

第一章 蛋白质化学

蛋白质与核酸是生命的物质基础。核酸携带遗传信息，从根本上控制生物表型；蛋白质是细胞中含量最多的生物大分子，大约占细胞干重的一半，数以万计的蛋白质分子不仅催化细胞中的各种代谢反应，而且参与组建细胞结构，细胞的任何活性几乎都离不开蛋白质。英文 Protein (蛋白质) 一词源于希腊文 Proteios 原意是“最重要”和“第一”体现了蛋白质生理功能的重要性。

蛋白质的组成元素主要有：C (约 50%)，H (约 7%)，O (约 23%)，N (约 16%)，S (1%~3%)，以及某些微量元素。蛋白质平均含氮 16% 是其元素组成的主要特征。

按照组分、形状或者功能，蛋白质可分成几种不同的类型。按照组分是否单一，蛋白质分为简单蛋白质和结合蛋白质两大类。简单蛋白质分子仅仅由 20 种普通氨基酸组成。而结合蛋白质分子除了蛋白组分以外，还有非蛋白组分——辅助因子，包括辅基、辅酶和金属离子。辅酶以非共价键与蛋白部分结合，可通过透析或超滤除去；而辅基以共价键与蛋白部分结合，不易与蛋白分离。按照分子外形，蛋白质分为球蛋白和纤维蛋白两大类。大多数蛋白为球蛋白，外形呈椭球状，能溶于水（除膜蛋白外），能结晶。纤维蛋白外形呈细棒状，有的可溶（如肌球蛋白），有的不可溶（如胶原蛋白、角蛋白和丝心蛋白），另外人们经常按照生物功能将蛋白质分为酶、运输蛋白、营养蛋白、运动蛋白和结构蛋白等。

蛋白质相对分子质量的范围变化很大。烟草花叶病毒外壳蛋白含有 2 000 多个亚基，相对分子质量高达几千万。最小的蛋白质是胰岛素和某些蛇毒蛋白，相对分子质量大约 6 000，小于该下限者叫做“肽”(peptides)，蛋白质通常称为多肽(polypeptides)。

本章将介绍蛋白质的基本组分——氨基酸的结构和性质，蛋白质的几个结构层次，以及分离分析蛋白质的几种常用方法。

第一节 氨基酸化学

氨基酸是蛋白质的基本组成单位。在生物体内发现的氨基酸已达 180 多种，但参与蛋白质生物合成的普通氨基酸只有 20 种，它们有相应的密码子，有特异的 tRNA 载体，以及专一的氨酰 - tRNA 合成酶。在蛋白质分子中，经常出现 20 种普通氨基酸的衍生物。

一、氨基酸的构型

20 种普通氨基酸都是 α - 氨基酸，其 α 碳原子是不对称的，分别连接 4 个不同基团：氨基 ($-\text{NH}_2$)、羧基 ($-\text{COOH}$)、氢原子和一个可变的侧链基团 (R)，形成四面体结构，因此氨基酸是手征性分子 (chiral molecules)，有两种互为镜像的立体异构体 (L- 和 D- 构型) 而蛋白质中的氨基酸几乎都是 L- 构型 (configuration)。

所谓“构型”是指由于不对称碳原子或者双键的存在造成有机分子的组成基团在空间按照一定顺序排列，形成立体异构体，这些异构体之间相互转换必须有一个或多个共价键发生断裂。以含不对称碳原子最少的糖分子——甘油醛为标准，可以确定普通氨基酸或其他手性分子的 L- 或 D- 构型。如果有机分子含有 n 个不对称碳原子，能够形成 2^n 种不同的构型。例如，苏氨酸和异亮氨酸均含有 2 个不对称碳原子，可形成 4 种不同的构型。与双键相连的基团不能自由旋转，可以形成顺式 (cis-) 和反式 (trans-) 两种立体异构体 (构型)。

L- 和 D- 构型的异构体具有相同的熔点、溶解度等物理性质和相同的化学性质，但旋光方向相反，故称之为光学异构体。在一定的温度 (25℃) 和光源波长 (589nm) 的条件下，比旋光度是光学异构体的物理常数，其数值和光振动面旋转的角度 (旋光度 α) 成正比，和光学异构体溶液的浓度以及溶液的厚度 (旋光管的长度) 成反比。

二、氨基酸的类型

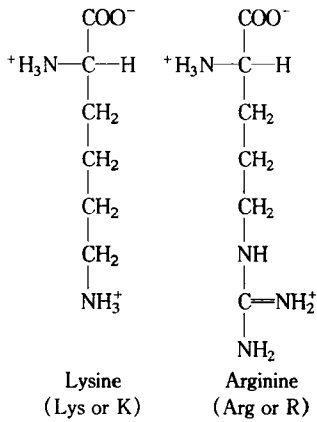
在生理条件下，游离的普通氨基酸含有偶极离子，因为它的氨基 ($-\text{NH}_3^+$) 和羧基 ($-\text{COO}^-$) 携带相反的电荷。20 种普通氨基酸的一般结构相同，其区别仅在于侧链基团的大小、形状、电荷、疏水性和反应活性。

根据侧链的性质和在水中的可溶性，20 种普通氨基酸分为三大类，

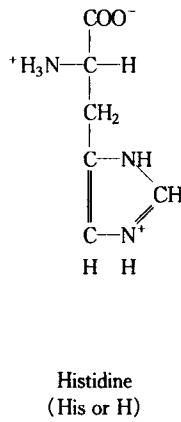
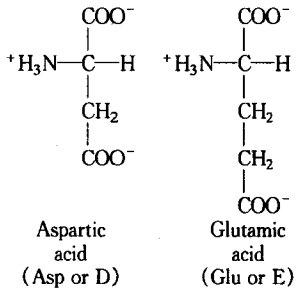
并分别用三字母或单字母符号表示 (图 1-1):

亲水性氨基酸

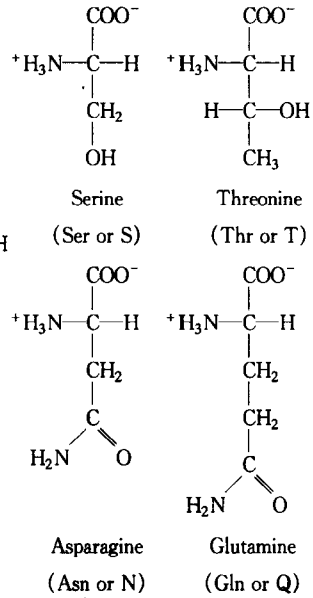
碱性氨基酸



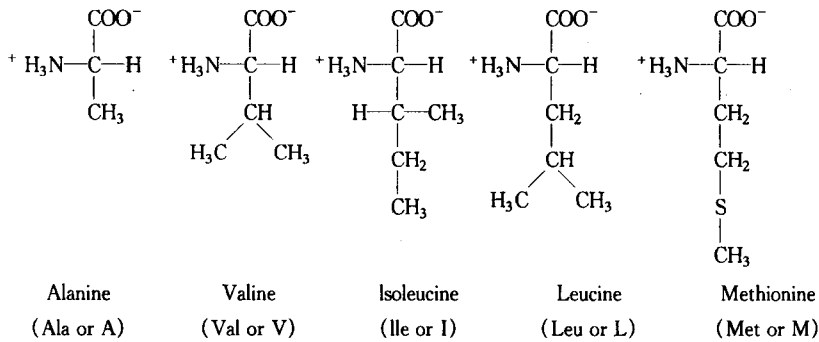
酸性氨基酸



极性氨基酸
(侧链不带电荷)



疏水性氨基酸



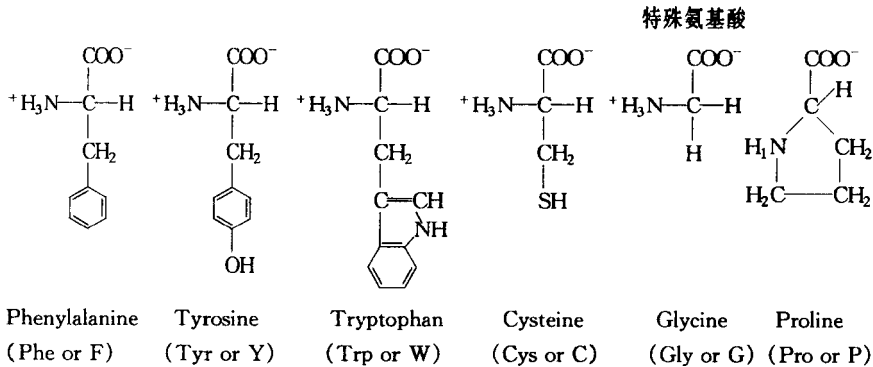


图 1-1 组成蛋白质的普通氨基酸的类型与结构

● 亲水性氨基酸：共 9 种，包括碱性氨基酸（赖氨酸、精氨酸、组氨酸）酸性氨基酸（天门冬氨酸、谷氨酸）和极性氨基酸（丝氨酸、苏氨酸、天门冬酰胺、谷氨酰胺）。

● 疏水性氨基酸：共 8 种，包括丙氨酸、缬氨酸、异亮氨酸、亮氨酸、蛋氨酸、苯丙氨酸、酪氨酸和色氨酸。

● 特殊氨基酸 共 3 种，包括半胱氨酸、甘氨酸和脯氨酸。

亲水的极性氨基酸经常出现在溶于水的球蛋白表面，以便保持其可溶性；而疏水的非极性氨基酸往往聚集在球蛋白的内部，形成疏水性的核心。

亲水性氨基酸含有离子型的或者极性的侧链。在中性 pH 条件下，Arg 和 Lys 带正电荷，Asp 和 Glu 带负电荷，这 4 种氨基酸对于决定蛋白质分子的净电荷起主要作用。His 的侧链是咪唑基，其 pK 为 6.8，所以细胞 pH 的微小变化将改变咪唑基的电荷。当 pH 为 5.8 时，90% 的 His 残基侧链被质子化，带正电荷；而 pH 为 7.8 时，90% 的 His 残基不带电荷。很多蛋白质的活性调节，需要借助于 pH 和 His 侧链质子化。Ser、Thr、Asn 和 Gln 的侧链均不带电荷，但含有极性的羟基或酰胺基团，因此能够同其他的极性分子形成氢键。这些带电荷的氨基酸和极性氨基酸通常出现在溶于水的球蛋白表面，它们不仅赋予球蛋白可溶性，而且形成和其他极性分子的相互作用位点。

疏水性氨基酸含有脂肪族侧链，不溶或微溶于水。Ala、Val、Ile、Leu 和 Met 的侧链完全由烃基组成（Met 的侧链含一个硫原子），所以是非极性的。Phe、Tyr 和 Trp 含有巨大的芳香族侧链。在溶于水性球蛋白中，疏水性氨基酸残基在熵的驱使下尽可能远离水相环境，聚集在球蛋白的内

部。然而，膜蛋白的表面经常出现疏水性氨基酸残基，因为膜蛋白插在疏水性的磷脂双层膜中。

Cys、Gly 和 Pro 侧链的性质颇为特殊，所以它们在蛋白质中能够显示独特的作用。Cys 的侧链含有活泼的巯基（—SH）通过氧化可以和另一个 Cys 的侧链形成二硫键（S—S）。借助二硫键，多肽链内部或多肽链之间可以发生共价交联。二硫键在细胞内的蛋白质分子中颇为罕见，但常见于细胞外蛋白质分子，它有助于维持其天然的折叠结构。Gly 是最小的氨基酸，它的侧链只有一个氢原子，使之适合于出现在多肽链折叠所形成的狭小空间。Pro 是唯一的亚氨基酸，它的 α -碳原子上的氨基和侧链基团之间共价相连，形成吡咯环，造成 C_{α} -N 键不能旋转。因此，Pro 具有很强的刚性，它的存在往往使多肽链的折叠打个固定的结（kink）。Gly 和 Pro 有时出现在蛋白质分子面的若干位点，多肽链在这些位点向后回折。

在蛋白质中，带电荷的和某些疏水性的氨基酸的含量比较丰富，例如 Arg、Lys、Asp、Glu、Ala、Val 和 Leu 的含量最高，每一种通常占氨基酸总数的 5%~10%。而 Cys、Trp 和 Met 在蛋白质中的含量较低，三者合计仅占氨基酸总数的 5%。20 种普通氨基酸的相对分子质量差异较大，考虑到它们在一般蛋白质中的丰度（abundance），其平均相对分子质量为 115。根据这一数值和蛋白质的相对分子质量，可以估计蛋白质所含有的氨基酸残基数。

三、氨基酸的酸碱性

在偏中性的水溶液中， α -氨基酸主要以偶极离子的形式存在，它的氨基（ $-\text{NH}_3^+$ ）可以起质子供体（donor）的作用，而羧基（ $-\text{COO}^-$ ）起质子受体（acceptor）的作用，所以氨基酸是两性电解质。

如果完全质子化，氨基酸就是多元酸，其侧链不解离者为二元酸，而酸性和碱性氨基酸为三元酸。例如，质子化的 Gly 是二元酸，分两步解离，首先羧基—COOH 解离，测得其 $\text{p}K_1 = 2.34$ （ K_1 为解离常数）；然后氨基— NH_3^+ 解离，其 $\text{p}K_2 = 9.60$ 。比较 Gly 和醋酸（ CH_3COOH ， $\text{p}K_a = 4.76$ ）的解离常数，Gly 羧基的酸性比醋酸高两个数量级以上（100 多倍）。显然在 Gly 分子中，邻近的氨基— NH_3^+ 对羧基—COOH 中质子的静电排斥作用，大大增加了后者的解离趋势。

氨基酸净电荷为零时，溶液的 pH 叫做氨基酸的等电点（isoelectric

point 缩写为 pI)。Gly 水溶液净电荷为零时的 pH 等于 5.97 这就是 Gly 的等电点 (pI) 它在数值上等于 Gly 的 pK_1 和 pK_2 的算术平均值。酸性或碱性氨基酸 (相当于三元酸) 的等电点, 等于净电荷为零状态前后两步解离的 pK 的算术平均值。小肽等电点的计算方法类似。

四、氨基酸的特性

氨基酸的 α -氨基、 α -羧基和某些侧链基团, 具有特殊的化学性质和物理性质。

α -氨基易于和酰化试剂反应。早期人们曾利用丹磺酰氯标记多肽链的 N 端, 进行氨基酸微量定量测定。有的酰化试剂在人工合成多肽时用于 α -氨基的保护剂。另外, α -氨基还可以发生烃基化反应。1945 年, 英国著名科学家 Sanger 首次利用 2,4-二硝基氟苯 (FDNB) 测定胰岛素的 N 端氨基酸残基。20 世纪 50 年代, Edman 等利用异硫氰酸苯酯 (PITC) 从 N 端测定多肽氨基酸顺序, 70 年代已发展成全自动的蛋白质氨基酸顺序分析仪。

α -羧基的性质类似于有机酸的羧基, 可以形成盐、酯、酰胺和酰氯, 也可以脱羧。在人工合成多肽时, 可利用成酰氯反应使 α -羧基活化, 以利于和另一个氨基酸的 α -氨基连接。

在氨基酸分析化学中, 茚三酮 (ninhydrin) 反应很重要。茚三酮可以和氨基酸的 α -氨基以及 α -羧基反应, 经过脱氨和脱羧, 生成特征性的紫色化合物。茚三酮与亚氨基酸 Pro 的反应生成黄色化合物, 不会发生脱氨。利用这一特性, 可进行氨基酸的微量定性或定量测定。

普通氨基酸在可见光区都没有光吸收, 在远紫外区 ($<220\text{ nm}$) 均有光吸收。而在近紫外区, 只有 Phe、Tyr 和 Trp 有光吸收, 其最大光吸收波长分别是 257 nm、275 nm 和 278 nm, 因为它们的苯环侧链具有共轭双键系统。由于这三种芳香族氨基酸的存在, 蛋白质具有紫外吸收能力, 最大光吸收波长为 280 nm。利用这一特性, 可测定蛋白质溶液的浓度。

第二节 蛋白质结构

一、肽键、肽平面、二面角和多肽链的构象

在蛋白质分子中, 肽键 (peptide bonds) 将一个个氨基酸残基连接起

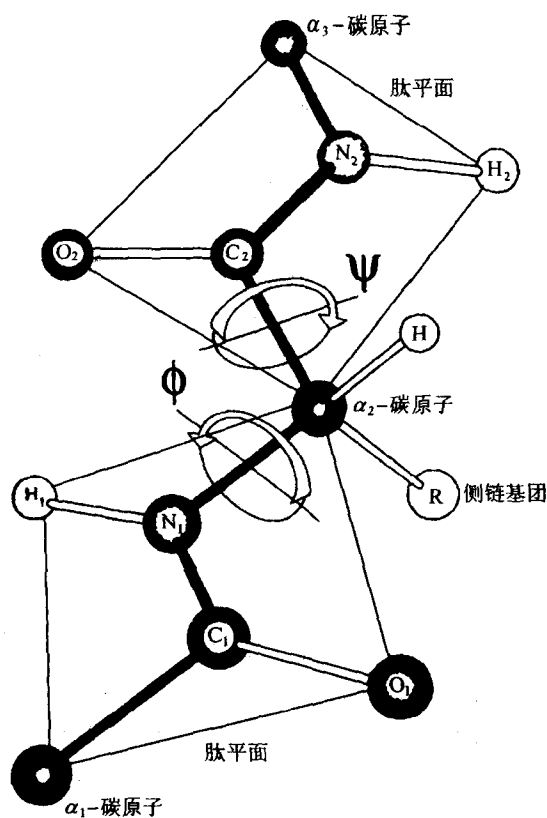


图 1-3 肽平面和二面角

肽链的组成，通常将 N 末端的氨基酸残基写在左首，而 C 末端残基写于右尾(图 1-2)。

二、蛋白质的结构层次

X 射线晶体衍射分析结果显示，蛋白质分子的结构有 4 个严格的层次 即：一级至四级结构。

蛋白质的一级结构 (primary structure)，是指多肽链的氨基酸残基的排列顺序。

二级结构 (secondary structure) 是指多肽链局部的空间结构 (构象)，特别是多肽链部分骨架的折叠方式。一条多肽链可能显示所有的二级结构类型。如果没有稳定因素，多肽链只能形成无规则卷曲 (random coil)。

然而，借助于某些氨基酸残基之间形成的氢键的稳定作用，多肽链骨架可以形成一定的几何形状，主要有以下三种类型：① α 螺旋(α helix)，多肽链骨架呈螺旋状；② β 折叠(β strand)，多肽链骨架呈伸展状；③转角(turn)，氢键作用使多肽链骨架转了一个U形弯(通常含4个氨基酸残基)，一般出现在球蛋白的表面，并将多肽链引向球蛋白内部。

三级结构(tertiary structure)是指整条多肽链的三维结构，包括骨架和侧链在内的所有原子的空间排列。蛋白质三级结构的稳定因素有共价键(二硫键)和次级键(氢键、疏水性相互作用、离子键和范德华力)。如果蛋白质分子仅由一条多肽链组成，三级结构就是它的最高结构层次。

四级结构(quaternary structure)是指在多个亚基组成的蛋白质分子中，亚基的种类、数目、在空间的相对位置及其相互作用，但不包括亚基本身的构象。维持蛋白质四级结构的力是次级键。亚基通常由一条多肽链组成，有时含两条以上多肽链(以二硫键相互连接)。亚基单独存在时一般没有生物活性。有的蛋白质分子的亚基种类相同，有的不同。在蛋白质分子的四级结构中，亚基的排列通常是有规律的，或者是对称的。

超二级结构(supersecondary structure)和结构域(domain)的水平，介于蛋白质的二级结构和三级结构之间。在蛋白质分子的三级结构中，经常发现氨基酸顺序相邻的二级结构单元形成有规律的聚集体——超二级结构，充当三级结构的构件(building block)。其基本形式有 $\alpha\alpha$ 、 $\beta\alpha\beta$ 和 $\beta\beta\beta$ 等。结构域是在超二级结构的基础上形成的，通常由100~300个氨基酸残基组成，其特点是在三维空间可以明显区分和相对独立，并且具有一定的生物功能。较大的蛋白质分子一般含有两个以上结构域，结构域之间以柔性的铰链(hinge)相连，以便相对运动。

在细胞中，蛋白质分子能够结合成巨型大分子组装体(macromolecular assembly)，如病毒的蛋白外壳，肌动蛋白纤维束以及核孔复合物等。然后，这些大分子结构再和脂类、糖类或核酸等生物大分子结合，形成复杂的细胞器。

三、研究蛋白质结构的几种方法

(一) 氨基酸组成分析

分析蛋白质的一级结构，需要利用化学方法测定蛋白质的氨基酸组成和氨基酸残基的排列顺序，这些方法都必须切割多肽链骨架上的肽键。

通过氨基酸组成分析，可以了解组成蛋白质的氨基酸的种类及其丰

度，但氨基酸残基的排列顺序尚不可知，获得的信息类似于分子的元素分析。

测定蛋白质的氨基酸组成的大致步骤是（图 1-4 A）：① 首先利用强酸（6mol/L HCl）将蛋白质彻底水解 24 h, 110℃），释放出一个个氨基酸；游离的氨基酸同茚三酮反应，形成特征性的紫色衍生物；③ 利用离子交换层析分离每一种氨基酸衍生物，监测其光吸收（450 nm）和洗脱时间，氨基酸衍生物的光吸收与其浓度有一定比例。

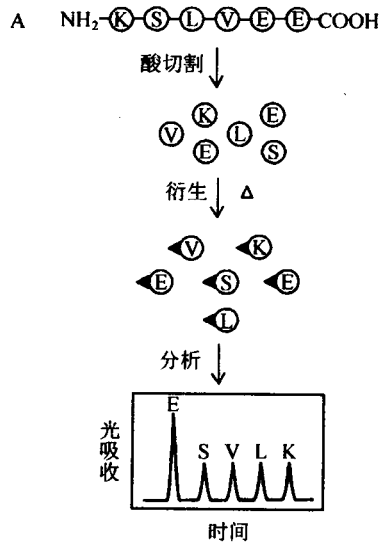
蛋白质的氨基酸组成可利用全自动的氨基酸分析仪进行。对于个别易于被强酸破坏的氨基酸（Trp）和亚氨基酸（Pro），需要某些特殊的方法进行分析。

（二）直接测定多肽的氨基酸顺序

测定肽链的氨基酸顺序必须解决两个关键问题，一是将氨基酸残基从肽链末端（N端或C端）依次一个个切割下来，通常使用化学法切割；二是将每次切割下来的氨基酸残基进行准确的鉴定，一般需要使氨基酸残基携带生色基团。

1. Edman 降解法

Edman 降解法是从 N 端测定肽链氨基酸残基顺序的经典方法（图 1-4 B）大致步骤为：偶联：PITC 异硫氰酸苯酯 结合于多肽链 N 末



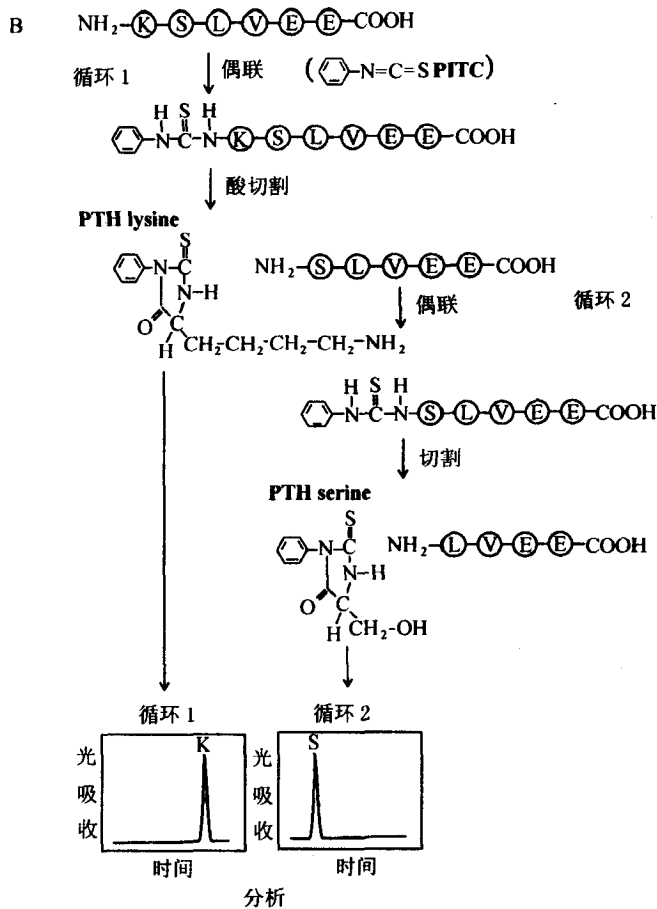


图 1-4 测定多肽的氨基酸组成和顺序

A. 蛋白质氨基酸组成的基本步骤：首先利用强酸水解，然后令氨基酸和茚三酮形成紫色衍生物，最后检测其通过离子交换柱的时间和光吸收

B. Edman降解法的每个循环由偶联、切割、转化和鉴定三步组成

端的游离 α -氨基，大大地削弱了连接第一和第二个氨基酸残基的肽键；

裂解：利用酸（无水三氟乙酸）将 N 末端氨基酸残基切割下来，形成环形的 ATZ-氨基酸（溶于有机溶剂）和少了一个残基的多肽链（溶于水），后者可重复进行下一个循环的偶联和切割，周而复始；③ 转化：ATZ-氨基酸不稳定，在酸性条件下可转化成稳定的 PTH-氨基酸④ 鉴定 通过高效液相层析（HPLC）、气相色谱、薄层层析（手工聚酰胺膜薄层双向层

析)和质谱等多种方法,均可鉴定每个循环产生的 PTH-氨基酸从而获得氨基酸残基的排列顺序。

N端蛋白质自动顺序分析仪是根据 Edman 降解原理设计的。20 世纪 70 年代的固相顺序分析仪是将肽链的 C 端共价固定在惰性载体上,使样品在洗涤、抽提过程中不会丢失,但 Glu、Asp 和 Lys 等残基的侧链基团也容易和载体结合,干扰测定。气相顺序分析仪利用四级铵盐聚合物 polybrene 与肽链极性基团之间的形成次级键,将样品有效地固定在玻璃纤维上,有的试剂如三氟乙酸在 Edman 降解反应过程中以气体方式输送,提高了灵敏度和分析速度,降低了样品和试剂的消耗。含活性基团的 PVDF 膜(聚偏二氟乙烯)出现后,蛋白质或多肽可共价固定在这种膜上进行测序。80 年代末问世的脉冲液相顺序分析仪,将三氟乙酸在气相分析仪中的气体输送方式改为液相脉冲方式,以适应分析的多种需要。

由于 Edman 降解反应有副反应,在自动顺序分析仪中每个循环的最佳产率为 95%~98%(手工法 90%~93%)经过 50 个循环以后,PTH 氨基酸分析图谱上出现的杂峰将严重地干扰鉴定。所以,利用蛋白质自动顺序分析仪最多只能测定蛋白质 N 端的 50 个氨基酸残基。

手工 DABITC/PITC 双偶合法比经典 Edman 降解法的灵敏度高 10~20 倍。因为 PTH-氨基酸在可见光下无色,必须在紫外光下检测,所以灵敏度不高。而 DABITC(4-N,N-二甲氨基偶氮苯-4'-异硫氰酸酯)本身呈橘红色,其工作原理同 PITC 相同,形成的 DABTH-氨基酸呈鲜红色,因而大大地提高了检测的灵敏度。但 DABITC 对多肽分子 N 端的偶合率不足 50% 其余分子的 N 端必须由 PITC 同时偶联,才能使样品分子的降解同步进行。

2. 质谱法

质谱法(mass spectrometry)是一种比较先进的物理方法能够精确地测定大分子的质量(误差不到 1),可用于测定长度小于 20 个氨基酸残基的肽的一级结构。其原理是,利用氦原子(He)随机碰撞并切割肽键,产生长度相差一个氨基酸残基的一系列肽(家族),再通过质谱法分析这些肽之间的质量差别,进而推导出原来的肽的氨基酸顺序。质谱法对于鉴定被修饰的氨基酸残基十分有用。

3. C 端顺序分析

蛋白质或多肽的 C 端顺序有时也需要测定,例如,鉴定重组基因的表达是否正确,基因表达产物大量制备时的质量控制, N 端封闭(如乙酰

化、焦谷氨酰环化或环形肽)的多肽或蛋白的分析,DNA探针的设计等。人们曾利用几种羧肽酶分析C端氨基酸残基,但羧肽酶对氨基酸残基(底物)有选择性,所以难于使用。有人先将C端标记,再分离出含有C端的酶解肽段,然后利用Edman降解法测序。近年来,化学法测定C端的技术得到进一步改进,1995年C端蛋白质自动顺序分析仪终于诞生,其原理和Edman降解法相似,主要步骤包括:①活化:令多肽C末端游离的 α -羧基同乙酸酐反应,生成环形衍生物,再转化成TH-氨基酸衍生物;②保护侧链羧基:侧链羧基同乙酸酐反应后难以成环,可将其转化成酰胺基团,以防干扰测序;③烷基化:C端的TH-氨基酸较稳定,不易被切割下来,将TH基团中的硫原子烷基化形成ATH基团后,裂解产率大大提高;④修饰S/T羟基:将多肽中的Thr和Ser的侧链乙酰化,以防干扰;⑤裂解:将C端的ATH-氨基酸在酸性条件下切割下来,通过HPLC鉴定,而新的C端自动形成TH基团,不必重新活化。

(三)测定蛋白质的一级结构

1955年,F. Sanger利用FDNB和纸层析完成了第一个蛋白质分子(牛胰岛素)的一级结构分析。1963年,S. Moore和W. Stein完成了第一个酶蛋白(RNase A)的氨基酸顺序分析,并且对Sanger的方法作了三点重要改进:发明并使用氨基酸自动分析仪测定样品的氨基酸组成;利用离子交换层析分离纯化酶解肽段;应用Edman降解法测定小肽的氨基酸顺序。另外,他们建立的测定蛋白质一级结构的基本战略沿用至今,即:使用两种方法将多肽链专一性裂解,逐一测定每个纯化的小肽段的顺序,然后根据这两套肽段氨基酸顺序中的重叠区确定小肽段的排列次序,从而完成整条多肽链的顺序分析。

在进行蛋白质顺序分析以前,需要做必要的准备工作,包括:样品纯化(一般纯度>97%),测定相对分子质量,确定多肽链的数目和大小,还原二硫键,测定氨基酸组成和末端分析等。

专一性裂解多肽链的方法有两类:①化学裂解法:溴化氰(CNBr)在Met残基羧基端将肽键(Met-X)特异断裂。另外,还有分别断裂Trp-X、X-Cys和Asn-Gly的化学试剂。②蛋白酶水解法:常用的特异性蛋白酶有胰蛋白酶(Arg-X或Lys-X)、糜蛋白酶(特异切割芳香族氨基酸形成的肽键的羧基端)、V8蛋白酶(Asp-X或Glu-X)、梭菌蛋白酶(clostripain, Arg-X)、凝血酶(Arg-Gly,近年来常用于重组融合蛋白裂解点),另外还有胃蛋白酶、枯草杆菌蛋白酶、木瓜蛋白酶和嗜热菌蛋白

酶等。

分离纯化肽段的方法很多，主要有层析和电泳两大类，详见本章第三节。

20 世纪 70 年代和 80 年代，重组 DNA 技术迅速发展，人们可以从 cDNA 或基因序列直接推导出蛋白质的氨基酸顺序，而且比利用化学方法测定蛋白质一级结构要快，目前这种方法已经成为最常用的测定蛋白质（特别是大分子蛋白质）一级结构的方法。

（四）人工合成肽

根据 20 世纪 60 年代 B. Merrifield 建立并发展的技术，现在人们可以在试管中合成长度在 50 个氨基酸残基以内的肽，其原理是反复在氨基酸之间进行缩合反应，以形成肽键。

肽的人工合成反应始于 C 端，大致步骤为：①通过酯化反应，首先将第一个氨基酸分子的羧基共价连接于不可溶的固相树脂，其氨基被 tBOC 基团（叔丁基氧羰基）所保护；②利用三氟乙酸去除 tBOC 基团；③在 DCC（二环己基碳二亚胺）的作用下，第一个氨基酸残基的氨基同第二个氨基酸分子的羧基缩合成肽键（后者的氨基也被 tBOC 基团保护）形成二肽；

重复进行上述过程，于是一个个氨基酸由 C 端依次连接到不断延伸的肽链的 N 端，直到被固定在树脂上的肽链的长度符合要求，最后利用氢氟酸 (HF) 将产物从树脂上切割下来，并除去氨基酸侧链上的保护基团。在合成过程中，必须封闭氨基酸侧链的氨基和羧基，否则在缩合反应中将发生副反应，使肽链产生分支。

在蛋白质和细胞研究工作中，人工合成肽是一个有力的工具。例如，人工合成的 10~15 个氨基酸残基的短肽可以作为抗原，诱使动物产生抗体，这些抗体同全长的天然蛋白抗原结合。抗体对于细胞中的蛋白质定位，以及从混合物中分离蛋白质，均十分有用。人工合成肽还有助于研究决定蛋白质的二级结构和三级结构的规律。通过系统地改变人工合成肽的氨基酸顺序，人们可以研究不同的氨基酸对蛋白质构象的影响。

（五）测定蛋白质的三维结构

经过几十年的辛勤努力，截止到 1995 年，800 多种蛋白质分子精密的三维空间结构分析已经完成。

根据蛋白质的状态，测定蛋白质分子构象的方法分为两大类：

测定晶体中的蛋白质分子构象，例如经典的 X 射线晶体衍射图谱法 (X-ray crystallography) 和近年来出现的中子衍射法；

测定溶液中的蛋白质分子构象 例如 核磁共振法 (nuclear magnetic resonance, NMR)、圆二色性光谱法、激光拉曼光谱法、荧光光谱法、紫外差光谱法以及氘同位素交换法等。

利用 X 射线晶体衍射法测定蛋白质分子的构象，结果比较可靠。20 世纪 80 年代中期，利用二维核磁共振法测定蛋白质三维结构获得令人满意的结果，应用越来越多。其他方法可以测定溶液中的蛋白质分子的局部构象，但很难获得蛋白质分子完整的三维结构，在应用上存在较大的局限性。

1. X 射线晶体衍射图谱

对于利用 X 射线晶体衍射法分析蛋白质分子的三维构象，M. Perutz 和 J. Kendrew 等作出了突出的贡献。

X 射线是波长很短的电磁波，波长范围 0.01~10 nm 用于结构分析的单色 X 射线的波长仅为 0.1~1 nm，相当于分子中原子之间的距离。所以当这种 X 射线穿过蛋白质晶体时，足以分辨并确定蛋白质分子中每个原子的位置，因为这些原子使 X 射线发生散射，可以在感光胶片上产生不连续的衍射斑点。解析 X 射线晶体衍射图谱非常复杂，一个小蛋白就能产生 25 000 个衍射斑点，必须经过精确的计算，才能获得蛋白质分子构象的信息。

X 射线晶体衍射结构分析，主要依据衍射线的方向（衍射斑点的位置——一定晶面对一定波长 X 射线所产生的衍射角）和强度（衍射斑点的黑度）前者通过 Bragg 方程进行计算，可以确定每个晶胞的大小和形状；后者需要测出晶胞体积、结构振幅和相角，再通过分布函数计算出晶胞中的电子云密度分布，画出电子密度图，才能确定晶胞中原子的分布，进而建立蛋白质分子的三维结构。

蛋白质晶体含水 30%~70%，蛋白质分子在晶体中所处的环境类似于溶液，可以保留生物活性。有证据显示，蛋白质在晶体和溶液中的构象相似。但是，与在溶液中的构象相比，蛋白质分子在晶体中的构象是静态的。所以，利用蛋白质晶体不能测定不稳定的过渡态的构象。而且，很多蛋白质很难结晶，或者很难得到用于结构分析的足够大的单晶。另外 X 射线晶体衍射的工作流程较长。

2 核磁共振法

近几年来，人们经常利用核磁共振法分析小蛋白的三维结构。这种方法的基本原理是，将蛋白质置于磁场中，测定不同的磁场强度（或射频

电磁波频率)对于蛋白质分子中不同原子的共振所产生的效应,任何原子的共振都要受到邻近氨基酸残基原子的影响,距离较远的残基的影响比较小。根据原子共振效应数值的大小,可以计算氨基酸残基之间的距离,然后利用这些距离构建蛋白质分子的三维结构模型。

近似球体、携带正电荷的原子核可以自旋,产生核磁矩。在外加磁场中,原子核不仅自旋,而且它的自旋轴以固定的夹角围绕磁场方向旋转,这种运动叫做“进动”。进动频率 ν_0)与外加磁场强度成正比,与每种原子核的特征常数磁旋比也成正比。核自旋轴与外加磁场的夹角是量子化的,具有 $2I+1$ 个能级,其中 I 是原子核的自旋量子数(与核内质子数和中子数有关),相邻的两个能级的能量差均为 ΔE 而 $\Delta E = h\nu_0$, h 为普朗克常数。如果照射原子核的射频电磁波频率等于原子核的进动频率 ν_0 ,原子核就吸收射频电磁波能量 $h\nu_0$ 从低能级跃迁到高能级,这一过程叫做核磁共振。自旋量子数为零的原子(如 ^{12}C 、 ^{16}O 和 ^{32}S)没有自旋现象,也不能产生核磁共振。自旋量子数不等于零的原子(如 ^1H 、 ^2H 、 ^3H 、 ^{13}C 、 ^{14}N 、 ^{31}P 和 ^{19}F)可以自旋产生核磁共振。

产生核磁共振波谱(NMR谱)和测定原子核共振频率的方法有两种:一是频率扫描,即固定外加磁场强度,改变射频电磁波频率;二是磁场扫描,固定射频电磁波频率,改变外加磁场强度。通常后者的应用较多,使用的仪器是核磁共振波谱仪。

NMR谱含有丰富的分子结构信息,但必须经过复杂的分析,才能建立分子的三维结构。在分子结构中,原子核不是孤立的,它的周围有电子流和其他原子核,而电子流可以产生与外加磁场方向相反的感生磁场,这种屏蔽作用叫做“化学位移”,其大小与核外电子云密度(化学环境)有关。乙醇分子有 $-\text{CH}_3$ 、 $-\text{CH}_2$ 和 $-\text{OH}$ 三种基团,其中的质子处于不同的化学环境,它们在不同的外加磁场强度下产生核磁共振,在 ^1H -NMR谱上出现3组特征性共振峰。因此,根据化学位移,能够确定分子的组成基团。影响化学位移(核外电子云密度)的因素较多,例如, π 电子云(Phe、Tyr、Trp和His等多元环侧链)的上方和四周的化学环境差别极大,位于蛋白质活性中心的含有未共用电子对的离子(Fe^{2+} 、 Zn^{2+} 、 Mg^{2+} 和 Mn^{2+})对周围的原子核的化学位移的影响也很大(顺磁效应),而氢键的形成将妨碍 $-\text{NH}_2$ 、 $-\text{OH}$ 和 $-\text{SH}$ 基团中的H原子与水溶液交换质子(化学交换效应),所以,NMR有利于测定氢核与多元环的相对位置,确定蛋白质的活性中心,以及研究氢键的形成。

另外 基团之间的核自旋磁矩发生相互作用(自旋偶合)可以使一条谱线分裂成多条,根据这些谱线的精细结构,可以了解基团之间的相互作用。再者,如果两个核在空间彼此靠近,第一个核被射频电磁波激发后,通过能量转移 第二个核的信号将增强(核的奥氏效应)其大小与核之间距离的 6 次方成反比,据此可以确定核间距。

核磁共振波谱有几种类型。 ^1H -NMR 谱和 ^{13}C -NMR 谱产生核磁共振的原子核不同。一维 NMR 谱是一个频率变量对吸收峰强度作出的平面图。二维 NMR 谱(2D-NMR)是两个频率变量对吸收峰作图,有两种表达方式 堆积图(看似三维立体图)和等高线图 通过计算机模拟 已成功地测定了蛋白质与核酸分子在溶液中的构象。三维 NMR 谱(3D-NMR)是 3 个频率变量的函数。

NMR 法不需要制备蛋白质晶体,但这种方法仅限于分析长度不超过 150 个氨基酸残基的小蛋白。

蛋白质分子的三维结构通过不同的图形表示,可以传递不同的结构信息,常见的图形有以下 4 种:①最简单的图形是利用一条实线表示多肽链骨架原子,显示除氨基酸侧链以外的多肽链整体结构。②最复杂的模型是将多肽链中每个原子的位置显示出来,能看到原子之间的相互作用,进而了解骨架的形成,以及蛋白质构象是如何稳定的。为了强调蛋白质分子中的二级结构组件及其各种组合,有人利用简单的符号表示二级结构 圆柱代表 α 螺旋,箭头代表 β 折叠,无规律的部分用线表示。通过电脑分析,发现蛋白质分子表面有哪些原子和环境中的水分子接触,进而图示分子表面不同区域的亲疏水性和电荷特征(酸碱性),这两点是配体结合位点的重要参数。而前 3 种方法都无法传递蛋白质分子表面的信息,这方面的信息很重要,因为配体结合于蛋白质分子表面。第 4 种图形,是从另一个分子的角度观察蛋白质分子的三维结构。

四、蛋白质的二级结构

在普通的蛋白质分子中,多肽链的 60% 以 α 螺旋和 β 折叠的形式存在,其余部分是无规则卷曲和转角。所以,两种有规律的二级结构形式—— α 螺旋和 β 折叠——是蛋白质分子内部主要的结构组件。胶原蛋白纤维由特殊的胶原蛋白三股螺旋(tropocollagen)组成。

(一) α 螺旋

20 世纪 50 年代早期,L. Pauling 和 R. Corey 发现并建立了蛋白质