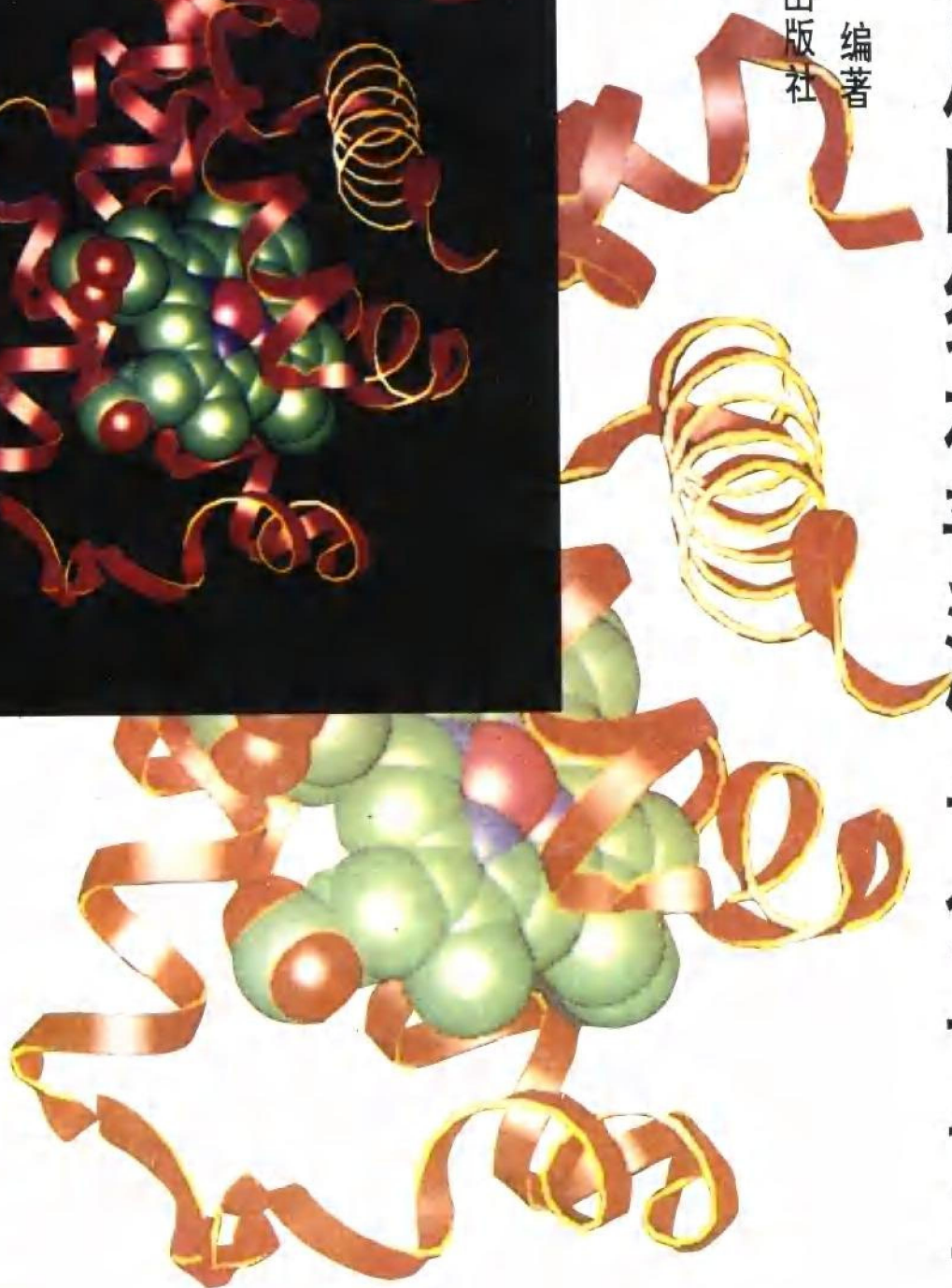
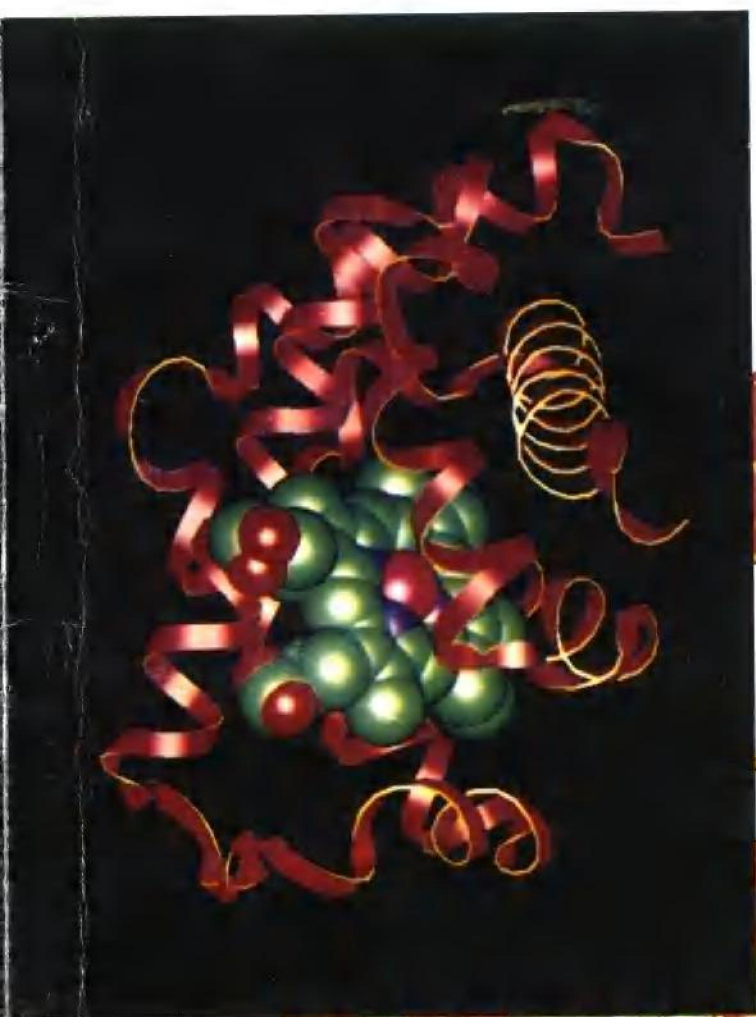


蛋白质的结构预测与分子设计

来鲁华等 编著

北京大学出版社



1
F10.1
11H

1
Q510.1
LLH

Y4128/08

蛋白质的结构预测与分子设计

来鲁华等 编著



A0098976

北京大学出版社

新登字（京）159号

图书在版编目(CIP)数据

蛋白质的结构预测与分子设计/来鲁华等编著. —北京
: 北京大学出版社, 1993. 12
ISBN 7-301-02406-1

I. 蛋…

Ⅰ. 来…

Ⅲ. ①蛋白质-生物化学 ②生物化学-蛋白

Ⅳ. Q510.1

出版者地址: 北京大学校内, 邮政编码: 100871

排印者: 北京大学印刷厂

发行者: 北京大学出版社

经销者: 新华书店

版本记录: 787×1092毫米 16开本 9.75印张 250千字

1993年12月第一版 1993年12月第一次印刷

印数: 0001—3,000册

定 价: 12.50元

内 容 简 介

蛋白质工程是生物技术的发展的新兴领域，其根本目的是要将自然界存在的蛋白质按造人们的设想进行改造，而这种改造的出发点是蛋白质的结构预测和分子设计。药物工业和发展要求能够更快更好地研制出新药，药物的作用对象多为蛋白质和核酸，也就使得药物分子设计成为与蛋白质结构预测和分子设计密切相关的研究领域。

本书的主要内容包括与生物技术及药物设计密切相关的蛋白质及药物的结构预测与分子设计的原理、方法和应用实例。作者以实际研究工作的经验为出发点，对大量有关文献和专著进行了分析、归纳和总结，除了对结构预测及分子设计中的常用方法及其原理进行了详细论述以外，还列举了不少实例，希望对于该领域的入门者有所启发。

本书可供从事蛋白质工程、药物设计及其它相关领域研究的科技人员参考，也可作为大学生物系、化学系及药学院的研究生和高年级本科生的参考用书。

序

本书论述关于蛋白质的结构预测与分子设计的原理、方法和应用。它代表一个方兴未艾的研究方向，也是一门正在发展中的学问。它将日益成为蛋白质工程和药物工程等分子工程中不可缺少的学科基础。

随着生物学之进入分子水平，人们越来越明确：蛋白质的功能是由其多肽分子折叠而成的高级结构所直接赋予的；蛋白质分子关于其折叠方式的信息又蕴藏在其氨基酸序列或一级结构中。实际上，蛋白质分子的合成，不论在体内或体外，都是从一级结构入手的；相应的DNA给出的也只是氨基酸序列。综上所述，这个领域的长远目标应该是从蛋白质的一级结构来预测高级结构以及从所需赋予蛋白质的功能出发来设计高级结构并进一步设计一个可据以合成的氨基酸序列。

无可讳言，我们现在离这个目标还相当遥远，但我们一定可以不断趋近这个目标，而且在这样的方向上要有任何进展都是最引人入胜的。更令人高兴的是我们这里这方面有不少志同道合的人。徐筱杰多年来一直带了研究生开展这方面的工作，成绩卓著。特别值得赞赏的是：他对新方向很敏锐，而且十分善于发挥大家的作用。本书主笔人来鲁华年富力强，学有专深，堪称后起之秀。我乐于作序如上。

唐有祺 于燕东园

1993年3月3日

前 言

自从 60 年代 Anfinsen 证明核糖核酸酶在体外条件下可以由失活状态自动折叠为活性状态以来,科学家们一直致力研究如何从蛋白质的一级结构预测其高级结构的问题。近年来随着分子生物学、X 射线晶体学以及计算机技术的发展,大量的蛋白质结构被测定,人们对蛋白质结构规律的认识逐步深入,使得蛋白质的结构预测与分子设计有了很大的进展,在酶的改性、免疫球蛋白及药物受体研究等方面取得了一系列的成功。由于蛋白质的结构预测及分子设计在生物技术及药物工业等方面的理论意义及潜在的经济效益,已经有越来越多的研究者开始涉足结构预测与分子设计领域。虽然在这个领域中尚存在众多的困难,目前的研究水平离预期目标还有相当的距离,但经过十几年的努力已经形成了不少行之有效的好方法,成功的蛋白质设计改造也已经有了实例。另外,合理药物设计的发展也使得人们更加重视药物受体蛋白质的结构预测与分子设计研究。作者所在的实验室在唐有祺教授的倡导下,由徐筱杰教授开始在 1986 年左右开展起来蛋白质结构预测和分子设计方面的工作。在“八六三”经费的支持下,经过几年来的努力做了一些有关的研究,对于结构预测和分子设计问题有了一些实际的体会。另外从 1991 年开始,我们实验室承担了攀登计划“生命过程中重要化学问题的研究”中全新蛋白质及功能域的设计、合成和结构研究的专题工作。在研究工作过程中迫切感到缺乏一本好的实用参考书,尤其对于刚刚涉及该领域的人更是困难。本书即是在我们的实际工作经验以及阅读大量有关文献和专著的基础上整理写成的,希望能够在分子设计日益得到广泛应用的今天对该领域的研究者及感兴趣者起到抛砖引玉的效果。

本书是在来鲁华的主持下由几个人分工写作而成的:其中第 1、2、9、11 章由来鲁华执笔,第 3、10 章由曲春旭执笔,第 4、5 章由罗宇执笔、第 6、8 章由邓巧临执笔,第 7 章由王彦力执笔,最后由来鲁华整理完成。由于这样一个新兴的领域里每天都会有新的进展,书中内容仅包括到 1991 年为止发表的文献。我们在书的最后列出了主要参考文献,以供感兴趣的读者进一步查阅。本书完稿时间较为仓促,加之作者水平有限,不足之处敬乞赐教。

唐有祺教授是作者的导师,他在很多年以前就已经提出了从蛋白质的一级结构预测高级结构的思想,在我们的研究过程及本书的写作过程中一直给与关心和指导,并且在百忙之中为本书作了序。徐筱杰教授给本书的写作以很大的鼓舞和帮助。北京大学化学系结构化学实验室的各位老师和研究生同学给予我们的科研工作及本书的写作以多方面的协助。北京大学出版社的赵学范老师作为本书的责任编辑做了大量认真细致的工作,使得本书得以顺利出版。在此谨表谢忱。

来鲁华 于燕园

1993 年 3 月 5 日

目 录

第1章 蛋白质结构预测的概况及发展历史

- 1.1 进行蛋白质结构预测和分子设计的必要性与可能性 (1)
- 1.2 蛋白质结构预测的发展过程 (1)
- 1.3 蛋白质结构预测的方法 (3)
- 1.4 蛋白质结构预测方法的应用条件与范围 (4)

第2章 蛋白质的结构原理

- 2.1 构成蛋白质结构的基本单元及构象限制 (6)
- 2.2 蛋白质的二级结构 (11)
- 2.3 不同氨基酸的构象特点 (20)
- 2.4 典型的蛋白质结构类型 (24)
- 2.5 蛋白质的功能与结构的关系 (31)

第3章 蛋白质的序列分析及蛋白质数据库

- 3.1 序列同源性分析 (32)
- 3.2 双重序列对比 (33)
- 3.3 多重序列对比 (37)
- 3.4 蛋白质进化研究 (40)
- 3.5 蛋白质数据库 (43)

第4章 蛋白质的二级结构预测

- 4.1 蛋白质二级结构预测概述 (49)
- 4.2 蛋白质二级结构预测的方法 (50)
- 4.3 二级结构预测存在的问题及发展前景 (60)

第5章 同源蛋白质结构预测

- 5.1 同源蛋白质结构预测的理论及实验基础 (61)
- 5.2 同源蛋白质结构预测的方法 (62)
- 5.3 应用实例 (67)
- 5.4 小结 (68)

第6章 理论计算方法在蛋白质结构预测中的应用

- 6.1 引言 (70)
- 6.2 力场 (71)
- 6.3 最小化算法 (73)
- 6.4 构象空间搜索方法 (75)
- 6.5 量子力学与分子力学相结合的方法 (79)
- 6.6 前景 (81)

第7章 蛋白质的静电性能研究

- 7.1 影响蛋白质静电性质的因素..... (82)
- 7.2 Poisson-Boltzmann 方程 (84)
- 7.3 Poisson-Boltzmann 方程的数值解法 (86)
- 7.4 Poisson-Boltzmann 方程的解析解——Tanford-Kirkwood 模型 (89)
- 7.5 TK 与 FDPB 方法的分析比较 (94)
- 7.6 应用举例..... (95)

第8章 蛋白质的分子设计

- 8.1 蛋白质分子设计的必要性 (100)
- 8.2 小改——少数残基的替换 (100)
- 8.3 中改——分子剪裁 (102)
- 8.4 大改——从头设计 (102)

第9章 基于蛋白质结构知识的合理药物分子设计

- 9.1 有关受体的结构知识极大地促进了药物分子设计的发展 (106)
- 9.2 合理药物设计概况 (107)
- 9.3 药物与受体分子的对接方法 (108)
- 9.4 基于受体结构模型的全新药设计 (108)
- 9.5 未知受体结构时的合理药物设计 (112)
- 9.6 HIV-1 蛋白水解酶抑制剂的合理设计 (114)

第10章 膜蛋白的结构预测

- 10.1 膜蛋白的结构特点..... (117)
- 10.2 膜蛋白的二级结构预测..... (118)
- 10.3 膜蛋白的三级结构预测..... (130)

第11章 北京大学蛋白质分子设计系统简介

- 11.1 研制蛋白质分子设计系统的必要性..... (132)
- 11.2 蛋白质分子设计系统设计描述..... (132)
- 11.3 蛋白质分子设计系统简介..... (133)

- 主要参考文献..... (141)

第1章 蛋白质结构预测的概况与发展历史

- 1.1 进行蛋白质结构预测和分子设计的必要性与可能性
- 1.2 蛋白质结构预测的发展过程
- 1.3 蛋白质结构预测的方法
- 1.4 蛋白质结构预测方法的应用条件与范围

1.1 进行蛋白质结构预测和分子设计的必要性与可能性

蛋白质所具有的功能在很大程度上取决于其空间结构,因此对蛋白质空间结构的研究在蛋白质工程中有着极其重要的意义。早在1959年C. B. Anfinsen在《进化的分子基础》一书中就写到“蛋白质化学家很自然地想到理解细胞行为最可能的方式在于研究蛋白质分子的结构与功能的关系”。X射线晶体学方法是至今为止研究蛋白质结构最有效的方法,所能达到的精度是任何其它方法所不能比拟的;它的缺点是蛋白质的晶体难以培养,晶体结构测定的周期较长。近年来发展起来的多维核磁共振方法可以直接测定蛋白质在溶液中的构象,但由于对样品的需要量大、纯度要求高,被测定的蛋白质的分子量一般不超过2万等,因此也受到很大限制。相比之下氨基酸序列已被测定的蛋白质数目由于DNA测序的实现而高速增长,生物工程的发展又迫切需要从结构上来认识所研究的对象,从而更有目的地对蛋白质进行改造。对蛋白质结构的研究是现代分子生物学的基础之一,可以说分子生物学所进行的一切研究都离不开结构信息做指导。因而迫切需要一种不依赖晶体培养,迅速、简便易行的测定蛋白质结构的方法。利用蛋白质的一级结构所提供的氨基酸序列信息来进行高级结构预测的方法,就是应这种需要而发展起来的。

实验表明,某些蛋白质在体外的一定条件下解聚失活后可以自动折叠而恢复其原有高级结构与活性,这也就意味着蛋白质的氨基酸序列及环境决定其三维构象。C. B. Anfinsen等人在1961年就用实验证实了这一点。虽然近年来对蛋白质在体内折叠过程的研究发现体内有“分子伴侣”参与其过程,但“分子伴侣”可能只是起到了稳定折叠中间体的作用,有关蛋白质高级结构的信息仍全部包含在其一级结构中。蛋白质的一级结构决定高级结构是进行蛋白质结构预测的理论基础。

1.2 蛋白质结构预测的发展过程

蛋白质按外形可以分为三大类:(1)球蛋白,(2)膜蛋白,(3)纤维蛋白。本章只拟介绍球蛋白的结构预测与分子设计情况,有关膜蛋白的结构将在第10章中单独加以介绍。蛋白质的结构预测是目前分子生物学研究中迫切需要解决的问题,已有许多科学家致力于这方面的工作。但由于蛋白质折叠及其结构本身的复杂性,至今尚未找到类似于DNA与蛋白质序列对应关系那样的密码子。目前的结构预测工作还是基于对已知蛋白质结构规律的总结、吸收、利用,即所谓的“基于结构知识的蛋白质结构预测”(knowledge-based protein modeling)。

球状蛋白质结构堆积紧密形成特定的空间结构,通常包含一个疏水核心。蛋白质的结构域由二级结构单元堆积在一起构成,另外还包括对蛋白质结构起稳定作用的二硫键。显然,为了研究蛋白质的结构,寻找进行结构预测的方法和途径,首先需要对蛋白质结构本身进行深入的研究,了解其结构规律和特点,然后很自然地就会想到蛋白质的天然构象是否处在“自由能最低状态”,或者在不考虑熵效应的情况下是否处在“能量最低状态”,这样就有可能利用计算化学中的手段对蛋白质的构象进行研究。如果只用理论计算不成功的话,可以将经验规则与之结合来进行研究。实际上蛋白质的结构预测研究基本上就是按照这样一个过程发展起来的。蛋白质结构预测研究的发展大体可分为3个阶段。

第一阶段属于对蛋白质空间结构进行消化的阶段。随着蛋白质晶体结构数据的积累,对蛋白质的空间结构进行了大量的研究和分析,得出了许多经验规律。例如,蛋白质的结构可分为4类: α , β , α/β , $\alpha+\beta$ 。典型 α 蛋白的结构有 α 螺旋束等;典型 β 蛋白质的结构由2个 β 折叠片构成,疏水残基侧链伸向2个 β 片的内部而构成疏水性核心; α/β 类蛋白的结构则是由2个平行的 β 折叠与1个 α 螺旋构成的结构单元而形成的; β 转角(β -turn)可以分为I, I', II, II', IV等类型。这就使得人们对蛋白质的高级结构有了较为深入的认识。这方面的工作在蛋白质晶体学发展的初期就开始了,现在仍在继续。随着新蛋白质晶体结构的测定以及对原有数据更深入的分析,不断有新的结构类型被发现,对蛋白质结构规律的认识也将更加深入。同时也有人对已有的晶体结构数据进行了统计处理,发展了从蛋白质的一级结构预测其二级结构的方法。在以后的20多年中,关于二级结构预测的方法又有了不少发展,但总的来说准确度均不超过65%,成为蛋白质空间结构预测的一个制约条件。

第二阶段主要是试图利用理论计算的方法得到蛋白质的高级结构模型。分子力学方法在处理小分子体系上的成功运用,激励着人们开始将其应用到蛋白质及核酸体系中,所根据的基本原理是蛋白质的活性构象对应于体系自由能最小的状态,在某种条件下也可认为是能量最小的状态。但由于蛋白质体系庞大,可变参数繁多,而数学处理上的整体极小问题难以解决,所以无论是能量优化方法还是随后发展起来的分子动力学方法都无法从理论上解决使蛋白质折叠成为正确构象的问题,只能进行在已有大体正确的结构参数后的局部结构优化或分子动力学模拟。就目前的情况来看,利用理论计算方法直接从一级结构得出蛋白质的空间结构的努力还有很大的困难。

在理论上遇到了一时难以跨越的障碍后,人们又回过头来在已有的晶体结构数据中寻找规律。近年来新发展起来的一些方法以及构建三级结构的方法都基于实验数据的积累。例如,二级结构预测的模式匹配方法就是在分类总结了已知结构蛋白质的结构规律的基础上而建立的。最近发展起来的利用神经网络算法进行二级结构预测的方法也是先对已有的结构进行学习,然后在此基础上进行预测。虽然仅靠一级结构的氨基酸序列信息的从头预测方法目前还存在着许多困难,但借助于一些其它信息还是可以成功地建立起立体结构模型的。这方面最为成功的一个例子就是T. Blundell等发展的利用同源蛋白质的结构进行结构预测的方法,其原理是同一家族中的蛋白质结构和功能类似,利用同族中已知的蛋白质结构就可以建立起未知结构蛋白质的模型。

近年来在蛋白质的结构预测方面又有了不少新的进展。蛋白质结构预测准确度不高的原因一般认为有两条:一是现有预测方法无法考虑氨基酸残基间的长程相互作用;二是已知的蛋白质晶体结构数据还不够多。Rooman及Wodak 1988年通过对蛋白质晶体结构数据库中

重复序列片段的结构进行分析后预言,独立的蛋白质晶体结构数据达到 1500 套之日就是蛋白质结构预测问题得以解决之时。按照当时的结构测定速度,几十年内恐怕难以达到。但由于分子生物学的飞速发展,所提供的可用于晶体结构培养的蛋白质样品越来越多,再加上晶体测定方法本身的发展,使得近 4 年来蛋白质晶体结构测定的速度大为加快。预计到 1993 年 8 月,蛋白质晶体结构数据库中可收集到 2300 套坐标,独立的坐标套大约在 200 套左右。但 1500 套晶体结构数据是否足以解决蛋白质结构预测中的全部问题,也还是一个迷。既然无法考虑长程相互作用是提高预测准确度的一大障碍,许多研究者就设法绕过这一卡脖子之处,直接将序列与三级结构相联系,办法之一就是寻找何种序列之间能够形成类似的空间结构折叠方式。这样对于一个新近测定的蛋白序列,即使找不到已经测定过立体结构的同源蛋白质,如果能够判断出其与何种蛋白质具有同种空间折叠方式,也就可以预测出其高级结构了。D. Eisenberg 小组利用三维参数表 (3D-profile) 法来处理这个问题,证实了两个低同源性的蛋白质具有类似的折叠方式。在二级结构预测中,利用与序列对比相结合的方法可以大为改善预测结果。Benner 及 Gerloff 利用已知多个蛋白激酶序列的优势,采用 79 个序列进行二级结构预测,与随后发表的蛋白激酶催化结构域的晶体结构吻合得相当好。蛋白激酶催化结构域二级结构的成功预测给了人们以极大鼓舞,也从一个侧面反映了数据积累的重要性。

在理论计算方面也发展了种种设法克服局部极小值问题的方法,例如分子动力学模拟退火及蒙特卡洛模拟退火方法在蛋白质结构预测方面的应用。1992 年 Desmet 在 *Nature* 上发表文章,介绍了“死路消除法”(dead-end-elimination)在蛋白质侧链堆积计算中的成功应用。虽然目前该方法还只限于同源蛋白质结构预测中侧链构象的确定问题,但由于此算法在广泛存在的多重极小值问题中的可能应用,引起了人们的极大关注。另外随着计算机技术的飞速发展,大容量、高速度计算机的出现和更新以及计算机软件设计技术的发展,也为以前无法实现的大计算量的结构预测及模拟提供了可能。

虽然到目前为止蛋白质结构预测中还存在许多问题,但从该领域蓬勃发展的情况来看还是有希望在不远的将来得以解决的。

1.3 蛋白质结构预测的方法

蛋白质结构预测的目的是利用已知的一级序列来构建出蛋白质的立体结构模型,从而在此基础上进行结构与功能的研究以及蛋白质分子设计工作。一般的结构预测可以分为:序列分析,二级结构预测,三级结构预测,分子力学及分子动力学模拟等步骤。

对蛋白质进行结构预测需要具体问题具体分析,在不同的已知条件下对于不同的蛋白质采取不同的策略。有关常用的蛋白质结构预测方法见图 1-1。

一般在给定蛋白质一级序列后,首先要做的是进行序列对比,在蛋白质序列库或晶体结构数据库中寻找与之同源的蛋白质。如果找出的同源蛋白质已经有了晶体结构,就可以利用同源蛋白质结构预测的方法构建出该蛋白质的结构模型。如果没有已知的同源蛋白质结构,经典的作法是先进行二级结构预测、超二级结构预测,再进行三级结构预测,由于二级结构预测准确度的限制,使得这种方法在多数情况下(已知很少的实验信息)难以完全成功,只能给出一些有用的参考信息。目前正在发展根据蛋白质一级序列判断其折叠模式的预测方法,该方法避开了二级结构预测,试图直接从一维结构到三维结构,是一种很有前途的发展方向。

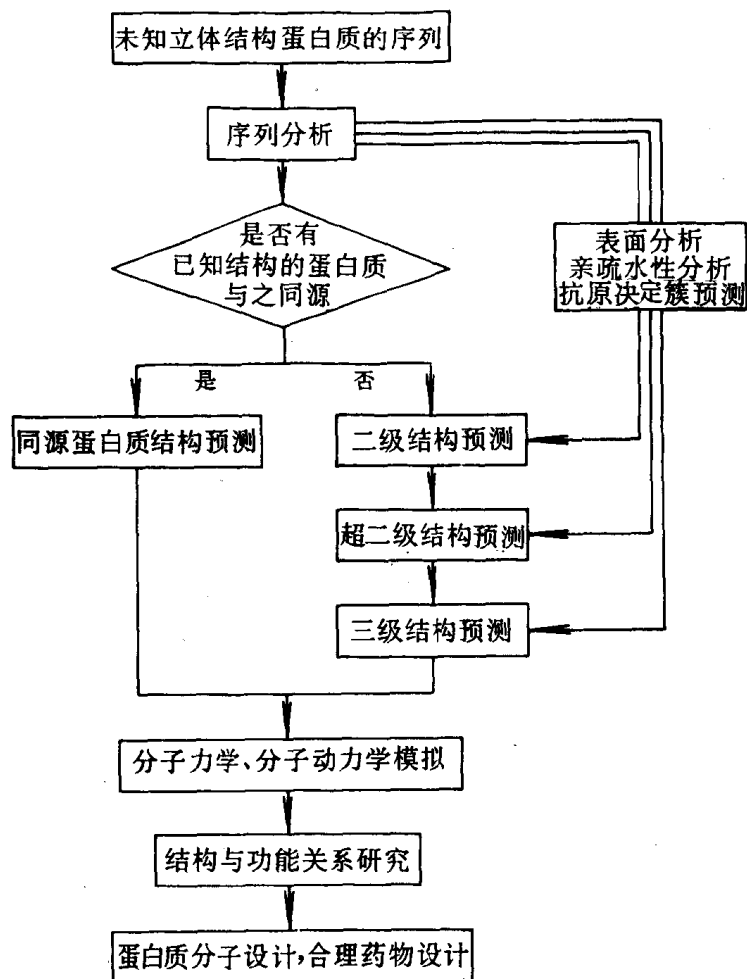


图 1-1 蛋白质结构预测方法示意图

1.4 蛋白质结构预测方法的应用条件与范围

目前的研究工作尚未达到直接从蛋白质的序列就可以预测出其高级结构的水平,还有待于进一步的深入研究,但在某些条件下已经可以开展有关的结构预测工作。可以进行的工作大体上分为以下几类。

(1) 在已知结构(如晶体结构)的基础上利用分子动力学方法研究蛋白质分子、核酸分子或复合物的动态性质。

(2) 利用能量优化或分子动力学方法对结构模型进行优化(以消除某些不合理之处),或者是在整体结构已知的情况下建立突变体的结构。

(3) 借助于同源物的结构参数建立未知蛋白质的结构,或利用序列模式识别方法判断可能的折叠方式。

(4) 结合多维核磁共振数据或 X 射线晶体粗结构数据或其它的实验数据建立蛋白质的整

体结构模型。

(5) 在以上 4 个条件均不符合的情况下,根据一级结构进行二级结构预测。

(6) 在已知二级结构的基础上(二级结构预测的准确度是目前一个难以逾越的障碍)进行片段堆积,根据从已知结构得出的一些规则进行筛选,挑选出可能的结构。往往得到多种可能性,尚不能得到唯一的正确堆积方式。

(7) 抗原决定簇预测。

(8) 建立复合物或多聚体的结构(在单体结构已知的情况下),参考复合物或聚合物本身的性质,利用几何匹配和最大接触面积规则进行。

由于蛋白质的结构原理是研究蛋白质结构与功能关系、进行结构预测及分子设计的基础,我们在以下的各章中将首先介绍蛋白质的结构原理,然后依次介绍序列分析,二级结构预测,同源蛋白质结构预测及分子力学、分子动力学等理论计算方法在蛋白质研究中的应用。由于蛋白质三级结构预测还远不成熟,在本书中只拟介绍有关同源蛋白质结构预测的内容。对于由二级结构片段堆积来预测三级结构的方法,由于方法本身还存在许多问题,成功率很低,本书中暂不作介绍,对此有兴趣的读者可以阅读有关的参考文献。在有关蛋白质结构预测的内容之后,我们将介绍研究蛋白质结构与性能关系的常用方法,蛋白质的分子设计以及在蛋白质结构基础之上的合理药物设计。由于膜蛋白的特殊性,我们专门设立了一章来讨论膜蛋白的结构预测问题。最后还将简要介绍作者所在实验室发展的北京大学蛋白质分子设计系统。

第2章 蛋白质的结构原理

2.1 构成蛋白质结构的基本单元及构象限制

2.1.1 构成蛋白质一级结构的 20 种标准氨基酸

2.1.2 蛋白质的多肽链状结构

2.1.3 多肽链间的共价连接：二硫键

2.1.4 蛋白质主链二面角的构象限制

2.1.5 影响蛋白质空间结构的主要因素

2.1.6 蛋白质的分级结构

2.2 蛋白质的二级结构

2.2.1 α 螺旋

2.2.2 β 折叠片

2.2.3 非重复性结构：转角、连接及紧密环

2.3 不同氨基酸的构象特点

2.3.1 甘氨酸

2.3.2 脯氨酸

2.3.3 半胱氨酸

2.3.4 丙氨酸

2.3.5 缬氨酸、异亮氨酸、亮氨酸和蛋氨酸

2.3.6 丝氨酸与苏氨酸

2.3.7 天冬酰胺与谷酰胺

2.3.8 天冬氨酸与谷氨酸

2.3.9 赖氨酸与精氨酸

2.3.10 组氨酸、苯丙氨酸、酪氨酸及色氨酸

2.4 典型的蛋白质结构类型

2.4.1 α 蛋白结构域

2.4.2 β 蛋白结构域

2.4.3 α/β 蛋白结构域

2.4.4 无规则的小蛋白

2.5 蛋白质的功能与结构的关系

2.1 构成蛋白质结构的基本单元及构象限制

2.1.1 构成蛋白质一级结构的 20 种标准氨基酸

天然蛋白质主要由 20 种标准氨基酸构成，这些氨基酸的不同是由相应的遗传密码子所决定的。20 种天然的氨基酸均属于 L 型的 α 氨基酸，中心碳原子的手性为 L 型，分别与 NH_2 、 COOH 、R 基团及 1 个 H 原子相连接；不同氨基酸的差异只是由于 R 基团的不同引起的。 α 碳原子的手性可按图 2-1 方式判断：当 H 原子垂直于纸面向外时，其余基团按 CORN 顺时针排列时为 L 型，反时针排列时为 R 型。 α 碳原子的手性对于蛋白质的各种结构形式起到了重要

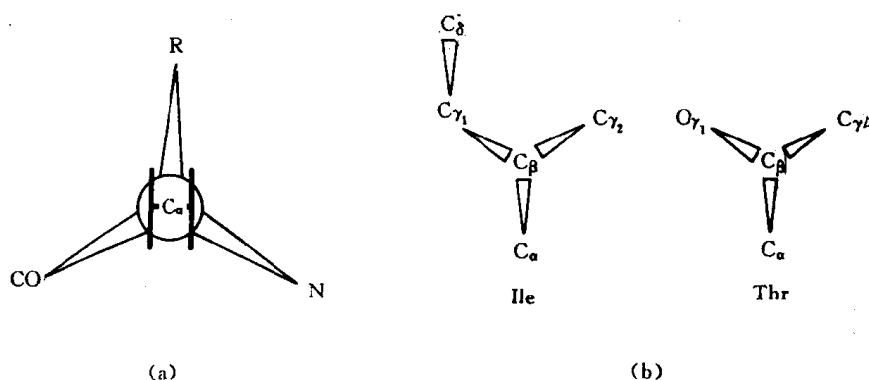


图 2-1 20 种标准氨基酸中手性碳原子的示意图

(a) α 碳原子的手性, 沿 H-C α 键俯视

(b) Ile、Thr 中 β 碳原子的手性, 沿 H-C β 键俯视

的决定作用。进化过程中自然界为何选择了 L 型氨基酸是宇宙形成及生物进化研究中一个重要的课题, 目前的研究趋向于用物理学中的弱相互作用理论来解释。

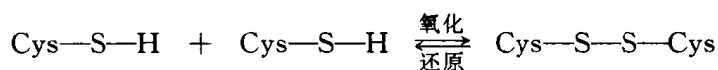
20 种标准氨基酸的不同侧链及氨基酸英文代码见图 2-2。氨基酸代码分三字母代码及单字母代码, 文献中常用的为单字母代码。由于氨基酸间的差异只是由于侧链不同引起的, 不同的侧链基团决定了不同氨基酸性质的差异, 因而对于单个氨基酸侧链性质的研究将大大有助于揭示蛋白质结构的规律性。按照侧链化学性质的不同, 氨基酸可以分为 3 类。第一类为疏水性的氨基酸, 包括: Ala, Val, Leu, Ile, Phe, Pro, Met; 第二类为带电氨基酸, 包括: Arg, Lys, Asp, Glu; 第三类为极性氨基酸, 包括: Ser, Thr, Cys, Asn, Gln, His, Tyr, Trp。甘氨酸的侧链只有 1 个氢原子, 是最简单的氨基酸, 具有独特的性质, 可以单独作为第四类, 也可以归入第一类。氨基酸中含有芳香性侧链的有: Phe, Tyr, Trp, His; 侧链为醇或酚的有: Ser, Thr, Tyr; 可以形成氢键的有: Arg, Lys, Asp, Glu, Ser, Thr, Asn, Gln, His, Tyr, Trp。最大的氨基酸为 Trp; 唯一的一个二级氨基酸 (N 上只有 1 个氢原子) 为脯氨酸。 β 碳原子上带有支链的氨基酸有 Val、Ile 及 Thr, 其中 Ile、Thr 的 β 碳子具有手性。按照有机化学中的命名法 Ile 中 β 碳原子的构型为 S, Thr 中为 R。实际上当垂直于 C α -C β 画 Newman 投影图时, Ile β 碳上的长侧链 (乙基) 与 Thr C β 上的 OH 位于同侧, 见图 2-1。

2.1.2 蛋白质的多肽链状结构

蛋白质是由氨基酸头尾以酰胺键相联而形成的多肽链构成的, 由于酰胺键中 C=O 的轨道与 N 原子的孤对电子存在共轭作用, 使得 C—N 键有部分双键性质, 以 C—N 键为中心构成一个酰胺键平面, 不能自由旋转。酰胺键通常为反式结构, 只有当酰胺键上的氮原子由脯氨酸供给时才可能出现顺式酰胺键。酰胺键平面的标准几何参数见图 2-3。

2.1.3 多肽链间的共价连接: 二硫键

蛋白质中除以酰胺键形式构成的主链共价连接以外, 同一多肽链或不同多肽链间还有以二硫键形式构成的侧链共价交联: 2 个在空间上相近的半胱氨酸被氧化时形成二硫键。



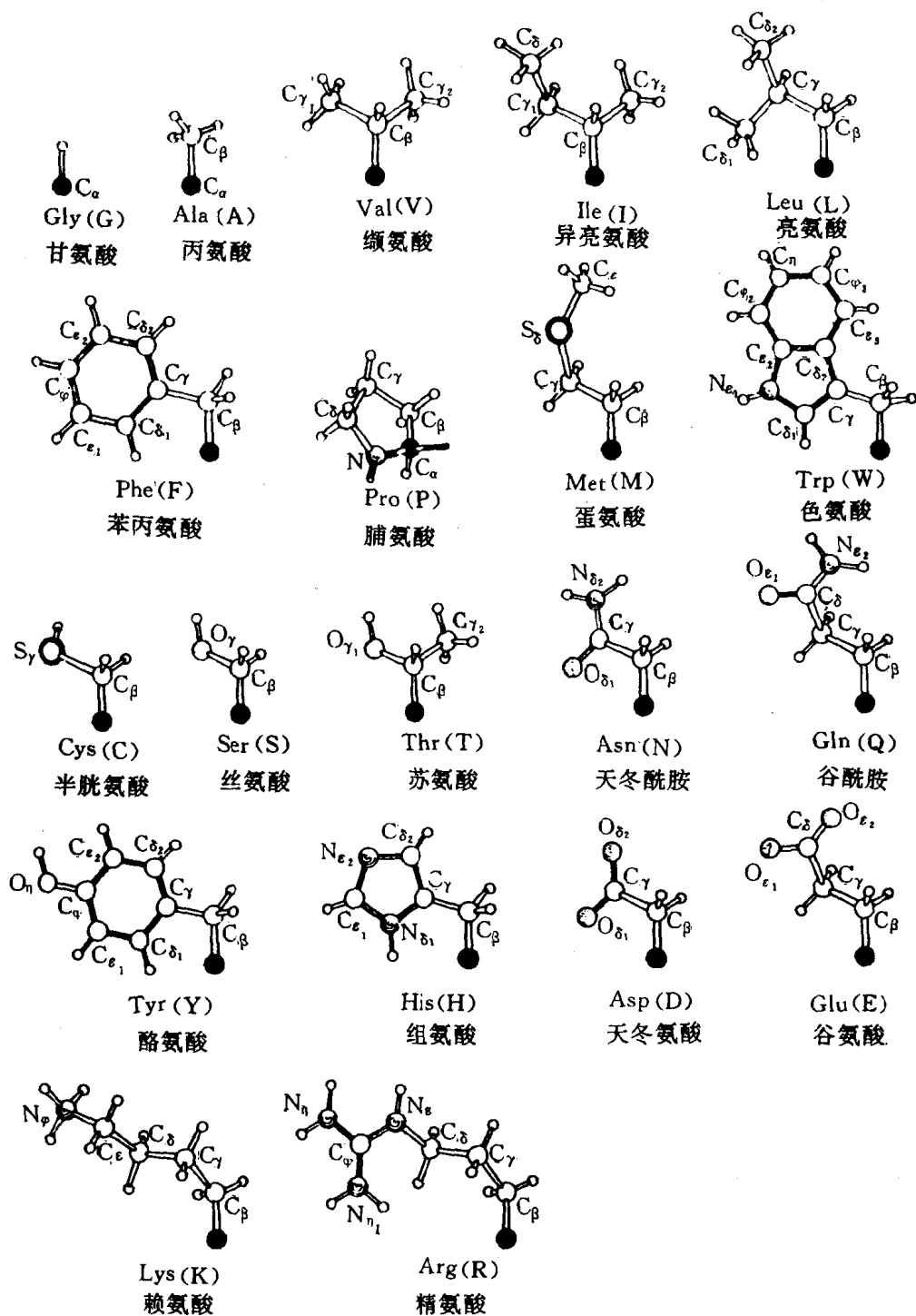


图 2-2 20 种标准氨基酸的侧链及英文代码

其中脯氨酸包括了部分主链，全黑的为 α 碳原子

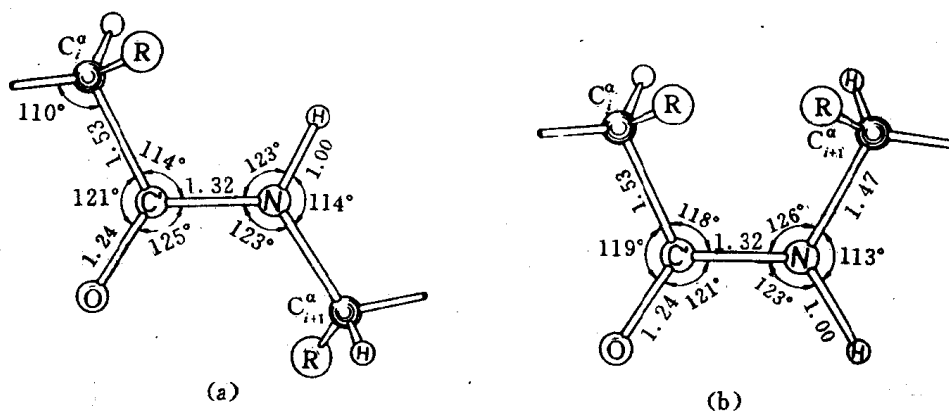


图 2-3 酰胺键平面的标准几何参数

(a) 常见的反式酰胺键 (b) 顺式酰胺键

由于二硫键形成所需要的氧化性环境，细胞内的蛋白质一般没有二硫键，而胞外蛋白则往往存在二硫键。二硫键对于蛋白质的结构可以起到稳定作用。蛋白质工程研究中常常试图引入新的二硫键以增加蛋白质的稳定性。 $C_\beta-S-S-C_\beta$ 二面角有一定的构象范围，通常在 $\pm 90^\circ$ 附近，在绘图及计算机图像显示中硫原子及二硫键常以黄色为标记。

2.1.4 蛋白质主链二面角的构象限制

多肽链由各个相邻的酰胺平面构成，酰胺键平面是刚性的，但酰胺键平面间的相对位置可以任意取向。主链中有 3 类不同的二面角（见图 2-4），分别为 ϕ （二面角 $C-N-C_\alpha-C$ ），

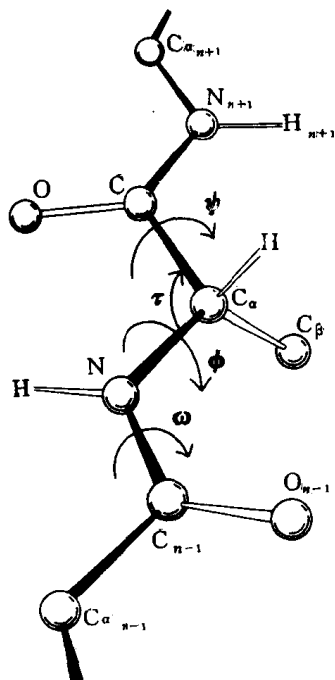


图 2-4 多肽链中原子及二面角的命名示意图

ψ （二面角 $N-C_\alpha-C-N$ ）及 ω 角（二面角 $C_\alpha-C-N-C_\alpha$ ）。侧链的二面角则称之为 χ_1, χ_2 等。

按照惯例将二面角的取值范围定义在 -180° 到 $+180^\circ$ 之间。对于一个由 A、B、C、D 原子相连构成的二面角，沿中心键由 B 往 C 看时（反向观察结果相同），A、D 的夹角为该二面角的数值，顺时针方向为正，见图 2-5。

由于酰胺键平面性的限制， ω 角只能

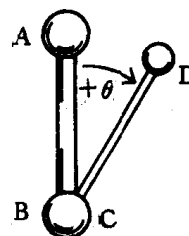


图2-5 二面角的定义
沿 B—C 键俯视