

生物大分子的 动力学分析和应用

——生命科学在定量化和
信息化研究中的理论核心问题

沈世镒/著

The Dynamic Analysis and
Application of
Biological Macromolecules

Life Science's Theoretical Core Issues
From the Perspective of Quantitative and
Information Research



科学出版社

生物大分子的动力学分析和应用

——生命科学在定量化和信息化研究中的理论核心问题

沈世镒 著

科学出版社

北京

内 容 简 介

生物大分子包括:糖、脂、蛋白质和核酸,也包括由它们的组合或混合所产生的细胞器、细胞等.其中的动力学是指它们的结构、功能、形成过程、相互作用中的各种问题.

本书由三部分,共二十二章组成.其中第一部分是围绕生命科学中的核心内容和多种不同学科有关的基本概念和基本知识,除了对它们作简单介绍外,还有一些发展和推广.在第1章中对有关理论核心问题作重点说明和讨论,有一些新观点的提出值得关注.

第二部分是动力学部分,其中包括信息动力学和分子动力学的内容.信息动力学是由观察数据来分析生物分子中的动力学规则,因此是动力学中的反问题.而分子动力学包括溶液中分子运动的动力学、原子与分子相互作用、产生生化反应中的动力学问题.

第三部分是应用部分,分别针对几个典型应用问题进行综合研究和讨论.

附录,对本书常用的数学公式、符号、物理量名称、量纲及多种不同类型物体的物理尺度作补充说明.

本书除了正文之外,还附数据资源,该资源给出一些原始数据、计算分析结果和彩色图像.请访问科学书店 www.ecsponline.com,检索图书名称,在图书详情页“资源下载”栏目中获取,数据资源名称为“本书光盘”.

图书在版编目(CIP)数据

生物大分子的动力学分析和应用:生命科学在定量化和信息化研究中的理论核心问题/沈世镒著. —北京:科学出版社,2018.6

ISBN 978-7-03-057436-7

I. ①生… II. ①沈… III. ①生物大分子-动力学分析 IV. ①Q71

中国版本图书馆CIP数据核字(2018)第101419号

责任编辑:李欣/责任校对:邹慧卿

责任印制:张伟/封面设计:无极书装

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>

北京虎彩文化传播有限公司印刷

科学出版社发行 各地新华书店经销

*

2018年6月第一版 开本:787×1092 1/16

2018年6月第一次印刷 印张:36

字数:726 000

定价:198.00元

(如有印装质量问题,我社负责调换)



谨以此书献给我的夫人——郭锡娟！感谢她在我
体弱多病时对我无微不至的照顾，使此书得以完成。

作者简介



沈世镒，1956 年进入南开大学学习，1961 年考取南开大学数学系研究生（导师胡国定）。1977 年由山西调入南开大学数学系任教。1986 年国务院学位委员会聘任为南开大学博士生导师。1997 年，任南开大学数学科学学院院长。1997 年，任天津市数学会第八届理事会理事长。曾任康奈尔大学访问学者。

主要研究方向：代数编码；信息的度量及其应用；组合密码学；神经网络系统理论及其应用；近代密码学。

主要研究成果及获奖：出版著作 5 部；发表论文 50 多篇，其中 SCI 收录 12 篇。1993 年，获“天津市自然科学技术领域中青年授衔专家”称号，天津市优秀教师。四次获天津市、教育部科技进步奖。

前 言

宇宙、万物和生命的起源及它们演化有许多研究和进展,但仍存在许多难解与未解的问题.它们同样神秘而有趣,而且有许多共同点,而生命科学与人们的日常生活有更加密切和直接的关系.

随着生物信息学的产生与发展,生命科学进入了定量化、信息化的分子水平研究阶段,这样又会产生许多新的疑难或未解问题,产生这些问题的原因最终都可归结为生物分子结构和功能的复杂性.每个生物分子包含数百到数亿个原子,因此在它们之间存在一系列复杂的动力学问题,我们把它归结为四大理论核心问题.

如果我们把系统看作研究对象,那么理论核心问题之一是三大基本要素问题,这就是物质、能量与信息是构成系统中的基本要素.它们也是在自然界、生命、万物中的基本要素,它们互不相同,但相互联系、可相互转化.

三大基本要素的存在与表现形式具有不同的物理尺度与不同的时空结构.其中物理量描述不同物体特征的指标,它们具有特定的量纲与单位.

物理尺度是指这些物理量大小变化的数量级.从基本粒子到宇宙空间,它们的物理量尺度指标的变化范围很大(有数十个数量级的差别).而在不同的物理尺度下它们又有各自的时空结构.相对论从根本上改变了人们传统的时空观,时空是相互联系的整体.同样在微观条件下也有多种不同时空结构类型.量子物理中的波粒二重性、海森伯的不确定性原理、宇宙大爆炸理论等都是在不同的时空结构中存在与表达.

除了宏观和微观的时空结构外,还有超宏观和超微观结构,一些未解和难解问题与这些超时空理论有关,但我们对其中的特征和规律还知道的很少,一些结果还只是在理论推导阶段.

三大基本要素,它们具有不同的物理尺度与时空结构,通过数学的理论、方法与模型对它们作综合研究与分析,由此形成一系列的规律与性质,这是自然科学,也是生命科学中的第一理论核心问题.它是其他理论核心问题的基础,也是本书的核心内容与主导思想.

理论核心问题二是对不同学科作综合研究.不同学科产生于不同的历史时期,分别针对不同的问题进行讨论与研究,其中的许多规律既有普遍性,但也有局部性与不完备性,它们都处在不断地发展、修正与完善中.

生命现象或生命过程实现了对这些学科的综合处理.它不仅与这些学科都有密切关系,而且使这些规律成为协调、有序、统一的整体,使它们在各种不同的条件下实现了有机的衔接.因此对它们作综合研究不仅是生命科学中的理论核心问题,也为这些学科提出了一系列新的问题.因此对生命现象的研究是推动这些学科作进一步研究的推动力与天然实验室.

理论核心问题三是关于能量的统一研究.在生命体中,从原子到分子、生物大分子及其他生命体都有各自的能量表现与运动的动力学关系,在宏观、微观与生物体中都存在能量问题.它们不仅有大小、强弱的区分,而且还有产生过程和存在方式的不同.太阳光、水和酶是能量的产生与转移过程中的三大关键因素.在不同的学科(如量子物理与热力学)中、在

分子内部结构(一级结构、空间结构、生物功能)中,能量有多种不同的存在或表现方式、特有的属性、可相互转化的关系,并发挥不同的作用。

能量不仅有强弱的区别,也有产生过程、存在和表达方式的不同,它们可以相互转化,在生物分子的结构和功能中有不同的效果。对能量的综合、统一研究也是生命科学中的理论核心问题。

理论核心问题之四是**选择性原理**。该原理是因果关系中的重要组成部分,万物的产生与演化(也包括物种进化)都与这个原理有关。因此宗教、哲学、科学都介入对该原理的讨论。许多自然现象与生命过程在实现这种选择性时具有三大基本特征,这就是大规模并行计算(或操作)、智能化的学习和训练过程及系统的最优化(优胜劣汰)的原则。它们在宇宙、万物和生命的演化过程都具有这些特征,因此是一种普遍的特征。

选择性原理中的三大基本特征是万物演变过程中的基本特征。但在另一方面,许多生命现象(分子和生物分子)所存在的许多特殊功能无法由这三大基本特征来说明。因此,我们认为一定还有第四个基本特征存在^①,并且猜测该特征和基本粒子、量子场论密切相关。这就是在原子核内部存在的,与各种不同类型的粒子和相互作用有关。正由于这些因素的存在,才有可能使一些特殊类型分子的形成和特殊的生物功能得以实现。对基本粒子、原子、分子、生物分子及对此四大理论核心问题的综合研究可能会使量子生物学这一新兴学科得以产生和发展。

在这些理论核心问题的讨论中,不可避免地会涉及在更广泛意义下的讨论,如宗教、哲学和科学问题的讨论。

例如在因果关系的讨论中,不同的宗教和哲学都有各自的学说和理论。由于在宇宙学、物理学、生命科学中,出现了许多新的、未解和难解问题,其中许多争论和讨论还会长期存在。

信息的概念是个哲学概念,在哲学中对信息的定义有数十种之多。我比较认同的说法是把信息定义为物质、能量结构和运动的表现形式,因此物质仍然是产生信息(也包括能量)的基础,但信息较物质和能量具有更高级、更复杂和更深入的表现特征。在本书中我们提出**信息实在论**的观点,这是和**关系实在论**比较的思想。关系是一种信息,但是信息的概念比关系具有更广泛、深入和丰富的内容。信息具有一系列重要特征,如可度量性的特征,这种度量性反映了系统运动的方向性,结构的复杂度、混乱度及不同系统之间的差异度等。这些特征在关系实在论中是无法说明的。

在生命科学中,信息有特殊的意义。生命过程就是分子运动在无序和有序之间的循环运动,各种生命体从产生到消亡和转化过程都符合这种运动循环的特征。

信息在生命中另一个重要意义是在生命体的各种系统中存在大量的信息处理问题。如神经、免疫、血液循环等所有系统中存在一系列信息处理过程问题,它们最后都可归结为分子信号、电磁信号(都是信息)的变化与运动的处理过程中。

由于生物信息学的出现,大量生物数据在原子或分子水平的测定,各种数据分析软件、

^① 该猜想已被量子生物学所证实。量子生物学说明了生物分子的许多功能和基本粒子的特殊运动有关(见文献[297])。

网络的产生. 它们是生命科学在定量化和信息化发展中的重要进展.

信息实在论仍然是以物质为基础的哲学, 但信息已成为客观世界的主体. 这种观点对我们认识世界, 尤其是在信息革命及信息社会发展的今天有特殊的现实意义. 现在所出现的各种信息处理问题 (如大数据、云计算、智能技术中所出现的问题) 已远远超过人们对信息的早期理解. 它们所涉及的领域、范围及对人们观念的影响也是关系实在论所不能说明的.

目前, 生命科学研究所涉及的物理尺度仍然是以微 (μ)、纳 (n) 尺度 (分子或原子) 级水平为主. 即使如此, 生物大分子所包含的原子数目可以多达数百亿个, 对它们的结构、功能等研究和分析已变得十分困难, 存在许多未解和无解之谜. 我们应该探求新的思路和方法.

如果物理尺度降低到皮 (p)、飞 (f)、阿 (a) (原子核和基本粒子) 级或更小的仄 (z)、幺 (y) 级的水平, 会出现什么样的结果值得讨论, 这种超微的时空结构会给我们带来新的困难, 但也可能出现新的机遇. 在物理学、生命科学中出现的这些困难和疑难问题也许能得到进一步的讨论与说明.

本书内容由三部分组成. 其中第一部分是预备知识, 这就是不同的学科中的与本书内容有关的一些基本知识. 除了一些基本概念和知识外, 对其中的一些特殊问题还进行专题讨论, 它们是这些基本知识的补充与推广.

第二部分是动力学部分, 其中包括信息动力学和分子动力学的内容. 信息动力学是动力学中的反问题, 这就是通过观察结果来分析分子运动的动力学规则. 大量生物信息的测量和记录是信息动力学的研究基础, 它们是生命语言的组成部分, 对它们的解读就是对生命语言的解读, 在物理、化学、生物中已积累的大量观察、测量数据也是信息动力学的组成部分.

信息动力学还处在刚刚起步的阶段, 面对如此复杂的生命现象、如何寻找其中的特征和规律还不充分, 因此有很大的发展空间.

分子动力学的内容包括分子键、分子碰撞、生化反应等动力学问题, 也包括各种生物大分子在结构、功能分析中的动力学问题. 它们的内容都很复杂, 涉及多种学科的不同理论和方法的讨论.

第三部分是从定量化、信息化的角度讨论一些应用问题. 其中包括神经网络系统和智能计算、蛋白质 (酶)、免疫系统和生物能源等问题. 这些问题都是不同领域中的大问题, 我们这里只能讨论与介绍其中的基本知识和关键问题.

对这部分内容补充说明如下.

(1) 以香农的信息论、维纳的控制论、冯·诺依曼的计算机为代表的 3C (Communication-Control-Computer) 理论是信息科学, 也是近代科学发展的起点. 它们几乎同时在 20 世纪 40 年代末产生. 在此短短的不到 70 年的时间内, 信息科学、IT 领域中的技术、产业、市场等领域实现一次巨大的革命过程, 由此形成了当今的信息社会.

(2) 回顾 3C 理论的发展过程, 前 25 年是理论的消化期, 这是对 3C 理论的学习、理解、消化和解读期, 20 世纪 70 年代的一系列专著和论文的发表标志这个阶段的完成. 之后的 20 年是理论和技术的发展、结合期, 其主要特点是理论体系研究的深化、电子技术的迅速发展、理论和技术结合产生了一系列的应用问题. 信息科学大规模进入产业和市场是近 20 ~ 30 年的事.

这一代科技革命的发展过程值得人们思考和借鉴。由此说明,这种原创性科学思想的产生、多学科的综合研究、理论和应用的结合,是造成这个科技革命的关键原因,其中也包括一系列科技、人才培养政策和措施的确立和推动。这些都是今后科技发展中的宝贵经验。

(3) 下一个科技发展的新爆发点是什么值得关注,以下几个方面值得关注。

信息科学和神经科学的结合。其中信息科学包括:3C 理论 + 大数据(其中包括网络和云计算),而神经科学包括脑科学和智能计算。

智能计算与生命科学的结合(简称生物智能计算)。这两个方向相互交叉,但层次略有不同。它们发展和更新周期可能会很快,规模和影响也会更大。

(4) 在生命现象中存在许多不可思议的现象。

以大肠杆菌为例,它的基因组长度是 3×10^{10} bp, 它的复制速率是 2000bp/s, 在 2000 bp 的核苷酸中包含原子数目约 6.0 万个(其中非氢原子约 4.0 万个)。如果按动力学的分析,它仍然是一个十分复杂的系统。但在生命现象中,能够如此快速地实现它们的分解、复制、组合和分离,并实现细胞核中的许多功能(如 RNA 的转录、蛋白质的转译和复制等,在大肠杆菌的 2000bp 的核苷酸序列可以复制 600 AA 规模的蛋白质)。因此无论是从动力学,还是计算机的角度来看,这种智能化的效率和水平都是无法想象的。

在量子理论中,即使对最简单的氢原子、氢分子,要确定它们中电子运动的规律,就需要对一些复杂的方程组求解。但在生命系统中对这种复杂系统的操作过程却显得十分轻而易举,而且能够如此快速、协调、有序、精确无误地实现这些功能。这正是生命科学的魅力所在,这也是十分神奇和不可思议的。

为实现这种生物智能计算,了解各种生命现象,如酶在温和条件下实现快速催化,免疫系统在应答过程中实现快速识别、激发、调控的过程,以及它们与神经、血液循环等系统之间的协调和一系列信息的处理过程。这些问题都与生物智能化密切相关,是其中的组成部分。

在生命现象中,除了我们所比较了解这些系统外,还有许多我们所不了解或很少知道的系统。例如在人体中存在 10^{14} 个细菌,它们和人体形成共生系统,正由于这种共生现象的存在,使人体生存所需要的各种功能才能得以实现(也会导致各种疾病的发生)。这种共生现象是如何存在的、各种生理功能是如何协调实现的、这些细菌有哪些类型,它们同样存在生物大分子的各种类型和结构。但我们对其中的情形很少了解。

由此可见,这种生物智能化的特征是十分奇妙与不可思议的,在本书中我们虽讨论了生命科学中存在的一系列理论核心问题,但是从根本上还没有解决其中的许多未解和难解问题。量子生物学可能成为我们作这种继续进行深入探讨的理论基础。

生命科学 + 智能计算 + 量子理论的综合研究可能是未来科技发展中的主流方向,也是这些理论核心问题中的最大问题。这是一个未经开拓、层次最高的新研究领域,值得我们去探索。

(5) 科学、技术的发展可以造福人类,但也存在风险和危害(巨大的风险已经存在,而且还在扩大中,新的风险还在不断产生)。科学、技术的发展进程势不可挡。如何解决这些问题是人类必须面对的新课题。通过行政、立法、条约等方法只能解决一些迫切性的问题,但不能从根本上消除这些风险。

实现世界大同等人类理想社会是克服这种双刃性的根本途径和最终出路。科学技术的这种新发展为实现这个目标带来支持和可能,但这个过程一定十分曲折与复杂,在不同的历史条件下有不同的理解和要求。高度发展的社会生产力、科学技术和精神文明,是实现这个目标的基本要求。

(6) 本人有幸见证并参与了信息论、信息科学的研究和发展的历程,早在 1962 年就证明了香农基本定理(第一定理和第二定理)的相互关系(见文献 [248], 在 [58] 等文献中被详细引用)。之后,又有幸见证并参与了生命信息学的确立过程,这是生命科学在定量化和信息化研究中的一个新起点,在 2002 年得到了序列比对 (Alignment) 的快速算法,该算法的计算复杂度与序列长度呈线性关系的比例增长(见文献 [163])。

随着对生物信息学研究的深入,发现其中存在的许多未解或难解问题,它们所涉及的内容是多学科中更深层次、综合性的关系,而且与智能化中的一系列问题有密切关系,如何建立它们沟通的桥梁也是考虑中的问题。因此希望能把这些观点表达出来,供读者考虑和讨论。

我国有数千年的优秀思想和文化发展的历史,也是能接受和实现中外文明融合的典范。这种精神在科技高度发达的今天尤其重要。实现思想、理论的创新和发展,这是生产力能得到持续发展的根本,也是人类未来文明发展的根本,也是实现中国梦的组成部分。

在本书的内容中还包括附录和数据资源。数据资源给出一些原始数据、计算分析结果和彩图。可访问科学书店 www.ecsponline.com, 检索图书名称,在图书详情页“资源下载”栏目中获取,数据资源名称为“本书光盘”。因为本书内容涉及多个不同的学科领域,对一些常用的概念和符号作统一的说明,这些符号与原来的学科大体保持一致,但同一符号在不同场合会有不同的含义。这些符号体系与文献 [164], [166] 中大体相同。

附录二给出了多种物理尺度指标,从这些指标中可以看到它们变化的状态分布情形及它们所对应物理尺度的数量级,尤其是多种不同类型的生物体、生物分子的运动和变化情形,为我们对定量化的研究中的一些内容和指标增加一些初步和直观的理解。

感谢吴忠华、胡刚、王奎、高建召在本书写作过程中给予的帮助。有些结果已在文献 [251], [252] 中给出。感谢清华大学夏树涛教授,他建议把时空理论作为重要内容写入,对该理论虽未详细展开讨论,但我们把它归结为核心理论中的重要组成部分。尺度和时空相互关联,在不同的尺度下有不同的时空结构和特征。

此书谨献给我的爱人郭锡娟,感谢她在我最困难的时候关心照顾、支持帮助我。

感谢中天联科有限公司、清华大学深圳研究生院为本书的出版提供资助。

由于作者水平所限,许多内容又涉及多学科的领域和专门性的课题,因此存在不当之处在所难免。有的论述或结论可能与他人重复,在此也未能全部进行说明,对此务请见谅。也请有关作者或读者给予批评指正,并希望能作进一步的讨论。

沈世铨

2017 年 3 月于南开园

目 录

前言

第一部分 基本原理和基础知识

第 1 章 生命科学概论	3
1.1 宇宙、地球与生命	3
1.1.1 宇宙学、宇宙大爆炸说	3
1.1.2 地球的早期演变	6
1.1.3 生命的出现和演变	8
1.1.4 一些重大事件和学说	9
1.2 理论核心问题	12
1.2.1 理论核心问题之一: 三大基本要素和它们的尺度指标	12
1.2.2 不同尺度下的时空观	14
1.2.3 时空结构中的数学和物理	15
1.2.4 理论核心问题之二: 不同学科的综合研究	17
1.2.5 理论核心问题之三: 能量的表达和转化	18
1.2.6 能量表达的类型	20
1.3 理论核心问题之四: 选择性原理	22
1.3.1 宗教和哲学中的讨论	22
1.3.2 自然科学中的选择性原理	23
1.3.3 选择性原理中的几个关键特征	24
1.3.4 选择性原理中存在第四特征的猜想	25
1.4 信息实在论	27
1.4.1 信息的含义和特征	27
1.4.2 信息实在论和关系实在论	28
1.4.3 不同尺度背景下的科学和哲学	29
第 2 章 量子理论概述	31
2.1 量子力学	31
2.1.1 从经典力学到量子力学	31
2.1.2 量子力学的数学模型	33
2.1.3 量子力学中的一些基本原理	35
2.1.4 不确定性原理和对它的讨论	37
2.2 原子结构	38
2.2.1 氢原子和类氢原子中的电子运动轨道	38
2.2.2 原子结构的一般理论	41

2.2.3	原子的光谱理论	44
2.3	元素结构周期表	45
2.3.1	元素周期表的排列方式	45
2.3.2	元素结构一览表	47
2.3.3	对表 2.3.4 中有关参数的分析	48
2.4	量子物理中的几个特殊问题	50
2.4.1	粒子在势能场中的运动	50
2.4.2	几种简单运动的量子特征	51
2.4.3	二态系统和量子的共振问题	54
第 3 章	量子化学和量子场	58
3.1	化学和量子化学的一些记号	58
3.1.1	化学键	58
3.1.2	基团和分子官能团	60
3.1.3	量子化学的模型和记号	62
3.2	量子化学中的一些计算问题	64
3.2.1	氢原子、氢离子和氢分子的量子化学模型概述	64
3.2.2	有关参数和解的讨论	67
3.2.3	杂化轨道理论	68
3.3	量子场简介	70
3.3.1	量子场的一些基本概念	70
3.3.2	经典动力场和量子动力场的基本运动方程	73
3.3.3	经典动力场的量子化	76
3.3.4	由光子和电磁场所产生的量子场	78
3.3.5	几个问题的补充说明	81
第 4 章	其他动力学问题	83
4.1	热力学及其基本定律	83
4.1.1	热力学体系	83
4.1.2	热力学第一定律	84
4.1.3	热力学第二、第三定律	86
4.1.4	统计力学对熵函数的计算	87
4.2	量子统计物理学概述	89
4.2.1	粒子系统的量子化	89
4.2.2	理想气体	91
4.2.3	其他微观粒子	93
4.3	几种重要类型的化学反应和电化学	95
4.3.1	酸碱反应	95
4.3.2	氧化和氧化还原反应	97
4.3.3	电化学概述	99

4.3.4 其他类型的化学反应	100
4.4 光电、光化和光合作用	101
4.4.1 太阳能概述	101
4.4.2 光电和光化作用	104
4.4.3 光合作用	106
4.4.4 代谢概述	109
第 5 章 分子点线图、分子空间结构和拓扑空间	111
5.1 分子点线图	111
5.1.1 点线图的一般定义和性质	111
5.1.2 分子点线图的表示	115
5.1.3 分子点线图的组合和分解及它们在化学反应中的表示	117
5.1.4 分子空间结构的点线图表示	118
5.2 分子空间结构的几何理论	120
5.2.1 分子空间结构的参数系	120
5.2.2 空间四原子点的结构和它们的参数系	122
5.2.3 多原子点结构的几何分析和它们的参数表达	124
5.2.4 不稳定参数的计数问题	126
5.3 坐标系和坐标变换	127
5.3.1 直角坐标系和它的变换	127
5.3.2 直角坐标系和极坐标系	129
5.3.3 欧拉角和旋转变换	131
5.4 拓扑空间和空间多面体	132
5.4.1 集合和拓扑空间	132
5.4.2 多边形和空间多面体	134
5.4.3 一般线性空间理论和生物大分子的空间结构	136
第 6 章 生物大分子的空间结构分析	139
6.1 总体结构和形态相似性分析	139
6.1.1 对总体结构的讨论	139
6.1.2 空间结构的相似性	140
6.2 生物大分子的空间结构分析-II (形态特征的分析)	143
6.2.1 深度分析和几何计算	143
6.2.2 其他类型深度的定义和性质	146
6.2.3 小球滚动法	147
6.2.4 多面体收缩法	151
6.3 多面体的拟合理论	153
6.3.1 多面体拟合的定义和要求	154
6.3.2 表面的凹、凸的特性指标	155
6.3.3 生物大分子空间结构讨论和分析的意义	157

第 7 章 随机分析	158
7.1 随机变量和随机系统	158
7.1.1 随机变量及其分布函数	158
7.1.2 主成分分析法和随机控制	162
7.1.3 随机变量的极限性质	163
7.2 几种重要的随机过程	165
7.2.1 随机过程概论	165
7.2.2 独立随机序列	166
7.2.3 泊松过程	167
7.2.4 可加过程	169
7.2.5 对偶过程和首达时间	170
7.2.6 马氏过程的定义	171
7.3 布朗运动和更新过程	174
7.3.1 布朗运动的定义和基本性质	174
7.3.2 布朗运动的极限性质和其他一些分布性质	176
7.3.3 布朗运动的函数变换性质	177
7.3.4 更新过程	178
7.4 复合随机过程	180
7.4.1 复合随机过程概述	180
7.4.2 复合更新过程	181
7.4.3 基因突变序列	181
7.4.4 排队过程	183
7.5 随机微分方程概论	184
7.5.1 随机积分 (伊藤积分) 和随机微分 (伊藤微分)	184
7.5.2 随机微分方程的定义和类型	185
7.5.3 随机微分方程的性质	186
第 8 章 随机分析应用和神经网络系统	187
8.1 朗之万方程及其应用	187
8.1.1 朗之万方程的定义和求解	187
8.1.2 关于朗之万方程解中参数的讨论	190
8.1.3 涨落分析	191
8.2 随机分析的其他应用	192
8.2.1 基因的突变和比对的随机模型	193
8.2.2 基因的突变和比对中的随机过程	194
8.2.3 基因突变的 A-空间理论和应用	196
8.3 神经网络系统概论	197
8.3.1 神经网络系统的一些基本知识	197
8.3.2 NNS 的信息处理过程	199

8.3.3	ANNS	200
8.4	ANNS 中的基本模型和它们的学习算法	202
8.4.1	感知器的模型和算法	202
8.4.2	HNNS 概述	204
8.5	其他类型的智能算法的分类	206
8.5.1	正向和反向的 HNNS	206
8.5.2	EM 算法及其理论分析	208
8.5.3	EM 算法的实例计算	209
8.5.4	其他算法和算法的分类	211
第 9 章	生物大分子的基本结构	213
9.1	糖和脂	213
9.1.1	糖类概论	213
9.1.2	单糖的结构和类型	214
9.1.3	脂类概论	216
9.2	氨基酸和蛋白质概论	218
9.2.1	氨基酸概论	218
9.2.2	氨基酸的结构和表示	221
9.2.3	肽链和蛋白质	224
9.3	核酸简介	227
9.3.1	核苷酸的分子结构	227
9.3.2	DNA 的双链结构模型	229
9.3.3	遗传密码子	231
9.3.4	染色体	232
9.4	细胞概论	234
9.4.1	细胞的类型和结构	234
9.4.2	细胞器概述	235
9.4.3	原核细胞和真核细胞的比较	236
第 10 章	基因组中的基因识别问题	239
10.1	生物信息学中的核酸序列	239
10.1.1	基因组概述	239
10.1.2	不同类型核酸序列的结构分析	242
10.1.3	基因组的其他重要功能特征	245
10.2	基因组和染色体的补充性质	246
10.2.1	DNA 和蛋白质的组合与混合结构	246
10.2.2	细胞的有丝分裂和 RNA 的结构分析	248
10.3	原核生物基因组的基因识别问题	249
10.3.1	原核生物的基因识别	249
10.3.2	ORF 序列的定义和生成算法	251

10.3.3 对基因识别问题的讨论	252
-------------------------	-----

第二部分 信息动力学和分子动力学

第 11 章 ID 概论和一级结构的 ID 分析	257
11.1 ID 概论	257
11.1.1 数据库的类型和特征	257
11.1.2 ID 的基本方法之一: 信息统计法	259
11.1.3 ID 的基本方法之二: 组合分析法	262
11.1.4 一级结构的词法和句法分析	263
11.2 蛋白质一级结构的 ID 分析	266
11.2.1 蛋白质一级结构数据库的选择和特征数的计算	266
11.2.2 蛋白质一级结构的 ID 运动方程	268
11.2.3 IDF 的频谱分析	271
11.2.4 有关频谱分析的讨论和思考	273
11.3 基因组的 ID 分析 (信息统计法)	274
11.3.1 概论	274
11.3.2 IDF 的计算结果和分析	276
11.3.3 由基因组 IDF 所产生的词和词法分析	279
11.3.4 基因组 ID 分析的组合法	281
11.4 原核生物的基因识别问题	284
11.4.1 CDS 和 ORF 序列的数据结构	285
11.4.2 原核生物的基因识别计算	286
11.4.3 基因识别的 IDF 法	288
第 12 章 蛋白质的三维结构分析	290
12.1 氨基酸的空间结构分析	290
12.1.1 氨基酸的空间结构模型	290
12.1.2 花盆和花架空间结构的参数分析	292
12.1.3 氨基酸花枝 (侧链) 的空间结构特征分析	295
12.1.4 氨基酸空间结构的其它计算方法	298
12.2 蛋白质主链的三维结构 ID 分析	300
12.2.1 蛋白质的主链结构	300
12.2.2 三角形拼接带中的一些其他性质	303
12.2.3 中位点曲线的一般性质	306
12.3 中位点曲线和蛋白质二级结构分析	309
12.3.1 在二级结构分析中的一般分析	309
12.3.2 对 α 螺旋结构的分析	310
12.3.3 对 β 折叠结构的分析	312
12.3.4 中位点曲线的应用	316

12.3.5 其他二级结构类型和性质	320
12.3.6 二肽的折叠系数	322
第 13 章 蛋白质的空间形态的 ID 分析	326
13.1 蛋白质的空间形态的 ID 分析方法	326
13.1.1 有关蛋白质的总体结构和相似性指标的计算	326
13.1.2 深度计算结果	327
13.1.3 深度倾向性因子	329
13.1.4 深度倾向性因子的拟合和预测	331
13.2 蛋白质形态特征的实例分析	335
13.2.1 蛋白质总体形态特征参数的实例分析	335
13.2.2 利用小球滚动法对蛋白质空间结构特征的分析	340
13.2.3 多面体拟合和多面体收缩法的实例分析	341
13.2.4 关于多面体的拟合问题	343
第 14 章 核酸序列的空间结构分析	345
14.1 RNA 的二级结构预测和空间结构分析	345
14.1.1 RNA 概述	345
14.1.2 二级结构的动力学分析	347
14.1.3 RNA 二级结构的性质和预测算法	348
14.2 核酸的空间结构	351
14.2.1 核酸序列分析概论	351
14.2.2 分子官能团 [OCCCOP] 的空间结构分析	353
14.2.3 核苷酸的其他空间结构分析	355
14.3 组蛋白的空间结构分析	357
14.3.1 DNA 的单链结构分析	357
14.3.2 DNA 双链的空间结构分析	360
14.3.3 混合结构分析	362
第 15 章 键动力学	364
15.1 键的分析	364
15.1.1 键的类型	364
15.1.2 键的作用	366
15.1.3 生物大分子中键的类型分析	367
15.2 对生物大分子中非固定键的搜索和分析	368
15.2.1 不同类型键的搜索和分析	368
15.2.2 不同分子的动力学倾向性因子	372
15.3 蛋白质分子结构的讨论	375
15.3.1 问题的由来	376
15.3.2 产生一些特殊肽链的可能性	377
15.3.3 有关蛋白质结构的讨论	378