

数学与现代科学技术丛书 4

信息动力学与生物信息学

——蛋白质与蛋白质组的结构分析

沈世镒 胡 刚 王 奎 高建召 著



科学出版社

数学与现代科学技术丛书 4

信息动力学与生物信息学

——蛋白质与蛋白质组的结构分析

沈世镒 胡 刚 王 奎 高建召 著

科学出版社

北 京

《数学与现代科学技术丛书》序

当代数学在向纵深发展的同时,被空前广泛地应用于几乎一切领域.一方面,它与其他学科交汇,形成了许许多多交叉学科(例如,信息科学、计算机科学、系统科学、数学物理、数学化学、生物数学、数学语言学、数量经济学、金融数学、复杂性科学、科学计算等);另一方面,它又被应用于高新技术的开发(例如,信息安全、信息传输、图像处理、语音识别、网络、海量数据处理、网页搜索、遥测遥感、交通管理、医疗诊断、手术方案、药物检验、商业广告等方面),成为一些高新技术的核心.应用数学的这种发展趋势急剧地扩展了数学的疆界,也深刻地改变了数学的面貌.

中国的经济正在迅猛发展,其中的科技含量也与日俱增.为了提高自主创新能力,我国已经有不少数学工作者投身于这类应用数学的研究中,还有更多的数学工作者则正在密切关注这方面的进展,看好它的前景.愈来愈多的人希望了解这类应用数学的现状,寻找入门之径.

《数学与现代科学技术丛书》是力图反映这个发展趋势的一套应用数学丛书,它将较全面地向我国读者介绍当今数学在现代科学技术各个领域中的应用的状况,通过必要的准备知识,逐步把读者引向相关的研究前沿.

从事交叉学科研究和高新技术开发的应用数学家,除了要精通所需的数学知识外,还必须深入了解其所研究问题的来龙去脉.“建模”是应用数学研究实际问题的关键.这也是一门数学艺术:从复杂的实际问题中抽象出关键的“量的关系”,使得既能反映出问题的基本特征,又能用现阶段的数学工具加以处理.有鉴于此,这套丛书的一个特点就是:不但要介绍有关的数学理论和方法,还必须介绍问题的来源与背景、数学建模以及如何运用数学工具来解决实际问题.

本丛书适用于数学及相关专业的大学生和研究生,以及与数学有关的各专业知识工作者.

张恭庆

2009年11月

前 言

信息动力学 (information dynamic, ID)是指由观察结果来分析观测对象内在规律的学说, 因此它是因果关系中的反问题, 是动力学研究的一个基本途径. 为了区别, 把目前已在物理、化学中形成的一些动力学理论 (如分析力学、分子动力学、量子力学与量子化学等) 称为常规动力学.

在本书中, ID 的讨论对象是各种不同类型的生物信息 (主要是蛋白质) 数据库, 这些数据库提供了大量的、多种类型的、在原子与分子水平下的观测与测量结果. 它们不仅规模庞大, 而且还几乎以每 2~3 年翻一番的速度呈爆炸式的增长. 这些数据库为 ID 的发展与研究提供了有力的基础与条件.

蛋白质结构分析包括蛋白质一级结构、空间结构与蛋白质组的结构分析, 其中一些问题是经典问题, 尽管有许多研究成果, 但也有许多问题没有得到完全解决. 其中许多问题正是目前讨论中的新问题与热点问题, 与生物、医学与医药卫生领域关系密切, 对这些问题的研究都还处在刚刚起步阶段, 许多问题还远远没有得到解决. 进一步探讨这些问题, 寻找其中的规律与方法, 是本书的基本目标.

在信息论中, 信息的度量有多种形式, 已成为电子计算机与通信领域理论与技术发展的基础. 信息与动力学发生联系的问题一直受到物理学与哲学界的关注, 信息与能量相互转化的关系早已成为物理学中的基本定律, 说明熵与能量相互关系的玻耳兹曼 (Boltzmann) 公式与说明质能相互转化的爱因斯坦 (Einstein) 公式同样完美与重要. 玻耳兹曼公式反映了物质运动在从无序到有序转化过程中所需要的能量, 这种关系在生命科学中有特殊的意义. 生命过程就是通过吸收能量, 使无序的分子运动变成各种不同类型的有序的生物分子运动. 它们可以通过生物信息数据库的内在结构, 以多种动力学规则与各种不同类型的结构形式反映出来, 虽然在许多情况下还无法知道这种规律的形成过程与原因, 但它们确实存在, 并可以观察得到, 这是 ID 产生的基础与依据. 本书的研究重点是:

(1) 建立 ID 的内容体系与研究方法, 确定其理论依据与存在意义. 针对不同的问题, 采用不同的数学方法, 如信息统计法、组合分析法、几何计算或其他数学方法, 使 ID 有一个较为完整的可以计算实现的理论体系. 本书的重点是研究蛋白质与蛋白质组中的问题.

(2) 对蛋白质结构的研究由两部分内容组成, 即一级结构与空间结构. 把蛋白质一级结构数据库看做一个带有句号的以氨基酸为基本单位的文库, 这样就可利用 ID 的方法作语义分析. 所谓语义结构就是指词、词法、句与句法的结构分析, 这些问题的研究的总目标是对生命语言的解读.

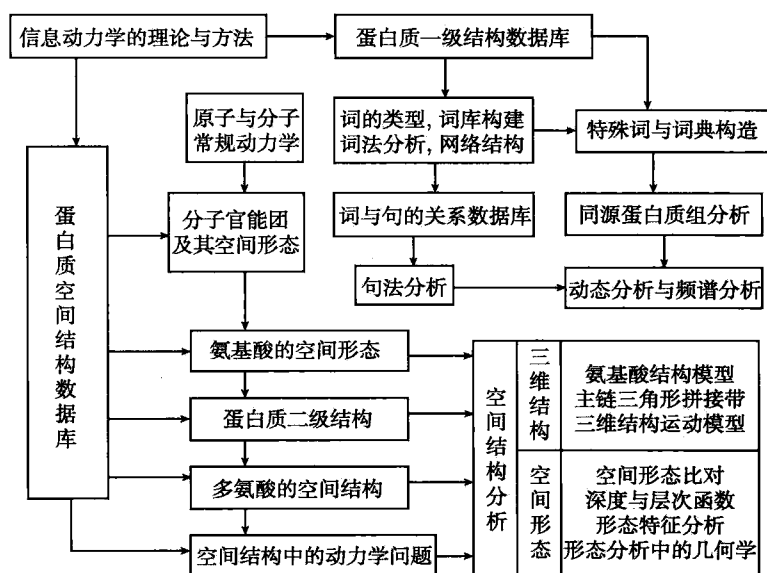
(3) 把蛋白质空间结构分为三维结构与空间形态结构两部分内容. 其中三维结构是以蛋白质主链为基础的空间结构, 而空间形态结构是蛋白质中各原子与分子在空间的分布结构特征. 对这两部分内容的研究目标相同, 但思路与方法不同, 因此所采用的几何模型与计算方法也不相同. 寻找其中的规律也是一种对生命语言的解读, 而且在生物、医学与医药卫生领域中有直接的应用意义.

本书由四部分组成. 沈世镒主著, 胡刚、王奎与高建召协助完成. 第一部分建立 ID 的一般理论与方法. 第二 ~ 四部分分别是蛋白质一级结构与空间结构研究. 附录是对物理、化学与生物学中的一些基本知识的介绍, 供不同专业的读者参考, 也为本书的其他各章直接引用. 本书涉及大量的计算, 这些结果规模较大, 已在光盘中给出, 书中的彩图也在光盘中. 该光盘中的各文件可作为正文的补充与参考.

本书系作者的研究成果, 部分思路与结果在文献 [124, 125] 中给出, 那里讨论了生物信息学中的几个具体问题, 如序列的比对与应用、蛋白质结构中的有关问题等, 本书则从更全面与深入的角度来考虑蛋白质结构分析中的问题. 有关思路与结果是探索性的, 由于所涉及问题的复杂性及作者水平所限, 不足与不妥之处在所难免. 所提出的观点、理论、方法与结果可供有关研究人员参考与讨论.

本书得到国家自然科学基金项目 (20836005) 与中国科学院科学出版基金的支持, 同时, 吴忠华老师为本书提供了一个英语文库, 特此表示感谢.

有关 ID 与蛋白质、蛋白质组结构的研究层次与本书内容的关系如下图所示.



目 录

《数学与现代科学技术丛书》序

前言

第 1 章 概论	1
1.1 常用符号	1
1.1.1 英文大、小写字母与希腊字母的表示	1
1.1.2 数学函数公式与符号的表示	4
1.1.3 一些重要的单位符号	5
1.2 主要内容与结果	6
1.2.1 ID 的理论要点	6
1.2.2 蛋白质一级结构的语义分析	7
1.2.3 蛋白质空间结构的 ID 分析	10
1.2.4 蛋白质空间结构的动力学问题	11
1.2.5 蛋白质的三维结构研究	13
1.2.6 蛋白质空间形态的研究	15
1.3 其他问题的说明	16
1.3.1 关于数据库与数据处理中的一些问题	16
1.3.2 对本书附带光盘的说明	17

第一部分 基本原理与方法

第 2 章 ID 与生物信息学	21
2.1 ID 概述	21
2.1.1 ID 的目的与意义	21
2.1.2 ID 与物理学	23
2.1.3 ID 的主要研究方法	25
2.1.4 ID 所存在的问题与注意事项	27
2.2 生物信息学简介	28
2.2.1 生物信息学的研究目标与特征	28
2.2.2 生物信息学的范畴与内容	29
2.2.3 后基因组研究中的一些问题	30
2.3 对光盘的说明	32
2.3.1 数据库的选择与预处理	32

2.3.2 对光盘中文件的说明	33
2.3.3 光盘阅读的注意事项	34
第 3 章 ID 的基本原理与方法	36
3.1 ID 的研究对象与数据库的类型与要素	36
3.1.1 数据库的类型与结构特征	36
3.1.2 数据库的基本要素	37
3.1.3 人类自然语言文字与生物信息语言的比较	40
3.2 信息统计分析法	42
3.2.1 IDF 的一般定义与性质	42
3.2.2 三种特殊类型的 IDF	43
3.2.3 由交互信息与条件概率产生的 IDF	45
3.2.4 IDF 的若干性质	47
3.2.5 局部词与局部词库	48
3.3 组合分析法与图论的概述	49
3.3.1 组合分析法的目的、内容与意义	49
3.3.2 有关图与树的基本知识	52
3.4 数据库的核心词与核心词库	54
3.4.1 有关数据结构的记号与说明	54
3.4.2 核心词的定义与性质	55
3.4.3 核心词库递推算法	57
第 4 章 语义分析概论	59
4.1 词与词法分析要点	59
4.1.1 词库的词法与句法分析的几个基本概念	59
4.1.2 有关词法分析的一些定义、记号与模型	60
4.1.3 词库中词的网络结构关系	62
4.1.4 有关极小树网图结构的一些定义	62
4.1.5 核心词的词法分析	64
4.2 词与句的结构关系与同源蛋白质组的类型分析	65
4.2.1 词与句的关系数据库	65
4.2.2 由关系数据库作有关语法问题的讨论	68
4.2.3 网络结构中的布尔代数	69
4.3 句或蛋白质的信息动力函数 (SIDF 或 PIDF)	70
4.3.1 引进 PIDF 的基本思想及其定义	70
4.3.2 PIDF 的随机分析概述	72
4.3.3 PIDF 的运动及其他类型的分析	75

第 5 章 英语文库词法与句法结构的 ID 分析	77
5.1 英语文库数据结构概述	77
5.1.1 英语文库的选择与说明	77
5.1.2 Ω_0 与 Ω_1 文库的基本状况	78
5.2 英语 Ω_0 文库词库搜索与分析	80
5.2.1 低阶词库搜索与分析	80
5.2.2 中、高阶词库搜索与分析	84
5.2.3 Ω_0 数据库的句法分析	86
5.3 英语文库 Ω_1 结构的信息统计分析法	89
5.3.1 Ω_1 文库的概况	89
5.3.2 IDF 阈值参数的选择与局部词的搜索	89
5.3.3 由 IDF 产生文库 Ω_1 中的局部词	90
5.3.4 由交互信息产生的 IDF 与高阶英语词的递推计算	96
5.3.5 关于标点符号的分析	99
5.4 数据库 Ω_1 结构的组合分析法	101
5.4.1 随机组合分析法的计算结果	101
5.4.2 数据库 Ω_1 中的词与核心词分析	102
5.4.3 由数据库 Ω_0 中的高阶词产生的核心词	105
5.5 英语数据库 Ω_1 的 M-PIDF 分析	107
5.5.1 Ω_1 数据库的 M-PIDF	107
5.5.2 M-PIDF 的定义与性质	107
5.5.3 利用 ID 对英语文库分析的综述	111

第二部分 蛋白质一级结构研究中的 ID 方法

第 6 章 蛋白质一级结构数据库的信息统计与组合分析	115
6.1 蛋白质数据库结构分析概论	115
6.1.1 蛋白质结构数据库与软件包	115
6.1.2 蛋白质结构分析概论	117
6.1.3 SP'06 数据库的一般性质	118
6.2 SP'06 数据库的信息统计分析	122
6.2.1 氨基酸的 IDF(AIDF)	122
6.2.2 由 SP'06 数据库产生局部词的词库	123
6.3 SP'06 数据库的组合分析	127
6.3.1 随机组合分析的计算结果	127
6.3.2 SP'06 数据库的核心词	128

6.3.3 核心词的收缩分析	130
第 7 章 蛋白质一级结构语义分析	133
7.1 SP'06 数据库词库的词法分析	133
7.1.1 词库的构建与词法分析要点	133
7.1.2 蛋白质中周期序列的分析	134
7.1.3 词库 $\mathcal{D}$ 的极小化与极大化网络结构	141
7.1.4 由短词组合所构成复合词	142
7.2 SP'06 数据库中词与句的关系分析	143
7.2.1 词与句的关系数据库的类型与记号	143
7.2.2 计算结果	144
7.2.3 词与句的“覆盖率”分析	146
7.3 同源蛋白质组的网络结构分析	147
7.3.1 对同源蛋白质组的搜索计算	148
7.3.2 同源蛋白质组的布尔网络结构	152
7.3.3 利用同源蛋白质组作人造蛋白质的构造设计	154
7.4 一些具有特殊结构的蛋白质	159
7.4.1 重要的小肽、寡肽与小蛋白	159
7.4.2 一些特殊分布类型的蛋白质	161
7.4.3 游程与周期序列所对应的蛋白质	164
第 8 章 蛋白质的 IDF 与 PIDF 分析	169
8.1 关于 PIDF 的计算与分析	169
8.1.1 关于 PIDF 记号的补充与计算结果	169
8.1.2 PIDF 的因子分析	173
8.2 关于 $\mathcal{K}(t)$ 与 $\mathcal{K}'(t)$ 阵列的分析	176
8.2.1 $\mathcal{K}(t)$ 与 $\mathcal{K}'(t)$ 阵列的统计计算	176
8.2.2 不同随机向量 ξ_t 与 $\bar{\xi}_t$ 的数据结构分析	182
8.2.3 关于互相关函数的计算的结果讨论	185
8.3 数据阵列 $\mathcal{K}_s$ 的运动分析	186
8.3.1 数据阵列 $\mathcal{K}_s$ 的运动方程	186
8.3.2 数据阵列 $\mathcal{K}_s$ 的运动特征	187
8.3.3 关于运动区域的性质	189
8.3.4 计算结果与结果分析	191
8.4 蛋白质 M-PIDF 的频谱分析	193
8.4.1 频谱分析概论	193
8.4.2 M-PIDF 的 Fourier 变换	195
8.4.3 不同长度蛋白质的频谱结构分析	197

8.4.4	蛋白质数据库的频谱分析	200
8.4.5	SP'06 数据库中蛋白质 M-PIDF 的频谱分析	200
8.5	M-PIDF 的其他分析问题	203
8.5.1	完全随机序列的选择与它们的 M-PIDF	203
8.5.2	完全随机序列 M-PIDF 的因子分解及其运动方程	206
8.5.3	蛋白质的判定问题	209
8.5.4	关于标点符号的讨论	211

第三部分 蛋白质的三维结构分析

第 9 章	四原子与多原子点的空间几何	215
9.1	四原子点的空间几何结构	215
9.1.1	空间四原子点的结构表示及其参数系	215
9.1.2	基本参数系的相互关系	217
9.1.3	空间四原子点的其他参数	219
9.2	四原子点一些特殊结构的计算	219
9.2.1	四原子点的位相分析	219
9.2.2	四原子点的平面展开	220
9.2.3	方程组 (9.2.3) 与 (9.2.4) 的求解计算	222
9.3	多原子点结构分析的几何理论	224
9.3.1	有关记号与类型	224
9.3.2	五原子点的参数系	226
9.3.3	若干特殊四原子或五原子点	227
第 10 章	氨基酸的一般性质及其分子官能团	230
10.1	氨基酸概论	230
10.1.1	氨基酸的化学成分与性质	230
10.1.2	氨基酸的相互连接	235
10.1.3	遗传密码子	236
10.2	氨基酸的分子结构特征	238
10.2.1	氨基酸的分子结构模型	238
10.2.2	氨基酸的空间形态特征	240
10.2.3	氨基酸侧链中非氢原子的组合形态分析	243
10.3	氨基酸中的氢原子与分子官能团	245
第 11 章	一般空间质点系研究中的几个理论问题	251
11.1	图论的补充知识	251
11.1.1	着色图的定义与类型	251

11.1.2 非氢原子空间结构的全着色图	253
11.2 空间结构的稳定性分析	254
11.2.1 稳定性的定义与类型	255
11.2.2 基本几何图形的组合	256
11.3 活动坐标系理论	259
11.3.1 活动坐标系的定义与性质	259
11.3.2 氨基酸的活动坐标系	260
第 12 章 氨基酸的空间结构分析	264
12.1 氨基酸中部分原子的结构分析	264
12.1.1 一般氨基酸原子的公共特征分析	264
12.1.2 甘氨酸中的原子结构	266
12.1.3 脯氨酸的原子结构	268
12.2 氨基酸侧链中非氢原子的空间结构	270
12.2.1 非氢原子空间结构的计算	271
12.2.2 按分子官能团稳定性结构模型计算	272
12.2.3 活动坐标系下第 2 层非氢原子点的运动状况	275
12.2.4 侧链中其他非氢原子点的扭动分析	277
12.3 氨基酸中分子官能团的运动	281
12.3.1 非氢原子绕侧链的运动问题	281
12.3.2 具有环形圈的侧链结构	282
12.3.3 带环形侧链的摆动与旋转	284
12.4 氢原子在氨基酸中的运动变化	286
12.4.1 基本分子官能团的表示与它们的形态	286
12.4.2 基本分子官能团的类型与表达	288
12.4.3 计算结果与分析	291
12.4.4 氨基酸空间结构分析小结	297
第 13 章 蛋白质主链的三角形拼接带	300
13.1 概论	300
13.1.1 蛋白质三维结构概述	300
13.1.2 主链三角形拼接带的类型与参数	301
13.1.3 三角形拼接带的有关性质	304
13.1.4 计算结果与分析	304
13.2 小三角形拼接带的结构分析	306
13.2.1 小三角形拼接带的基本特征	306
13.2.2 扭角分布与氨基酸的关系分析	309
13.3 三角形拼接带的其他性质	312

13.3.1 三角形拼接带的一些其他性质	312
13.3.2 三角形拼接带的平面展开	314
13.3.3 计算公式与计算结果	315
第 14 章 主链的中位点曲线分析	318
14.1 中位点曲线的定义与性质	318
14.1.1 中位点曲线的定义记号与一般性质	318
14.1.2 中位点曲线的有关参数的性质	319
14.2 关于中位点曲线的计算与分析	321
14.2.1 计算模型与结果	321
14.2.2 中位点曲线与蛋白质二级结构关系问题	323
14.2.3 关于 β 折叠结构的讨论	329
14.3 中位点曲线的平面展开	333
14.3.1 关于转角问题的讨论	333
14.3.2 中位点曲线的平面展开	335
14.3.3 中位点曲线的平面展开图的拟合与计算	337
第 15 章 部分原子的空间结构分析	342
15.1 二肽链中部分原子的空间结构	342
15.1.1 部分已知结论的回顾	342
15.1.2 关于 A, C, O, H', N', A' 六原子点的特性讨论	344
15.1.3 部分原子点位置的预测	346
15.1.4 二肽链双底座原子集合空间结构的讨论	348
15.1.5 三肽链中部分原子的计算	350
15.2 二肽链的中位点曲线计算	350
15.2.1 二肽链的中位转角的计算结果	350
15.2.2 计算结果的分析	357
15.3 三肽链的中位点曲线的计算分析	361
15.3.1 关于三肽链中位点曲线参数的定义与计算	361
15.3.2 三肽链相间双氨酸中位点曲线参数的分布计算	364
15.3.3 计算结果的分析	367
15.4 四肽链的部分分子官能团分析	372
15.4.1 四肽链的部分分子官能团的结构特征	372
15.4.2 四肽链的部分分子官能团的参数表示	373
15.4.3 四肽主链大三角形拼接带的空间形态特征分析	374
15.4.4 关于计算结果的分析	378
15.4.5 一些特殊的四肽链	380

第 16 章 蛋白质三维结构动力学问题	382
16.1 有关动力学中的一些概念与问题	382
16.1.1 作用力的类型与意义	382
16.1.2 蛋白质空间结构的形成过程	383
16.2 蛋白质三维结构的运动方程	385
16.2.1 决定蛋白质三维结构的参数系	385
16.2.2 蛋白质三维结构的运动方程	387
16.2.3 参数系 Ψ^* 的随机分析	388
16.2.4 计算结果与分析	390
16.3 自由能与收敛性	394
16.3.1 自由能的定义与性质	394
16.3.2 自由能的优化计算与收敛性讨论	395
第 17 章 三维结构中其他动力学问题的分析	398
17.1 氢键与二硫键的分析	398
17.1.1 氢键与二硫键的分析概论	398
17.1.2 计算结果	399
17.1.3 计算结果分析	400
17.1.4 远程折叠的结构类型	409
17.2 B 原子点结构的讨论	410
17.2.1 B 原子点与主链关系的结构模型	410
17.2.2 计算结果	412
17.2.3 中位点曲线上的 B 原子结构	414
17.2.4 五肽链计算分析中的重要参数	417
17.3 氨基酸侧链非氢原子的梢点	418
17.3.1 氨基酸中非氢原子的梢点	418
17.3.2 单氨基酸梢点的计算结果与分析	419
17.3.3 单氨基酸非氢原子梢点的极坐标计算与分析	421
17.3.4 二肽链的梢点距离计算与分析	424
17.4 以中位点曲线为基础的计算结果与分析	428
17.4.1 对 B 原子的一般计算结果与分析	428
17.4.2 计算结果的其他分析	430
17.4.3 关于 16 参数的因子分析	432
17.5 蛋白质三维结构中的其他 ID 指标	435
17.5.1 蛋白质中位点曲线的折叠系数	435
17.5.2 蛋白质一级结构与空间结构的比较	436
17.5.3 一些具有特殊成分的蛋白质	437

第四部分 蛋白质的空间形态

第 18 章 蛋白质空间结构的相似性分析	441
18.1 蛋白质结构的相似性问题	441
18.1.1 蛋白质一级结构序列的比对	441
18.1.2 蛋白质的空间形态概述	442
18.1.3 拓扑空间中的一些基本概念	443
18.1.4 蛋白质总体结构的参数指标	444
18.1.5 对凸闭包的形态分类	445
18.2 蛋白质空间形态的比对	447
18.2.1 蛋白质内部结构比对的基本思路与方法	447
18.2.2 离散化的蛋白质序列比对	449
18.3 计算结果与分析	450
18.3.1 对长度在 300 AA 左右的蛋白质的总体相似性分析	450
18.3.2 内部结构的相似性的比对结果与分析	452
18.3.3 长短蛋白质的比对与定位	458
18.4 其他实例计算	459
18.4.1 离散情形下的实例计算	459
18.4.2 同源蛋白质族的实例比对计算	463
18.4.3 其他同源蛋白质的实例比对计算	466
18.4.4 连续情形下的实例计算	467
第 19 章 空间质点系的深度理论	470
19.1 空间质点系的深度分析	470
19.1.1 深度分析概论	470
19.1.2 深度的定义与计算	472
19.2 深度的一些基本性质	474
19.2.1 T 深度的一般性质	474
19.2.2 L 深度的性质	475
19.3 深度的统计分析	476
19.3.1 T 深度与 L 深度的关系分析	477
19.3.2 氨基酸深度的倾向性因子计算	478
19.3.3 深度倾向性因子的其他分析	480
19.3.4 肽链在蛋白质中的深度预测	483
19.4 空间质点系的层次函数	485
19.4.1 层次函数的计算步骤与结果	486
19.4.2 计算结果分析	486

19.4.3 一般多面体的层次函数定义与计算	488
第 20 章 超图与晶格	489
20.1 超图的定义与性质	489
20.1.1 蛋白质空间形态与空间质点系	489
20.1.2 关于超图的定义与性质	491
20.1.3 一般多面体的超图表示与它的表示要素	493
20.1.4 对多面体的形态结构的分析	495
20.1.5 实例计算	496
20.2 晶格理论	497
20.2.1 晶格群与晶格的定义与类型	497
20.2.2 超晶格群的生成原理	499
第 21 章 蛋白质空间形态的特征分析	501
21.1 蛋白质的空间形态概论	501
21.1.1 蛋白质的空间形态中的复杂特征	501
21.1.2 复杂结构分析中的几个基本概念	502
21.1.3 几个基本计算关系	504
21.1.4 空间质点系的 γ 到达半径的定义与性质	506
21.2 由空球产生的网络结构	507
21.2.1 空球网络结构图	507
21.2.2 一些特殊的算法与它们的主要思路	509
21.2.3 实例计算	510
21.3 γ 半径计算中的几个几何问题	511
21.3.1 到达与穿越固定的三角形的定义与记号	512
21.3.2 关于 γ 到达半径计算中的基本定理	514
21.3.3 γ 函数的递推计算	516
21.3.4 实例计算	518
21.4 有关定理的证明与计算公式的推导	519
21.4.1 若干定理的证明	519
21.4.2 若干几何计算公式的推导	522
第 22 章 蛋白质空间结构的凹槽的计算	526
22.1 空间质点系的圆柱形凹槽与棺形凹槽的结构分析	526
22.1.1 概论	526
22.1.2 棺形凹槽的计算分析	528
22.2 不规则凹槽的计算	531
22.2.1 不规则凹槽的一般算法	531
22.2.2 步骤 22.2.2 与步骤 22.2.3 的补充计算	533

22.3 蛋白质三维结构与空间形态的综合研究	535
22.3.1 综合研究的基础与方法	536
22.3.2 对计算结果的综合分析	537
附录 A 原子、分子与分子生物官能团	540
A.1 原子与分子	540
A.1.1 原子结构	540
A.1.2 元素结构的周期律	542
A.1.3 分子中原子的相互作用	543
A.1.4 分子间的相互作用力	545
A.2 生物分子官能团概述	547
A.2.1 研究生物分子官能团的意义	547
A.2.2 分子官能团的构造原理与最基本的生物分子官能团	548
A.2.3 分子官能团的形态结构	549
A.2.4 生物分子官能团的构型的理论计算	552
A.3 分子官能团的相互作用	553
A.3.1 分子官能团的相互作用的基本形式与规则	553
A.3.2 分子官能团相互作用类型	554
A.4 水分子的动力学性质	555
A.4.1 水分子的形态特征	555
A.4.2 水分子与其他分子官能团的相互作用	557
A.4.3 有关水分子的动力学理论	557
A.5 蛋白质与肽链	561
A.5.1 蛋白质的合成与分解	563
A.5.2 关于蛋白质组学的一些讨论	564
A.5.3 几种重要的肽链	564
A.5.4 蛋白质的功能、分类与命名	564
A.5.5 几种重要蛋白质的简介	566
附录 B 光盘文件目录表及其说明	571
B.1 光盘中的数据文件	571
B.1.1 DATA0 文件包	571
B.1.2 DATA1 与 DATA2 文件包	571
B.2 对光盘阅读的注意事项	572
参考文献	573
索引	583

第1章 概 论

为使读者对本书内容与结果的全貌有一初步了解,本章先对有关方面作一说明,其中包括对一些常用符号的说明.

1.1 常用符号

由于本书的内容涉及多学科领域(数学、生物学与计算机科学,也与物理、化学中的一些问题有关),在不同学科中所采用的记号与名称往往不同,所以有必要先对它们作统一的表达与说明,尽量使这些定义与记号与原学科的表达保持一致.

1.1.1 英文大、小写字母与希腊字母的表示

在本书中,除了不同学科,即使在同一学科中,同一变量也会有不同的表达方式,同一符号也可表示不同的变量,因此务请读者注意有关符号的定义与说明,注意其中的内在含义与区别.

1. 英文大写字母的表示

在本书中,英文大写字母有4种不同的表达方式,分大写斜体、正体、黑体与花体,它们在不同的场合使用.

(1) 英文大写斜体字母,如 A, B, C, D, X, Y, Z 等表示集合、向量或空间中的点的一般记号.几个英文大写斜体字母的连写有其特殊含义,如 AB 为线段, $\overrightarrow{AB}$ 为有向线段或向量, $|AB|$ 为线段长度, R 为实数空间等.

(2) 英文大写正体字母,如 A, R, N, D 等表示氨基酸一字符, C, N, O, S 等为原子记号.几个英文大写正体字母的连写有其特殊含义,后面还会做特别的说明.

(3) 英文大写黑体字母,如 A, B, C, D 等表示矩阵.有时也表示向量的集合或集合的集合,如 $Z = \{Z_1, Z_2\}$ 表示由两个向量(或集合) Z_1, Z_2 所组成的集合.

(4) 英文大写花体字母,如 A, B, P, V 等表示集合的集合、数据库或参数系,一般表示由本书所设计产生的特殊类型的数据库.另外较为常用的记号如 $\mathcal{E}$ 表示坐标系, $\mathcal{P}$ 表示某种类型的参数系, $\mathbb{Z}, \mathbb{Z}_0, \mathbb{Z}_+$ 分别表示全体整数、全体非负整数与全体正整数.

2. 英文小写字母的表示

英文小写字母的类型也分小写斜体、正体、黑体三种,它们在不同的场合使用.