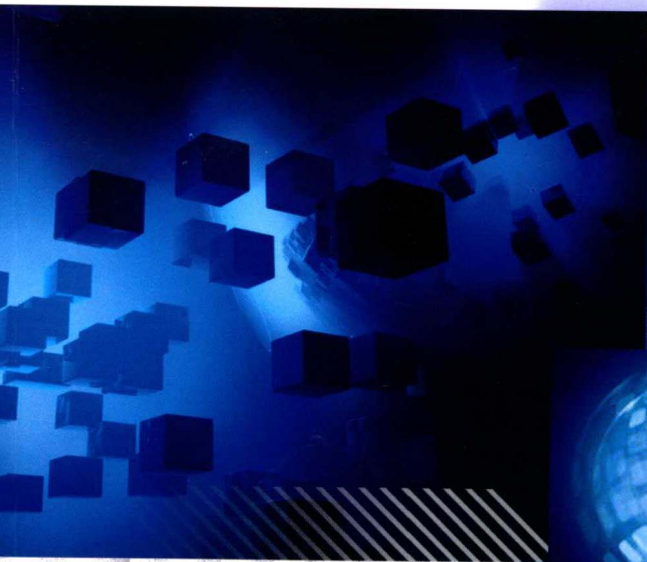




Protein Science

—Theory, Technology and Applications



蛋白质科学

——理论、技术与应用

张艳贞 宣劲松 主编



北京大学出版社
PEKING UNIVERSITY PRESS



蛋白质科学

—— 蛋白质 结构与功能

2000 年第 1 期

第 1 卷第 1 期

蛋白质科学

理论、技术与应用

主 编 张艳贞 宣劲松



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

蛋白质科学：理论、技术与应用/张艳贞,宣劲松主编. —北京：北京大学出版社，
2013.1

ISBN 978-7-301-21719-1

I. ①蛋… II. ①张… ②宣… III. ①蛋白质—高等学校—教材 IV. ①Q51

中国版本图书馆 CIP 数据核字(2012)第 290410 号

书 名：蛋白质科学——理论、技术与应用

著作责任者：张艳贞 宣劲松 主编

责任编辑：黄 炜

标准书号：ISBN 978-7-301-21719-1/Q · 0132

出版发行：北京大学出版社

地 址：北京市海淀区成府路 205 号 100871

网 址：<http://www.pup.cn> 新浪官方微博：@北京大学出版社

电子信箱：zpup@pup.cn

电 话：邮购部 62752015 发行部 62750672 编辑部 62752038

出版部 62754962

印 刷 者：北京大学印刷厂

经 销 者：新华书店

787 毫米×1092 毫米 16 开本 13 印张 296 千字

2013 年 1 月第 1 版 2013 年 1 月第 1 次印刷

定 价：28.00 元

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容

版权所有，侵权必究

举报电话：010-62752024 电子信箱：fd@pup.pku.edu.cn

前 言

蛋白质是生物体最重要的组成成分之一,几乎参与了生命活动的全过程。随着人类基因组计划的实施、推进和深入,蛋白质研究逐步受到关注和重视,并被提升到前所未有的高度,正如人类基因组计划(Human Genome Project, HGP)主要负责人 Francis Collins 在人类基因组草图公布之际所强调的那样:真正的竞赛才刚刚开始!竞赛的核心内容,就是要研究基因的产物——蛋白质。随即,美国、欧洲和亚太地区联合启动了国际性大合作——人类蛋白质组计划。2006年1月我国国务院颁布的《国家中长期科学和技术发展规划纲要》(2006—2020年)及时把握科技发展机遇,将蛋白质研究列为今后15年全国重点部署的四项重大科学研究计划之一,充分体现了蛋白质科学与技术引领未来科学技术发展、实现重点跨越式发展、建设创新型国家中的重要战略地位和重要战略意义。

经过近十年的发展,蛋白质科学与技术已经成为21世纪生命科学与生物技术的重要战略前沿,蛋白质研究相关的基础理论和实验技术已成为各大高校生物、医学、化学、药学类专业必修的重点课程和难点课程。可以说,近些年蛋白质科学取得了很多人振奋的技术进步和研究成果,也有不少蛋白质研究领域相关的专著刊出,但这些新进展还没有及时转化为教材内容,所以目前很难找到一本适合本科生教学和学习的、难度适中、系统全面、新颖实用的介绍蛋白质科学研究的教科书。

本书由教学科研一线教师汇集五年来教学实践经验和教学研究成果编写而成,广泛参考现有的教材、著作和文献并进行了吸收和归纳。全书注重理论与技术相衔接,语言严谨且简明易懂。以蛋白质研究的发展历程和领域层次为主线,依次介绍了蛋白质的基本结构单位——氨基酸、氨基酸的连接——肽键与肽、蛋白质的表征(包括蛋白质的定量、纯度的测定、相对分子质量的测定、等电点的测定)、蛋白质的分子构象、蛋白质构象解析技术策略、蛋白质的结构与功能、蛋白质的翻译后加工与修饰、蛋白质的理化性质与分离鉴定以及蛋白质组学等内容,是一本系统介绍蛋白质的理论研究和发展的教科书。

本书不仅有蛋白质研究一般知识的普及,还特别注重介绍蛋白质研究各个领域的最新进展和发展趋势,比较准确地归纳和体现了蛋白质研究的现状,尤其是每一章末增加了思考题以及相关研究热点问题的拓展阅读,可极大地提高读者学习蛋白质科学的兴趣,有助于其创新思维的启发和培养。

本书不仅可作为生物类、医学类、化学类、药学类、环境生态类、食品科学与技术类专业的本科生、专升本学生的教材,也可作为相关专业研究生和从事蛋白质研究的科研人员的参考用书,还适合于有一定分子生物学和生物化学知识,希望尽快了解蛋白质研究领域的科技工作者或管理者阅读。

本书第一、二、三、五、六、七、八、十章主要由张艳贞(北京联合大学应用文理学院)编写,第四、九章主要由宣劲松(北京科技大学)编写,郭俊霞、陈文(北京联合大学应用文理学院)参与各章部分节次的补充和修订;拓展阅读部分由宣劲松、张艳贞、郭俊霞、陈文分

工编写;此外,张静、米生权、常平(北京联合大学应用文理学院)等参与讨论、查阅文献、校稿和订正等工作;全书由张艳贞、宣劲松策划统稿。

在本书编写过程中,参考了国内外众多作者的著作、文章以及相关教材,还通过网络获得了很多有用的知识,特别是中国科学网、北京蛋白质组研究中心的“蛋白质组天空”及一些个人的学术 blog,在此表示衷心的感谢。

本书由“北京联合大学人才强校计划人才资助项目”资助,北京大学出版社黄炜担任责任编辑。编辑出版过程中,黄炜老师做了大量细致工作并给予很多编写启发,北京联合大学应用文理学院各级领导和各位同事也给予了各方面的配合和支持,在此一并表示感谢。

由于蛋白质科学发展迅速、知识内容涉及广泛,加之编者水平有限,尽管本书已经历多次审核和校订,但肯定还会有错误和问题存在,恳请同行专家以及使用本书的广大师生和读者批评指正。

张艳贞 宣劲松

2012 年 10 月

目 录

第一章 绪论	(1)
一、什么是蛋白质	(1)
二、蛋白质在生命体中的重要地位	(1)
三、蛋白质的元素组成和结构组成	(2)
四、蛋白质的分类	(3)
五、蛋白质研究史上的重要人物和事件	(4)
六、我国国内蛋白质研究现状	(5)
七、当前蛋白质研究的热点和特点	(6)
第二章 蛋白质的基本结构单位——氨基酸	(11)
第一节 氨基酸的结构及分类	(11)
一、常见的蛋白质氨基酸	(11)
二、不常见的蛋白质氨基酸	(14)
三、非蛋白质氨基酸	(15)
第二节 氨基酸的理化性质与应用	(16)
一、氨基酸的物理性质	(16)
二、氨基酸的化学性质	(18)
第三节 氨基酸的制备、分离与分析	(23)
一、氨基酸的制备	(23)
二、氨基酸的分离与分析	(24)
第三章 氨基酸的连接——肽键与肽	(29)
一、肽键和肽的基本概念	(29)
二、肽键的结构	(30)
三、肽的物理和化学性质	(31)
四、生物活性肽	(32)
五、多肽组与多肽组学	(33)
第四章 蛋白质的表征	(37)
第一节 蛋白质的定量	(37)
一、紫外吸收法	(37)
二、Folin-酚法	(39)
三、BCA 法	(40)
四、考马斯亮蓝 G-250 比色法 (Bradford 法)	(42)
五、利用微波快速蛋白质定量法	(44)
第二节 蛋白质纯度的测定	(45)

一、光谱法	(45)
二、层析法	(46)
三、电泳法	(47)
四、N-末端氨基酸测定法	(47)
五、生物活性测定法	(48)
六、纯度测定的原则	(48)
第三节 蛋白质相对分子质量的测定	(48)
一、SDS-聚丙烯酰胺凝胶电泳测定法	(49)
二、凝胶过滤层析法	(53)
三、生物质谱法	(58)
第四节 蛋白质等电点的测定	(62)
第五章 蛋白质的分子构象	(68)
第一节 蛋白质分子中的化学键及蛋白质一级结构	(68)
一、蛋白质分子中的化学键	(68)
二、蛋白质的一级结构	(69)
第二节 蛋白质的二级结构	(69)
一、 α 螺旋	(69)
二、 β 折叠	(70)
三、 β 转角	(71)
四、 β 凸起	(72)
五、无规卷曲	(72)
六、超二级结构和结构域	(72)
第三节 蛋白质的三级结构和四级结构	(75)
一、蛋白质的三级结构	(75)
二、蛋白质的四级结构	(76)
第四节 蛋白质折叠异构酶和分子伴侣	(77)
第六章 蛋白质构象解析技术策略	(81)
第一节 蛋白质一级结构测定	(81)
一、蛋白质序列测定的基本战略	(81)
二、蛋白质序列测定的一般步骤	(81)
三、测序过程中关键步骤的测定方法	(82)
四、蛋白质序列测定的新方法——生物质谱	(91)
第二节 蛋白质高级结构解析技术	(92)
一、X射线晶体衍射技术	(93)
二、核磁共振技术	(96)
三、高级结构解析的其他方法	(101)
第三节 蛋白质高级结构预测技术	(103)
一、蛋白质结构预测的实验理论基础	(103)
二、蛋白质结构预测方法	(105)

第七章 蛋白质的结构与功能	(112)
第一节 蛋白质结构的价值及其与功能的关系	(112)
一、蛋白质的功能是蛋白质结构的延伸	(112)
二、蛋白质结构是蛋白质功能的基础	(112)
三、蛋白质的功能在与其他分子相互作用的互动中实现	(114)
第二节 蛋白质一级结构与功能的关系	(114)
一、活性蛋白质与前体	(114)
二、一级结构与物种进化的关系	(118)
三、一级结构改变与分子病	(120)
第三节 蛋白质空间结构与功能的关系	(120)
一、血红素辅基的结构	(121)
二、肌红蛋白和血红蛋白的结构与功能	(121)
第四节 蛋白质构象病及其治疗策略	(127)
一、蛋白质构象病的概念	(127)
二、构象病的发生、发现及蛋白质构象致病假说	(128)
三、其他的蛋白质构象病	(131)
四、蛋白质构象病治疗的新思路	(131)
第八章 蛋白质的翻译后加工与修饰	(135)
第一节 蛋白质的翻译后加工	(135)
一、氨基端的加工	(135)
二、蛋白质肽链的加工	(136)
第二节 蛋白质的翻译后修饰	(137)
一、蛋白质的磷酸化修饰	(137)
二、蛋白质的糖基化修饰	(143)
三、蛋白质的泛素化修饰	(146)
四、蛋白质的乙酰化修饰	(148)
五、蛋白质的甲基化修饰	(149)
第九章 蛋白质的理化性质与分离鉴定	(152)
第一节 理化性质	(152)
一、蛋白质的酸碱性	(152)
二、蛋白质的胶体性质	(152)
三、蛋白质的变性、沉淀和凝固	(153)
四、蛋白质的紫外吸收	(154)
第二节 蛋白质的分离和纯化	(154)
一、蛋白质分离提纯的一般步骤	(154)
二、蛋白质分离、纯化的方法	(155)
第十章 蛋白质组学	(172)
第一节 蛋白质组学产生的历史背景	(172)
一、大规模生物学的兴起	(172)

二、从基因组、转录组到蛋白质组	(173)
三、蛋白质组学的基本概念及意义	(173)
第二节 蛋白质组学的研究范畴	(174)
第三节 蛋白质组学研究的技术策略与技术平台	(176)
一、蛋白质组分离鉴定方法的发展	(176)
二、蛋白质组学研究的技术策略	(180)
三、大规模蛋白质组的制备、分离与鉴定	(181)
第四节 蛋白质组学的应用	(186)
一、在基础研究方面的应用	(186)
二、在医学与药物学中的应用	(186)
三、在农业科学研究中的应用	(187)
第五节 蛋白质组学的挑战	(188)
参考文献	(191)

第一章 绪 论

一、什么是蛋白质

“蛋白质”(protein)一词是 1938 年由瑞典化学家 Jöns Jacob Berzelius 首先提出来的,用来描述生命有机体中富有的、由氨基酸链组成的一类特殊的大分子,该词源自希腊文“Protos”,是“第一的,首要的,重要的”的意思。随着对这种大分子的不断了解,人们愈加觉得这个术语是非常贴切的。

蛋白质是一切生物体中普遍存在的,由 20 种天然 α -氨基酸^①,通过肽键相连形成的高分子含氮化合物。它们种类繁多,具有一定的相对分子质量、复杂分子结构和特定生物学功能,是表达生物遗传性状的一类主要物质。

二、蛋白质在生命体中的重要地位

蛋白质在生物体内占有特殊重要的地位。蛋白质和核酸是构成细胞内原生质的主要成分,而原生质是生命现象的物质基础。早在 1878 年,恩格斯就在《反杜林论》中指出:“生命是蛋白体的存在方式,这种存在方式本质上就在于这些蛋白体的化学组成部分的不断自我更新。”

蛋白质是生命活动的物质基础,它参与了几乎所有的生命活动过程。在细胞结构、生物催化、物质传输、运动、防御、调控以及记忆、认知等各方面都起着极其重要的作用。蛋白质是生物体内必不可少的重要成分,是构成生物体最基本的结构物质和功能物质,是生物结构和功能的载体。实际上,每种细胞活性都依赖于一种或几种特定的蛋白质。归纳起来,蛋白质的生物学功能主要有以下几个方面:

(1) 催化功能。蛋白质的一个最重要的生物功能是作为生物体新陈代谢的催化剂——酶。绝大多数的酶都是蛋白质,生物体内的各种化学反应几乎都是在酶的参与下进行的。如萤火虫的荧光就是由荧光素和 ATP 在荧光素酶催化下产生的。

(2) 结构成分。蛋白质另一重要功能是建造和维持生物体的结构,这类蛋白称为结构蛋白,它们给细胞和组织提供支撑和保护。结构蛋白的单体一般聚合成长的纤维或纤维状排列的保护层,如构成毛发、角、蹄、甲的 α 角蛋白,存在于骨、腱、韧带、皮的胶原蛋白等。

(3) 储藏功能。蛋白质是氨基酸的聚合物,是含有氮素的高分子有机物。氮素通常是生物体生长的限制性养分,所以在必要时,生物体利用蛋白质提供充足的氮素,例如,卵清蛋白为鸟类胚胎发育提供氮源,乳中的酪蛋白是哺乳类幼子的主要氮源。许多高等植

^① 近年发现硒代半胱氨酸和吡咯赖氨酸也是蛋白质生物合成的原材料,是由具双重功能的无义密码子 UGA 和 UAG 以上下游序列决定的方式插入的。

物的种子含高达 60% 的贮存蛋白,为种子的发芽准备足够的氮素。蛋白质除为生物体发育提供 C、H、O、N、S 元素外,像铁蛋白还能贮存 Fe,用于含铁蛋白(如血红蛋白)的合成。

(4) 运输功能。有一些蛋白是转运蛋白,其功能是转运特定的物质。如血红蛋白、血清蛋白是通过血流转运物质的,前者将氧气从肺转运到其他组织(第七章);后者将脂肪酸从脂肪组织转运到各器官。另一类转运蛋白是膜转运蛋白,它们能通过渗透性屏障(细胞膜)转运代谢物和养分(葡萄糖、氨基酸等),如葡萄糖转运蛋白。至今研究过的所有天然膜转运蛋白都是在膜内形成通道,被转运的物质经过它进出细胞。

(5) 运动功能。某些蛋白质赋予细胞以运动的能力,如肌肉收缩和细胞游动都是由相应的蛋白质参与完成的。作为运动基础的收缩蛋白和游动蛋白具有共同的性质:它们都是丝状分子或丝状聚集体。例如,形成细胞收缩系统的肌动蛋白(actin)和肌球蛋白以及作为微管(microtubule)主要成分的微管蛋白(tubulin)都属于这一类蛋白。另一类参与运动的蛋白质称为发动机蛋白质(motor protein),如动力蛋白和驱动蛋白,它们可驱使小泡、颗粒和细胞器沿微管轨道移动。

(6) 调节功能。许多蛋白质具有调节其他蛋白质执行其生理功能的能力,这些蛋白质称为调节蛋白,最著名的例子是胰腺兰氏小岛及 β 细胞分泌的胰岛素,它是调节动物体内血糖代谢的一种蛋白质类激素。

(7) 基因表达调控功能。还有一些蛋白质参与基因的表达调控,它们可激活(正控制因子)或抑制(负控制因子)遗传信息转录为 RNA。

(8) 免疫和防御功能。与一些结构蛋白的被动性防护不同,一类确切地称为保护蛋白的蛋白质在细胞防御、保护方面的作用是主动的。此类蛋白中最突出的是脊椎动物体内的免疫球蛋白或称抗体。抗体是在外来的蛋白质或其他高分子化合物,即所谓抗原的影响下由淋巴细胞产生,并能与相应的抗原结合而排除外来物质对生物体的干扰的一类蛋白。另一类保护蛋白是血液凝固蛋白、凝血酶和血纤蛋白原等。南极鱼和北极鱼含有抗冻蛋白,能防止在深海低于 0℃ 水温下血液冷冻。此外,起防卫作用的还包括蛇毒和蜂毒中的溶血蛋白和神经毒蛋白,还有植物毒蛋白和细菌毒素。

(9) 信号传导功能。某些蛋白质在细胞应答激素、生长因子等信号途径中起作用,参与对激素和其他信号分子的胞内应答的协调和通讯。如视网膜上的视紫质蛋白可通过 G-蛋白将光信号转换成化学信号和神经冲动。

(10) 电子传递功能。代谢物脱下的还原型辅酶经呼吸链传递伴随磷酸化生成 ATP 的过程是需氧生物获得能量的主要途径。呼吸链中电子传递体都和蛋白质结合存在或起作用。

(11) 特殊功能。某些蛋白质具有除上所述以外的特殊的功能,如应乐果甜蛋白有着极高的甜度;昆虫翅膀的铰合部存在一种具有特殊弹性的蛋白质,称为节肢弹性蛋白;某些海洋生物(如贝类)分泌一类胶质蛋白,能将贝壳牢固地黏在岩石或其他硬表面上。

三、蛋白质的元素组成和结构组成

1. 蛋白质的元素组成及特点

根据蛋白质结晶元素分析,发现它们的元素组成与糖和脂质不同,除含有碳、氢、氧外,还有氮和少量的硫。有些蛋白质还含有磷和其他一些金属元素,如铁、铜、锌、锰、钴、

钼,个别蛋白质含有碘。不同元素在蛋白质中的组成百分比约为:碳 50%,氢 7%,氧 23%,氮 16%,硫 0~3%,其他微量。

一切蛋白质都含 N 元素,且各种蛋白质的含氮量很接近,平均为 16%,这是蛋白质元素组成的一个特点,也是凯氏(Kjedahl)定氮法测定蛋白质含量的计算基础。蛋白质的含量=蛋白氮 $\times$ 6.25,式中 6.25 即 16%的倒数,表示任何生物样品中 1 克 N 的存在,就意味着有大约 $100/16=6.25$ 克蛋白质的存在,6.25 常称为蛋白质系数。

2. 蛋白质的结构组成

一百多年前关于蛋白质的化学研究就开始了。在早期的研究中,水解作用提供了关于蛋白质组成和结构的极有价值的资料。蛋白质可以被酸、碱或蛋白酶催化水解,在水解过程中,逐渐降解成相对分子质量越来越小的肽段,直到最后成为氨基酸的混合物。蛋白质分子是由成百上千个氨基酸分子首尾相连组成的共价多肽链,但是天然的蛋白质分子并不是走向随机的松散多肽链。每一种天然蛋白质都有自己特有的空间结构或称三维结构,这种三维结构通常被称为蛋白质的构象。指导蛋白质多肽链折叠成能量上有利的构象的规律至今仍不完全清楚,寻找这些规律是当前着力研究的课题之一。

四、蛋白质的分类

蛋白质的分类有不同的依据,依据不同就有不同的分类结果。

1. 依据蛋白质的外形分为球状蛋白质和纤维状蛋白质

蛋白质按其分子外形的对称程度可以分为球状蛋白质(globular protein)和纤维状蛋白质(fibrous protein)两大类。球状蛋白质,分子对称性佳,外形接近球状或椭球状,溶解度较好,能结晶;大多数蛋白质属于这一类。纤维状蛋白质,对称性差,分子类似细棒状或纤维状。它又可分成可溶性纤维状蛋白质,如肌球蛋白(myosin)、血纤维蛋白原(fibrinogen)等以及不溶性纤维状蛋白质,包括胶原蛋白、弹性蛋白、角蛋白以及雌心蛋白等。

2. 依据蛋白质的组成为简单蛋白和结合蛋白

许多蛋白质仅由氨基酸组成,不含其他化学成分,例如,核糖核酸酶、肌动蛋白等,这些蛋白质称为单纯蛋白质(simple protein)。许多其他蛋白质含有除氨基酸外的各种化学成分,这些化学成分成为其结构的一部分,这样的蛋白质称为缀合蛋白质(conjugated protein),其中非蛋白质部分称为辅基(prosthetic group)或配体(ligand)。

(1) 单纯蛋白质还可以根据其物理性质(如溶解度)分类如下:

清蛋白:溶于水及稀盐、稀酸或稀碱溶液,可以被饱和硫酸铵所沉淀。清蛋白广泛存在于生物体内,如血清清蛋白、乳清蛋白等。

球蛋白:为半饱和硫酸铵所沉淀。不溶于水而溶于稀盐溶液的称优球蛋白;溶于水的称假球蛋白。球蛋白普遍存在于生物体内,如血清球蛋白、肌球蛋白和植物种子球蛋白。

谷蛋白:不溶于水、醇及中性盐溶液,但易溶于稀酸或稀碱。如米谷蛋白和麦谷蛋白等。

醇溶谷蛋白:不溶于水及无水乙醇,但溶于 70%~80%乙醇中。组成上的特征是脯氨酸和酰胺较多,非极性侧链远较极性侧链多。这类蛋白质主要存在于植物种子中。如玉米醇溶蛋白、麦醇溶蛋白等。

组蛋白：溶于水及稀酸，但为稀氨水所沉淀。分子中组氨酸、赖氨酸较多，分子呈碱性。如小牛胸腺组蛋白等。

鱼精蛋白：溶于水及稀酸，不溶于氨水。分子中碱性氨基酸特别多，因此呈碱性。如鲑精蛋白。

硬蛋白：不溶于水、盐、稀酸或稀碱。这类蛋白是动物体内具有结缔及保护功能的蛋白质。例如，角蛋白、胶原蛋白、网硬蛋白和弹性蛋白等。

(2) 缀合蛋白质可以按其非氨基酸成分进行如下分类：

核蛋白：辅基是核酸，如脱氧核糖核蛋白、核糖体、烟草花叶病毒等。

脂蛋白：与脂质结合的蛋白质。脂质成分有磷脂、固醇和中性脂等。如血中的血清脂蛋白、卵黄球蛋白。

糖蛋白和黏蛋白：辅基成分为半乳糖、甘露糖、己糖胺、己糖醛酸、唾液酸、硫酸或磷酸等。如卵清蛋白、 γ -球蛋白，血清类黏蛋白等。

磷蛋白：磷酸基通过酯键与蛋白质中的丝氨酸或苏氨酸残基侧链相连。如酪蛋白、胃蛋白酶等。

血红素蛋白：辅基为血红素，它是卟啉类化合物，卟啉环中心含有金属。含铁的如血红蛋白、细胞色素 c，含镁的有叶绿蛋白，含铜的有血蓝蛋白等。

黄素蛋白：辅基为黄素腺嘌呤二核苷酸。如琥珀酸脱氢酶、D-氨基酸氧化酶等。

金属蛋白：与金属直接结合的蛋白质。如铁蛋白含铁，乙醇脱氢酶含锌，黄嘌呤氧化酶含钼和铁等。

3. 根据多肽链的数目分为单体蛋白质和寡聚蛋白质

单体蛋白质(monomeric proteins)是只含有一条肽链的蛋白质，如肌红蛋白；具有两条或两条以上独立三级结构的多肽链通过非共价键相互结合而形成的蛋白质为寡聚蛋白(oligomeric protein)，如血红蛋白。

4. 根据功能分为活性蛋白质和非活性蛋白质

生物体内很多蛋白质，尤其是酶和激素，大多以非活性的前体或酶原形式存在，在激活物或激活剂的刺激下才转变成具有生物活性的蛋白质。

5. 根据营养价值分为完全蛋白和不完全蛋白

营养价值较高的蛋白质称为完全蛋白，较低的称为不完全蛋白。营养价值的高低主要取决于其氨基酸组成，含有全部必需氨基酸且必需氨基酸比例符合或接近人体需求的蛋白质，其营养价值较高。

五、蛋白质研究史上的重要人物和事件

1878年，恩格斯就在《反杜林论》中指出：“生命是蛋白体的存在方式”；

19世纪末，从蛋白质水解产物分离得到了13种氨基酸，基团学说销声匿迹；

1902年，Fischer和Hofmeister同时提出肽键理论；

1924年，瑞典Theodor Svedberg利用他首创的超离心技术得知蛋白质分子是均一的，并具有固定的相对分子质量；

1950年，Pauling提出蛋白质二级结构的基本单位： α 螺旋和 β 折叠；

1953,英国剑桥 Frederick Sanger 利用纸电泳和纸层析测出了第一个蛋白——牛胰岛素的一级结构,开创蛋白质一级结构研究的新纪元,并由此获得 1958 年诺贝尔奖;

50 年代末,美国学者 Stanford Moore 等改进了 Sanger 的方法,完成了第一个酶蛋白——核糖核酸酶的序列分析;

1961 年,Anfinsen 用核糖核酸酶变性复性实验证明蛋白质的一级结构决定其高级结构;

1963 年,John Kendrew 和 Max Perutz 利用他们首创的重原子同晶置换技术测出了第一个蛋白——肌红蛋白晶体的三维结构;

1965 年,中国科学院与北京大学在世界上首次人工合成具有天然活性和构象的牛胰岛素;

1969 年,美国 Merrifield 等首次人工合成酶——牛胰核糖核酸酶;

80 年代初,在人类基因组计划提出之前,美国科学家 Norman G. Anderson 提出 Human Protein Index 计划,旨在分析细胞内的所有蛋白质,但没有受到认可和重视,被搁浅下来……

1994 年 Marc Wilkins 在意大利锡耶纳(Siena)的一次 2-DE(双向电泳)会议上首次提出“Proteome(蛋白质组)”概念: Proteome=protein+genome,指一个生物系统在特定病理或生理状态下表达的所有种类的蛋白质;同时,其导师澳大利亚 Macquarie 大学的 Keith Williams 向澳政府建议开展蛋白质组研究;

1995 年悉尼大学与 Williams 等 4 家实验室合作,对一种最小的自我复制生物(一种支原体)的蛋白质进行了大规模分离鉴定;

1996 年 APAF(Australia Proteome Analysis Facility)世界第一个蛋白质组研究中心成立,Wilkins 师徒二人分别负责蛋白质组技术平台建立并担任蛋白质组学学科带头人;

其后,丹麦、加拿大、日本也先后成立蛋白质组研究中心;瑞士成立 GeneProt 公司,专注于应用蛋白质组研究;

2001 年,美国成立国际人类蛋白质组组织(Human Proteome Organization, HUPO),欧洲、亚太地区分组组织相继开展工作;

2002 年人类蛋白质组计划(Human Proteome Project, HUPP)(www.hupo.org)启动:中国主持肝蛋白质组研究,德国主持脑蛋白质组研究,美国主持心脏、肺、血浆蛋白质组研究,越来越多的国家和地区参与进来,形成了真正的国际性大合作……

六、我国国内蛋白质研究现状

1997 国家自然科学基金委员会设立重大项目“蛋白质组学技术体系的建立”,中科院生化所(夏其昌)、军事医学科学院(贺福初院士)、湖南师范大学(梁宋平)迅速启动,成立蛋白质组学研究中心;

其后,《国家中长期科学与技术发展规划纲要》(2006~2020)将“蛋白质研究”列为四项重大科学研究计划之一;

科技部将疾病蛋白质组研究列入我国“973”计划项目和“863”计划项目;国家自然科学基金委员会也将“蛋白质组研究”列为重点项目;

近十年来,我国在鼻咽癌、白血病、肝癌和肺癌蛋白质组研究方面取得了很大进展,受

到国际同行的赞誉。

七、当前蛋白质研究的热点和特点

人类基因组计划于 1985 年正式提出,1990 年启动,预计用 20 年左右的时间完成,被誉为生命科学领域的阿波罗登月计划。人们设想其完成之日就是生命奥秘被揭示之日,人类基因组框架图的发表使大众对科学家将揭示人类的遗传奥秘充满了幻想与信心。但是,真实的图景远不像人们想象的那样简单。

表 1-1 是已经完成基因组测序的生物及基因预测结果:

表 1-1 已完成基因组测序的生物

生物	基因组大小/Mb	完成时间/年	预计基因数目/个
酵母	12.1	1996	6 034
线虫	97	1998	19 099
果蝇	180	2000	13 061
拟南芥	125	2000	25 498
人类	3000	2001	30 000
水稻	460	2001	44 000~65 000

就人类而言,其基因组测定预计基因约 3 万个,但远比人类低等得多的线虫预计基因约 2 万个,拟南芥 2.5 万个,看来,基因数目并不能完全解释生命现象的复杂和差异;而且如果把翻译后修饰考虑进去,人类蛋白质组所包含的蛋白质预计超过 100 万个。也许,只有真正利用蛋白质组水平上所增加的复杂性和多样性,才可以解释人类与其他生物相比所增加的生物学复杂性。正如国际人类基因组计划主要负责人 Francis Collins 在 2001 年 6 月一次会议上所强调的那样:真正的竞赛才刚刚开始!竞赛的核心内容,就是要研究基因的产物——蛋白质。同时,国际权威期刊 Nature 和 Science 专门配发了鼓励和引导蛋白质组研究的述评,真正掀起了国际蛋白质研究热点的新动向。

概要地说,当前蛋白质研究已在传统的基于单个蛋白质研究的基础上转向了对全部蛋白质组的研究,相应产生了蛋白质组学的概念。所谓蛋白质组学(proteomics),就是指从整体的角度分析细胞内动态变化的蛋白质组成成分、表达水平与修饰状态,了解蛋白质之间的相互作用与联系,揭示蛋白质功能与细胞生命活动规律的一个新的研究领域。那么,为什么组学研究的理念会受到普遍重视并成为基因组计划完成之后必然的趋势呢?这是因为:生命现象的发生是多因素影响的,必然涉及多个蛋白质;其次,多个蛋白质的参与是交织成网络的,或平行发生,或呈级联因果;而且,在执行生理功能时,蛋白质的表现是多样的、动态的,并不像基因组那样固定不变。如果不从组学的角度去研究和分析,势必产生盲人摸象的效果。

与我们大家都熟悉的人类基因组计划研究相比,蛋白质组的研究呈现出其独有的特点:

(1) 同一性与多样性。对一个细胞或组织来说,它的基因组组成是不变的,代表了它的全部遗传特征。它在生物的不同组织和部位,在个体发育与细胞生长的不同阶段都是恒定的,而它的蛋白质组成则会随着时间、环境变化、发育阶段等因素发生变化。这种变化不仅仅表现在细胞表达的种类不同,同种蛋白质的表达量上也存在着很大的变化。

(2) 有限与无限。随着人类基因组计划的进行,我们已经知道了各种生物基因组的大小,基因组不论大小,其核苷酸数量是明确的;而对于蛋白质组来说,蛋白质组蛋白数量无法把握,因为同一种蛋白质前体经过不同的加工可以成为不同的蛋白质产物,发挥不同的作用,而体内的蛋白质的修饰又是随着时间、环境等因素而不停变化的,因此,蛋白质组的研究工作似乎会成为一个无限的工作。

(3) 静态与动态。一个个体的基因组自个体诞生到死亡,始终保持不变。而作为新陈代谢主要执行者的蛋白质组,在个体的生命活动中却总是变动不停。人们可以通过确定变化的蛋白来理解其功能。

(4) 时间与空间。DNA 通常位于细胞核内,且保持稳定,因此测定基因组序列不受时空的影响。对于转录的 mRNA 来说,时间是主要的参考因素,在发育的不同阶段或细胞的不同活动时期,mRNA 的表达是不一样的。而在蛋白质组的研究中,不仅要考虑时间因素,更要考虑空间因素。因为,不同的蛋白分布在细胞的不同部位,它们的功能与其空间定位密切相关;而且,许多蛋白在细胞里不是静止不动的,它们常常通过在不同的亚细胞环境里的运动发挥作用。

(5) 孤立行为与相互作用。基因组基因表达的各种 mRNA 是彼此孤立的,互不干扰;但是 mRNA 的产物——蛋白质正好相反,蛋白质与蛋白质之间、蛋白质与其他生物大分子之间有着广泛的相互作用,蛋白质复合体也通过结构型的变化来调整其功能型。

(6) 单一手段与多种技术。对于核酸和蛋白质这类生物大分子来说,测定其序列是首要任务。测定 DNA 的核苷酸序列要远比测定蛋白质的氨基酸序列容易得多。前面已经说过,基因组既无量的变化,也没有质的变化,在基因组研究中,DNA 测序技术是最基本、最重要的工具;但是在蛋白质组研究中,需要的技术远远不止一种,而且技术的难度也远远大于基因组研究技术。简单地说,蛋白质组研究技术可分为两大类:分离技术和鉴定技术。分离技术的难点是不同蛋白质的量和质是不同的,鉴定技术则依赖于已知的蛋白质或基因序列为基础,因此,蛋白质组学的发展是受技术限制的,又是受技术发展推动的。

(7) 互补与互助。蛋白质组研究和基因组研究是形影相随的两个重要领域,蛋白质组的许多工作离不开对基因组的研究。

让我们再来看看更为精辟的论述,Raj Parekh 这样说:基因组(Genome)告诉你理论上能够发生什么;转录组(Transcriptome)告诉你可能发生什么;蛋白质组(Proteome)告诉你正在发生什么。这就是蛋白质组,正在发生的事情,不了解它就不可能了解和认识生命现象的本质。



思考题

查阅以下文献,选择其中你认为最重要的一段翻译成汉语,并谈谈你对蛋白质科学研究的认识。

(1) Nature: And now for the proteome. 2001, 409:747

(2) Science: Proteomics in genomeland. 2001, 291:1221

(3) Science China: Proteomics in China: Ready for prime time. 2010,53(1):22—33

(张艳贞)