

蛋 白 质 化 学

第六届全国高级生化训练班

补 充 讲 义

1988.9—12

第一章 蛋白质是一类重要的生物高分子

十九世纪中叶，随着化学这门科学的兴起，蛋白质也成了科学工作者的研究对象，并且一开始人们就注意到它在生物体中的重要作用。在十九世纪中叶以后约半个世纪中，曾用过不同的词汇来称呼它，如英文中的“Protein bodies”，“Albuminoids”与“Proteids”，德文中的“Eiweisskörper”，“Proreinstoff”及“Proteinkörper”等。1838年，德国科学家G.J. Müdler在文章中提出了Protein这个字。到本世纪初，生物化学家统一采用“Protein”一词。该词系从希腊文 $\pi\rho\omicron\tau\omicron$ 转化而来，意思是“最原始的”。中文译为蛋白质。曾根据Protein的原义建议设新字“朊”表示，但因蛋白质一字沿用已久，“朊”字未被广泛接受。

1820年，Braconnet证明在明胶中存在着甘氨酸。开始了发现氨基酸的历史，弄清了蛋白质的基本组成成分是氨基酸，1840年，F.L. Hünefeld报道，将蚯蚓的血液挥发浓缩后，血红蛋白即以结晶析出。这是第一个结晶的蛋白质。1877年，从巴西核桃中抽提的球蛋白也出现结晶。1926年SUMMER报道了脲酶的结晶，Abel报道了胰岛素的结晶，这分别是第一个酶蛋白质和第一个蛋白质激素的结晶。1902年，E. Fisher与F. Hofmeister提出了蛋白质中氨基酸间以肽键相联的学说。与此同时，从胶体化学的角度研究了蛋白质溶液的性质。Svedberg设计了超离心机，测定了不同蛋白质的分子量，发现与一般由果糖溶解不同，纯的蛋白质是分子量均一的大分子，蛋白质是单分散体系。Ostwald研究了蛋白质的电泳性质，设计了电泳仪，证明蛋白质具有两性性质，并证明纯的蛋白质有自己特定的等电点。从此以后蛋白质逐步进入了研究纯分子的阶

段。本世纪三十年代和四十年代研究蛋白质的中心问题是蛋白质的纯化，蛋白质分子的大小和形状以及带电性质。

从五十年代开始，F. Sanger 测出了胰岛素的全部化学结构，说明蛋白质的化学本质就是氨基酸通过肽键联成的线状肽链。与此同时，英国学派对纤维状蛋白质的 X-衍射分析，及 L. Pauling 关于 α -螺旋结构模型的提出，是蛋白质立体结构研究的重要突破。接着 M. F. Perutz 与 J. C. Kendrew 解出了血红蛋白及肌红蛋白在晶态中的立体结构，解决了蛋白质基本立体结构的问题。经这样长时间的历史发展，解决了关于蛋白质结构的基本概念。

蛋白质的种类非常之多。为了研究的方便，就出现了分类问题。曾经出现过众多的分类方法，它们各依据蛋白质分子的某一特征而进行分类。这些分类法大多因发现大量新的蛋白质而产生困难。为了有助于了解蛋白质的全貌，在此介绍几种分类方法。一种是按分子对称性分类的方法，如表 1. 1。这种分类方法在蛋白质研究热衷于分子大小与分子形状的时期，是十分简便有用的，但太笼统。表 1. 2 介绍的是一种按溶解度与结构内容混合考虑的分类法。20 世纪初，蛋白质的研究工作主要集中在利用蛋白质的各种不同的溶解性质而将它们分离纯化。表 1. 2 前一半的分类法是这一历史的反映，表 1. 2 后一部分是蛋白质研究进入化学结构时期的反映。

表 1. 1 按分子对称性划分的蛋白质分类表

球状蛋白质 (Globular protein): 分子接近球状或椭球状，能

结晶。包括大多数蛋白质。

纤维状蛋白质 (Fibrous protein): 分子很不对称，类似细棒或纤维。

可分成:

可溶性纤维状蛋白质: 如许多肌肉的结构蛋白, 血纤蛋白原等等。

不溶性纤维 一词：包括胶原、弹性蛋白、角蛋白及丝心蛋白等。

表1. 2 按溶解性质与结构内容混合考虑的分类法

一、简单蛋白质

1. 白蛋白 (Albumins)：溶于水及稀盐、稀酸或稀碱溶液。硫酸铵盐析时，50%饱和度以上开始析出。普遍存在于生物体系之中。

2. 球蛋白 (Globulins)：一般在等电点时不溶于水，但加少量盐、酸或碱后可以溶解。50%饱和度硫酸铵可以析出。这类球蛋白又名优球蛋白 (Euglobulins)。

另有一些球蛋白盐析的性质相同，但溶于水，称为拟球蛋白 (Pseudoglobulins)，普遍存在于生物体系之中。

3. 醇溶谷蛋白 (Prolamines)：可溶于70-80%乙醇中，但不溶于水或无水乙醇。在化学组成上有一定特点，如含脯氨酸及酰胺较多，非极性侧链远较极性侧链多。这类蛋白质在植物种子中较多。典型的例子如玉米醇溶蛋白 (Zein)、麦醇溶蛋白 (Gliadin)、大麦醇溶蛋白 (Hordein)。

4. 谷蛋白 (Glutelins)：等电点时不溶于水和稀盐溶液中，但易溶于稀酸和稀碱。如麦谷蛋白 (Glutenin) 和米谷蛋白等 (Oryzenin)。

5. 鱼精蛋白 (Protamines)：溶于水及稀酸，但为稀氨水所沉淀。分子中碱性氨基酸特别多，所以分子呈强碱性。分子量很小，没有特定的空间结构。一般多自鱼精中抽提。似

应归于多肽。

6.组蛋白 (Histones)：溶于水及稀酸，但为稀氨水所沉淀。分子中精氨酸与赖氨酸特别多，分子呈弱碱性。一般存在于动物体中，与酸性物质如核酸等相结合。

二、硬蛋白 (Scleroproteins)

这类蛋白质是动物体中作为结缔及保护功能的蛋白质，不溶于水、盐溶液、稀碱和稀酸溶液。

1.角蛋白 (Keratins)

2.胶原 (Collagens)

3.网硬蛋白 (Reticulins)

4.弹性蛋白 (Elastin)

三、缀合蛋白 (Conjugated proteins)

1.磷蛋白 (Phosphoproteins)：丝氨酸或苏氨酸侧链上连结了磷酸的蛋白质，如酪蛋白 (Casein)、胃蛋白酶。

2.粘蛋白 (Mucoproteins)：与氨基多糖结合的蛋白质。一般多含有 α -氨基葡萄糖，与下列一种或多种物质结合——半乳糖、甘露糖、鼠李糖、葡萄糖醛酸、硫酸或磷酸。这类蛋白质可分为粘蛋白 (Mucins)——如唾液腺及消化道所分泌的粘蛋白——和类粘蛋白 (Mucoid)——如卵类粘蛋白 (Ovomucoid)、血清粘蛋白 (Seromucoid) 等。

3.糖蛋白 (Glycoprotein)：与粘蛋白类似，但含糖量在4%以下，例如卵白蛋白。

4.核蛋白 (Nucleoprotein)：与核酸结合的蛋白质。

如小牛胸腺核蛋白、简单的植物病毒等。

5.脂蛋白 (Lipoproteins) 和蛋白脂 (Proteolipids) : 与脂类相结合的蛋白质。脂类成分有磷脂、固醇等。脂类包于分子内部而呈水溶性, 即是脂蛋白, 例如卵黄球蛋白 (Lipovitellin), 凝血致活酶 (Thromboplastin), 血清中 α 和 β -脂蛋白等。脂类包于外而溶于有机溶剂的, 就是蛋白脂, 如脑中的蛋白脂, 也有两类溶剂都溶解的, 如牛脑磷脂蛋白 II。

6.血红素蛋白 (Hemoproteins) : 与辅基血红素相结合的蛋白质, 主要功能为参与呼吸或运氧。血红素是卟啉类化合物。高等动物的血红蛋白含铁, 低等动物的血蓝蛋白含铜。

7.金属蛋白 (Metalloproteins) : 与金属直接结合的蛋白质, 如铁蛋白 (Ferritin) 含铁。许多蛋白质都含有少量的金属。

8.黄素蛋白 (Flavoproteins) 及氮苯蛋白 (Pyridinoproteins)

二十年代来, 蛋白质研究已发展到深入了解结构功能间关系及蛋白质-蛋白质 (或其它生物高分子) 相互关系的阶段, 因此相继有人提出根据蛋白质的功能进行分类的方法。表 1. 3 就是这样一种分类法。其实, 这无非也是各种蛋白质的列举。随着时间的转移, 表的内容还将有大的发展。

表 1. 3 蛋白质按功能的分类表

活性蛋白质 (Active proteins) 包括在生命运动过程中一切有活性的蛋白质, 以及它们的前体。绝大部分蛋白质都属于

此。可以分属下列各种种类：

酶：司催化活性的蛋白质，种类繁多。生物体内几乎所有的化学反应都由酶所催化。酶的催化活性比人工制备的催化剂催化效率高得多。不少蛋白水解酶的抑制蛋白也归在这里。

激素蛋白质：司调节机体各种代谢过程的蛋白质和相应的前体。激素蛋白的经典概念认为它们是由内分泌腺所分泌的。近年来这一概念越来越不严格，例如肠胃道就分泌一系列的激素蛋白。蛋白质类激素中包括了大量的肽。

运输和贮存蛋白质：生物体内有专一输送各种小分子和离子（也包括电子）的蛋白质。如血清白蛋白、血红蛋白等。有许多专一结合激素或其它小分子的蛋白质也属于这一类。

运动蛋白质：司生物体或有机体运动的蛋白质。如一系列肌肉蛋白，细菌的鞭毛或纤毛蛋白等。近年来发现，在非肌肉的运动体系中普遍存在着运动蛋白。

防御蛋白质和病毒外壳蛋白：在这一类蛋白质中，有一类是司防御异物侵入体内的功能的，如各种免疫球蛋白。防御病毒侵入后致病的干扰素也可属于这一类。血纤维蛋白和它的前体是另一类起防御功能的蛋白质，它们的功能是防止体液大量损失和异物侵入。此外，病毒的外壳蛋白不能简单地认为是防御，但有其相当相似之处，归于此。

受体蛋白：接受和传递调节信息的蛋白质。如一系列专一地接受各种激素的受体蛋白。蛋白质类激素的受体在细胞

膜上，甾体激素的受体在细胞内。接受外界刺激的蛋白质——感觉蛋白也可归于此，如视网膜上的视色素，味蕾上的味觉蛋白都是。此外还有许多药物的受体蛋白。

控制生长与分化的蛋白质：参与生长与分化的调节的各种蛋白质。各种生长因子归属于此。

毒蛋白：侵入动物体引起各种中毒症状甚至死亡的异体蛋白质。细菌毒素种类很多，有过详细的研究。蛇毒及虫蝎等毒蛋白结构相似，分子量一般都比较小，硫-硫键含量高。植物毒蛋白种类很多，是另一类毒蛋白。

参与基因表达与调控的蛋白质：如组蛋白、阻遏蛋白等。

膜蛋白：存在于膜上的具各种生物活性的蛋白质。这类蛋白质大多是脂溶性而水不溶的。结构上具有自己的特征。

非活性蛋白质 (Passive proteins) 主要包括一大类担任生物的保护或支持作用的蛋白质，从现有的了解看，都是不具生物活性的物质。主要包括：

胶原：哺乳类动物皮肤的主要成份，一般不溶解。从化学组成看，脯氨酸及羟脯氨酸的含量特别高。有非硫-硫键的侧链间的化学键使肽链联结起来。鞣制成革，部分水解成明胶。这类蛋白质是工业原料。

角蛋白：一类化学性质十分稳定的、含二硫键特别多的蛋白质。一般条件下都不溶解。功能是保护作用，或加强机

械强度用，包括毛发、羽毛、蹄角等。都是工业上的重要原料，如毛纺工业用的羊毛等即是。

弹性蛋白：存在于韧带、血管壁等处，性质与胶原类似。主要作用是有机体的支持与润滑。

丝心蛋白：一般不溶，含丝氨酸特别多。蚕丝去丝胶后的纤维就是它。

此外还有网硬蛋白等等。

所有的蛋白质归纳起来，参与了生物体的下列功能：新陈代谢（包含了酶、激素及运输蛋白），运动，感觉讯号加工，防御，及基因表达与调控。从中可以看到蛋白质的种类之多，所具的功能之广。总起来说，可以认为，活性蛋白质是所有生命活动过程的操纵者和调节者，而非活性蛋白质是构成生物体各种结构的主要成分。这就是蛋白质的重要性所在。

关于具体蛋白质的命名，没有统一的原则。英文中除酶外常以字尾-in, -en, -on等组字。也有复合词，用factor, protein等组成。中文命名有时候并不与英文相对应，具体蛋白质常以“素”，或“蛋白”置于词尾而名之，也有用“因子”，“酶”，“激素”等置于词尾者。近年来对“蛋白质”和“蛋白”两个词常被弄混，引起了诸多不便。对这两个字已规定如下：

蛋白质：是一类物质的总称，或者在不具体指某一蛋白质时的泛指。在指某一类蛋白质时，经常应该用“蛋白质”这个词，如球状蛋白质，收缩蛋白质、缀合蛋白质等等。

蛋白：是具体蛋白质名称的字尾或字部。在上述情况下，不得以“蛋白”代替“蛋白质”。

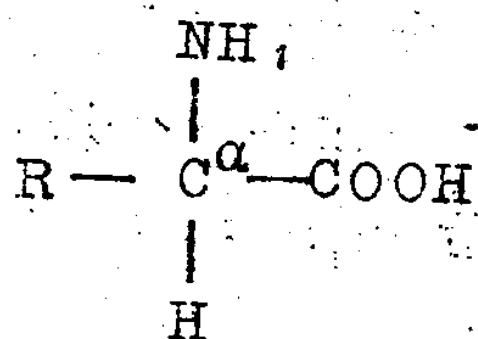
关于具体蛋白质的命名，没有统一的原则，因此相当混乱。概括起来，大体上有几种。有许多种蛋白质是根据它们的出处来命名的，如天花粉蛋白和半夏蛋白，是根据它们的物种来源命名的，胰岛素、水晶体蛋白是根据器官来源命名的。有许多蛋白质是根据它们的生理功能来命名的，如促长激素，运铁蛋白等就是。根据蛋白质的性质来命名的也不在少数，如白蛋白、冷球蛋白等等。也有不少蛋白质是将上述三种情况中的两种混合使用而命名的，如血红蛋白，既根据它的出处在血，又根据它的性质是红色的。免疫球蛋白既因其功能是免疫，又因其分子是球状的。胰高血糖素，是出处与功能结合命名。

有一些蛋白质的名词是比较少见的，它们以系列名加序号形成。如补体的经典途径，以C1, C2……C9来命名九个蛋白质；凝血系统也以第五因子，第六因子等来命名一部分蛋白质。

中文译名原则与英文命名原则相同。有极个别蛋白质音译。

第二章 氨基酸

含有氨基及羧基的有机化合物都叫氨基酸。但在生物体中出现的氨基酸种类不多，在蛋白质中出现的更只是一类特殊的氨基酸，它们的氨基和羧基都在 α -C上，称为 α 氨基酸，可表达成下式：



R称侧链。R不同，就是不同的氨基酸。

出现在蛋白质中的氨基酸

作为蛋白质组成成分的氨基酸一共有二十种，它们的名称及结构列在表2.1中。为了表达蛋白质或肽结构的需要，每个氨基酸都有符号表示，常用符号由氨基酸英文名字的前三个字母组成。一个蛋白质有时可以由上千个氨基酸组成，为了表达的方便，又设计出单字符号。这种符号节省印符，因此节省结构处理量，如输入计算机就有很大的好处。但是因过于精简，读结构时不如三字符号那样，有足够时间让脑子中反映出残基结构的特征。因此至今这两种符号都在继续使用。但是近年来单字符号使用者越来越多。特别是研究蛋白质氨基酸排列顺序规律的工作，几乎一律用单字符号。中文的符号用字头表示，但是几个用双字符号的氨基酸常常引起困难，看来是需要加以改进的。为了有利于建立空间结构的概念，表中附上了侧链的空间结构。

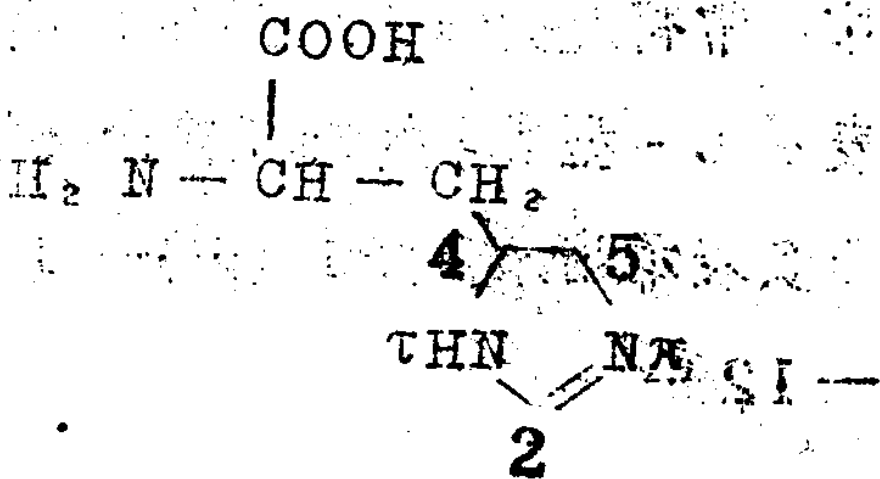
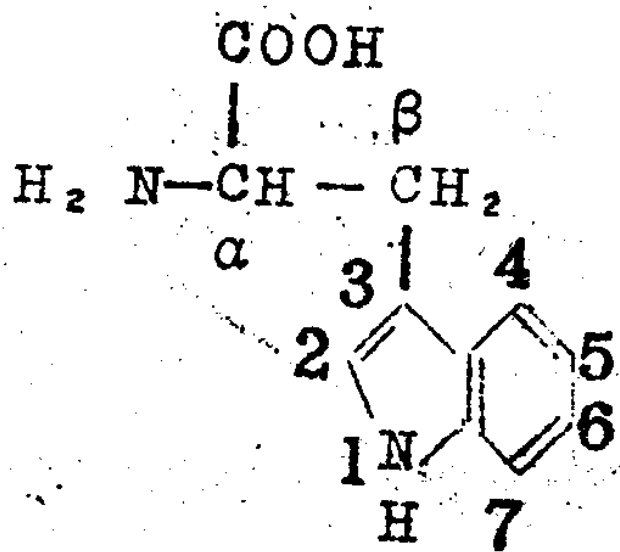
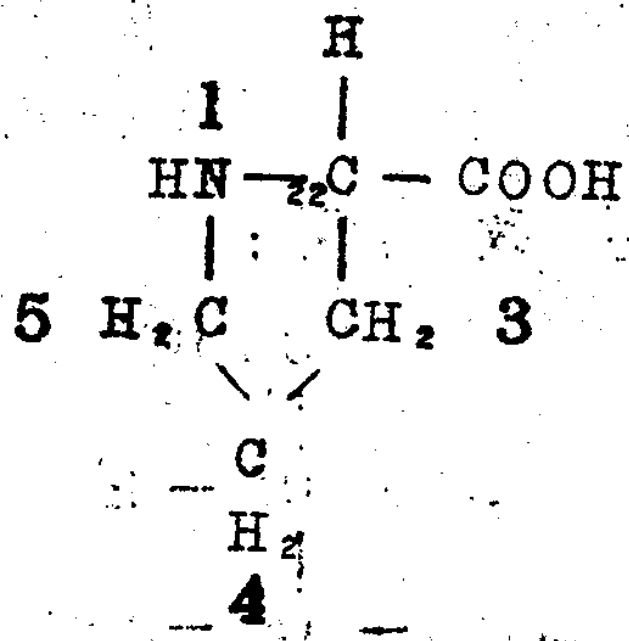
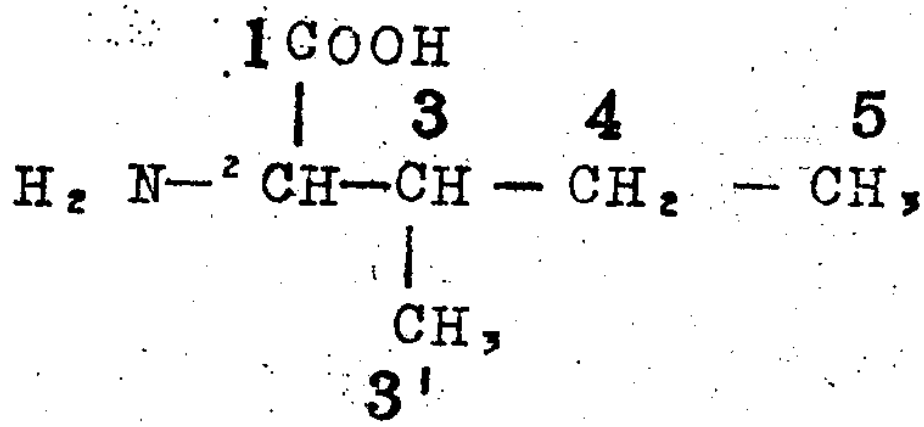
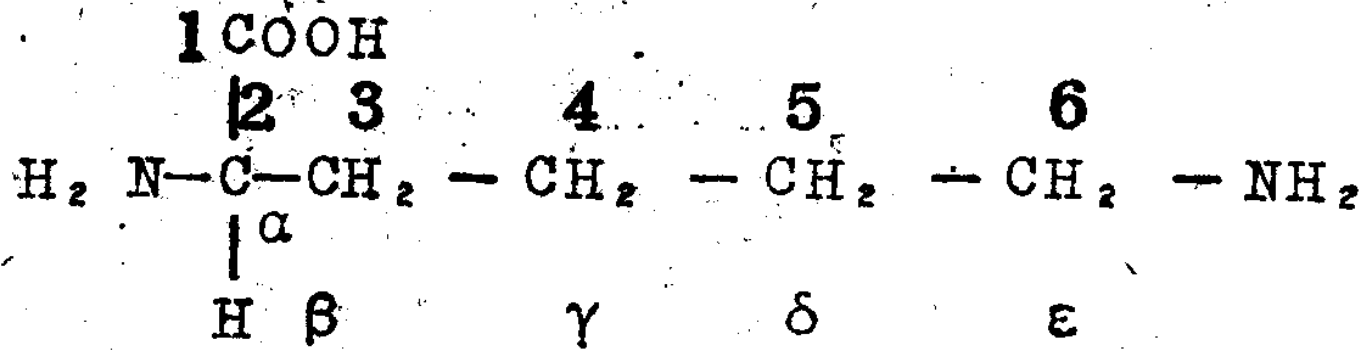
早年认为组成蛋白质的氨基酸一共有二十一种。但是发现羟脯氨酸是在蛋白质生物合成以后经专一的酶的作用，将脯氨酸羟基化而成的，因此不计入基本氨基酸之列，而认为蛋白质由二十种氨基酸组成。后来经三联密码证明，DNA只密码上述20个氨基酸。但是在具体的蛋白质中，许多氨基酸经受了后加工而被修饰，因此要比表中的复杂。(表2.1见后)

氨基酸的构型

上面谈到，组成蛋白质的氨基酸都是 α -氨基酸。从结构上看，除甘氨酸以外，所有这些氨基酸的 α -C上的四个基团都不相同。脯氨酸虽是环状结构，其 α -C上的四个基团也不相同。这种碳原子就是不对称碳原子。根据立体化学的研究，凡是有这种碳原子的分子都有光学活性，或者说有旋光。分子的手性(Chirality，或译手征性)指的就是光学活性，近年来用这一词汇的人越来越多。一个化合

物因化学结构的不对称引起分子的手性，这一类问题就是构型问题。

氨基酸上碳原子位置的标号有两种，一种用希腊字母，一种用数字序号，如下：



一般 α -氨基酸的构型有两种，用图形表达如图 2. 1 (见后)。经与乳酸的构型相比，发现蛋白质中的丙氨酸与 L-乳酸的构型相同：

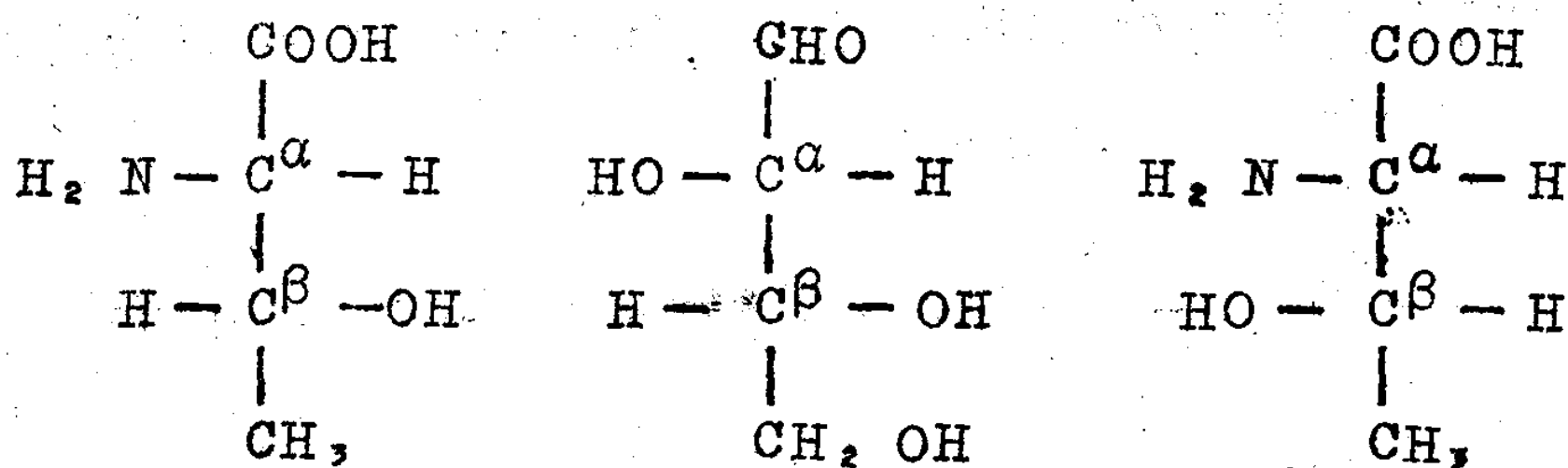


L-乳酸

丙氨酸

还发现，存在于蛋白质中的氨基酸的构型除甘氨酸外，都是 L-型的，甘氨酸例外，是非光学活性的。这一点颇有意思，到现在为止还找不出满意的解释来说明，为什么蛋白质选择了全部 L-型的氨基酸。

从表 2. 1 中可以看到苏氨酸与异亮氨酸的 β -碳也是不对称碳原子。因此，它们各有四种构型。 α -C 是 L-型的苏或异亮氨酸，可因 β -C 的构型不同而出现两个对映体。经对比，天然蛋白质中的苏氨酸的 β -C 构型与 D-苏丁糖的构型一致，如下：



L-苏氨酸

D-苏丁糖

L-别苏氨酸

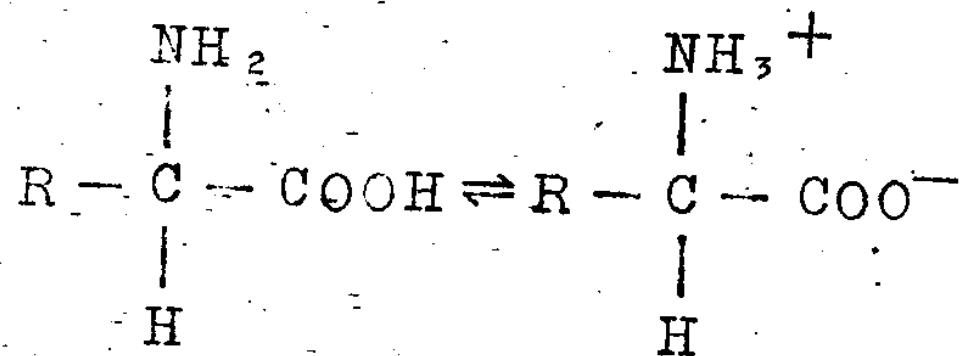
天然的苏氨酸称 L-型， C^β 与此相对映的称 L-别苏氨酸，立体地表示在图 2. 2 中 (见后)。

需注意，L-与 D- 只表示空间四个基团的相互位置，与旋光值的正负值是两码事。在有机化学中经常将旋光值的正 (+) 或负 (-) 表示出来。如 L (+)-乳酸，表示 L-型乳酸的旋光值是正值。与此相应，可以将氨基酸的旋光值正负表示出来，如 L (+)-Ala, L

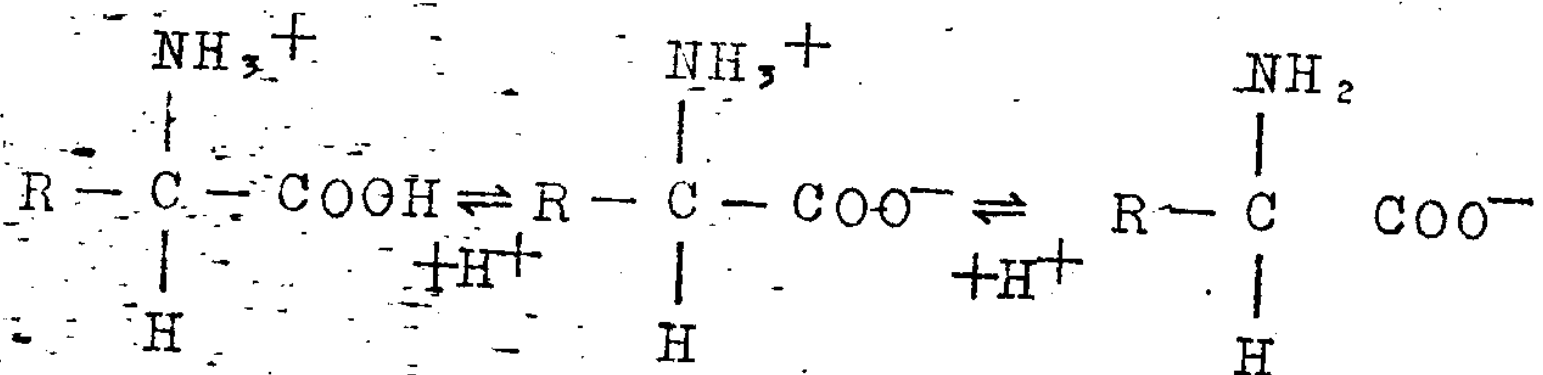
(-)-Thr 等。但按生化工作者的习惯，不作如此表示。在日常使用中，因蛋白质所含氨基酸都是 L-型的，因此连 L-型也不标出。

氨基酸的电离性质

氨基与羧基都是可电离的基团，因此氨基酸可形成两性离子：



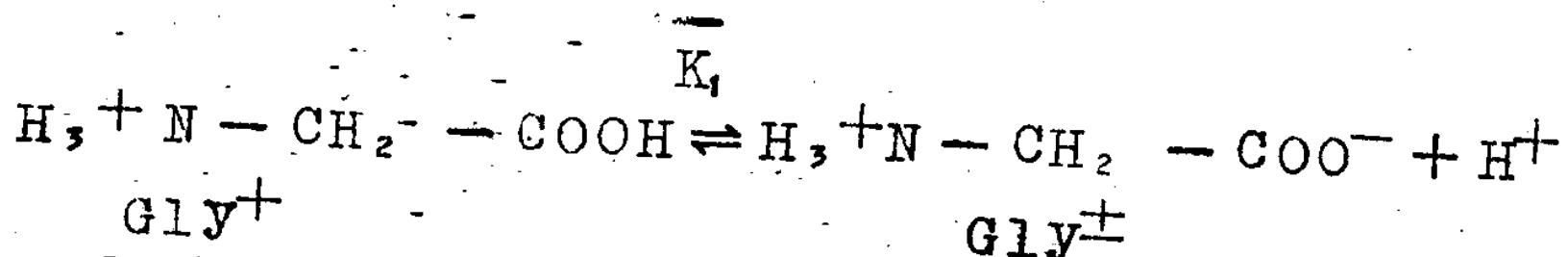
经实验证明，不同条件时带电情况如下：



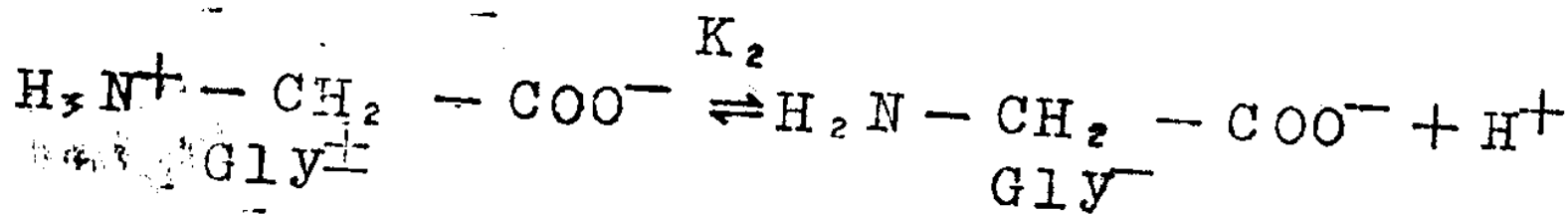
pH1时的主要形式 pH7时的主要形式 pH11时的主要形式

不电离的形式在溶液中只以少量存在。

依照 Brønsted 的酸碱理论，酸是质子的供体，碱是质子的受体。氨基酸是酸又是碱。如以甘氨酸为例：



$$K_1 = \frac{[\text{GLY}^\pm][\text{H}^+]}{[\text{GLY}^+]}$$



$$K_2 = \frac{[\text{GLY}^-][\text{H}^+]}{[\text{GLY}^\pm]}$$

(1) 中 GLY^{\pm} 接受一个质子，所以是碱。(2) 中 GLY^{\pm} 是质子的供体，所以是酸。按定义，所谓等电点是某氨基酸所带净电荷为零时溶液的 pH 值。对于甘氨酸，此时

$$[\text{GLY}^+] = [\text{GLY}^-]$$

因此有

$$\frac{[\text{GLY}^{\pm}][\text{H}^+]}{K_1} = \frac{K_2 [\text{GLY}^{\pm}]}{[\text{H}^+]}$$

$$\therefore \frac{[\text{H}^+]}{K_1} = \frac{K_2}{[\text{H}^+]}$$

$$[\text{H}^+] = \sqrt{K_1 K_2}$$

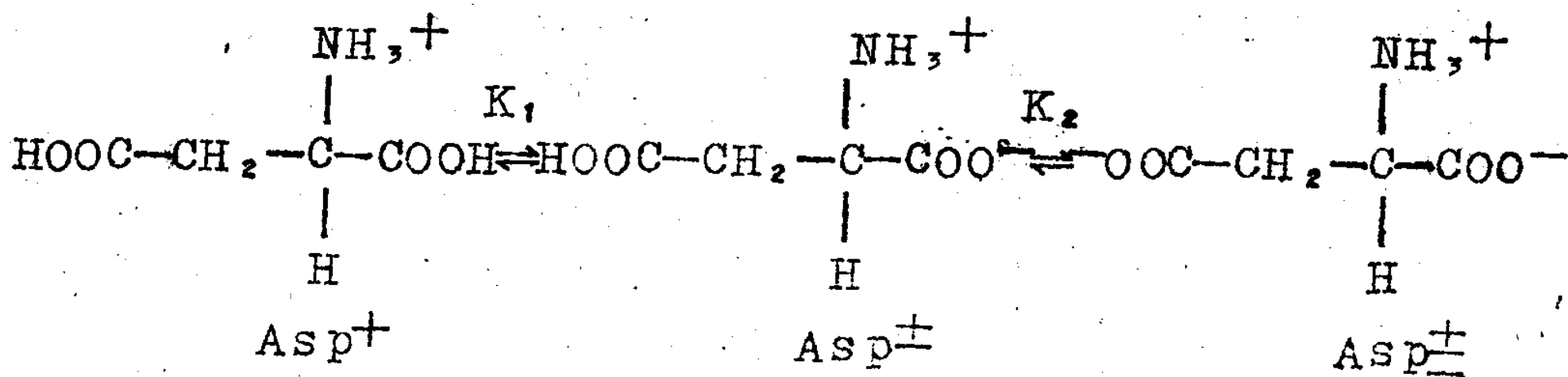
令 pI 表示等电点时的 pH，则

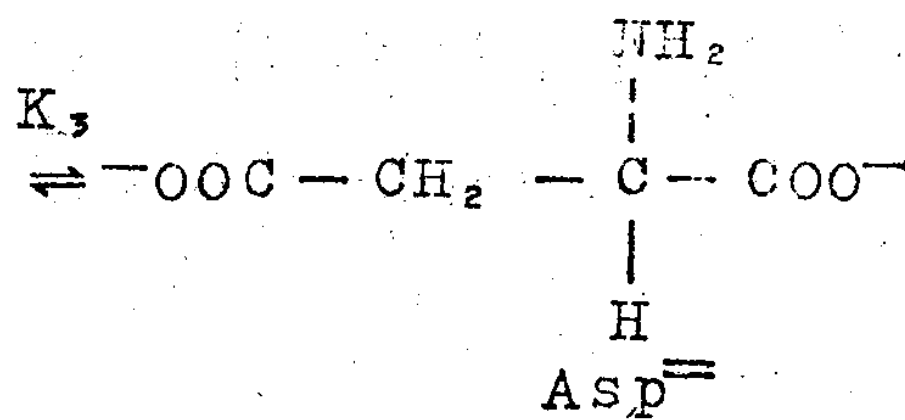
$$pI = \frac{pK_1 + pK_2}{2}$$

根据实验，甘氨酸的两个 pK 值是 2.34 与 9.60，所以它的等电点是

$$pI = \frac{2.34 + 9.60}{2} = 5.97$$

对于侧链上没有解离基团的氨基酸来讲，上述处理完全适用。侧链上有解离基团的氨基酸的 pI 如何计算呢？试以天冬氨酸为例来说明。天冬氨酸解离时：

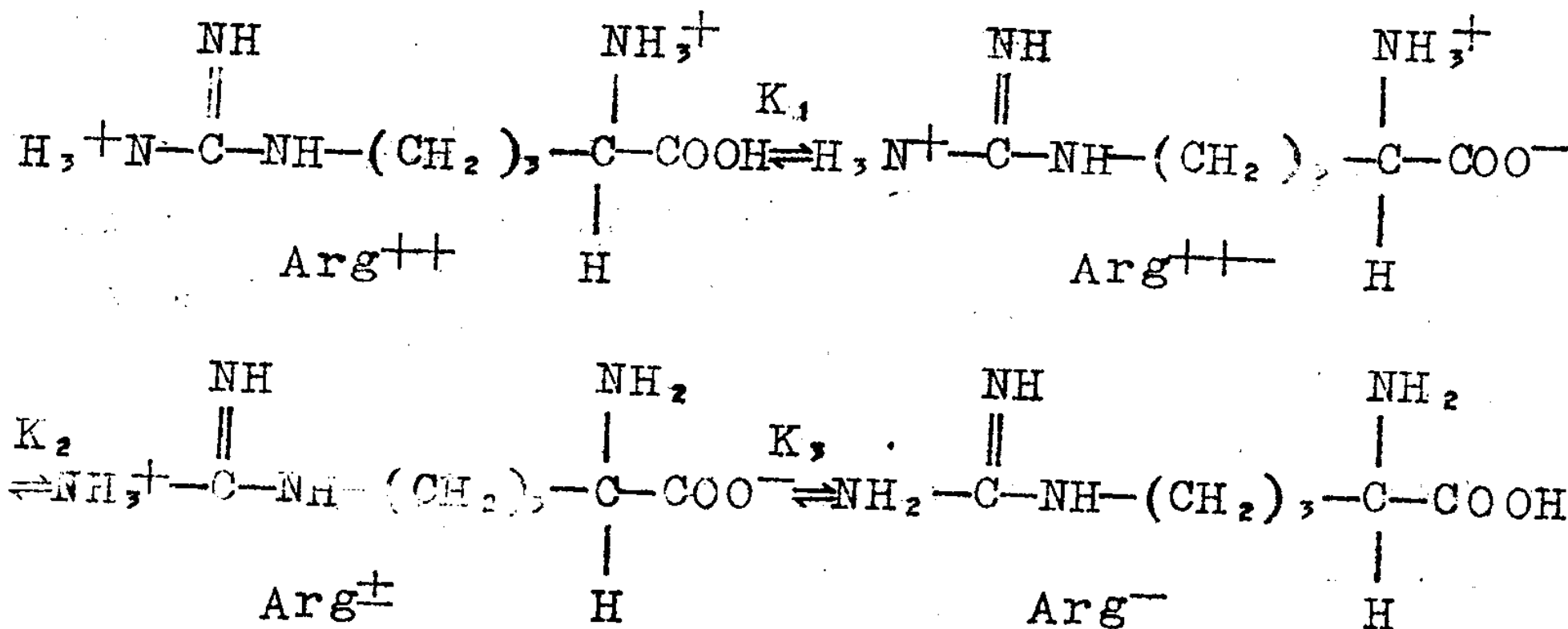




等电点时是 Asp^{\pm} 。此时 $[\text{Asp}^{\ominus}]$ 的量可忽略不计，因此天冬氨酸的等电点是

$$\text{pI} = \frac{\text{pK}_1 + \text{pK}_2}{2}$$

又如精氨酸，解离情况如下：



与天冬氨酸同， Arg^{\pm} 是等电点， $[\text{Arg}^{++}]$ 的影响极小，因此

$$\text{pI} = \frac{\text{pK}_2 + \text{pK}_3}{2}$$

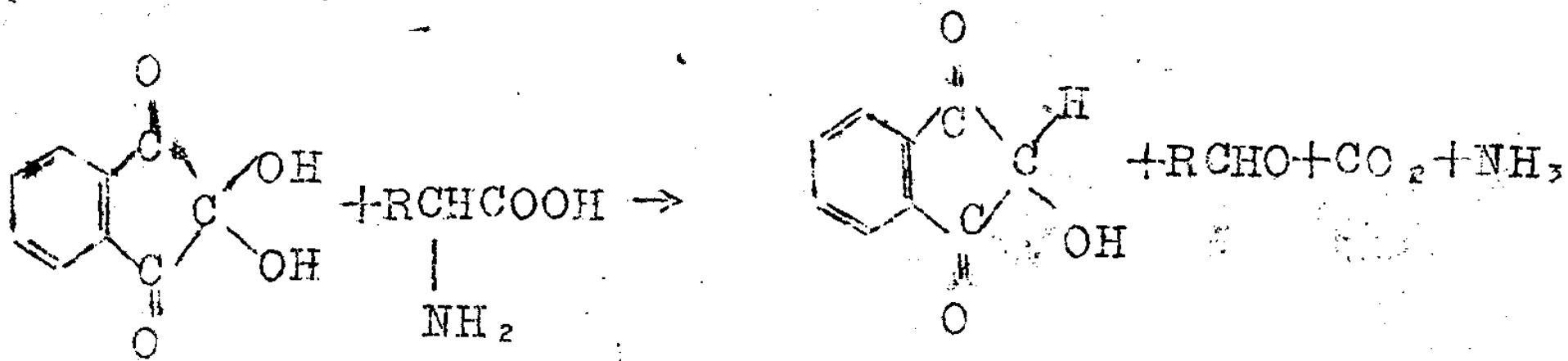
其它含有多功能团的氨基酸的解离情况及等电点的计算，与此相似。

氨基酸的化学性质

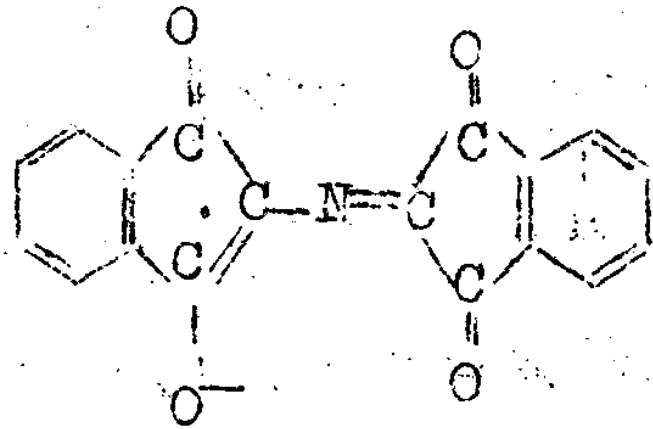
烷基与芳香基在化学上是不活泼的，因此氨基酸的化学性质主要由其它的功能团决定。这些功能团有：氨基、羧基、胍基、咪唑基、

羧基、羟基、酚基及吲哚基。下面将讨论一些生化研究工作经常遇到的化学反应，主要是氨基的化学反应。

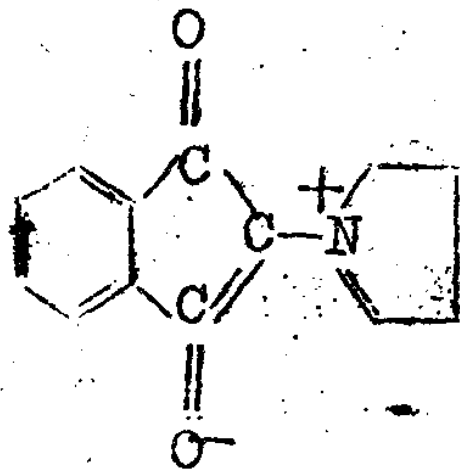
氨基的反应 氨基酸与茚三酮的反应是非常有用的反应，可以定性检出氨基酸或定量测定氨基酸。反应第一步是：



这步反应的产物 RCHO、CO₂ 及 NH₃，是定量转变来的。所产生的 NH₃，又定量地与一分子茚三酮及一分子还原型茚三酮合成 Ruhemann 紫：



这一产物显示紫蓝色，吸收光谱峰在 570nm。茚三酮与天冬氨酸反应释放二分子的 CO₂，与脯氨酸及羟脯氨酸反应不释放 NH₃，而直接生成黄色产物，结构是



氨基的酰化是比较方便的，乙酰氯、醋酸酐、苯甲酰氯、邻苯二