



A Guide

to Protein Structure Prediction

蛋白质结构预测 实验指南

岳俊杰 冯华 梁龙 主编

10. 1-33
18



化学工业出版社

Q510.1-33
Y948

生物实验室系列

30

蛋白质结构预测实验指南

岳俊杰 冯 华 梁 龙 主编

Q510.1-33

Y948



化学工业出版社

· 北京 ·

图书在版编目(CIP)数据

蛋白质结构预测实验指南 / 岳俊杰, 冯华, 梁龙主编.
—北京: 化学工业出版社, 2010.4
(生物实验室系列)
ISBN 978-7-122-07755-4

I. 蛋… II. ①岳… ②冯… ③梁… III. 蛋白质—生物结构—实验—指南 IV. Q510.1-33

中国版本图书馆 CIP 数据核字 (2010) 第 024203 号

责任编辑: 傅四周
责任校对: 陈 静

文字编辑: 周 佩
装帧设计: 关 飞



出版发行: 化学工业出版社 (北京市东城区青年湖南街 13 号 邮政编码 100011)

印 装: 大厂聚鑫印刷有限责任公司

720mm×1000mm 1/16 印张 8½ 字数 188 千字 2010 年 4 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888 (传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书, 如有缺损质量问题, 本社销售中心负责调换。

定 价: 38.00 元

版权所有 违者必究

编者名单

主 编：岳俊杰 冯 华 梁 龙

编 者：（按姓氏汉语拼音排序）

冯 华	创腾科技有限公司
郝格非	华中师范大学化学学院
李北平	军事医学科学院生物工程研究所
梁 龙	军事医学科学院生物工程研究所
刘 健	苏州大学材料与化学化工学部
刘 明	嘉和生物药业有限公司
吕 炜	创腾科技有限公司
田 然	创腾科技有限公司
徐 涛	创腾科技有限公司
岳俊杰	军事医学科学院生物工程研究所

出版者的话

21 世纪是生命科学的世纪，这已成为人们的共识。

生命科学随着人类对自身和自然的认识、探索而萌芽，随着人类生产和科学实践的进步而发展。现代生命科学包括生物学、医学、农学等传统学科领域以及生物学、生物技术与环境科学乃至社会科学等其他学科相互渗透、交叉而产生的新型学科体系。20 世纪后叶现代生物科学尤其是分子生物学取得了一系列突破性成就，使得生命科学在自然科学体系中的位置发生了革命性的变化，成为 21 世纪的带头学科。人们对生命科学也寄予了无限的期望，希望能够解决人类社会所面临的人口膨胀、资源匮乏、疾病危害、环境污染和生态破坏等一系列重大问题。

回顾生命科学的发展历程，实验技术一直起着非常重要的促进作用。如 17 世纪 Leeuwenhoek 等人发明并应用显微镜技术，直接催生了“细胞学说”的建立和发展；1973 年 Cohn 和 Boyer 完成了 DNA 体外重组实验，标志着基因工程的肇始；1988 年 Kary Mullis 发明的 PCR 技术甚至使生命科学产生了飞跃性的发展。可以说，生命科学无时无刻离不开实验，实验是开启神奇的生命王国大门的钥匙。没有实验技术的不断进步，也就没有生命科学今天的巨大发展；同时，生命科学的发展又对实验技术提出了更高的要求，进一步刺激了后者的不断进步。生命科学正是在“实验催生和验证着基础理论，理论指导和发展了实验技术”的不断循环中从必然王国走向自由王国。

工欲善其事，必先利其器。为了有助于生命科学工作者更多地了解相关实验技术和仪器设备，更好地设计实验方案，更有效地开展实验过程，更合理地处理实验结果，化学工业出版社组织出版了“生物实验室系列”图书。系列图书在整体规划的基础上，本着“经典、前沿、实用，理论与技术并重”的原则组织编写，分批出版。

在题材上，系列图书涵盖综合实验技术和单项实验技术两个方面。其中综合实验技术既有以实验目的为题，如“蛋白质化学分析技术”，内容纵向覆盖多项实验技术；也有以某一生命学科领域的综合实验技术为题，如“发酵工程实验技术”、“生物化学实验技术”等。而单项实验技术则以深入介绍某一专项技术及其应用为主，在阐述其基本原理的基础上，横向介绍该项技术在多个领域的应用，如“双向电泳技术”、“流式细胞术”等。

在内容上，系列图书主要有以下两个显著特点。一是强调先进性——除了系统介绍常用和经典实验技术以外，特别突出了当前该领域实验手段的新理论、新技术、新发展，为国内专业人员起到借鉴和引导作用。二是强调可操作性——对于每一项实验技术，系统介绍其原理方法、设备仪器和实验过程，让读者明了实验的目的、方案设计以及具体步骤和结果处理，以期起到实验指南的作用。

本系列图书坚持质量为先，开拓国内和国际两个出版资源。一方面，约请国内相关领域兼具理论造诣和丰富实验室工作经验的专家学者编著；另一方面，时刻关注国际生命科学前沿领域和先进技术的进展，及时引进（翻译或影印）国外知名出版社的

权威力作。

《生物实验室系列》图书的读者对象设定为国内从事生命科学及生物技术和相关领域（如医学、药学、农学）的专业研究人员，企业或公司的生产、研发、管理技术人员，以及高校相关专业的教师、研究生等。

我们殷切希望《生物实验室系列》图书的出版能够服务于我国生命科学的发展需要，同时热忱欢迎从事和关心生命科学的广大科技人员不仅能对已出版图书提供宝贵意见和建议，也能对系列图书的后续题目设计贡献良策或推荐作者，以便我们能够集思广益，将这一系列图书沿着可持续发展的方向不断丰富品种，推陈出新。

谨向所有关心和热爱生命科学，为生命科学的发展孜孜以求的科学工作者致以崇高的敬意！

祝愿我国的科技事业如生命之树根深叶茂，欣欣向荣！

化学工业出版社
生物·医药出版分社

前 言

蛋白质是一切生物借以表现生命的最重要的基本单元，是自然界存在的小型自动机器。任何一个生命体的繁衍、新陈代谢、运动等，都需要数十亿个蛋白质分子的协调行动才能得以顺利进行。随着人类功能基因组研究的展开，科学家对于基因的研究焦点，已由基因测序转移到基因表达产物——蛋白质上。为了执行特定的生物功能，每一个蛋白质分子都有一个独特的三维结构。研究蛋白质的结构，有助于了解蛋白质的功能，了解蛋白质的作用机制以及了解蛋白质与其他分子之间的相互作用。对于全新或功能未知的蛋白质分子，通过结构分析，可以进行功能诠释，指导设计生物学实验来进行功能研究。通过分析蛋白质的结构，确认功能位点，为设计新的蛋白质或改造蛋白质提供可靠的依据。到目前为止，确定蛋白质三维结构的方法主要分为两大类，其一是利用实验的方法来测定；其二则是利用计算机技术，根据现有理论和已知的序列等信息进行蛋白质的结构预测。测定蛋白质三维结构的方法分为三大类：（1）X射线晶体衍射图谱法，（2）核磁共振方法，（3）电镜方法。

近年来分子生物学的快速发展使蛋白质序列数据库的数据量得到快速增长。而利用实验方法测定蛋白质三维结构则由于耗时长、成本高以及方法在技术上的困难，而使解析的蛋白质结构比已知的蛋白质序列要少得多。另一方面，随着DNA测序技术的发展，人类基因组及更多的模式生物基因组将被完全测序，DNA序列数量将会急增，由于DNA序列分析技术和基因识别方法的进步，从DNA可以很方便地推导出大量的蛋白质序列。这意味着已知序列的蛋白质数量和已测定结构的蛋白质数量的差距将会越来越大。因此通过实验方法获得蛋白质三维结构越来越无法满足人们对蛋白质结构和功能关系研究的要求，迫切需要一种不依赖实验而又有一定准确性的获取蛋白质三维结构信息的手段，蛋白质结构预测在后基因组时代就显得格外重要。

目前国内外关于蛋白质结构预测的专著已有不少，但是大都偏重于理论方面。对于从事生命科学与生物技术研究的广大科研工作者来说，需要的是实用性强、具有可操作性的技术类著作，而这方面的资料几乎是空白。为了使在生物医学领域的科技人员及研究生能够循序渐进地掌握蛋白质结构预测的方法，我们组织了一批在一线从事相关工作的年轻博士出版了这本小册子。主要向广大读者详细介绍蛋白质结构建模、蛋白质的分子对接、蛋白质-蛋白质相互作用的模拟、蛋白质结构的分子动力学模拟等操作。每一种分析技术都选择两种软件，一是在业界声誉最好的商用软件，二是应用最普遍的免费软件。计算分析的对象都是各位编者个人在实践工作中的具体的分析对象。期望通过阅读本手册，可以使从事生命科学与生物技术研究的广大科研工作者能够掌握相关的技术要领，学会掌握这些方法，为他们的科研工作服务。

本书的第一章由岳俊杰撰写，第二章、第四章由岳俊杰、梁龙、李北平撰写，第三章由岳俊杰、冯华、徐涛撰写，第五章由冯华、刘明、田然、吕炜撰写，第六章由冯华、刘明、徐涛、吕炜撰写，第七章由冯华、田然、刘明、郝格非、刘健撰写。

编写过程得到了军事医学科学院生物工程研究所和创腾科技有限公司的大力协助，在此表示感谢。

限于编者的水平和精力，再加上时间仓促，书中可能存在许多不妥之处，敬请广大读者批评和指正。热情期盼把您的宝贵意见函至 yue_junjie@126.com。

编者
2010年3月

目 录

第一章 蛋白质结构概述	1
第一节 蛋白质的结构层次	1
一、一级结构	1
二、二级结构	1
三、三级结构（折叠模式）	2
四、四级结构	2
第二节 蛋白质结构的测定方法	2
一、X 射线晶体学	2
二、核磁共振	4
三、结构分析的其他方法	4
第三节 蛋白质结构预测的意义	5
第二章 蛋白质二级结构预测	7
第一节 蛋白质二级结构预测概述	7
第二节 蛋白质二级结构预测实例	8
第三章 利用同源模建法预测蛋白质结构	10
第一节 同源模建的基本原理及基础	10
第二节 同源模建预测蛋白质结构的基本步骤	11
一、参考蛋白搜索	11
二、确定结构保守区	11
三、蛋白主链的模建	11
四、侧链安装	12
五、优化处理	12
六、合理性检测	12
第三节 利用 SWISS-MODEL 服务器进行蛋白质结构同源模建操作实例	12
一、登录服务器，输入目标序列	12
二、模建结果与分析	13
第四节 利用 Discovery Studio 结构模拟系统进行同源模建操作实例	15
一、识别模板，比对模板	16
二、将模板进行比对	19
三、将目标序列与模板比对	23
四、使用 MODELER 构建目标序列的 3D 模型	25
五、模型评估	26
第四章 利用折叠模式识别法预测蛋白质结构	32
第一节 折叠模式识别法的基本原理	32

第二节 折叠模式识别法预测蛋白质结构的操作实例	33
一、程序的安装	33
二、Threader 程序的运行	33
三、结果分析	34
四、过滤假阳性结果	36
第五章 蛋白质分子对接技术	37
第一节 蛋白质分子对接概述	37
第二节 利用 LibDock 进行分子对接的基本操作方法及实例	37
一、准备分子对接体系, 执行分子对接计算	38
二、分析对接结果	43
第三节 利用 Flexible Docking 进行受体和配体柔性的分子对接实例	46
一、准备对接的分子体系, 完成分子对接计算	46
二、分析配体对接结果	49
第四节 利用 AutoDock 程序进行分子对接的操作及研究实例	50
一、AutoDock 的安装	50
二、AutoDock 的基本应用	52
三、实例分析	68
第六章 蛋白质-蛋白质复合物的结构模拟	70
第一节 蛋白质与蛋白质对接的意义	70
第二节 利用 ZDOCK 来研究蛋白质-蛋白质的对接	70
一、设定一次 ZDOCK 计算	71
二、分析 ZDOCK 结果	72
三、利用 RDOCK 优化对接构型	75
四、使用 RMSD 分析结合界面的氨基酸	76
第三节 利用 Hex 研究大分子与大分子之间的相互作用	78
一、Hex 的安装	79
二、Hex 的基本应用	79
三、对接实例	84
第七章 蛋白质结构的分子动力学模拟	85
第一节 蛋白质分子动力学模拟概述	85
第二节 Discovery Studio 平台中分子动力学操作实例	86
一、准备分子动力学体系, 执行分子动力学计算	86
二、分析分子动力学计算结果	90
第三节 利用 NAMD 进行分子动力学计算的操作及研究实例	95
一、NAMD 简介	95
二、NAMD 程序的安装与使用	95
三、构建模型	97
四、分子动力学模拟	102
五、可视化与数据处理	108
第四节 利用 GROMACS 进行分子动力学模拟操作	113

一、GROMACS 介绍.....	113
二、GROMACS 的下载及安装.....	113
三、GROMACS 的基本操作.....	114
本书主要参考文献.....	123

第一章

蛋白质结构概述

蛋白质是生命活动的主要承担者，一切生命活动无不与蛋白质密切相关。蛋白质所具有的生物学功能在很大程度上取决于蛋白质分子的结构。蛋白质是由一条或几条多肽组成的大分子，每条多肽链都是一个线性的氨基酸链。蛋白质是生物大分子，具有明显的结构层次性，由低层到高层可分为一级结构、二级结构、三级结构和四级结构。一级结构也称为初级结构，二级结构、三级结构和四级结构被称为高级结构。

第一节 蛋白质的结构层次

一、一级结构

蛋白质的一级结构是指肽链的氨基酸组成及其排列顺序，包括二硫键的连接关系。氨基酸序列是蛋白质分子结构的基础，它决定蛋白质的高级结构。氨基酸通过肽键连接在一起，肽键一般采取反式构型，也就是说羰基氧原子和临近氨基酸的氨基氢原子彼此指向相反的方向。肽键本身是刚性的，但是其他键则具有很大的柔性，从而使多肽主链能够在空间折叠。尽管反式构型依然是优先的，但脯氨酸的残余基团结合到多肽链的主链上，因而这样的残基可以形成顺式的肽键，使羰基氧原子和邻近氨基酸的氨基氢原子指向同一个方向。这对肽链主链的折叠有很重要的影响，而且用其他氨基酸残基取代脯氨酸后不可避免地会对整体结构有重大的影响。正是因为这个原因，脯氨酸通常在蛋白质结构中是高度保守的。与此类似的是，甘氨酸也是很重要的残基，因为它具有很小的残余基团（只有一个氢原子），这样使它比其他氨基酸残基具有更大柔性度。半胱氨酸残基也是高度保守的，因为它有形成二硫键的能力，可以将分离的多肽链连接在一起。

二、二级结构

蛋白质的二级结构是由分子内氢键产生的有规则的和重复的局部构型。二级结构有时候涉及极性侧链（如丝氨酸和苏氨酸残基的侧链），但是多肽主链本身也是极性的，因为NH基团可以作为氢原子供体，C=O基团可以作为氢原子受体。肽键贯穿多肽主链的有规则分布有利于重复有序的结构形成。 α 螺旋和 β 折叠片是两种最普通的结构。 α 螺旋通常是右手螺旋，当两个相距四个残基的肽单元之间形成氢键时就会出现 α 螺旋。 α 螺旋使肽键排成了一行。 α 螺旋的大小变化是4~40个残基，对应于1~11圈螺旋。相反， β 折叠片是在多肽链的键角完全扩展的区域（这些区域称为 β 折叠股）形成的。几个 β 折叠股可以排列成平行（parallel）、反平行（antiparallel）或者混合型的阵列，相邻折叠股之间的肽单元形成了氢键。 α 螺旋和 β 折叠片可以通过一些连接区域结合在一起，这些连接区域采取的是它本身的二级结构，这种二级结构称为回折（turn）结构。如果连接区域没有出现氢键，这样的连接区域成为连接回环（loop）。一个蛋白质的内核通常是富含二级结构的，而环区一

般出现在表面。

三、三级结构（折叠模式）

一个多肽的三级结构，或者折叠模式是它的总体外形，反映了二级结构和模体堆积在一起形成致密结构域的方式。一个结构域可以被看成是一个多肽链中的一部分，这部分可以独立折叠成一个稳定三维结构。结构域也可以被定义为功能的单元。一个蛋白质可以包含一个结构域，也可以包含几个结构域。在后面的这一种情况下，不同的结构域可以按照与蛋白质的整体生物学功能有关的前后联系来执行完全不同的生物学功能。

四、四级结构

很多蛋白质只有一条多肽链，但是其他的很多蛋白质则包含多条多肽亚基。这些亚基的装配方式决定了蛋白质的四级结构。一个多结构域蛋白质和具有几个不同多肽亚基的蛋白质之间没有功能上的差别，很多蛋白质以这两种形式存在。例如，大多数转录因子都是一个具有 DNA 结合和转录激活结构域的单条多肽，另外一个则是由不同的亚基装配成的。实际上，相互作用的亚基形成的转录因子的装配正是酵母双杂交系统检测二元蛋白质作用的基础。蛋白质亚基可以通过非共价键相互作用，也可以通过多肽链间二硫键结合在一起。

第二节 蛋白质结构的测定方法

在没有任何形式的结构数据来源的情况下，从一个蛋白质的序列开始，从头预测这个蛋白质的三级结构在目前依然是不可能的。因而，在所有可信程度的情况下都可以有把握地确定一个在某种程度上特征完全未知的蛋白质的结构的唯一方法就是通过实验手段来测定。有两种主要的技术可以用于这个目的，这些技术就是 X 射线晶体学（X-ray crystallography, XRC）和核磁共振（nuclear magnetic resonance, NMR）谱。蛋白质结构库（PDB）中超过 98% 的结构是采用上述两种方法中的一种测定的，余下的 2% 中的绝大多数是基于 X 射线晶体学结构和 NMR 结构的理论模型。全部结构中有不到 100 个是利用其他方法，如中子衍射、电子衍射和电子显微镜得到的。

X 射线晶体学和 NMR 两种方法都有非常苛刻的要求，对于每一个蛋白质来说，必须根据经验来确定准确的实验条件。X 射线晶体学中蛋白质晶体的制备被认为是一门“艺术”，测定蛋白质结构的很多尝试都是因为拿不到合适的晶体而失败。对于 NMR 谱，则是蛋白质以溶液状态测定结构，蛋白质必须是在高浓度下是可溶而且稳定的，而且在这样的条件下不能聚集或变性。这两种技术都需要对大量的数据进行收集和处理，然后是艰苦的模型装配以便于产生原子坐标，装配的模型要与实验得出的结果相一致。

一、X 射线晶体学

X 射线晶体学利用的是这样一个原理，即当 X 射线经过蛋白质晶体时，它会以一种可预测的方式被散射或衍射。X 射线在遇到电子时会发生衍射，因此散射的特征依赖于出现在每个原子中电子的数量和原子在空间的排列。与其他的波一样，衍射的 X 射线彼此之间会发生正向的或负向的干涉。当蛋白质分子有规则地排列在一个晶体中时，由不同分子的等价原子散射产生的在同一方向的 X 射线会发生相互作用，这将在探测仪上生成一个斑点图案，这种斑点图案称为反射。这些衍射图案可以用于构建分子的电子云的三维图像，这

种电子云的三维图像被称为电子密度图。蛋白质的结构模型就是搭建在这样的电子密度图中。

精确的结构测定需要一个高度有序的晶体，这样的晶体能够使 X 射线产生强烈衍射。蛋白质晶体生长特别困难，这是结构测定中一个不可忽视的瓶颈。疏水蛋白质或含有疏水结构域的蛋白质是最难结晶的，这也就是 PDB 数据库中只有很少几个完整膜蛋白结构的原因，因为这些蛋白质都有疏水的跨膜结构域。自动化结晶工作台的开发有助于扩大 X 射线晶体学通量，这样就可以在同一时间尝试试验数千个不同的参数条件，比如不同的蛋白质浓度、盐浓度、温度和 pH 等，以确定最佳结晶条件。少量的样品也可以采用，这样可以进行那些含量不丰富的蛋白质的结晶研究。

获得蛋白质晶体之后，X 射线晶体学面对的下一个问题是从衍射图案计算一个电子密度图。这需要三方面的信息：入射 X 射线的波长、散射的 X 射线的振幅（可以通过反射的强度测定出来）和衍射的相位。不幸的是，相位不能从反射的图案中测定出来，这就带来了相位问题。有时候可以从已经解析的、存放在蛋白质数据库中的相关结构中来“借”相位，这种处理手段称为分子置换法。不过，多数时候需要进行进一步的实验来测定衍射相位。标准的处理过程是制备包含重金属原子的同形晶体，也就是在相同总体结构的晶体中结合上较重的原子，这样可以产生另外一种不同的衍射图案。包含重金属原子的同形晶体可以通过在重金属的盐溶液中浸泡来制备，经过浸泡，重金属原子能够扩散到原先由溶剂分子占据的空间，并与蛋白质分子中某些确定的位置结合。重金属原子比那些蛋白质分子中正常出现的原子更强烈地使 X 射线发生衍射。通过比较多种不同的同型晶体产生的反射（称为多对同晶置换法，multiple isomorphous replacement, MIR），就可以确定这些重原子的位置，这样就可以推导出未置换晶体中衍射的相位。每一个反射的全部描述即波长、振幅、相位，它们就是结构因子。

测定结构因子的相位也可以用异常散射法来实现。当蛋白质分子中的重金属原子在遇到波长接近其天然吸收边限的 X 射线照射时，就会导致这些重金属原子会以附加 X 射线的形式重新发射出其中的部分能量，这就是发生了异常散射。异常散射的幅度随着入射 X 射线的波长而变化，因此一种含有重金属原子的晶体可以被几种不同波长的 X 射线照射，而产生几种不同的衍射图案，从这些衍射图案中，散射的相位就可以计算出来。这就是单对同晶置换加异常散射（single isomorphous replacement with anomalous scattering, SIRAS）和多波长异常散射（multiple wavelength anomalous dispersion, MAD）等技术的基础。一种改进的处理手段可使蛋白质晶体无须浸泡，这种方法是将蛋白质在细菌中表达时加入一些重金属原子取代的氨基酸类似物。

最后，需要在电子密度图中搭建一个结构模型。这需要一些更有决定性作用的信息——氨基酸序列，因为通过 X 射线照射不可能明确地区分出碳、氧和氮等原子，所以识别氨基酸侧链很困难。结果得到的模型是一套原子 XYZ 坐标，赋予了除氢以外的所有原子。电子密度的数据越多，原子位置的确定程度越高，模型的分辨率也越高。虽然如此，仍然可能会有蛋白质分子中某些区域的原子位置不能精确测定。每个原子都被赋予一个所谓的温度因子，这是确定性的一个度量。温度因子越高，则确定性越低。高温因子说明了两种情况，一种是无序的一个度量；另外一种是动力学状态。

二、核磁共振

核磁共振 (nuclear magnetic resonance, NMR) 是因为某些原子核具有磁的性质而发生的一种现象。在 NMR 谱中, 这些性质可以用来获得化学的信息。可以认为亚原子的微粒都在绕着它们的轴旋转, 并且在很多原子中, 这些旋转彼此平衡抵消, 所以原子核本身没有总旋转。在氢原子 (^1H) 和某些天然存在的碳、氮的同位素 (^{13}C 、 ^{15}N) 中, 这些旋转不能抵消, 这类原子核会拥有一个所谓的磁矩。这样的原子核可以采取两种可能的取向中的一种, 在正常的情况下, 这两种取向都有相同的能量。但是在一个外加磁场的情况下, 这些能级就会分裂, 这是因为原子核磁矩的一个取向是与外加磁场平行的, 而另外一个取向则不是。如果这种能量间隔存在, 而且原子核暴露在某一特定频率的电磁辐射中时, 那么这个原子核就会被诱导发生跃迁, 从低能量的磁旋转状态跃迁到不优先选择的高能量旋转状态。因为电磁辐射的频率与原子核旋转的频率是一致的, 所以这种吸收称为共振。当原子核回到它们的初始定向时, 这些原子核会发射出电磁波, 这些电磁波可以被测量出来。质子 (^1H) 能够给出最强的信号, 这一点正是利用 NMR 谱进行蛋白质结构分析的基础。

NMR 谱可以用于测定溶液中蛋白质的结构, 但需要蛋白质是非常可溶而且高度稳定, 需要溶液浓度接近 1mmol/L , 并且需要的体积比较大 ($1\sim 5\text{mL}$)。因为所有成键原子附近的电子都会产生自己的磁场, 所以成键原子附近的电子会影响到每个原子核的磁共振频率, 因此 NMR 谱可以用来测定结构。外加的磁场的强度必须提高到能够克服电子产生的磁场的对抗或屏蔽, 扰动的幅度或化学位移取决于每一个原子核的化学环境。通过这种方法, 可辨别不同的氢原子, 如甲基和芳香基团的氢原子的辨别。

一维核磁共振实验可以检测到化学位移以及其他的类似于自旋-自旋耦合 (spin-spin coupling) 这样的屏蔽效应, 但是在通常情况下, 这些对于描述像蛋白质这样复杂的分子是不够的。可以采用一些由不同时间间隔区分的系列脉冲代替使用单个脉冲。这样可以得到一个二维的 NMR 谱, 这种谱携带了额外的峰, 这种峰可以显示具有相互作用的原子核对。

当把相互作用的原子核对的各种效应都考虑进去之后, NMR 分析的结果就是一些距离制约的组合, 通过这些距离约束, 可以用来推算某些特定的原子对之间的距离 (包括成键的和没有成键的)。如果足够的距离约束计算出来, 那么符合这些数据的蛋白质的结构数量就会变得非常有限。NMR 分析产生的是 $10\sim 50$ 个模型的组合体, 而不是一个单一的结构。高精度的 NMR 谱依赖于蛋白质分子在溶剂中的快速翻转, 这限制了可以分析的蛋白质的大小, 目前可以分析的蛋白质都少于 300 个氨基酸残基。如果质子的 NMR 谱过于密集, 分析也可以扩展到其他的原子核 (如 ^{13}C 及 ^{15}N) 以产生多维和异核的 NMR 谱, 这样可以降低数据密度。

三、结构分析的其他方法

除 X 射线晶体学和 NMR 外, 有些其他方法也可以为解析的结构提供补充信息, 或者分析那些在 X 射线晶体学和 NMR 谱两种情况都不能分析的蛋白质结构。例如, 圆二色谱 (circular dichroism spectrophotometry, CDS) 就是一个很有用的测定蛋白质二级结构的方法。圆二色是一种光学现象。不对称的分子如蛋白质, 在左旋圆偏振化光和右旋圆偏振化光中有不同的吸收谱。圆二色采用的是波长 $160\sim 240\text{nm}$ 的光, 这个波段的光能够分别对含 α 螺旋和 β 折叠的蛋白质产生明显不同和具有特征的谱。圆二色谱不能测定蛋白质的三

级结构，但是对于结构生物学中的 X 射线晶体学和 NMR 谱是个有益的补充。

其他的可以用于研究蛋白质三级结构的方法包括中子衍射和电子衍射。中子衍射比 X 射线衍射 (XRD) 使用的机会少得多，这是因为中子源不能广泛地得到，而且中子束的流量约比 X 射线束的流量低 10 个数量级。电子衍射用来研究那些能够结晶形成或天然装配成二维排列但不形成有序的三维晶体的蛋白质的结构。电子显微镜的优势就是可以用分析结晶排列的蛋白质相同的方式分析那些单个分子，在不用结晶的情况下测定出大蛋白复合物的结构。电子显微镜的分辨率常常没有 X 射线衍射的分辨率高，但是由于不需要结晶，在只有少量而且不太纯的样品的情况下也可以进行测定。

第三节 蛋白质结构预测的意义

尽管解析蛋白质结构的技术已经取得了非常重大的进展，获得了大量的蛋白质结构，但解析蛋白质结构依然是一个劳动密集型和花费昂贵的过程，而且目前各种实验方法都存在缺陷或限制。X 射线单晶衍射最大的缺点是要测定的蛋白质必须能够形成晶体，而许多蛋白质不能结晶，这就限制了测定的范围。核磁共振方法虽然不需培养蛋白质晶体，但需要蛋白质能够在较高浓度下可溶、稳定、不聚集或变性，更重要的是，核磁共振技术对于较大的蛋白质分子的结构还无能为力。电镜三维重构需要培养二维晶体，而且目前测定的结构的分辨率还不够高，还没有成为测定蛋白质结构的常规方法。

蛋白质空间结构的测定速度还远远赶不上蛋白质序列增长的速度，并且随着基因组测序的进行，二者差距有越来越大的趋势。截止 2009 年 10 月，已知序列的蛋白质已超过 100 万，而已知结构的蛋白质仅有 6 万多。因此蛋白质结构预测成了目前了解蛋白质结构信息的一个非常有效的实际手段。随着基因组和蛋白质组计划的研究进展，大量的具有特殊功能的蛋白质将被发现，对蛋白质空间结构预测的要求也越来越迫切。

除了获得蛋白质结构信息之外，蛋白质结构预测还有重大的理论意义。生命遗传信息存储传递及表达的认识是 20 世纪生物学所取得的最重要的突破，其中的关键问题是由 3 个相连的核苷酸顺序决定蛋白质分子肽链中的 1 个氨基酸，即“三联遗传密码”（“第一遗传密码”）的破译。蛋白质必须有特定的三维空间结构，才能表现其特定的生物学功能。20 世纪 60 年代 Anfinsen 就提出蛋白质分子的一级序列决定其空间结构的论断，这一论断得到多次实验证实并被人们广泛接收，大量实验充分说明蛋白质的氨基酸顺序与其空间结构之间存在着确定的关系。国际上将蛋白质的氨基酸序列与其空间结构的对应关系称之为“第二遗传密码”。但是直到现在，“第二遗传密码”依然没有得到破译，这个问题是蛋白质研究最后几个尚未揭示的奥秘之一。蛋白质结构预测实际上是去从理论上最直接地解决蛋白质的折叠问题，即破译“第二遗传密码”。

近年来随着结构生物学的发展，解析的蛋白质结构越来越多，这就为研究和总结蛋白质结构的规律打下了很好的基础，也为蛋白质的结构预测提供了参考。另外计算机科学与技术快速进步也大大地促进了蛋白质结构预测的发展。蛋白质结构预测问题虽然还没有最终解决，但是已经取得了一些令人欣喜的成就。目前，蛋白质空间结构预测的方法有 3 种，即从头预测法、比较建模法和折叠识别法。

1. 从头预测法

从头预测 (ab initio prediction) 也称为理论计算预测，是指从蛋白质的一级结构出发，

根据物理化学、量子化学、量子物理的基本原理，利用各种理论方法计算出蛋白质肽链所有可能的构象的能量，然后从中找到能量最低的构象，就是蛋白质的天然构象。这种方法不需要已知结构信息，能够产生全新结构。但是由于计算的难度，这种方法只能用于计算很小的分子或蛋白质分子的局部结构，目前还不能作为一种常用的预测蛋白质结构的方法。现在从头预测主要用作其他预测方法的补充或作为一种优化结构的手段。

2. 比较建模法

比较建模法 (comparative modeling) 也称为同源建模法 (homology modeling)，是基于知识的蛋白质结构预测方法。根据对 PDB 中的蛋白质进行结构比较的定量研究得知，任何一对蛋白质，只要它们序列的长度达到一定程度，序列相似性超过 30%，则可以保证它们具有相似的三维结构。因此，对于一个未知结构的蛋白质，如果找到一个已知结构的同源蛋白质，就可以该蛋白质的结构为模板，为未知结构的蛋白质建立结构模型。同源建模通常包括下列主要步骤：模板搜寻、序列比对、结构保守区寻找、目标模型搭建、结构优化和评估等。在目前的三种预测蛋白质结构的方法中，比较建模法是最简单、最为成熟的。很多实验室都可以开展同源建模预测蛋白质结构的工作。

3. 折叠识别法

折叠识别法 (fold recognition) 也称为反向折叠法 (inverse fold method)、threading 等。该方法基于这样一个事实，即很多序列没有相关性的蛋白质具有相似的折叠模式。因此可以基于序列结构比对 (sequence structure alignment) 的预测方法，通过目标蛋白质的氨基酸序列和已知折叠模式的逐一比对，根据特定的计分函数 (scoring function) 找出最适合目标序列的折叠模式。折叠识别法可以弥补同源建模方法只能依赖序列相似性寻找模板的不足，是目前三种预测蛋白质结构的方法中发展最快也是最有前途的方法。