

目 录

第一章 序 言	1
一、蛋白质顺序分析的意义	1
二、历史回顾与现状	2
三、展 望	10
第二章 蛋白质顺序测定的一般方法、步骤和基本战略	14
一、蛋白质的肽链结构	14
(一) 蛋白质的结构单位——氨基酸	14
(二) 蛋白质顺序的含意	17
二、样品的纯度要求	18
三、顺序测定的一般方法和步骤	19
(一) 顺序测定前的准备	19
(二) 二硫键的断裂	19
(三) 顺序测定的一般方法和步骤	21
四、顺序测定的基本战略	22
第三章 氨基酸组成分析	25
一、引 言	25
二、氨基酸自动分析的原理及有关仪器简介	25
(一) 氨基酸自动分析仪的一般构造及功能	25
(二) 氨基酸自动分析的原理	28
三、蛋白质样品的水解处理	33
(一) 盐酸水解法	34
(二) 磷酸水解法	37
(三) 碱水解法	39
(四) 酶水解法	39
四、氨基酸组成分析的误差纠正	39
五、氨基酸残基数的计算	40
第四章 蛋白质的裂解	44
一、引 言	44
二、蛋白质的酶裂解	44
(一) 简 述	44
(二) 几种常用的肽链裂解酶	45
(三) 近年来发现的几种高特异性的肽链内切酶	50
(四) 实验方法	52
三、蛋白质的化学裂解	54
(一) 溴化氰裂解	55

(二) 羧肽裂解	57
(三) 色氨酸位点的裂解	59
(四) 部分酸水解	61
第五章 肽段的分离纯化	65
一、引言	65
二、凝胶过滤层析	65
(一) 基本原理	65
(二) 凝胶的类型	67
(三) 实验方法	69
三、离子交换层析	74
(一) 一般原理	74
(二) 离子交换剂的种类	75
(三) 实验方法	80
四、肽的薄膜层析分离	83
(一) 基本原理和一般情况介绍	83
(二) 实验方法	85
五、SDS 凝胶电泳	86
第六章 末端氨基酸残基的测定	89
一、N 末端测定	89
(一) FDNR 分析法	89
(二) DNS 分析法	95
(三) Edman 降解法	97
(四) DABITC 法	97
(五) 氨肽酶法	97
(六) N 末端氨基受封闭时的测定	98
二、C 末端测定	102
(一) 肼解法	102
(二) 乙内酰硫脲法	104
(三) C 末端选择性氘标记法	104
(四) 减数 C 末端测定	105
(五) 还原法	106
(六) 羧肽酶法	106
第七章 手工微量顺序测定技术	111
一、手工 Edman 化学降解法	111
(一) 原理	111
(二) 实验方法	113
(三) PTH-氨基酸的鉴定	114
二、DNS-Edman 测顺序法	120
三、减数 Edman 测顺序法	121
四、DABITC/PITC 双偶合法	122
(一) 引言	122

(二) 原理	123
(三) 实验方法	124
五、氨基酸降解法	129
第八章 固相顺序分析技术	133
一、载体的选择及制备	133
(一) 载体的选择	133
(二) 载体的制备	136
二、固定蛋白质或多肽的原理及方法	137
(一) 对苯二异硫氰酸 (DITC) 法	137
(二) 高丝氨酸内酯法	141
(三) 水溶性碳二亚胺 (EDC) 法	144
三、双偶合法在手工固相顺序测定中的应用	146
(一) 用 EDC 方法固定胰岛素 A 链	146
(二) 双偶合手工固相法顺序测定	147
(三) TLC 鉴定及结果	147
四、固相法测顺序的优缺点	149
(一) 固相法的优缺点	149
(二) 固相法测顺序的战略运用	149
第九章 蛋白质自动顺序分析仪	155
一、自动液相顺序分析仪	155
(一) 引言	155
(二) 液相顺序仪的一般构造原理	156
(三) 蛋白质和小肽样品在液相顺序仪中的降解	157
(四) 携带体 Polybrene 引入反应杯	158
二、自动固相顺序仪	159
(一) 引言	159
(二) 固相顺序仪的一般构造原理	159
(三) 固相顺序仪的微量顺序分析	161
三、气相顺序仪	162
(一) 引言	162
(二) 气相顺序仪的一般构造原理	165
第十章 高效液相色谱在蛋白质顺序分析中的应用	167
一、引言	167
二、高效液相色谱仪的一般构造原理	167
三、利用高效液相色谱技术分离纯化蛋白质和多肽	172
(一) 一般说明	172
(二) 应用示例	175
四、Edman 降解产物的高效液相色谱	180
(一) PTH-氨基酸高效液相色谱	180
(二) DAETH-氨基酸高效液相色谱	185
(三) DNS-氨基酸的超微量高效液相色谱	186

第十一章 顺序分析技术在异常血红蛋白结构研究中的应用	189
一、引 言	189
二、异常血红蛋白化学结构分析	194
(一) 血液标本的采集和溶血液的制备	194
(二) 珠蛋白的制备	195
(三) 异常肽链的分离及其氨乙基化	195
(四) 胰蛋白酶水解	196
(五) 酶解片段肽的分离	199
(六) 异常肽的二次酶解和分离	203
(七) 不溶性胰蛋白酶酶解肽的处理	203
(八) 异常肽的氨基酸组成分析和氨基酸顺序测定	204
附录一 常用英文名字缩写	205
附录二 部分试剂和溶剂的纯化	206
附录三 氨基酸的一些物理常数	207
附录四 缓冲溶液的配制	209
(一) 几种特殊缓冲液的配制	209
(二) 常用缓冲液的配制	209
(三) 标准缓冲液的配制	214
附录五 常用数据表	215
(一) 某些有机溶剂的主要物理常数	215
(二) 实验室中常用酸碱的比重和浓度的关系	216
(三) 常用固态化合物的当量浓度(或 mol 浓度)配制参考表	216
(四) 常用酸碱的强度	217
附录六 硫酸铵饱和度常用表	218
(一) 调整硫酸铵溶液饱和度计算表	218
(二) 调整硫酸铵溶液饱和度计算表	218
(三) 不同温度下的饱和硫酸铵溶液	219
附录七 冷却剂和干燥剂	220
(一) 冷却剂	220
(二) 干燥剂	220

第一章 序 言

一、蛋白质顺序分析的意义

蛋白质是构成生物体的一类十分重要的有机含氮化合物,是生命的物质基础。蛋白质是多种多样的,并且行使不同的功能。例如,新陈代谢中的各种化学反应就是在多种特异的蛋白质酶的催化下进行的;作为调节代谢过程的激素、防御外来物侵袭的抗体以及与遗传控制有关的核蛋白等,也都是由蛋白质或它们的衍生物构成的;生命现象的各种活动,如呼吸、运动、营养输送、神经传导、记忆思维等,也是通过蛋白质来实现的。蛋白质的不同功能是以其特定的结构为基础的。所有蛋白质不论功能和来源如何,均由二十种基本氨基酸组成。氨基酸按不同的顺序排列,构成蛋白质的一级结构,在此基础上建立起相应的二级、三级以至四级结构。不同的蛋白质行使不同的功能,这是通过其特定的结构来实现的。一级结构决定高级结构,蛋白质的功能归根结底是它们一级结构的反映。因此,弄清不同蛋白质的氨基酸顺序,阐明每种顺序与蛋白质的功能及其与生物进化、遗传变异的关系,这是当代生物化学的重要组成部分^[1]。事实证明,通过蛋白质顺序分析的研究,加深了对生命科学的了解,推动了分子生物学的飞跃发展,同时对国计民生也作出了一定的贡献。

(1) 蛋白质氨基酸顺序的研究使古老的进化论研究获得了新生。从某种意义上,蛋白质一级结构可作为“超微化石”。Margoliash, E. 及其同事^[2]从五十多种不同生物种属中提取细胞色素 c,全面研究了它们的氨基酸顺序变异,揭示了蛋白质分子在生物进化中的变化。他们根据这些细胞色素 c 的氨基酸变异,绘制了种属发生树,显示出这些物种之间的进化关系(图 1-1)。

(2) 蛋白质氨基酸顺序研究揭示了某些蛋白质在进化上可能来自一个共同的祖先。如只有单条多肽链的肌红蛋白与有四条多肽链的血红蛋白具有相当多的顺序同一性,两种蛋白质在它们的亚铁血红素基团上都有可逆地与氧结合的能力,只是前者来自肌肉,后

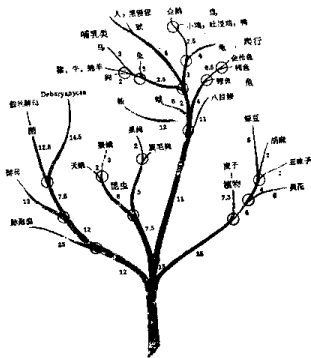


图 1-1 表示细胞色素 c 进化过程的种属发生树。这是用计算机从许多种属的细胞色素 c 氨基酸排列顺序计算绘制的。每一个圆圈表示一种细胞色素 c 的顺序是从该属发出的各分支种属的祖先。每一分支旁边的数字表示从祖先的每 100 个氨基酸中的变异残基数。因此绿豆的细胞色素 c 与它的祖先比较,每 100 个氨基酸中有 5 个不同,而芝麻细胞色素 c 则每 100 个中只有 2 个不同。引自 A. L. 伦宁格著,任邦哲等译:《生物化学》细胞结构与功能的分子基础,上册,科学出版社, p. 99, (1981)。

者来自红细胞,因此它们有可能来自同一祖先。这个祖先很可能具有一条单链及一个单纯的亚铁血红素基团^[4]。以上例证还揭示了蛋白质在进化上的另一个重要现象——基因复制。

(3) 在医学上,分子病就是在研究蛋白质氨基酸顺序中发现并提出来的,从而在蛋白质分子水平上了解病变的机理。如镰刀状贫血患者的血红蛋白与正常人的血红蛋白分子比较,仅在 β 链上存在一个氨基酸的差异,即第六位的谷氨酸突变成缬氨酸,导致功能上的缺陷^[5]。最近报道,美国国立癌症研究所 Barbacid 和麻省理工学院的 Weinberg 发现,膀胱癌细胞中分子量为 21,000 的一种 P_{21} 蛋白,与正常细胞的蛋白质相比,也仅仅是一个氨基酸的差异,即正常的第 12 位甘氨酸被缬氨酸所取代,因而发生癌变。现已清楚,氨基酸残基的置换是基因突变造成的,这一重大发现将为癌研究开辟新的前景。

(4) 蛋白质顺序研究在医学上的意义还在于为合成某些生物类药物提供依据。有了一级结构的分析基础,我们就有可能在多肽药物方面设计合成新的药物。我国在一种多肽性激素即促黄体释放激素一级结构阐明的基础上,利用多肽合成技术解决了该激素的生产技术,在淡水养殖上推广应用发挥了很大经济效益。

以上事例说明,在生物学和医学研究中,蛋白质顺序分析是揭示生命本质、阐明结构与功能的关系、研究酶的活性中心和蛋白质的高级结构的基础,也是当代前沿科学——基因工程中研究基因表达、克隆和 DNA (或 RNA) 顺序分析的重要内容。

二、历史回顾与现状

在一个世纪前,有人曾提出蛋白质结构肽的假设,但长期以来,人们无从着手研究并真正了解它,直到 1955 年英国著名科学家 Sanger, F. 发表了胰岛素的全部氨基酸排列顺序^[6],从而开创了研究蛋白质一级结构的新纪元, Sanger 的这一贡献是分子生物学发展进程中的一个重要突破,因此获得诺贝尔化学奖。但由于受实验技术的限制, Sanger 当时采用的是比较简单的方法:如氨基酸组成分析是用纸层析法, N 末端及 N 端顺序测定是用 DNP 法,不仅费时而且要消耗过多样品,尤其是非特异性裂解法,肽段的分离相当困难。 Sanger 和他的同事用了上百克的胰岛素,整整花了十年时间才完成了胰岛素的一级结构测定。尽管目前分析方法已完全改观,但 Sanger 当时的一些基本战略仍被广泛采用。

在 Sanger 开拓性工作的直接影响下,蛋白质顺序分析研究于六十年代达到高峰。1963 年美国洛克菲勒研究所的 Moore, S. 和 Stein, W. 完成了第一个酶蛋白——核糖核酸酶的顺序分析。这是一条有 124 个氨基酸残基的单链,含有四个二硫键。他们对 Sanger 的方法有如下的发展:(1)发明了氨基酸自动分析仪;(2)用离子交换柱层析分离酶解后的肽段;(3)采用 Edman 顺序降解法。他们卓有成效的工作,大大加快了顺序分析的进程,推动了蛋白质顺序测定。最明显的例子是六十年代初,美国和法国的两个实验室成功地测定了血红蛋白两种多肽链的氨基酸顺序。1968 年 Edman 和 Begg 研制的自动液相顺序仪^[7]和由 Laursen (1971 年)创建的自动固相顺序仪^[8]的问世,使蛋白质氨基酸顺序研究工作的面貌焕然一新。

图 1-2, 1-3 和表 1-1, 表 1-2 示出蛋白质顺序分析的进展情况。

从下述图表中不难看出,进入七十年代后,蛋白质顺序分析又获得迅猛的发展。据

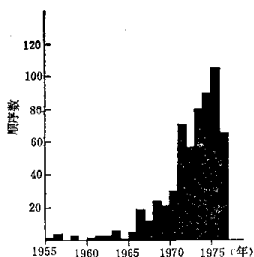


图 1-2 1955—1976 年测定的新的蛋白质(大于 50 残基)顺序数

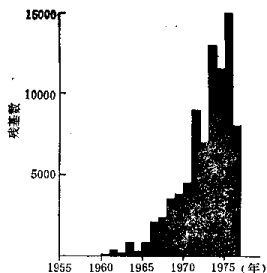


图 1-3 1955—1976 年测定的蛋白质顺序残基数

表 1-1 过去三十年中测定一条 50 个残基肽所需的样品量和时间

年 份	需 样 品 量	花 费 时 间
1950	100g	10 年
1960	10g	3 年
1970	1g	1 年
1980	1mg	1 周

表 1-2 过去三十年中蛋白质顺序测定的进程

年 份	被测定的蛋白质顺序	残基数
1955	胰岛素	51
1961	血红蛋白的 β 链	145
1962	血红蛋白的 α 链	141
1962	细胞色素 c	105
1963	核糖核酶	124
1963	溶菌酶	129
1965	肌红蛋白	153
1965	病毒外壳蛋白	159
1966	胰蛋白酶原	216
1967	甘油醛 3-磷酸脱氢酶	340
1969	免疫球蛋白 γ 链	446
1973	免疫球蛋白 μ 链	537
1975	血清白蛋白	585
1977	兔肌肉磷酸酶	841
1978	β -半乳糖苷酶	1,021
1981*	大肠杆菌 RNA 聚合酶	1,407

* 是用 cDNA 方法推导的。

Edman 估计, 1976 年已定出的顺序总数为 8 万氨基酸残基, 三年以后达到 16 万个残基, 完成近 1,100 个蛋白质全部顺序。在这期间, 主要因为自动氨基酸分析仪、自动液相顺序仪、自动固相顺序仪和高效液相色谱仪等仪器的出现, 以及不断创造的新方法和新技术在顺序分析中得到广泛应用。

从图 1-2 和 1-3 也可以看到, 到七十年代后期, 顺序分析的发展速度开始慢下来了^[7]。

蛋白质结构研究进展迟缓的原因很多, 其中非常重要的一点, 是目前采用的研究技术和方法跟不上生物学日益发展的需要。现在测蛋白质氨基酸顺序常规的基本方法仍然是 Edman 降解法。尽管有现代化顺序仪、氨基酸分析仪和高效液相色谱仪等可供利用, 但是效率依然不高, 特别是测定大分子量(5 万以上)的蛋白质(见图 1-8, 1-9)。又因为用 Edman 降解法一次连续降解次数有限, 而未知的蛋白质样品分离纯化又相当困难, 对于较大的蛋白质样品, 用酶法或化学方法裂解成肽段后, 往往遇到不溶甚至聚集等问题, 给分离纯化带来很大困难。还有样品量的问题, 有的蛋白质如受体, 要想提取上百微克量已相当困难。因此, 如何使顺序分析快速化、微量化, 已成为各国学者共同关心的问题。

为了促进顺序测定工作的开展, 从 1975 年开始, 每两年举行一次世界性专题会议, 交流、探讨有关蛋白质顺序分析中的新技术、新方法和它的战略思想问题。本节里只简要介绍有关微量分析的新技术、新方法上的进展情况。其它一些内容, 诸如顺序分析中的有关战略思想、高特异性蛋白质水解酶以及肽段的高效率分离纯化上的进展等, 读者可参见后面的有关章节。

(1) Edman 降解试剂和方法上的改进

过去的减法 Edman 方法, Dansyl-Edman 方法都是在 Edman 方法基础上加以改进。1976 年外籍华人张瑞耀提出一种新的 Edman 反应试剂——DABITC。利用 DABITC/PITC 双偶合法测顺序, 无论是手工的还是自动的, 固相方法或是液相方法都很有效, 灵敏度比常规 Edman 试剂提高 1—2 个数量级(详见第七章)。

(2) 顺序仪设计上的改进和创新

液相顺序仪, 由于在萃取时样品损失及机械损失等原因, 一次连续降解步数只能在 20—40 步, 所用样品量也大(通常 > 50 nmol), 后来在机械上作了改进, 如提高反应杯的真空度, 但效果不明显。1978 年, 在反应杯中引入 Polybrene (一种四级铵盐聚合物) 后, 使一次连续降解步数增至 60 甚至 80 步, 且样品需要量减少为 1—10 nmol。1981 年 Hewick, R. M. 和 Hunkapiller, M. W.^[8] 研究了一种新型的顺序仪——气相顺序仪 (Gas-Phase Sequencer), 只要 5 pmol 的样品量就能上机测顺序, 试剂和溶剂的消耗量仅是液相仪的十分之一, 给微量化、快速化的分析提供了有效的手段(详见第八章)。

(3) GC-MS 联用仪进入了多肽顺序分析行列

质谱法不同于 Edman 化学降解法的物理化学分析法。其原理如图 1-4 所示。

固体或液体试样的蒸汽或气体分子受一定能量的电子冲击时, 将形成带正电荷的离子, 在电场和磁场的作用下, 离子会按照试样组物质的分子质量 m 与电荷 e 的比值 (m/e 称质荷比) 大小而被分开, 形成一定的谱线排列, 这叫做质谱, 若用照相等方法将质谱记录下来, 这种仪器叫质谱仪; 若借助于电子学方法将质谱中表现的各项谱线强度显示出来, 这种仪器称质谱计。目前商售的质谱仪, 尤其是质谱计的种类繁多。它们的主要部件可分为离子源、电场、磁场和收集器或记录显示系统四部分。从质谱图可知, 被测分子的单

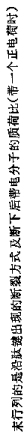


表 1-3 甲基化衍生后各种氨基酸的质谱值 (m/e)

氨基酸	衍生的结构	A N端的质谱值	B 残基的质谱值	C c端的质谱值
甘氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}_2-\text{CO}- \end{array}$	114	71	102
丙氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_3 \end{array}$	128	85	116
缬氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}(\text{CH}_3)_2 \end{array}$	156	113	144
亮氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2 \\ \\ \text{CH}(\text{CH}_3)_2 \end{array}$	170	127	158
丝氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2 \\ \\ \text{O}-\text{CH}_3 \end{array}$	158	115	146
苏氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}-\text{O}-\text{CH}_3 \\ \\ \text{CH}_3 \end{array}$	172	129	160
天冬氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2 \\ \\ \text{CO}-\text{OCH}_3 \end{array}$	186	143	174
谷氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CO}-\text{O}-\text{CH}_3 \end{array}$	200	157	188
精氨酸	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2 \\ \\ \text{CO}-\text{N}(\text{CH}_3)_3 \end{array}$	199	156	187
谷氨酰胺	$\begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2 \\ \\ \text{CO}-\text{N}(\text{CH}_3)_2 \end{array}$	213	170	201

附表 1-3

氨基酸	衍生的结构	A N端的质谱值	B 残基的质谱值	C c端的质谱值
苯丙氨酸	$ \begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2-\text{C}_6\text{H}_5 \end{array} $	204	161	192
酪氨酸	$ \begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2-\text{C}_6\text{H}_4-\text{OCH}_3 \end{array} $	234	191	222
色氨酸	$ \begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2-\text{Indole} \end{array} $	257	214	245
赖氨酸	$ \begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ (\text{CH}_2)_4-\text{N}(\text{CH}_3) \\ \\ \text{CO}-\text{CH}_3 \end{array} $	241	198	229
组氨酸	$ \begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ \text{CH}_2-\text{Imidazole} \end{array} $	208	165	196
鸟氨酸	$ \begin{array}{c} \text{CH}_3 \\ \\ -\text{N}-\text{CH}-\text{CO}- \\ \\ (\text{CH}_2)_3 \\ \\ \text{N}(\text{CH}_3) \\ \\ \text{CO}-\text{CH}_3 \end{array} $	227	184	215
脯氨酸	$ \begin{array}{c} \text{H} \\ \\ \text{C} \\ / \quad \backslash \\ \text{CH}_2 \quad \text{CH}_2 \\ \quad \quad \\ -\text{N}-\text{CH}-\text{CO}- \end{array} $	140	97	128

位质量数就是相应的分子量或它的倍数。肽链中的不同氨基酸残基，其侧链基团(R)的质量数是不同的，因此在质谱上显示不同的图谱，即可推算出不同的氨基酸残基及它的排列。以七肽为例，若采用碘代甲烷作甲基化试剂，以氧化银为催化剂，将肽键上的 N—H 全部甲基化，上述肽在质谱值 m/e 是 43, 114, 275, 436, 627 及 756 的地方出现主要的谱

线，且都与相应的一 $\text{C}=\text{O}$ —N 间的键断裂相联系。对照表 1-3，即可确认该肽 N 端前五

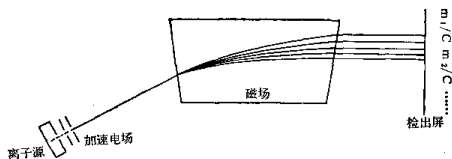


图 1-4 质谱仪的示意图

个氨基酸排列顺序为 Gly-Phe-Phe-Tyr-Thr。

为什么要将质谱 (MS) 和气相色谱 (GC) 联用? 因为气相色谱适合于分离复杂物质, 无论该物质含的多种分子是具有极性的, 或是沸程范围宽广的, 几乎都能完全分离, 尤其对一些挥发性好或热稳定的化合物分子效果更好。质谱仪不仅灵敏, 而且非常适合于分析高度挥发性的或热稳定的物质, 它刚好与气相色谱分离组分的数量级大小和分子性质相吻合, 因此将这两种仪器联用能使各自的优点充分发挥出来。

GC-MS 联用仪测定肽链顺序的要点是:

- ① 将待测的蛋白质样品裂解成大肽段, 并分离这些肽段。
- ② 将各个大肽段裂解成更小的肽段。
- ③ 再次分离肽段, 并进行部分水解。这时形成的肽段一般较小 (几个或十几个残基), 适合 GC-MS 分析。
- ④ 将这些小肽段转变成挥发性的衍生物 (通常是 N 末端烷基化), 然后送入 GC-MS 联用仪分析。
- ⑤ 利用电脑 (仪器配套) 对测得的质荷比 (m/e) 参数进行大量分析后, 就可确定其部分或全部结果。
- ⑥ 最后弄清楚小肽顺序。

图 1-5 是测定顺序的示意图。

用 GC-MS 法测多肽顺序的优点很多, 自动化程度很高, 只要数秒钟就可测出含十个氨基酸残基的小肽顺序, 样品用量很小。现已证明, 用这个方法测 N 末端被封固特别是酰化的肽段顺序非常有效。二氢叶酸还原酶是第一个用质谱法测定的未知蛋白的全顺序, 于 1979 年完成。用质谱法测蛋白质氨基酸顺序是有大前途的, 有人预言^[9,10], 将来此方法有可能代替传统的 Edman 降解法, 估计 1 个 2 万分子量的蛋白质只需要 5—30nmol 样品量, 两个操作熟练的工作人员, 用 1—2 个月时间就能完成整个顺序分析。但目前看来还有它的不足之处, 如 m/e 的测量值有限, 质谱图很复杂, 解释相当困难, 再加上仪器价格昂贵, 因此要想广泛应用, 还有许多困难。

(4) 用测定核酸顺序推断蛋白质顺序

已经知道, 遗传信息贮存在核酸顺序中, 这种信息能指令每个细胞, 按照分子遗传密码产生一种蛋白质的特征性结构, 所以, 在每种蛋白质的多肽链中, 氨基酸的排列顺序归根到底是由核酸的核苷酸残基顺序所规定或编码的。人们把编码一条完整多肽链的核酸分子片段称为基因。在核酸中, 每三个相邻的核苷酸残基称作一个三联体密码, 每个三联

体密码编码一个氨基酸。基因通常存在于染色体中。在核糖体进行蛋白质生物合成时,它并不直接作为模板,而是基因中的遗传信息首先通过酶的作用转录而成特殊类型的核糖核酸——信使核糖核酸 (mRNA)。它的核苷酸顺序与基因 DNA 的核苷酸顺序是互补

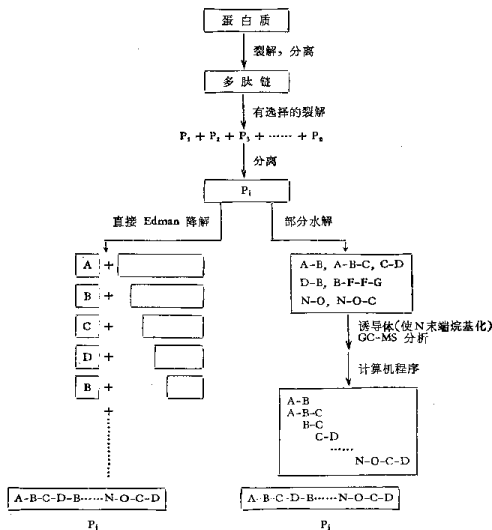


图 1-5

的,它们的三联体密码也是互补的。这就是说, DNA 通过RNA 中间模板而与它编码的多肽链的氨基酸顺序——对应 (见图 1-6)。遗传信息遵循 $\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{蛋白质}$ 的方向传递。因此,通过测定 DNA 的核苷酸顺序可以推测出相应的蛋白质氨基酸顺序。

其方法是:先测定未知蛋白质的肽段的氨基酸排列顺序,然后利用免疫或分子杂交分离出相应蛋白质的 mRNA,通过 mRNA 反转录成为 cDNA。经过克隆扩增分离 cDNA,然后分析测定 cDNA 顺序。同时测知该蛋白质的N端和C端的部分顺序。最后依据密码编码规律寻找与 cDNA 顺序相对应的部分,即可推测出蛋白质的氨基酸顺序。这是一个很有发展前途的方法^[1]。目前测定 DNA 顺序的技术方法已经基本成熟。一个熟练的工作者,一天就可测定250个以上核苷酸的 DNA 链。在一般的实验室,测定几千甚至上万个核苷酸顺序已经不是高不可攀了。

从 cDNA 的顺序推出蛋白质的顺序不仅明显地提高了分析速度,而且对于用传统的蛋白质化学方法难以测定的大分子量的(十万道尔顿以上)或含量很低的蛋白质,此方

法将是十分有效的。尽管如此,测 cDNA 顺序方法也不能完全代替传统的蛋白质化学方法,即使对研究核酸顺序的人来说,有时也需借助于经典方法,用测定蛋白质顺序来验证核酸顺序。

DNA	mRNA	多肽链
3' 末端	5' 末端	氨基末端
G	C	精氨酸
C	G	甘氨酸
A	U	酪氨酸
C	A	苏氨酸
C	C	苯丙氨酸
T	U	丙氨酸
A	G	缬氨酸
T	A	丝氨酸
G	C	
A	U	
A	U	
A	U	
C	G	
G	C	
G	G	
C	U	
A	U	
A	U	
G	C	
A	U	
5' 末端	3' 末端	羧基末端

图 1-6 DNA、mRNA 的核苷酸顺序与多肽链的氨基酸顺序之间的并行关系。DNA 的核苷酸单元三联体通过 RNA 的中间形成而决定蛋白质的氨基酸顺序;RNA 的核苷酸三联体(密码子)与 DNA 的核苷酸三联体是相对应的。图中英文字母表示 DNA 与 RNA 的核苷酸残基的特殊碱基成分: A. 腺嘌呤, T. 胸腺嘧啶, G. 鸟嘌呤, C. 胞嘧啶, U. 尿嘧啶。

引自 A. L. 伦宁格著,任邦哲等译,《生物化学:细胞结构与功能的分子基础》,上册,科学出版社,p.58(1981)。

此外,还出现一些新的方法,如通过测定蛋白质三级结构推断一级结构,它的前提是要求晶体结构的分辨率不低于 $2\text{\AA}$ 。然而这是一种笨拙的办法,因为测高级结构比测一级结构通常难度更大,个别的有可能做到,但要普遍推广,至少目前还不现实。利用核磁共振研究蛋白质顺序是新近开始尝试的另一方法,测很单纯的小肽片段顺序虽有成功的,但对大分子量蛋白质尚有很多困难。

三、展 望

综上所述,自胰岛素氨基酸顺序发表以来的三十年间,蛋白质氨基酸顺序分析取得了飞速的发展。这些成就的获得与自动化仪器的出现以及技术、方法的不断改进密切相关,当前顺序研究总的趋向是向快速化、微量化发展,这是生命科学日新月异发展的需要。

目前测定蛋白质顺序最基本和最常用的方法仍然是 Edman 降解法,但此法有其致

命的弊病,即一次连续降解步数(循环数)受 Edman 每步化学反应(主要是偶合和裂解二步)的产率限制。反应产率与降解步数之间的关系是:

99%→120 步

98%→60 步

97%→40 步

按公式计算 $\left(\frac{99}{100}\right)^{120} \approx 30\%$

这就是说,前 119 步反应残留下来的各种杂质的量,已远远超过被鉴定的物质,势必难以将结果分析清楚。据 Wittmann, Abt. 估计^[32],目前测定蛋白质顺序,手工技术每步产率为 90—93%,固相顺序仪产率为 93—95%,改进后的液相仪产率为 98%。再加上鉴定方法的局限性和化学处理中的损失等,一次连续测定氨基酸残基的个数将很有限。因此,测定一个分子量上万的蛋白质顺序,事先需将它裂解成若干肽段,分子量越大需裂解的次数越多,每次裂解产生的肽碎片也越多。它们的分析效率将随着肽链长度的增加而明显下降。图 1-7 是蛋白质顺序分析在效率上的进展情况,也示出了不同大小的蛋白质分子与测顺序效率之间的关系。

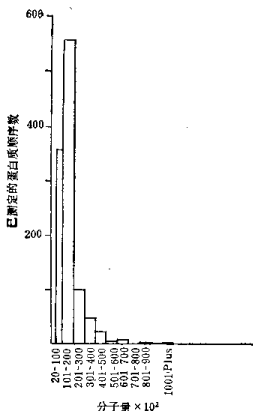


图 1-7 1952—1980 年蛋白质顺序分析在效率上的变化

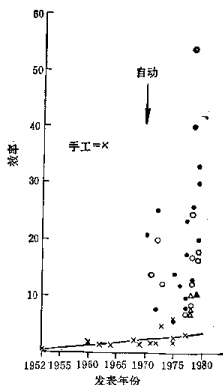


图 1-8 已测定的蛋白质顺序按分子量大小分布
○ <125 残基 ● 125—250 △ >260 残基

图 1-8 示出了蛋白质顺序与分子量之间的关系。在已测定的上千种蛋白质全氨基酸顺序中,95% 以上蛋白质分子量低于 40,000。这并不意味着自然界中的大分子蛋白质稀少,而是人们不愿为了测定一个大分子蛋白质顺序而耗费过多的时间和资源。

近年来随着生命科学的飞速发展和蛋白质分析、分离纯化技术的不断改进与提高,蛋

白质工作者的研究对象转向膜蛋白、受体蛋白、病毒蛋白以及代谢过程中出现的某些不稳定的中间等蛋白质。这些蛋白质一般都比较复杂,很难提纯,有的含量极低,有的分子量很大,因此,对于顺序分析所遇到的困难,不仅需要技术方法上的改进与创新,而且也需要战略思想上的突破。在有关学者的努力下,这几年也确实出现不少新动向,如气相顺序仪的问世使测量灵敏度提高了两个数量级,达到 pmol 水平。双向凝胶电泳和银染色法^[1],以及灵敏、快速、高分辨的高效液相色谱仪用于分离、制备和分析鉴定,为生物微量样品分析开辟了道路。将 Polybrene 引入液相顺序仪的反应杯后,使一次连续降解步数由原来的 40—60 步提高到 80 步^[2],可以相信,通过努力,一次连续测定上百个氨基酸残基必将成为现实。

在裂解技术上,更加高效专一的内切酶和化学试剂的发现,使测顺序的战略设计更具有针对性,测试效率也随之提高,这将为大分子量蛋白质顺序分析创造条件。从 DNA 顺序推测蛋白质顺序法,虽有某些弊病,但已预示它很有发展前途。质谱计和气相色谱仪或液相色谱仪与计算机联用,应用于蛋白质顺序分析,达到快速化、微量化。如果能进一步改进质谱计装置,使 m/e 的测量值成倍提高,那末必将给顺序分析带来福音。不难预料,在蛋白质顺序分析中,若将质谱法和 DNA 顺序法有机地结合起来,不仅使目前的分析效率明显提高,而且将改变我们现有的蛋白质化学测顺序的战略思想,最终将导致我们放弃传统的测顺序方法——Edman 降解法^[3]。Gray, W. R.^[4,5]就曾提出一种独特的设想,即将蛋白质裂解以后不再分离肽,直接进行 Edman 降解反应,然后逐个测定降解下来的氨基酸。这样便可从不同裂解方法裂解的二组混合肽的测量结果中,找出相同的氨基酸排列顺序。因为两组混合肽都来自同一个蛋白质,所以应包含相同的先后顺序。如果裂解次数很多,那末就可以从多套(每套二组)混合肽测得的结果中排出完整的蛋白质顺序。作者以核糖核酸酶为材料,选择 Arg、Met、Tyr、Lys、Cys、His 等为不同的裂解点,按照上述方法从六套混合肽的十次降解结果,排出了核糖核酸酶的氨基酸顺序,其中只有第 24 位谷氨酰胺不能确定。按理说这是很成功的,但自文章发表以来一直没有被人采用,可能是这个方法要求太苛刻,既要求肽链裂解的产率很高,又要求每个残基的降解产率很高,这是难于达到的。还有人曾想从蛋白质的大小、组成、溶解度等因素通过一定的程序设计,用电脑计算来完成顺序分析。但这一尝试一直无法确定未知蛋白质的一级结构,所以能否单独用电脑来完成顺序分析,至今仍是谜。尽管如此,他们想从战略思想上去改变以前的方法,对大家是有启发的。可以相信随着时间的推移,在蛋白质氨基酸顺序分析中,会有许多新的技术、新的方法以及新的战略思想问世。

参 考 文 献

- [1] A. L. 伦宁格著,任邦哲译:《生物化学》细胞结构和功能的分子基础,上册,科学出版社, p.57 (1981).
- [2] Dayhoff, M. O., Park, C. M., and McLaughlin, P. J., in Dayhoff, M. O., (ed): Atlas of Protein Sequence and Structure (1972), National Biomedical Research Foundation, Washington, D. C. (1972).
- [3] Lehman, H. and Kynoch, P. A. M.: Human Haemoglobin Variants and their Characteristics, North-Holland Pub. Co. (1976).
- [4] Sanger, F., et al.: The Structure of Insulin, *Biochem. J.*, 60, 541—565(1955).
- [5] Edman, P. and Berg, G.: A Protein Sequencer, *Eur. J. Biochem.*, 1, 80—91(1967).
- [6] Larsen, R. A.: Solid-Phase Edman Degradation: an Automatic Peptide Sequencer, *Eur. J. Biochem.*, 20, 89—102(1971).