邵伟文 译
刘继凤 审校

Mathematics and 21st Century Biology

数学与21世纪生物学

Committee on Mathematical Sciences Research for DOE's Computational Biology
美国能源部计算生物学项目数学科学研究委员会
National Research Council of The National Academies
美国国家学术院国家研究委员会 著

清华大学出版社

Mathematics and 21st Century Biology

数学与21世纪生物学

美国能源部计算生物学项目数学科学研究委员会
美国国家学术院国家研究委员会 著

清华大学出版社
北京

北京市版权局著作权合同登记号 图字:01-2014-1705

图书在版编目(CIP)数据

数学与21世纪生物学/美国能源部计算生物学项目数学科学研究委员会等著;邵伟文译. --北京:清华大学出版社,2015
书名原文:Mathematics and 21st century biology
ISBN 978-7-302-39548-5

Ⅰ. ①数… Ⅱ. ①美… ②邵… Ⅲ. ①数学－应用－生物学－研究 Ⅳ. ①Q-332

中国版本图书馆 CIP 数据核字(2015)第 041383 号

责任编辑:汪 操 王 华
封面设计:傅瑞学
责任校对:刘玉霞
责任印制:沈 露

出版发行:清华大学出版社
网 址:http://www.tup.com.cn, http://www.wqbook.com
地 址:北京清华大学学研大厦 A 座
邮 编:100084
社 总 机:010-62770175 邮 购:010-62786544
投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn
质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn
印 装 者:北京密云胶印厂
经 销:全国新华书店
开 本:148mm×210mm **印张**:7.75 **字 数**:158 千字
版 次:2015 年 4 月第 1 版 **印 次**:2015 年 4 月第1次印刷
定 价:39.00 元

产品编号:055300-01

译 者 序

20 世纪 70 年代后，生物科学的新进展如雨后春笋，层出不穷。从总体上看，当代生物科学主要朝着微观和宏观两个方面发展：在微观方面，生物学已经从细胞水平进入到分子水平去探索生命的本质；在宏观方面，生态学的发展正在为解决全球性的资源和环境等问题发挥着重要作用。许多科学家常说：如果 20 世纪是物理学的世纪，那么 21 世纪就是生物学的世纪。20 世纪末到 21 世纪初的十几年中，广泛的科学技术大融合和生物学内部的发展触发了生物学领域发展的完美风暴。自动化仪器的开发与利用，使高通量生物学数据呈爆炸式增长；互联网高性能计算为生物学研究带来巨大的影响；人类基因组计划的成功奠定了生物学研究的核心资源基础；分子和细胞生物学达到一定的成熟阶段使生物学家对生物体的生命机制有所了解。总的来说，这些发展正将生物学转变成为一门定量的、数据

密集型的科学。这种转变要求数学与生物学有更加密切的合作。

21世纪以来，数学科学的新思想和新应用大量涌现，如千禧年问题之一庞加莱猜想的证明，朗兰兹纲领基本引理的证明，孪生素数猜想的重大突破，不确定性系统的量化建模，复杂系统和复杂网络的建模和分析方法，蛋白质折叠，基因测序，云计算，海量数据挖掘，大数据系统分析方法的发展，等等。随着数学与其他研究领域之间的相互交叉与融合，数学科学发挥了巨大的作用。

数学与生物学的交叉研究早在20世纪20年代R. A. 费舍尔应用数理统计帮助建立种群生物学模型就开始了，R. A. 费舍尔对生物学问题的研究同时也促进了统计学的发展。近年来随着生物学的迅猛发展，现代生物学依赖于数学科学方法来测定和处理数据成为常规方法，而数学研究在生物学中的应用范围也越来越广泛和深入，包括各种生物学模型的建模、DNA测序、蛋白质序列比对、种群动力学等。中国科学院院士马志明曾说："在生物领域需要用数学来解决的问题有很多。比如，在DNA序列的研究中就涉及许多数学问题，包括概率、统计、运筹、图论等，还涉及大量的计算问题。不仅要用到已有的高深的数学工具，比如统计推断、扩散过程、分枝过程等，而且还正在呼唤新的数学工具。例如，如何处理生物领域里的海量复杂数据，就是对数学工作者的一个挑战。"

但是总体来说，数学与生物学的融合还处于初级阶段，其中还有很多有待解决的问题。生物学中应用的还仅仅是数学中的一些基本原理，并不能反映数学中的最新进展。数学家所了解的生物学也只

停留在比较浅显的层面，还不了解生物学的最高成就。正如马志明院士所说："在数学交叉科学研究上，我提倡合作。比如，我和我的学生不可能像生物学家那样对生物学有那么透彻的理解，反之亦然。因此，要更鼓励合作，这样双方就有知识的互补。"诚然，两个学科间的隔阂不能阻止它们越来越紧密地联系在一起，数学家和生物学家们已经看到学科间更广泛的密切合作是多么的重要，而对于生物学和数学交叉领域研究进行资助的机构来说，也应该认识到这一点，鼓励两个学科间研究的密切合作项目，那么必将促进两个学科达到共同的新的繁荣。

本书为美国国家学术院国家研究委员会的能源部计算生物学项目数学科学研究委员关于《数学与21世纪生物学》的报告的中译本。该报告的目的是深入研究数学与生物学的相互影响，展现数学在各级生物组织层面生物学中的应用情况、存在的机遇和挑战、未来的发展方向等。本书共分8章，第1章介绍了生物学与数学交叉领域的现状，第2章回顾了数学与生物学成功互动的历史，第3～7章阐述了各级生物组织层面上数学在生物学中的应用情况、存在的机遇和挑战、未来发展方向等，最后一章列举了一些跨级别的横切数学应用主题，以强调生物学中数学应用研究的综合性、通用性。

关于美国能源部计算生物学项目数学科学研究委员会、美国国家学术院国家研究委员会、美国工程与自然科学部、数学科学及其应用委员会的简介与人员组成可以参见以下网址：www. national-academies. org。

本书翻译自始至终得到清华大学出版社汪操老师和王华老师的大力支持和帮助，以及刘继凤老师的审校协助，在此表示衷心感谢，同时感谢家人的大力支持和理解。

由于译者的水平有限，翻译中难免出现欠缺之处，还望各位读者批评指正。

译　者

2015年3月

前　言

本报告是受美国能源部(Department of Energy,DOE)高级科学计算研究办公室(Office of Advanced Scientific Computing Research,OASCR)的委托撰写的。该办公室肩负着多方面的责任,将数学和计算应用于对能源部很重要的各个科学领域。该办公室负责向委员会寻求指定的建议:

本研究将为美国能源部的数学科学研究活动提供建议,使科学研究能够有效地利用现有的大量基因组信息,并有效地收集利用更大量和更多样化的即将生成的结构性和功能性基因组信息。这些建议的研究活动应该既涵盖目前的科研需要,也包括一些可能通向未来的创新方法的高风险研究。

在与OASCR官员们的多次讨论中,其意图变得很明显,即要赞助广泛的、基于科学角度的机遇,这些机遇现在就在数学科学和生物

学的交叉领域。这里的“数学科学”是广义的定义，包括统计学、计算科学和所有应用数学领域[①]。尽管在将数学科学应用到物理科学方面能源部是一个有深厚根基的机构，并且在选定的生物学应用如蛋白质结构测定和基因组测序方面是一个先驱，但是这并不表示委员会只分析具体的能源部计划或者是把自己限制在能源部现有项目的界限之内。因此，这些建议只是一般性的说明，并且适用于任何一个资助机构的项目，包括但并不限于美国能源部，这些资助机构的任务围绕数学科学、生物学以及这些领域的相互作用实施。委员会一直非常努力地工作，为这些机构组织正准备支持的科学研究提供相关的有事实依据的指导。

按照NRC(美国国家研究理事会)下属报告审查委员会批准的程序，本报告已经以草稿形式通过了那些因其不同的视角和技术专长而被选出的人员的审查。这种独立审查的目的是为了提供坦诚的批评性意见，这将有助于机构发表的报告理由尽可能充分，并保证本报告对该研究委托的客观判断、证据和反馈符合机构标准。对这些审查意见和草稿保密是为了保护审议过程的完整性。我们要感谢以下人员对本报告的审查：

James Collins，波士顿大学；

Terry Gaasterland，洛克菲勒大学；

David Haussler，加州大学圣塔克鲁斯分校；

---

① 计算机科学与通信委员会即将完成的一份美国国家科学院报告将论述计算机科学和生物学交叉领域。

Douglas Lauffenburger，麻省理工学院；

Simon Levin，普林斯顿大学。

虽然以上所列的审查人员提供了许多建设性的意见，但是并没有要求他们支持本报告中的结论或建议，他们也没有看到报告发布之前的最终稿。本报告的审查是由得克萨斯州农工大学的 Ronald Douglas 监督的。在美国国家研究理事会任命下，他负责确保这份报告的独立审查是按照机构程序进行的，并确保认真参考了所有的审查意见。本编写委员会和机构对本报告的最终内容拥有完全的解释权。

此外，委员会感谢 Mark Daly、Avner Friedman 和 Alan Perelson 在本研究过程中给出的意见和建议。

目　录

执行摘要 …… 1

建议 …… 2

建议的理由 …… 5

生物学的首要地位 …… 10

不可预测性 …… 12

未来前景 …… 13

1　领域的性质 …… **15**

引言 …… 15

数学与生物学的交叉 …… 16

近年来有什么变化？ …… 19

计算生物学问题困难的原因是什么？ …… 25

数学与生物科学成功互动的常见因素 …… 27

为改善两个领域的利益协作打好基础 …… 31

本报告的结构 …… 37
参考文献 …… 38

2 历史上的成功 …… 40
种群生物学的开端 …… 40
通过同源性推断基因功能 …… 42
种群进化过程 …… 45
建模 …… 47
医学和生物学成像 …… 48
总结 …… 50
参考文献 …… 50

3 了解分子 …… 53
引言 …… 53
数学-生物学关系 …… 54
分子的数学应用领域 …… 57
序列分析 …… 57
结构分析 …… 60
动力学 …… 64
相互作用 …… 65
未来方向 …… 67
参考文献 …… 69

4 了解细胞 …… 71

引言 …… 71

这些问题的范例 …… 73

细胞结构 …… 76

细胞网络及其功能的发现 …… 79

从网络到细胞功能 …… 83

从细胞到组织 …… 90

数据整合 …… 94

生物学方面的考虑 …… 96

未来方向 …… 99

参考文献 …… 101

5 了解生物体 …… 117

心脏生理学 …… 119

循环生理学 …… 123

呼吸生理学 …… 124

信息处理 …… 125

内分泌生理学 …… 127

形态发生和模式生成 …… 128

运动 …… 132

癌症 …… 133

针对靶向肿瘤细胞的治疗输送 …… 134

药物作用的机制 …… 134

细胞群体的生长和分化 …… 134
抗性的发展 …… 135
HIV-1 感染的体内动力学 …… 136
未来方向 …… 138
参考文献 …… 139

6 了解种群 …… 148
种群遗传学 …… 148
种群的生态方面 …… 155
生态学与进化论的综合 …… 158
参考文献 …… 160

7 了解群落和生态系统 …… 165
计算 …… 174
未来方向 …… 176
参考文献 …… 184

8 横切主题 …… 195
“小 n,大 P”问题 …… 195
发现基因表达数据中的模式 …… 197
有监督学习 …… 200
无监督学习 …… 203

有序系统的分析 …… 205
隐马尔可夫模型在 DNA、RNA 和蛋白质序列分析中的应用 …… 206
序列谱隐马尔可夫模型 …… 207
基因发现中的隐马尔可夫模型 …… 209
蒙特卡罗方法在计算生物学中的应用 …… 211
模体发现中的吉布斯采样 …… 212
调控网络的推断 …… 213
蛋白质构象采样 …… 214
目前引入的数学主题的经验教训 …… 215
低级数据的处理 …… 216
结束语 …… 220
参考文献 …… 221

执行摘要

在所有生物组织尺度上，呈指数增加的大量生物学数据，连同可与之相匹配的计算能力的进步，为科学家们创造了构建定量的、预测性的生物系统模型的潜力。广泛的成功将改变基础生物学、医学、农学和环境科学。在未来几十年中，生物学发展的主要推动力将是对生物功能越来越多的定量理解；取得进展的速度将取决于对各种定量方法更深入、有效的执行和一种生物科学中的定量性视角。

这一成功转变在一定程度上取决于在生物学和数学之间创建和培养一个稳健的交叉领域，这应该成为最优先考虑的科学政策。政策挑战将是大量且多方面的。生物学与数学交叉领域是一个跨学科的前沿，横跨一片广袤的知识地带，这一知识地带是非常多样的、没有明确标记的，并会越来越广阔。委员会将在后面的章节中探讨这个前沿领域。虽然不可能在一个单一的研究中捕捉该领域的所有部分，但是委员会力图确定出一些显著特征，并举例说明机遇和挑战。

建议

委员会提出了5个建议，分别如下：

建议：支持生命科学相关数学研究的资助机构应该乐于接受那些涉及任一级别生物组织（分子、细胞、生物体、种群和生态系统）的研究提案。虽然当前大量的研究在某一特定的生物级别上可获得富有成效的成果，但是在分析各级别生物组织间相互作用方面还有大量的挑战。

生物科学已经变得越来越定量化和数据密集化；事实上，爆炸式的数据生产和充满估计精度的定量分析是21世纪生物科学最显著的特点。生物科学的进步将越来越依赖于数学分析在各级生物组织研究中的深入和广泛集成。没有哪一个级别的组织特别为数学应用提供极具吸引力的机会。在不同级别上面临的挑战各具有与众不同的特色，但也有统一的主题。本报告的一些章节是围绕不同级别生物组织而编写的，但其他的章节，包括“领域的性质”、“历史上的成功”和“横切主题”，则更广泛地着眼于过去和现在数学对生物学的应用的共性方面。

建议：支持生命科学相关数学研究的资助机构应该优先考虑下面这样的提案，即表明对研究的具体生物学对象有一个清晰的认识，并包括一个数学家和生物学家将如何合作以实现目标的切实可行的计划的提案。