

内附光盘

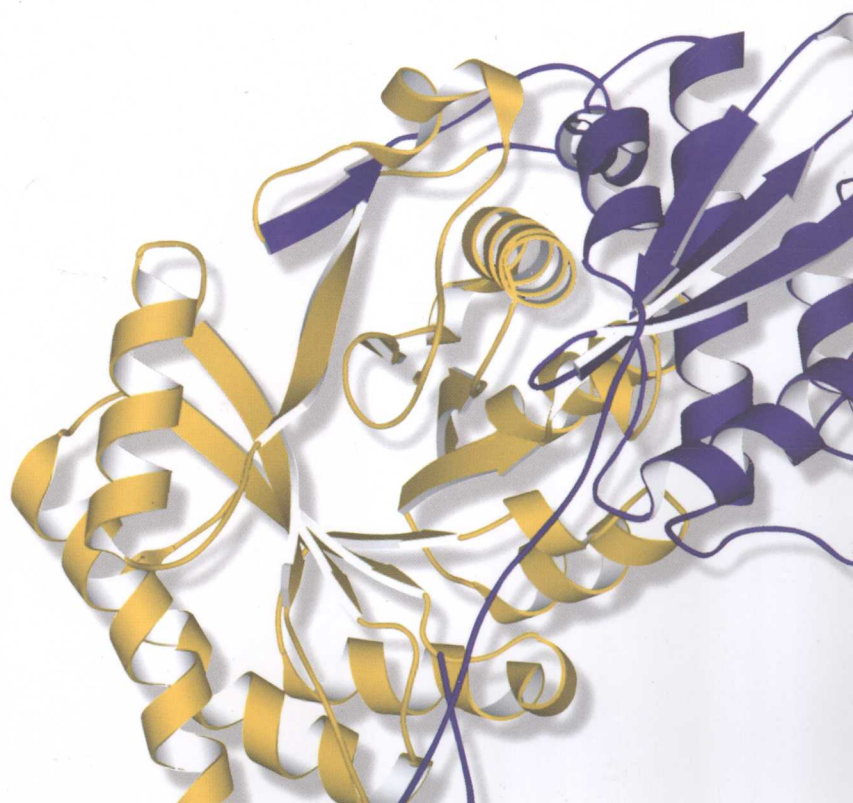


蛋白质结构与功能入门

〔英〕 G.A. 佩特斯科 〔美〕 D. 林格 著
葛晓春 等 译

Proteins
Structure and
Function

 科学出版社
www.sciencep.com



蛋白质结构与功能入门

Proteins Structure and Function

〔英〕 G. A. 佩特斯科 著
〔美〕 D. 林格
葛晓春 等 译

科学出版社

北 京

图字:01-2008-0361 号

内 容 简 介

本书是 Gregory A. Petsko 和 Dagmar Ringe 编写的 *Proteins Structure and Function* 的中文翻译版。全书由五章组成,第一章主要介绍蛋白质各个结构层次上的基本知识,按照一级结构到四级结构的顺序讲述;第二章讲述蛋白质行使功能的结构基础,包括配体结合位点、活性位点行使功能的结构与功能机制;第三章则具体介绍了各种调控蛋白质功能的机制,包括蛋白质磷酸化、糖基化、甲基化、乙酰化、脂质化等各种修饰和蛋白质降解等对蛋白质功能的调控;第四章着重介绍如何利用各种生物信息学手段从序列推测功能,并举例说明如何对一个未知功能的蛋白质展开研究;第五章简要介绍解析蛋白质三维结构的技术,包括 X 射线晶体衍射法和核磁共振法。

本书为从事蛋白质结构与功能研究和教学的教师以及研究人员提供参考,同时还可作为研究生和高年级本科生的学习用书。

© 2007 New Science Press Limited

“First published in Great Britain in 2007 by New Science Press Ltd.”

本书正文中彩图请见随书所配光盘。

图书在版编目(CIP)数据

蛋白质结构与功能入门/(英)佩特斯科(Petsko, G. A)等著,葛晓春等译. —北京:科学出版社, 2009

ISBN 978-7-03-023349-3

I. 蛋… II. ①佩…②葛… III. 蛋白质-研究 IV. Q51

中国版本图书馆 CIP 数据核字(2008)第 172806 号

责任编辑:李 悦 刘 晶/责任校对:李奕莹

责任印制:钱玉芬/封面设计:王 浩

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>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2009 年 4 月第 一 版 开本:890×1240 1/16

2009 年 4 月第一次印刷 印张:13 3/4

印数:1—3 000 字数:412 000

定价:58.00 元(含光盘)

(如有印装质量问题,我社负责调换〈环伟〉)

译者序

蛋白质的研究自 18 世纪以来,涌现出了许多辉煌的成果。蛋白质“protein”一词,源于希腊文“proteios”,这个单词是“头等重要”的意思,显然人们很早就意识到蛋白质对于生命的重要性了。随着研究的深入,大量的事实已经证明蛋白质是细胞功能的执行者,它对于维持生物的正常生存必不可少。目前,大规模基因组测序的热潮逐渐沉寂下来,随后的主要任务就是解析这些海量的顺序信息中编码蛋白质的基因的功能,从基因组时代跨入蛋白质组时代,解析每一个蛋白质甚至是蛋白质复合体的结构和功能,是生物学发展的必然趋势。

在基因组测序所得到的大量信息激励之下,越来越多的科研工作者逐步转向蛋白质研究,分别从个别蛋白质和蛋白质组学等不同水平上开展蛋白质的功能研究工作。蛋白质与生物遗传性状的关系、蛋白质与重大疾病的关系以及蛋白质药物的开发已成为生物学、医学和药学领域的研究热点。但是,完全阐明蛋白质行使功能的机制通常需要它的结构信息,这些结构信息不仅仅为其单独行使功能提供依据,也为蛋白质-蛋白质相互作用提供了结构学上的证据。

目前将蛋白质的结构和功能结合起来讲述的参考书非常之少,其中一个原因是结构和功能两方面都研究得较为清楚的蛋白质并不多。《蛋白质结构与功能》一书从蛋白质结构和功能方面的基础知识开始,逐层深入描述,并且为帮助读者理解,作者还提供了多个典型研究范例,对其中每一个蛋白质行使功能的结构基础,与功能相关的一些构象变化过程都阐述得非常详细。阅读本书不仅可以帮助读者较为全面地了解蛋白质结构与功能方面的基础知识,同时还有助于了解蛋白质各种作用方式的结构基础或行使功能过程中的结构变化等。本书为蛋白质结构与功能研究领域的科研工作者及研究生提供了很好的启示和借鉴作用,对高年级本科生也可起到拓展思路、启发研究兴趣的作用。

与一般的书籍不同,本书的编排非常有特色,每一节的篇幅正好就是一本打开的书左右两页,不用翻到反面就能阅读完整一节的内容。作者还为每一主题配备了大量精美的彩色图片,使读者对所述及的结构细节一目了然,故而在阅读译本时如果同时参照随书所附的彩色图片光盘将更有助于理解。

本书的翻译凝结着多人的辛勤劳动。本书的翻译人员为:葛晓春(序言、第一章和第二章)、李园梨(第三章)、汪海(第四章)、唐茜子(第五章);校译人员为:曹凯鸣(第一章、第二章及第三章)、孙崇荣(第四章)、葛晓春(第五章)。另外,我们实验室的研究生金越、张海磊以及全向荣和尚洁两位老师也参与了书稿的编辑工作。尽管参译人员尽了最大努力,但由于水平有限,欠妥或错误之处在所难免,希望广大读者提出宝贵意见,以便将来能适时修正。

最后,要特别感谢科学出版社及出版社的编辑李悦同志的辛勤劳动和大力支持。

葛晓春
复旦大学
2008 年 12 月

前言

顺序决定结构,结构决定功能——生物学的真正中心法则

过去二十年中,两大突破改变了生物学家思考科学问题的思路,影响了研究实验设计的方法,这两大突破尽管有非常不同的实验基础,但具有相关性。

第一个突破是结构学的突破。观察大分子三维结构模型的技术和计算机操作模型技术的发展,为预测实验结果以及解释大量的实验结果提供了强有力的工具,也使得设计一些原先不可能做到的实验成为可能。结构的解析使得我们的视角可以从酶动力学实验扩展到信号途径的机制。随着结构酶学的出现,现在详细地描述酶的催化过程,甚至包括描述原先可能被忽视的中间物的结构已经变得可能。在此水平上,化学和生物学的珠联璧合既强大又深入,由此形成了结构辅助药物设计的基础。

第二个突破是基因组学的突破。从细菌到人类基因组的所有基因顺序以及居间顺序的测定,为描述细胞整体以及其最基本组成部分的生命活动提供了“准入证”。功能细胞中基因之间的相互关系为阐明细胞生存和死亡的机制提供了基础。现在,我们可以从一个初始信号直到最终的死亡机制来确定细胞的命运,器官和机体中细胞彼此之间的关系最终也都将按照这种机制来了解。

本书的宗旨不是详细描述许多复杂的解析蛋白质结构和功能的实验方法,而是讲述现今已经获得的知识,如何获得这些知识则不是撰写本书的目的。为了方便大家了解一些常用的方法,我们在第四章中也简短地总结了目前用来探索蛋白质功能的一些最重要的技术,在第五章中则概括了解析结构的一些实用方法,这些技术和方法可以帮助找出本书中所阐述问题的答案。例如,蛋白质的化学性质如何与它的三维结构相关?三维结构如何在化学反应中及识别另一个分子时起作用?基因顺序如何不仅仅编码了蛋白质的顺序,更重要的是还决定了蛋白质的三维结构以至最终的功能?我们在本书中试图简明扼要地介绍目前用于解答这些问题的方法。

如果您所喜闻乐见的蛋白质和途径没有在此提到,我们表示歉意。本书抛砖引玉的性质使得我们无法包罗万象。我们试图选择一些有代表性的例子来阐述普遍的规则,但仍然意识到某些情况下是存在例外的,多个例子将有助于解释某一特殊情况。

在此我们对以前的学生 Ezra Peisach 表达谢意,正是由于他孜孜不倦的努力,才有了这些漂亮且意义明确的蛋白质结构图。我们要特别感谢 Eleanor Lawrence 和 Matthew McClements,以及 New Science Press 的工作人员 Amy Austin、Vivien Chen、Karen Freeland 和 Liam Lynch,感谢他们对这本书的耐心帮助和热心支持。最后,向我们的出版商 Miranda Robertson 致以特殊的感谢,正是由于他对这个出版计划的精心呵护才使我们最后得以完成本书。

Gregory A. Petsko
Dagmar Ringe

本书使用说明

《蛋白质结构与功能》一书是按照模块化的原则编写的,其目的是使本书更加适合于作为教学或参考用书,而又不失将其作为一部有教益的好书所必备的综合性的。下面的示意图展现了这些书的模块化结构和特点。每一章分为若干节,每一节以两页纸为单位,涵盖了一个独立的主题以及与之相关的文字、插图、概念和参考文献。每一节又分为若干小节,每一小节由一句标题所引导,这些标题反映了行文的思路和全章的整体逻辑结构。

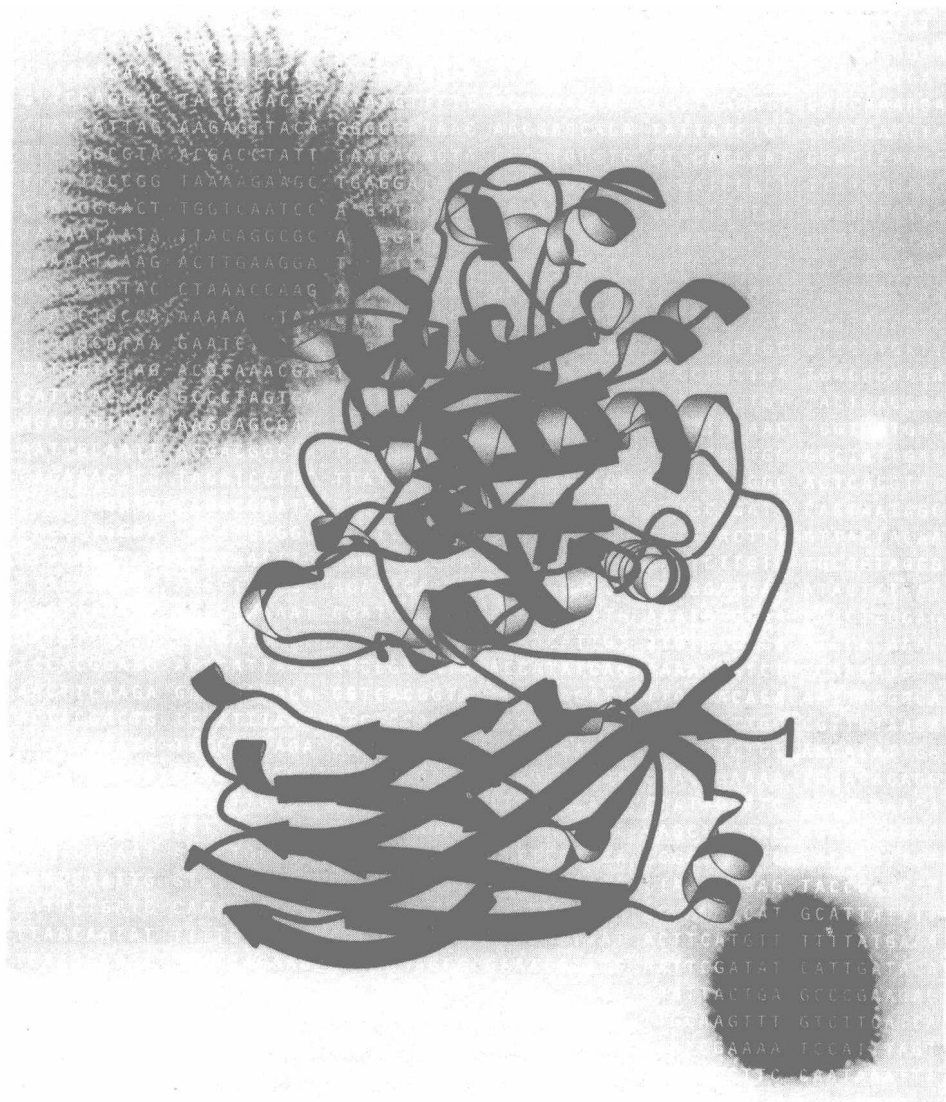
本书模块化的结构和清晰的组织方式使教师很容易选择他们授课时的思路,使学生更容易复习,也使科研工作者更容易在本书中找到他们感兴趣的最新内容,以及尽可能多的关于他们所需的某个主题概念的来龙去脉。

书末集中列出了所有的定义和参考文献,并且注明了它们出现在哪一节或哪几节。词汇表中的定义有时可能会对文中出现的定义作一些有助于理解的详细阐述。参考文献中列出了完整的作者名单。



从序列到结果

激素原加工蛋白酶 Kex2。本图片展示的是酵母菌蛋白酶 Kex2 的氨基酸序列, Kex2 是高等真核生物中二元激素原加工酶的范例。这个序列来自于对编码这个蛋白质的基因所含信息的翻译。图的中间部分是这个蛋白质的三维结构。这一结构的测定工作是由 Todd Holyoak、Mark Wilson、Dagmar Ringe 和 Robert Fuller 合作完成的。蛋白质的结构由氨基酸序列所包含的信息决定。这个蛋白质具有两个域, 位于上方的一个域含有切割激素原的催化位点, 位于下方的这个域功能未知。背景中是病原真菌 *Candida albicans* 的两个株系的照片(经旧金山加利福尼亚大学 Nina Agabian 许可)。这个例子说明如果生物体基因组中没有这种蛋白质, 将会有什么样的结果。上方是野生型株系, 含有有功能的 Kex2 蛋白, 该株系形成了菌丝——这种线状凸起对于菌株的毒性是必需的。下方株系是缺失了两个 *Kex2* 基因拷贝的突变体。它们不能形成菌丝, 且毒性急剧减弱, 说明这个蛋白质的细胞和生物化学功能对于致病性是非常重要的(据 Timothy Fenn 图)。



目 录

译者序

前言

本书使用说明

从序列到结果

第一章 从序列到结构	1
1-0 概述:蛋白质的功能和构造	2
1-1 氨基酸	4
1-2 基因和蛋白质	6
1-3 肽键	8
1-4 稳定折叠蛋白质的键	10
1-5 二级结构的重要性和决定因素	12
1-6 α 螺旋的特征	14
1-7 β 折叠的特征	16
1-8 二级结构的预测	18
1-9 折叠	20
1-10 三级结构	22
1-11 膜蛋白的结构	24
1-12 蛋白质稳定性:弱相互作用和柔韧性	26
1-13 蛋白质稳定性:翻译后修饰	28
1-14 蛋白质的域	30
1-15 蛋白质结构通则	32
1-16 蛋白质基序	34
1-17 α 域和 β 域	36
1-18 α/β 、 $\alpha+\beta$ 和交联域	38
1-19 四级结构:一般原则	40
1-20 四级结构:分子间界面	42
1-21 四级结构:几何学	44
1-22 蛋白质的柔韧性	46
第二章 从结构到功能	49
2-0 概述:蛋白质功能的结构基础	50
2-1 识别、互补和活性位点	52
2-2 柔韧性和蛋白质功能	54
2-3 结合位点的位置	56
2-4 结合位点的性质	58
2-5 结构蛋白的功能特性	60
2-6 催化:概述	62
2-7 活性位点的几何学	64
2-8 邻近性和基态去稳定	66
2-9 稳定过渡态和排斥水	68
2-10 氧化反应	70

2-11	加成/消除、水解和脱羧	72
2-12	活性位点的化学	74
2-13	辅助因子	76
2-14	多步反应	78
2-15	多功能酶	80
2-16	具有通道的多功能酶	82
第三章	蛋白质功能的调控	85
3-0	概述:调节的机制	86
3-1	蛋白质相互作用域	88
3-2	定位调节	90
3-3	pH 和氧化还原环境对蛋白质功能的调控	92
3-4	效应物配体:竞争性结合与协同效应	94
3-5	效应物配体:构象变化与别构效应	96
3-6	基于核苷酸水解反应的蛋白质开关	98
3-7	GTP 酶开关:小信号 G 蛋白	100
3-8	GTP 酶开关:异源三聚体 G 蛋白酶转导的信号	102
3-9	GTP 酶开关:蛋白质合成	104
3-10	马达蛋白开关	106
3-11	受降解作用的调节	108
3-12	蛋白质功能的磷酸化调节	110
3-13	信号蛋白质激酶的调节:激活机制	112
3-14	信号蛋白质激酶的调节:Cdk 的激活	114
3-15	细菌的双组分信号转导系统	116
3-16	蛋白酶水解的调控作用:蛋白质前体的激活	118
3-17	蛋白质剪接:通过内蛋白子自身酶解	120
3-18	糖基化	122
3-19	脂质修饰介导的蛋白质靶向	124
3-20	甲基化、N-乙酰化、SUMO 化和亚硝基化	126
第四章	从顺序到功能:结构和功能基因组学中的案例分析	129
4-0	概述:在基因组学的时代背景下从顺序到功能	130
4-1	序列对位排列和比较	132
4-2	蛋白质特征集	134
4-3	从顺序推测功能	136
4-4	研究蛋白质功能的实验手段	138
4-5	趋异进化和趋同进化	140
4-6	从顺序到结构:同源模建	142
4-7	从顺序到结构:Profile-based threading 方法和“Rosetta”	144
4-8	从结构预测功能:蛋白质超家族	146
4-9	确定结合位点的策略	148
4-10	确定催化残基的策略	150
4-11	TIM 桶:一种结构,多种功能	152
4-12	PLP 酶:多种功能,一种结构	154
4-13	兼职:多功能蛋白质	156
4-14	变色龙序列:一条序列,多种折叠方式	158
4-15	朊病毒、淀粉状蛋白和丝氨酸蛋白酶抑制剂:亚稳定的蛋白质折叠	160

4-16	从未研究过的基因的功能:半乳糖酸脱水酶	162
4-17	从零开始:一个功能未知的基因产物	164
第五章	结构测定	167
5-1	结构信息的解读	168
5-2	利用 X 射线晶体分析法和 NMR 法进行结构测定	170
5-3	晶体和 NMR 结构的质量及表述	172
参考文献		175
定义		183
索引		191

第一章 从序列到结构

基因组学的突破带来以指数增长的基因序列,将这些序列信息变成由它们所编码产物的功能信息是后基因组生物学所要面临的艰巨任务。这个过程的第一步通常是从蛋白质序列推断折叠而成的三维结构。本章总结了序列和结构之间关系的基本概念,并概述了蛋白质的构造。

- 1-0 概述:蛋白质的功能和构造
- 1-1 氨基酸
- 1-2 基因和蛋白质
- 1-3 肽键
- 1-4 稳定折叠蛋白质的键
- 1-5 二级结构的重要性和决定因素
- 1-6 α 螺旋的特征
- 1-7 β 折叠的特征
- 1-8 二级结构的预测
- 1-9 折叠
- 1-10 三级结构
- 1-11 膜蛋白的结构
- 1-12 蛋白质稳定性:弱相互作用和柔韧性
- 1-13 蛋白质稳定性:翻译后修饰
- 1-14 蛋白质的域
- 1-15 蛋白质结构通则
- 1-16 蛋白质基序
- 1-17 α 域和 β 域
- 1-18 α/β 、 $\alpha+\beta$ 和交联域
- 1-19 四级结构:一般原则
- 1-20 四级结构:分子间界面
- 1-21 四级结构:几何学
- 1-22 蛋白质的柔韧性

1-0 概述:蛋白质的功能和构造

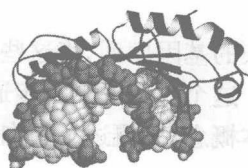
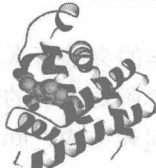
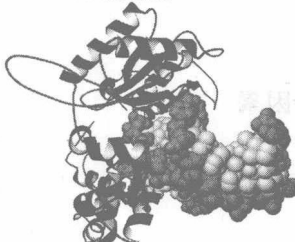
结合 特异性地识别其他分子是蛋白质功能的关键。结合分子(配体)可大可小,小的如与肌红蛋白血红素基团协同结合的氧分子;大的如结合在 TATA 结合蛋白上并被弯曲的特定 DNA 顺序(TA-TA box)。特异性的结合受形态互补或者如氢键那样的极性相互作用支配。	TATA结合蛋白  TATA结合蛋白结合特定的DNA顺序,并为负责起始遗传信息转录的复合物的结合提供平台。(PDB 1tgh)	肌红蛋白  肌红蛋白的血红素基团(灰色)上的铁原子(绿色)可逆地结合一分子氧,它可以在肌肉组织中储存氧以备使用。(PDB 1a6k)
催化 基本上活细胞中的每一个化学反应都是被催化的,大多数的催化剂是酶蛋白。酶的催化效率是显著的,与仅有缓冲液相比,反应可以加快 17 个数量级。酶的许多结构特性都为催化力做出了贡献:将反应基团按利于反应的方向聚到一起(相邻性);更紧密地结合反应的过渡态而不是基态(稳定过渡态);酸-碱催化等。	DNA聚合酶  DNA复制由一个特殊的聚合酶催化,它可以拷贝遗传物质,并校正拷贝中错误的地方。(PDB 1pbx)	HIV蛋白酶  艾滋病病毒HIV的复制有赖于一个被称为HIV蛋白酶的蛋白质剪切酶的作用。这个酶是蛋白酶抑制剂药物的靶点(灰色)。(PDB 1a8k)
开关 蛋白质是柔性分子,其构象可以随着 pH 的变化或者结合配体而改变。这种变化可以作为控制细胞过程的分子开关。对许多肿瘤发生的分子基础而言十分重要的一个例子是,当 GTP 水解为 GDP 时,小 GTP 酶 Ras 发生的构象变化:结合 GTP 的构象是促使细胞生长的“开”信号;而结合 GDP 的构象是“关”信号。	“关”  Ras的GDP-结合状态(“关”, PDB 1p1l)与GTP-结合状态(“开”, PDB 121p)有本质的不同。这种差异使得这两种状态可以被信号转导途径中不同的蛋白质所识别。	Ras “开” 
结构蛋白 蛋白质分子可作为活体中一些主要结构的元件。这种功能来自于蛋白质亚基自身之间,以及与其他蛋白质、碳水化合物等的特异性结合,使得像肌动蛋白丝这样的复杂系统也可以自发组装。结构蛋白也是生物材料如丝、胶原、角蛋白的重要来源。	丝  丝的强度和柔性来自于它的结构——一个由反平行 β 折叠片层组成的巨大堆叠物。它的强度来自于每个片层之间的共价键和氢键;柔性来自于将片层聚在一起范德华力。(PDB 1slk)	F-肌动蛋白  肌动蛋白丝对于肌肉收缩以及细胞骨架都是非常重要的,它们是肌动蛋白和肌动蛋白结合蛋白的螺旋状组装物。(Ken Holmes提供)

图 1-1 蛋白质行使生物化学功能的 4 个例子

蛋白质是细胞中最具多功能性的大分子

本书主要讲述蛋白质的各种功能以及蛋白质结构如何决定这些功能。“蛋白质功能”既可能意味着分子单独存在时的生物化学功能;也可能意味着它作为一个组装体,或者与其他分子结合的复合体的一部分时所行使的细胞学功能;或者意味着它在细胞或生物体中产生的表型。

蛋白质生物化学功能的主要例子包括结合、催化、作为分子开关以及作为细胞和机体的结构组分(图 1-1)。

蛋白质可能结合到其他大分子上,如 DNA 是 DNA 聚合酶和基因调控蛋白的结合对象;或者是结合到蛋白质分子上,如载体蛋白或受体与信号分子的结合。这种功能利用了蛋白质呈现的多种结构和表面的化学性质不同,这些表面可以同其他分子进行高度特异性的结合;催化不仅仅需要与底物或调节分子进行特异性结合,还需要特异的化学反应性;受调控的酶和开关,例如,起信号作用的 G 蛋白(它是被调控的催化 GTP 水解的酶),需要大尺度的构象变化,这种变化取决于结构稳定性和柔韧性的微妙平衡;结构蛋白可能像丝一样强健,或像毛发、角质和皮革中的角蛋白组分一样坚韧和耐久,还可能像肌动蛋白和微管蛋白那样具有依赖核苷酸水解的复杂动力学性质。这种非同寻常的功能多样性来自于构成它的氨基酸侧链的化学性质的多样性、多肽链的柔韧性以及具有不同氨基酸组成的多肽链多种多样的折叠方式。

蛋白质结构有 4 个水平

蛋白质是由 20 种不同的氨基酸经肽键连接而成的多聚物。在生理温度的水溶液中,蛋白质多肽链多半会折叠成为一个球形[图 1-2(c)]。蛋白质中不同的氨基酸顺序,是直接由基因的核苷酸顺序编码的,称之为一级结构[图 1-2(a)]。一级结构同时也决定了蛋白质如何折叠成更高级的结构。多肽链的二级结构可以是 α 螺旋或者 β 折叠,它们都是通过多肽链主链(**backbone**)氨基酸上恒定部分的 N—H 和 C=O 基团之间有规律的氢键相互作用形成的[图 1-2(b)]。在球状蛋白质中, α 螺旋、 β 折叠或者二者一起,连同本身无二级结构的环和连接(loop and link)折叠成三级结构[图 1-2(c)]。许多蛋白质由一个以上折叠好的多肽链结合而成,构成了蛋白质的四级结构[图 1-2(d)]。

行使蛋白质功能的多肽链,通常必须在生理条件下可以形成(或折叠成)稳定的三级结构。另外,蛋白质起作用时需要折叠的蛋白质不能太刚硬。可能由于这些限制,蛋白质采用的折叠数目,尽管很大,却是有限的。有限的折叠数究竟是反映物理束缚力,还是仅仅反映从一个已存在的稳定折叠进行趋异进化是有利的,目前还不得而知,但这一点却是具有应用价值的:如果自然状态下许多可能的稳定折叠没有表现出来,那么利用这点为工业和医用生产全新的蛋白质将是可能的。

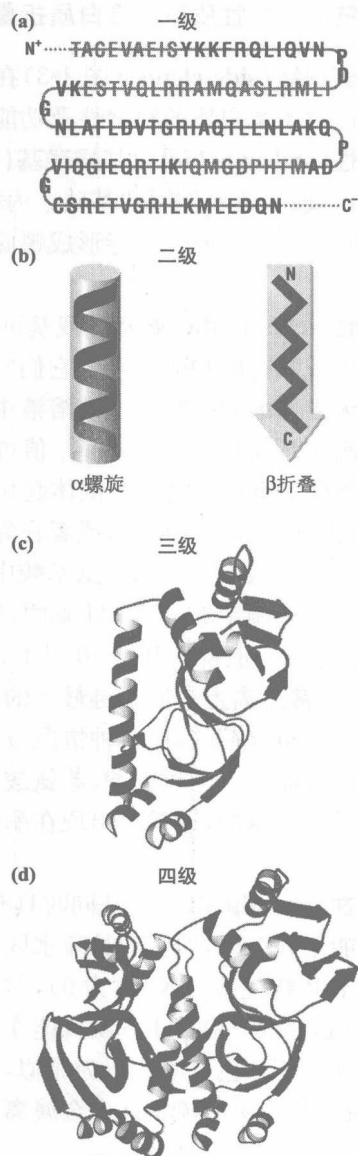


图 1-2 分解物活化蛋白的蛋白质结构水平示意图。(a)蛋白质的氨基酸顺序(一级结构),包含了所有特定信息;(b)形成氢键的主链构象的有规则的重复方式(二级结构),如 α 螺旋(红)和 β 折叠(蓝);(c)这些元件包装在一起形成蛋白质总体折叠(三级结构)的方式(PDB 2cgp);(d)两条或更多条多肽链相对排列称为四级结构。(PDB 1cgp)

定义:

个氨基酸的酰胺—N—H、 α 碳—C—H 和羰基—C=O

主链:多聚物的规律性重复的部分。在蛋白质中它包含了每 基团。

1-1 氨基酸

氨基酸侧链的化学性质对于蛋白质折叠和功能具有重要影响

氨基酸侧链(side chain)(图 1-3)在参与氨基酸间相互作用以及和水的相互作用时有不同的倾向性,这些不同特性对于蛋白质的稳定性及功能都造成了复杂的影响。

疏水性(hydrophobic)氨基酸残基(residue)仅仅参与范德华相互作用。它们倾向于聚在一起,避免和水分子接触,构成了疏水作用的基础。丙氨酸和亮氨酸强烈倾向于形成螺旋,而脯氨酸则很少出现在螺旋中,这是因为它的主链氮原子不能形成螺旋所需要的氢键。苯丙氨酸的芳香族侧链有时可以参与弱的极性相互作用。

亲水性(hydrophilic)氨基酸残基可以在彼此之间,或者与肽链主链、极性有机分子和水分子之间形成氢键。这种形成氢键的趋势主导了它们所参与的相互作用。有些氨基酸还可以根据 pH 条件和微环境改变电荷状态。天冬氨酸和谷氨酸在水溶液中的 pK_a 值接近 5,所以在 pH 7 时通常去质子化并带负电荷,但在蛋白质内部疏水环境中,它们的 pK_a 值可能变化到 7 甚至更高(负电荷在附近也会造成相同的效应),这样在生理 pH 条件下就可作为质子供体起作用。赖氨酸也有类似的情况,在水中它的 pK_a 值大于 10,通常是带正电荷的,但在非极性环境中,或者在邻近正电荷的影响下,它的 pK_a 值可能低于 6,导致中性化并可能成为质子受体。组氨酸可能是所有氨基酸中最具多功能性的分子了,这就是它常在酶的活性位点出现的原因。它具有两个可被滴定的一N—H 基团,每个的 pK_a 值都在 6 左右。当其中一个一N—H 基团丢失一个质子时,另一个的 pK_a 值将上升到 10 以上;当两个都被质子化时,氨基酸整体上就是带正电荷的;当只有一个被质子化时(通常是离蛋白质主链最远的那一个),侧链是中性的,兼备质子供体和质子受体的特性;完全去质子化时则是带负电荷的,但这种情况极少发生。精氨酸在中性 pH 条件下总是完全质子化的,其正电荷主要位于胍基的碳原子上。丝氨酸、苏氨酸、谷氨酰胺和天冬酰胺不会电离但可以同时成为氢键的受体或供体。半胱氨酸类似组氨酸,通常也出现在酶的活性位点,这是因为在天然的氨基酸中巯基阴离子是最强的亲核基团。

两亲性(amphipathic)残基同时具有极性和非极性,是形成界面的理想分子。将赖氨酸的带电侧链视为两亲性可能令人惊讶,但它的长疏水区常常参与和其他疏水侧链的范德华相互作用。酪氨酸在生理 pH 条件下一般不电离(它的 pK_a 值为 9),但在某些酶的活性位点它可能参与酸碱反应,因为其所处环境可以降低它的 pK_a 值,它的一O—H 基团既能作为氢键供体也能作为氢键受体起作用,而且它的芳香环也可以形成弱的极性相互作用。色氨酸的行为相似,但吲哚的一N—H 基团不会电离。甲硫氨酸是两亲性氨基酸里极性最弱的,硫醚基团中的硫是许多金属离子极佳的配体。

定义:

两亲性:同时具有极性和非极性,故易在疏水性和亲水性分子间形成界面。

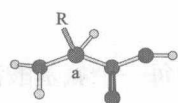
亲水性:易同水发生相互作用。亲水性分子是极性或带电的,因而水溶性强。在多聚物中,亲水侧链倾向于通过氢键同其他亲水侧链或者水分子结合。

疏水性:倾向于避开水。疏水性分子是非极性、不带电的,故

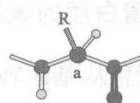
而在水中是相对不溶的。在多聚物中,疏水性侧链倾向于相互结合以减少同水或极性侧链的接触。

残基:是多聚物的基本组成单位,是连接多聚物组分的键被打断时释放出的片段。在蛋白质中,残基是指氨基酸。

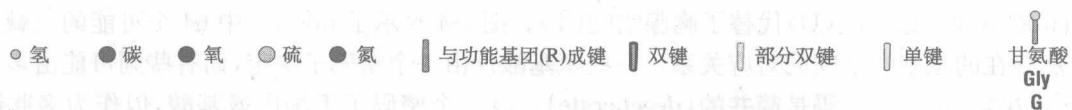
侧链:是指从多聚物重复的主链上伸展出去的化学基团。在蛋白质中,侧链与主链的 α 碳原子键合,赋予了 20 个氨基酸中每个氨基酸独特的化学性质。



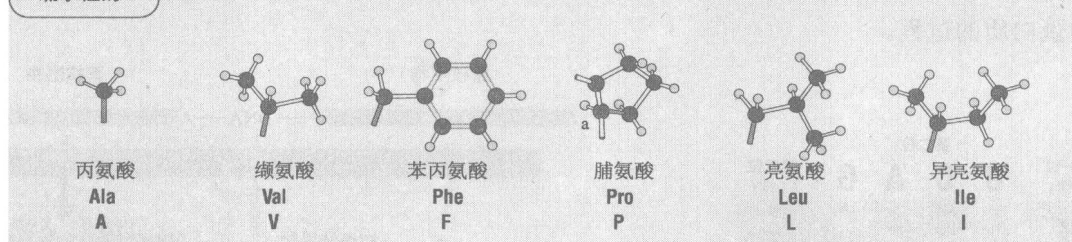
氨基酸的化学结构。所有氨基酸的主链都是一样的，由氨基基团($-NH_2$)、 α 碳原子和羧基基团组成。不同氨基酸的侧链R不一样。图示一个氨基酸的中性形式：在pH7的溶液中氨基和羧基电离为 NH_3^+ 和 COO^- 。除了甘氨酸的 $R=H$ 外，氨基酸是手性分子(就是说，它们是左右不对称的)。图中显示的是最常见的L构型。



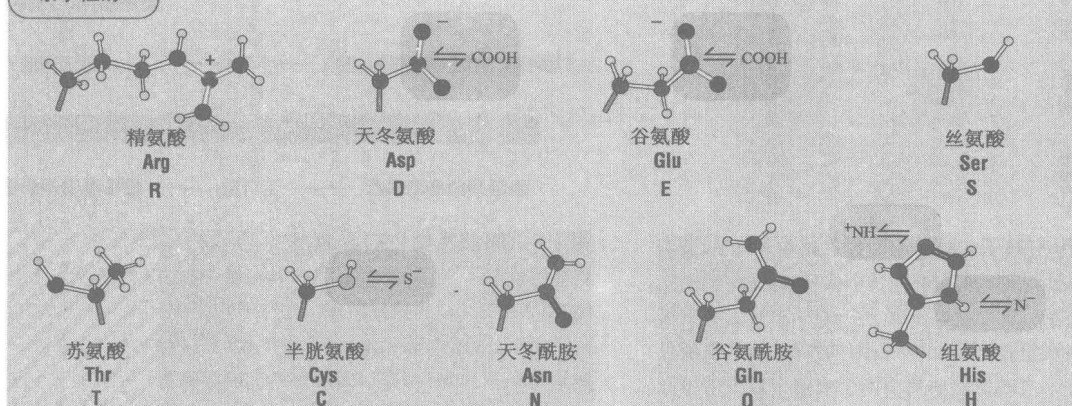
多肽链中的氨基酸残基，R基团就是侧链。下图示意的是蛋白质中20种不同氨基酸侧链。脯氨酸的侧链回折过来结合到主链氮原子上。对蛋白质中大多数氨基酸而言， α 碳原子的构型是L构型。



疏水性的



亲水性的



两亲性的

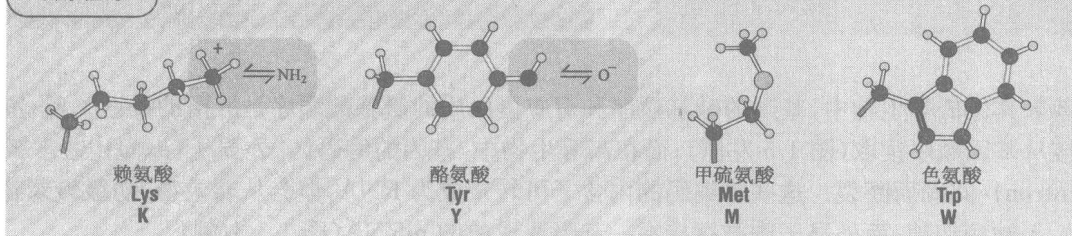


图 1-3 氨基酸结构以及氨基酸侧链的化学性质。对于带电侧链，显示 pH 7 情况下的主要存在形式。对于脯氨酸，图中显示了氮原子和 α 碳原子，因为侧链与氮原子结合形成的环中包含了这两个原子。

1-2 基因和蛋白质

基因上的 DNA 碱基顺序与它所编码的蛋白质的氨基酸顺序呈线性相关

遗传密码(genetic code)是将遗传信息从基因转变为蛋白质的公式。蛋白质中的每一个氨基酸都由基因上的三个连续核苷酸所组成的密码子(codon)表示。DNA 具有 4 种不同的核苷酸(nucleotide),分别含有 4 种不同的碱基(base):腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),基因中这些碱基的顺序决定了蛋白质中的氨基酸顺序,即蛋白质的一级结构(primary structure)。DNA 的核苷酸顺序转录(transcribe)成为信使 RNA(mRNA)时,尿嘧啶(U)代替了胸腺嘧啶(T)。图 1-4 显示了 mRNA 中 64 个可能的三碱基密码子和 20 个自然存在的氨基酸之间的对应关系。一些氨基酸仅由一个密码子决定,而有些则可能由多至 6 个不同的密码子所决定,即遗传密码是简并的(degenerate)。有三个密码子不编码氨基酸,但作为多肽链的终止信号[终止密码子(stop codon)]。图 1-5 概述了 DNA 上的核苷酸顺序首先转录成为 RNA 然后翻译(translate)成为蛋白质的过程。

第二位的:					第三位 (3'端)
第一位 (5'端)	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Tyr	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

图 1-4 遗传密码子。64 个可能的三碱基密码子,分别编码氨基酸或者基因编码部分的结束信号(终止密码子)。粉红色的氨基酸具有非极性(疏水性)侧链;蓝色的具有极性或者带电荷的侧链;浅紫色的具有两性侧链;甘氨酸没有侧链。几乎所有的氨基酸都由两个或两个以上的不同密码子所决定,这些密码子只在第三位有所不同。密码子其他位置的单碱基变化通常产生具有相似物理化学特性的不同氨基酸。

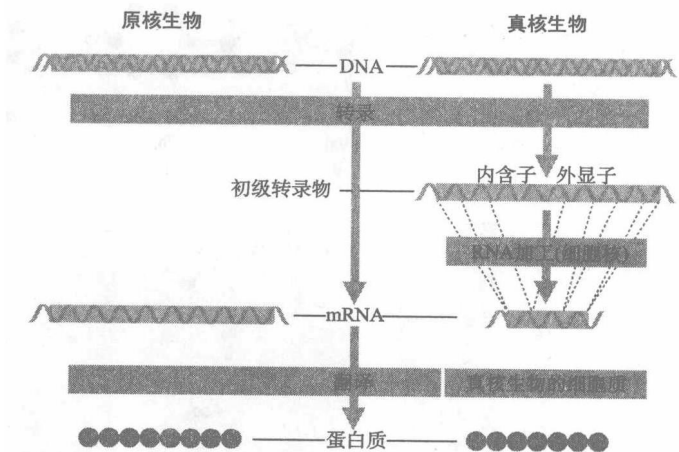


图 1-5 原核生物(左)和真核生物(右)的遗传信息流。蛋白质的氨基酸顺序由 DNA 的碱基顺序所编码。遗传信息先转录成为互补的信使 RNA(mRNA)的碱基顺序。在原核生物中,mRNA 通常直接由 DNA 顺序(图的左侧)转录而来;而真核生物的基因(右侧)通常被一个或多个被称为内含子的非编码居间顺序所间断,内含子和外显子一起转录成为初级转录物,然后在细胞核中内含子被切除,编码区域即外显子被连在一起形成 mRNA。最后,mRNA 的碱基顺序在核糖体中被翻译成为相应的氨基酸顺序,这一步发生在真核生物的细胞质中。(图中不代表比例)

在细菌和其他低等生物中,基因的碱基顺序和对应蛋白质的氨基酸顺序是严格的线性关系:蛋白质的顺序可以直接从基因顺序读取(图 1-5 左侧);而在高等生物中,基因的编码区[外显子(exon)]通常被非编码区[内含子(intron)]间隔而断裂。这些非编码的内含子可转录成为 RNA,但会从转录物[初级转录物(primary transcript)]上被酶切掉,而外显子则被拼接在一起形成成熟的 mRNA(图 1-5 右侧)。

在进化的过程中出现了可变剪接(alternative splicing),通过初级转录物上内含子和外显子不同的剪切和连接方式,可以形成一种以上的 mRNA,并进而翻译成一种以上的蛋白质。依据内含子的不同排列方式,可变剪接可以形成截短蛋白质、中间成段氨基酸顺序不一样的蛋白质,以及由于移码造成的大部分顺序完全不同于可读框正确时的蛋白质。编码顺序也可以通过 RNA 编辑(RNA editing)被改变。在 RNA 编辑过程中,有些核苷酸被变换成其他的核苷酸;成段的核苷酸可在翻译之前插入到 mRNA 中。以这些方式通过 RNA 加工对编码顺序造成的改变,使得阐明基因顺序所编码的蛋白质结构更为复杂化,但这种复杂化并不适用于 cDNA 顺序,因为 cDNA 顺序只是人为的 mRNA 的逆转录拷贝。

遗传密码的构成反映了氨基酸的化学分类

氨基酸根据其物理化学性质可分为几类(图 1-3)。遗传密码子的构成反映了这种分类,如图 1-4 所示。注意密码子的第三位的单碱基变化[**单核苷酸多态性 (single-nucleotide polymorphism)**]通常产生同样的氨基酸。密码子其他位置的单碱基变化通常产生不同的氨基酸,但具有相同的物理化学特性:如第二位的碱基决定了氨基酸是极性还是非极性的,这种变化被称为**保守性替换 (conservative substitution)**,在比较蛋白质顺序时通常用保守性替换来表示两个被比较蛋白质间结构上的保守性。在一个大的进化距离上调查相同基因产物的蛋白质顺序就可以证实这一规律(图 1-6)。蛋白质顺序中某一特定位置的氨基酸在不同物种中常常变化为具有相似物理化学特性的氨基酸残基,正如通过遗传密码子的构成所预见的一样。

	Gly	Val	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly																				
Ala	58																			
Val	10	37																		
Leu	2	10	30																	
Ile		7	66	25																
Met	1	3	8	21	6															
Cys	1	3	3		2															
Ser	45	77	4	3	2	2	12													
Thr	5	59	19	5	13	3	1	70												
Asn	16	11	1	4	4			43	17											
Gln	3	9	3	8	1	2		5	4	5										
Asp	16	15	2		1			10	6	53	8									
Glu	11	27	4	2	4	1		9	3	9	42	83								
Lys	6	6	2	4	4	9		17	20	32	15		10							
Arg	1	3	2	2	3	2	1	14	2	2	12	9		48						
His	1	2	3	4			1	3	1	23	24	4	2	2	10					
Phe	2	2	1	17	9	2		4	1	1					1	2				
Tyr		2	2	2	1		3	2	2	4			1	1		4	26			
Trp				1				2							3		1	1		
Pro	5	35	5	4	1		1	27	7	3	9	1	4	4	7	5	1			

图 1-6 不同生物体中同一蛋白质中的氨基酸被其他氨基酸所替代的频率表。数字越大,表明替换越频繁。例如,甘氨酸通常被丙氨酸所取代,反之亦然;由于这两种氨基酸的侧链最小,故而这种取代具有化学意义。相似地,天冬氨酸和谷氨酸这两种带负电的氨基酸,频繁地相互取代。也有一些意外,例如,丝氨酸和脯氨酸经常相互取代,谷氨酸和丙氨酸也是一样。丝氨酸取代脯氨酸是因为其侧链的 OH 可以与自身主链上的 NH 形成氢键,类似于脯氨酸的闭合环。

定义:

可变剪接:在 RNA 加工过程中,通过移去含有或影响编码顺序的部分 RNA 从而从基因中选择不同的编码顺序。

碱基:核苷酸上与糖基相连的芳香族基团。

密码子:DNA 或 RNA 链上表示一个特定氨基酸或终止基因转录物翻译信号的三个相连的核苷酸。

保守性替换:一个氨基酸被另一个具有相似化学和(或)物理性质的氨基酸所代替。

简并的:一个氨基酸具有一个以上的密码子。

外显子:一个基因的编码顺序(与内含子相比较而言)。

遗传密码:A、U(或 T)、G 和 C 中的三个字母组合而成的 64 种可能性中,每一种与组成蛋白质的 20 种天然氨基酸之间的对应关系。

内含子:基因内的非编码 DNA。

信使 RNA(mRNA):由基因转录而来,去掉了内含子并编辑

过后的 RNA 分子。

核苷酸:核酸多聚物的基本重复单位。它由碱基[A、U(RNA 中)、T(DNA 中)、G 或 C]、核糖(RNA 中是核糖, DNA 中是脱氧核糖)和一个磷酸基团组成。

一级结构:多肽链的氨基酸顺序。

初级转录物:从基因直接转录而来还未经过加工的 RNA 分子。

RNA 编辑:通过酶对 RNA 碱基顺序进行改变。

单核苷酸多态性(SNP):密码子中一个碱基的突变。

终止密码子:意味着编码顺序结束的密码子,通常终止翻译。

转录:通过依赖 DNA 的 RNA 聚合酶从 DNA 的编码链上合成 RNA。

翻译:遗传信息从 mRNA 上的密码子顺序转换成多肽链上的氨基酸顺序。