观点来思考问题, 采用非线性动力学理论和方法, 对工程科学、生命科学、社会科学等领域中的非线性系统建立数学模型, 预测其长期的动力学行为, 揭示内在的规律性, 提出改善系统品质的控制策略. 一系列成功的实践使人们认识到: 许多过去无法解决的难题源于系统的非线性, 而解决难题的关键在于对问题所呈现的分岔、混沌、分形、孤立子等复杂非线性动力学现象具有正确的认识和理解.

近年来, 非线性动力学理论和方法正从低维向高维乃至无穷维发展. 伴随着计算机代数、数值模拟和图形技术的进步, 非线性动力学所处理的问题规模和难度不断提高. 已逐步接近一些实际系统. 在工程科学界, 以往研究人员对于非线性问题绕道而行的现象正在发生变化. 人们不仅力求深入分析非线性对系统动力学的影响, 使系统和产品的动态设计、加工、运行与控制满足日益提高的运行速度和精度需求, 而且开始探索利用分岔、混沌等非线性现象造福人类.

在这样的背景下, 有必要组织在工程科学、生命科学、社会科学等领域中从事非线性动力学研究的学者撰写一套非线性动力学丛书, 着重介绍近几年来非线性动力学理论和方法在上述领域的一些研究进展, 特别是我国学者的研究成果, 为从事

非线性动力学理论及应用研究的人员,包括硕士研究生和博士研究生等,提供最新的理论、方法及应用范例.在科学出版社的大力支持下,我们组织了这套《非线性动力学丛书》.

本套丛书在选题和内容上有别于郝柏林先生主编的《非线性科学丛书》(上海教育出版社出版),它更加侧重于对工程科学、生命科学、社会科学等领域中的非线性动力学问题进行建模、理论分析、计算和实验.与国外的同类丛书相比,它更具有整体的出版思想,每分册阐述一个主题,互不重复.丛书的选题主要来自我国学者在国家自然科学基金等资助下取得的研究成果,有些研究成果已被国内外学者广泛引用或应用于工程和社会实践,还有一些选题取自作者多年的教学成果.

希望作者、读者、丛书编委会和科学出版社共同努力,使这套丛书取得成功.

胡海岩

2001年8月

前 言

本书主要是想用非线性科学和复杂性科学所提供的思想来研究和分析生命科学问题。事实上,这样的研究方法是 20 世纪数理科学和生物科学发展的必然结果。

20 世纪 70 年代,人们在研究两个客体相互作用时发现了混沌动力学行为,从而推动了非线性科学迅速发展。在此基础上,人们又进一步考虑许多客体发生相互作用时会发生什么现象的问题。模型研究表明这种系统存在着不同于定态、周期态和混沌态的更复杂的动力学行为,于是人们对这种现象产生了兴趣,提出了复杂性问题。在 20 多年对复杂性问题的研究中,科学家碰到了前所未有的困难,当然也总结了不少经验。这些经验中最重要的一条是复杂性的研究首先要基于复杂系统分析,从具体的复杂系统研究中找出规律性东西。进一步的工作又认识到生命系统是复杂系统最典型的例子。这样就自然地复杂性理论角度提出了研究生物问题的要求。

另外,20 世纪中期对分子生物学的研究使人们认识到生命现象产生的本质是大量生物分子(基因、蛋白质等各种生物分子)相互作用的结果。为了加速生物学发展和尽快理解生命本质,科学家认识到应该在系统层面上全面研究这种相互作用。事实上,现在认识到这种生物分子在作用中会表现出服从中心法则,即有规律而言。同时在作用过程中,生物分子会表现出转录、翻译、分子组装、代谢反应、输送和物理运动等动力学过程。上述认识使我们意识到可以用在数理科学中具有普遍意义的动力学和控制方法加以处理,这样生物学就自然地提出用非线性科学和复杂系统理论进行研究的要求。

两方面的互相需要使得非线性科学和生物科学最终走到一起,开启了用复杂网络来研究生物现象的时代。这是因为一方面由复杂性理论发现复杂网络是研究大量相互作用客体的有效建模工具;另一方面生命现象从系统层面研究需要考虑大量生物分子相互作用,从而由数据出发构建出各种生物网络。这样就自然迫切需要发展出研究生物复杂网络的方法,本书就是在此基础上完成的。

由于这方面研究涉及多方面内容,作者个人知识有限,所以邀请了从东京大学和加利福尼亚大学洛杉矶分校从事博士后研究工作后回国的王瑞琦、杨凌和赵兴明共同合作编著本书。我们尽可能给出这方面研究的较全面材料,希望给有志从理论上开展生命科学研究的年轻学者提供入门的渠道。

作为组织者,作者要声明三位合作者是按其姓名英文字母排序排列。每个人对本书的具体贡献如下:

刘曾荣: 绪论、第 1 章、第 5 章及 10.2、10.3、10.5 节;

王瑞琦: 第 3 章、第 8 章、9.2、9.3、9.4 节及 10.4 节部分和 10.6、10.7 节;

杨凌: 第 2 章、9.1 节、10.1 节及 10.4 节部分;

赵兴明: 第 4 章、第 6 章及第 7 章.

然后由作者在郝军军博士的协助下通读全稿, 并加以整理.

最后作者对于国家自然科学基金 (编号: 10832006, 10672093) 以及上海市重点学科 (编号: S30104) 和上海市教育委员会重点学科建设项目 (编号: J50101) 的支持表示感谢, 同时也要感谢力学学科动力学和控制二级学科同仁们的一贯支持.

刘曾荣

2011 年 3 月

目 录

《非线性动力学丛书》序

前言

绪论	1
0.1 分子生物学简介	1
0.1.1 DNA	1
0.1.2 蛋白质	2
0.1.3 从基因到蛋白质	2
0.1.4 生物大分子在生命活动中的作用	5
0.2 分子系统生物学研究的基本特征	5
参考文献	9
第 1 章 复杂网络	11
1.1 复杂系统研究的概况	11
1.1.1 耗散结构理论	11
1.1.2 协同论	13
1.1.3 复杂系统中的自适应性	17
1.1.4 非线性共性 —— 混沌	20
1.1.5 混沌控制、混沌同步与混沌边缘	27
1.2 生物系统是复杂系统	31
1.3 网络基础理论	33
1.4 复杂系统建模 —— 复杂网络	37
1.5 生物网络特征的简介	43
参考文献	45
第 2 章 化学反应动力学	50
2.1 质量作用定律	50
2.2 酶动力学 (enzyme kinetics)	51
2.2.1 Michaelis-Menten 法则	52
2.2.2 可逆的催化反应	55
2.2.3 Hill 公式	56
2.2.4 Species-Reaction (SR) Graph 简介	57

参考文献	58
第 3 章 单调动力系统	59
3.1 单调动力系统基础	59
3.2 单调动力系统的动态与分解	61
参考文献	65
第 4 章 基于生物数据的网络推导	66
4.1 概述	66
4.2 转录调控网络	67
4.2.1 构建转录调控网络的常用数据	68
4.2.2 转录调控网络重建	70
4.3 转录后调控网络	79
4.3.1 线性规划模型	79
4.3.2 动力学模型	82
4.3.3 分级网络模型	84
4.4 转录调控网络模体	85
4.5 常用生物网络分析和可视化软件	87
4.5.1 CentiScaPe	88
4.5.2 SNOW	88
4.5.3 VisANT	88
4.5.4 ChiBE	89
4.5.5 Cytoscape	89
4.5.6 Pajek	89
4.5.7 FANMOD	89
4.5.8 POINeT	89
4.5.9 Osprey	90
4.5.10 BioLayout Express3D	90
4.6 小结	90
参考文献	90
第 5 章 由进化论构建网络的方法	94
5.1 用 DD 方法研究生物网络的度负关联性	94
5.1.1 随机复制模型	95
5.1.2 偏爱复制模型	96
5.1.3 节点删除变异模型	98
5.1.4 删边变异模型	100
5.1.5 加边变异模型	102

5.1.6 重组边变异模型	104
5.2 DD 方法构建具有综合生物网络特征的模型	106
5.2.1 第一种方案	107
5.2.2 第二种方案	109
5.2.3 第三种方案	110
5.2.4 第四种方案	111
5.2.5 新的混合 DD 模型	115
5.3 用 DD 方法构建平均场意义下的生物网络	116
参考文献	129
第 6 章 基于生物分子网络的知识发现	133
6.1 网络模体	133
6.1.1 转录调控网络中的网络模体	133
6.1.2 转录后调控网络中的网络模体	137
6.1.3 蛋白质相互作用网络中的模体和模块	141
6.2 生物网络中的分层结构	143
6.3 基于生物网络的知识挖掘	148
6.3.1 人类疾病网络	148
6.3.2 药物-靶蛋白网络	153
6.3.3 小结	156
参考文献	156
第 7 章 细胞信号转导通路: 建模和识别	158
7.1 细胞信号转导	158
7.2 信号转导网络建模	161
7.2.1 微分方程模型	162
7.2.2 Petri 网络模型	164
7.2.3 布尔网络	165
7.3 基于蛋白质相互作用网络的信号通路识别	166
7.3.1 统计方法	166
7.3.2 概率图模型	168
7.3.3 优化方法	170
7.4 信号途径的应用	186
7.5 总结	188
参考文献	188
第 8 章 具有典型动力学性质的生物分子网络介绍	193
8.1 具有平凡动力学特征的网络	193

8.2 具有开关效应的网络	195
8.2.1 单基因双稳系统	196
8.2.2 双基因双稳系统	197
8.2.3 高维生物系统中的多稳态	200
8.2.4 更多平衡点的多稳态系统	201
8.3 具有振荡效应的网络	203
8.4 具有可激发性质的网络	206
参考文献	210
第 9 章 生物分子网络中的调控现象	213
9.1 基因表达调控	213
9.2 小 RNA 调控	217
9.3 积分控制	225
9.4 单调控制系统	226
参考文献	228
第 10 章 一些生命现象的网络分析例子	232
10.1 细胞周期	232
10.1.1 一类简单的通用模型	233
10.1.2 一些常见的模型	240
10.2 酵母细胞周期网络的离散和连续模型	240
10.3 酵母细胞周期网络的随机模型	244
10.4 生物节律	249
10.4.1 果蝇的生物钟	249
10.4.2 蓝藻生物钟	252
10.4.3 哺乳动物生物钟	253
10.5 哺乳动物节律的网络模型	258
10.6 细胞通信和细胞同步行为研究	268
10.7 适应性模型	273
参考文献	281

绪 论

0.1 分子生物学简介

分子生物学是一门内容丰富并对生物学发展有重大影响的学科. 从本书的需要出发, 在这里我们仅就一些分子生物学基本概念与相关知识做一简要介绍. 详细的分子生物学内容可从各种参考书中得到^[1~6].

0.1.1 DNA

生物的遗传性表现为子代与亲代的相似. 亲代传递给子代的是带有大量信息的遗传物质. 研究表明对绝大多数生物而言这种遗传物质是脱氧核糖核酸 (DNA), 只有很少一部分为核糖核酸 (RNA).

DNA 和 RNA 是由核苷酸构成. 每一个核苷酸由一个磷酸分子、一个糖分子和一个碱基构成, 具体成分可见表 0.1.

表 0.1 核苷酸的化学组成

核苷酸	DNA	RNA
碱基	腺嘌呤 (adenine, A) 鸟嘌呤 (guanine, G) 胞嘧啶 (cytosine, C) 胸腺嘧啶 (thymine, T)	腺嘌呤 鸟嘌呤 胞嘧啶 尿嘧啶 (uracil, U)
戊糖	脱氧核糖 (deoxyribose)	核糖 (ribose)
磷酸	磷酸	磷酸

DNA 的一级结构是由 A、C、G、T 四种脱氧核苷酸通过 3', 5'- 磷酸二酯键连接而成的长链高分子多聚体. 一级结构下, DNA 分子是直链分子, 一端是 5' 端, 一端是 3' 端, 它是极性分子, 因而在相关的生物理论研究中, 可以简化地把 DNA 看成由 A、C、G、T 四个字母构成的序列.

1953 年 Watson 和 Crick 在 DNA 的一级结构基础上提出了 DNA 的双螺旋结构模型. 他们提出的 DNA 双螺旋模型是由两条多核苷酸链成反向平行, 即一条链由 5' 端到 3' 端, 另一条由 3' 端到 5' 端, 呈双螺旋状结构. 在双螺旋结构中, 两条链中核苷酸排列要求 A 与 T、C 与 G 配对. 这种配对要求在维持 DNA 分子结构方面起重要作用, 而且在 DNA 复制、转录过程中起着决定性作用. 当然这种 DNA 双螺旋的二级结构可以进一步扭曲、折叠, 形成 DNA 超螺旋的三级结构.

DNA 是可以复制的. 复制过程为原先的 DNA 双链解螺旋, 以解开的双链中的每一条链作为合成模板, 生成一条与该模板配对的新链, 最后形成两个相同的 DNA 双螺旋分子. 由此可见, 在复制的 DNA 双链中, 一条是母链, 另一条是子链, 故一般称 DNA 复制为半保留复制. 这个复制过程相当复杂, 现有研究表明 DNA 聚合酶、解旋酶、拓扑异构酶和连接酶都参与了此过程. 详细情况读者可以进一步参阅有关分子生物学的书籍.

0.1.2 蛋白质

现在我们都知道蛋白质是由氨基酸构成, 自然界有 20 种氨基酸, 它们在结构上的共同之处是与一个羧基相连的 α - 碳原子、一个氨基、一个质子和一个随氨基酸不同的侧链. 除甘氨酸外, 所有氨基酸都有手性. 在水溶液中, 它们可呈酸性、中性和碱性. 氨基酸的侧链在大小、形状、电荷和化学活性上不同.

按侧链不同, 20 种氨基酸可分成下面四类.

- (1) 带电荷的侧链: 天冬氨酸、谷氨酸、组氨酸、赖氨酸、精氨酸;
- (2) 不带电极性的侧链: 丝氨酸、苏氨酸、天冬酰胺、谷氨酰胺、半胱氨酸;
- (3) 非极性脂族侧链: 甘氨酸、缬氨酸、亮氨酸、异亮氨酸、甲硫氨酸、脯氨酸;
- (4) 芳香族侧链: 苯丙氨酸、酪氨酸、色氨酸.

蛋白质的基本结构单位是氨基酸, 由此形成非常复杂的空间结构. 从形状上来说有球蛋白和纤维蛋白两种. 事实上, 每一个蛋白质都有其复杂的三维空间结构, 这种结构对其在生物学上的功能有很大影响. 在生命过程中, 蛋白质所表现的主要功能有:

- (1) 酶可以催化绝大多数生化反应, 底物的结合取决于特异的非共价相互作用;
- (2) 当与酶结合时, 膜受体蛋白将信号传入细胞内部;
- (3) 有转运与储存作用, 如血红蛋白在血液中转运氧, 铁蛋白在肝脏内储存铁;
- (4) 胶原和角蛋白是重要结构蛋白, 肌动蛋白和肌球蛋白构成有收缩力的肌肉纤维;
- (5) 酪蛋白和卵清蛋白为生长提供氨基酸的营养蛋白;
- (6) 免疫系统依赖抗体蛋白抵御感染;
- (7) 调节蛋白, 诸如转录因子, 具有结合并调节其他分子 (如 DNA) 的功能.

0.1.3 从基因到蛋白质

生物学上最重要的问题是 DNA 分子上的遗传信息是如何通过蛋白质转化为控制生物生长发育、分化和死亡的指令. 现有的研究结果表明这些功能都是通过中心法则来执行. 因而中心法则是在分子层次上理解生命现象的最重要法则.

图 0.1 给出了中心法则的示意图. 其中 DNA 通过自我增殖, 将遗传信息维持并传递下去, 这是由复制来完成的; 另外, 以 DNA 双链分子中的一条链为模板, 合成 RNA 的过程在中心法则中称为转录. 然后, 再将 RNA 的核苷酸信息转换成蛋白质氨基酸序列, 这个过程在中心法则中称为翻译, 其中转录和翻译是中心法则的主线. 除此主线外, 对中心法则还存在两点补充.

(1) 在 RNA 生物中, RNA 本身成为遗传物质, 故它也可以自我复制;

(2) 对 RNA 肿瘤病毒研究表明, 可以以 RNA 为模板形成单链 DNA, 再形成双链 DNA, 即存在由 RNA 到 DNA 的逆转录现象.

接下来, 对中心法则所描述的转录和翻译过程作一些简单说明. 这些过程在事实上非常复杂, 详细细节的了解需阅读更多参考资料, 在这里由于不是本书关心的重点, 所以就省略了.

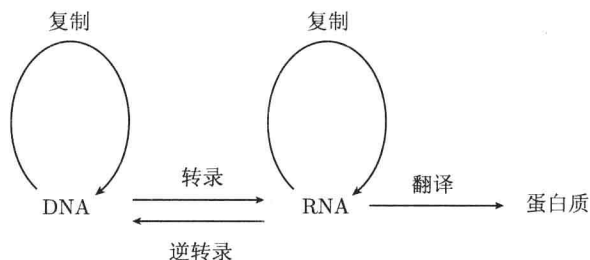


图 0.1 中心法则

在转录开始发生时, 首先是一些转录因子和 DNA 结合, 促进了 RNA 聚合酶、一些蛋白因子和 DNA 双链结合, 使双螺旋解开. 然后, 以 3' 到 5' 的 DNA 单链作为模板, 用四种 5' 脱氧核糖核苷三磷酸作底物, 按碱基配对原则接上第一个核苷酸, 这个核苷酸的位置就是基因转录的起始位点. 接着, 在延伸阶段核苷酸逐个加到前一个核苷酸 3' 端, 形成 RNA-DNA 杂合分子. 随着 RNA 聚合酶向前移动, 双螺旋逐渐打开, 模板链不断露出. 最终合成的 RNA 从模板 DNA 链上解离并脱出, DNA 仍恢复原双螺旋状. 延伸结束时的特定部位, 核苷酸加不上去了, 成为终止位点. RNA 聚合酶和 RNA 脱离模板, 形成独立 RNA, 转录过程完成, 同时 DNA 也完全恢复原状.

转录后得到的 RNA 对于原核生物一般不需修饰, 但对于真核生物一般需要修饰, 常见的修饰方式有: ① 加上帽子结构; ② 加上 poly A 尾巴; ③ RNA 拼接. 通过这些修饰可以使 RNA 成为有活性的 RNA.

下一步要考虑的是, 通过转录得到 RNA 中由核苷酸所包含的遗传信息是如何通过翻译传递给蛋白质的. 这方面的建议最早是由 Gamow 提出三联体密码或密码子的观点, 这种观点认为可由三个核苷酸的碱基来决定一个氨基酸编码. 经过人们

长期的研究, 终于得到全部 20 种氨基酸密码子, 详细结果见表 0.2.

表 0.2 氨基酸密码子

5'-OH 末端的碱基	中间的碱基				末端的碱基 3'-OH
	U	C	A	G	
U	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	U
	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	C
	亮氨酸	丝氨酸	终止子	终止子	A
	亮氨酸	丝氨酸	终止子	色氨酸	G
C	亮氨酸	脯氨酸	组氨酸	精氨酸	U
	亮氨酸	脯氨酸	组氨酸	精氨酸	C
	亮氨酸	脯氨酸	谷氨酰胺	精氨酸	A
	亮氨酸	脯氨酸	谷氨酰胺	精氨酸	G
A	异亮氨酸	苏氨酸	天冬酰胺	丝氨酸	U
	异亮氨酸	苏氨酸	天冬酰胺	丝氨酸	C
	异亮氨酸	苏氨酸	赖氨酸	精氨酸	A
	甲硫氨酸	苏氨酸	赖氨酸	精氨酸	G
G	缬氨酸	甘氨酸	天冬氨酸	甘氨酸	U
	缬氨酸	甘氨酸	天冬氨酸	甘氨酸	C
	缬氨酸	甘氨酸	谷氨酸	甘氨酸	A
	缬氨酸	甘氨酸	谷氨酸	甘氨酸	G

由表 0.2, 我们可以从核苷酸的碱基三联密码子中找到对应的氨基酸, 如碱基三联密码子为 GAA 就表示谷氨酸. 通过这种规则可把 RNA 中核苷酸序列翻译为氨基酸序列, 从而得到相应的蛋白质. 此外, 由表中看到密码子有简并性, 即同一氨基酸有好几个密码子. 比如, UCU、UCC、UCA 和 UCG 均为丝氨酸的密码子. AUG 和 GCG 即甲硫氨酸和缬氨酸的密码子, 同时兼作起始密码子; 而 UAA、UAG、UGA 都是无义密码子, 代表了终止密码和终止密码子.

RNA 通过翻译成为蛋白质的具体过程是一个极其复杂的生物过程. 涉及不同分子的合成和降解反应, 事实上还有不少细节目前尚不清楚, 故在这里我们不可能用很短的篇幅把具体生物过程讲清楚. 要想了解详细过程, 读者必须进一步学习有关的参考文献和专业书籍.

此外, 在这里我们要指出, DNA 中所有核苷酸序列并不是全部可以通过翻译编码而成为蛋白质. 通常我们把 DNA 中可以编码为蛋白质的片断所对应的 DNA 序列称为基因. 在 DNA 转录过程形成的 RNA 中, 有些可以翻译成蛋白质, RNA 中不能翻译为蛋白质的片段称为小 RNA, 小 RNA 分成四类, 分别为 siRNA、miRNA、piRNA 和 hsRNA. 现有的认识是小 RNA 的功能研究将使人类可以更加完整地理解基因表达的调控方式.

0.1.4 生物大分子在生命活动中的作用

细胞是生命现象产生的基本单元。从分子生物学可知,生物大分子是细胞的基本成分,这些生物大分子既有含有遗传信息的基因,又有传递遗传信息与其他各种生物功能的蛋白质,还有完成转录、翻译以及其他生物功能的各种生物分子。从理论上讲,这些生物分子在完成生物功能所表现的过程都可以通过化学反应来表达,而且我们可以依据化学动力学原理把每一个这样的化学反应用数学模型来描述。为此,本书也专门加上关于化学动力学内容的一章,使得读者按照化学反应来写出对应的方程。

从分子生物学理论上讲,生物系统所表现的任何生命现象(在分子水平上,如果我们了解了生物过程每一个详细细节)都应当可以通过建模进行理论水平上的研究。当然,由于一个生命现象的发生涉及许多生物分子,所建模型是巨大的,在以往科学研究中是很少碰到的。因而如何建立一整套办法,来处理所面临的新的充满挑战的问题,成为从理论上开展对生物学研究的一个紧迫的任务。

0.2 分子系统生物学研究的基本特征

分子生物学的研究成果表明对生命现象本质的理解是来自于对生物分子的理解;同时,科学界又普遍地认识到生命现象是不可能通过分析单个生物分子(基因、蛋白质和生化代谢物)而彻底解释清楚的。因此,人们逐步认识到在后基因组时代,为了深刻地理解生命现象,生物学面临的一个重大挑战就是如何从整体层面上揭示生物系统中DNA、RNA、蛋白质和各种生物小分子通过相互作用而产生的生命现象,为此需要通过从组成系统基本单元的相互作用和它们的动态行为入手来进行系统性研究,才能最终回答物种形成和执行生命功能的科学问题。系统生物学(systems biology)这门新学科就是在这种背景下诞生的。不少科学家认为系统生物学是生物学发展中第五个伟大思想(其他四个分别为基因、细胞、生物化学和适者生存的进化论),它代表了21世纪生物学的发展方向^[7]。

相对于传统生物学集中于分析构成生物系统的基本单元生物功能,系统生物学更强调在生物体的系统层面上开展对生命现象的研究。生命系统从本质上来讲是多尺度的,即一般来说,它涉及生物分子、细胞、组织、个体和群体的各个层次,因而从广义上讲系统生物学的研究按所研究的问题可以涉及上述各个层次。不管是研究一个层次还是多个层次的系统生物学研究,其核心思想是研究该系统基本单元之间如何通过相互作用所产生的动态过程来实现生物功能,从而达到解释和理解各种复杂生命现象的目的^[8~13]。

事实上,由于近代分子生物科学的迅速发展,使得科研工作者逐渐意识到,生

命的复杂程度与基因个体数目没有直接的关系. 生命在从简单到复杂、从低级到高级的进化过程中, 起决定作用的不是基因个体数目, 而是生命系统中基因、蛋白质和生物分子的相互作用而形成的复杂性. 因此, 客观上当前最迫切的课题是需要系统生物研究工作者从生物分子整体水平上解释生命现象. 另外, 随着生物学的发展, 人们利用各种高通量技术已经采集到海量的生物数据, 可以把这些数据直接用来建模; 同时, 科学家们还可间接利用来源于不同角度所积累的各种高通量数据中的信息, 获得更为完整的生物分子相互作用的模型. 总之, 从目前来看, 我们既有在分子水平进行系统生物学的理论需要, 又有构建进行研究的模型的实验基础, 因而分子系统生物学首先得到了发展, 这也是 *Nature* 杂志把其系统生物学的子刊定名为 *Nature-Molecular Systems Biology* 的原因. 当然, 有了模型后, 就需要科学家们对模型的整体性态和动态过程进行更深入的研究. 从目前来看, 大致要用到计算机科学、动力系统、控制论和图论等各方面知识. 如上所述, 可看出分子系统生物学研究是涉及多学科的研究, 国外已经非常重视这种交叉研究.

已经开展的分子系统生物学的研究结果表明这样的研究方式不仅可以帮助生物学家理解复杂的生物功能, 也可以从系统的广度来理解生命系统的根本机制或者本质规律. 因此, 这种系统分子生物学研究方法在短短几年中得到迅速发展, 成为生物学研究中的热点. 本书主要集中于分子系统生物学的内容, 主要是介绍建模、动力学分析和控制机制研究, 希望能推动国内在这方面的交叉研究. 从这个角度出发, 本节主要介绍与分子系统生物学有关的问题.

分子系统生物学的研究工作除了有关的生物实验外, 大致也要经历所研究问题的建模、模型的动力学行为和控制机制的分析以及用所得结果给出生物功能的讨论和对如何理解生物现象的解释.

首先讨论分子生物系统的建模. 在分子系统生物学研究中, 一个重要的任务是如何整合大量的生物数据来正确地建立所需要研究问题的模型. 高通量生物实验手段得到的数据通常有两个主要问题: 一是相对的数量和信息量不足, 只能观测到复杂生命现象中的一个部分; 二是数据的质量不够可靠, 通常有大量的假阳性数据存在. 因而必须发展各种计算方法, 针对这些问题来切实有效地推断或解决模型构建问题.

这里我们先以基因调控关系模型建立为例来作一个简单说明. 基因调控关系预测的研究是一个重要方面, 大量积累的微阵列 (microarray) 基因表达数据为从生物实验数据中学习基因间调控关系, 进而构建基因调控网络和信号传导通路提供了有效的途径. 这类数据的特点是一次微阵列实验能获得细胞在某一条件下的全基因组表达数据, 包含成千上万的基因在细胞中的相对或绝对丰度. 但是由于生物实验条件所限, 一次实验只能获得少数的样本点, 因此基因数据维数远远高于样本个数, 这就导致了由于基因表达数据的“维数灾难”所造成的数据相对于构建数学模