

第五章 蛋白质



1. 氨基酸

2. 肽键和多肽

3. 蛋白质

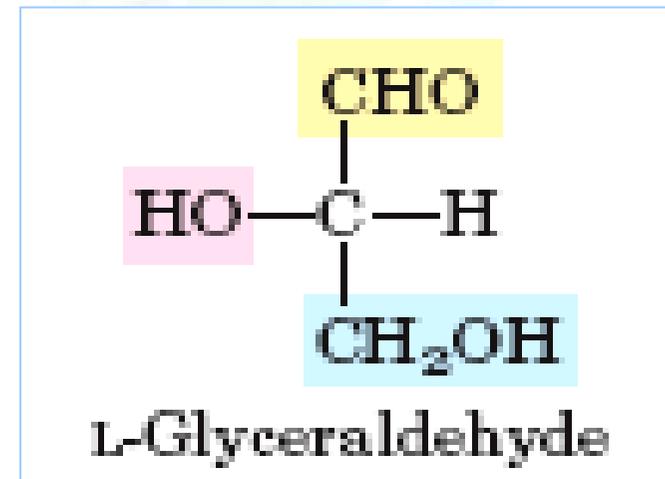
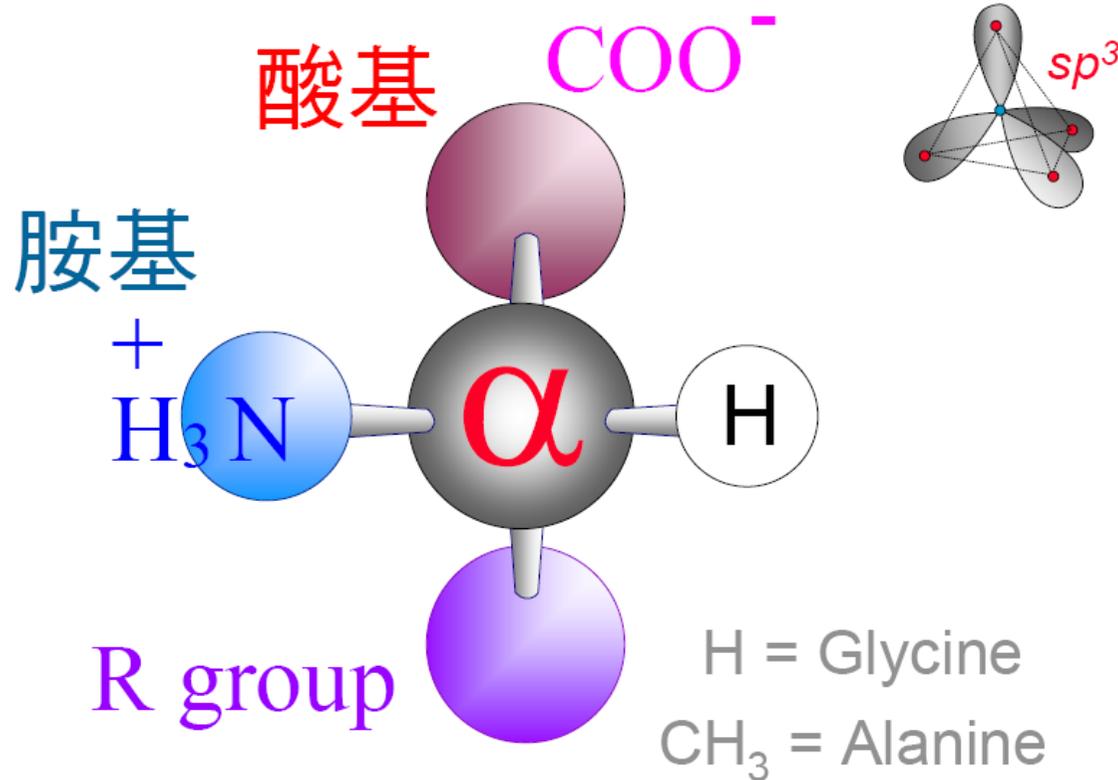


1.1. 氨基酸的分类、结构和名称

1.1.1. 氨基酸是蛋白质的基本组成单位

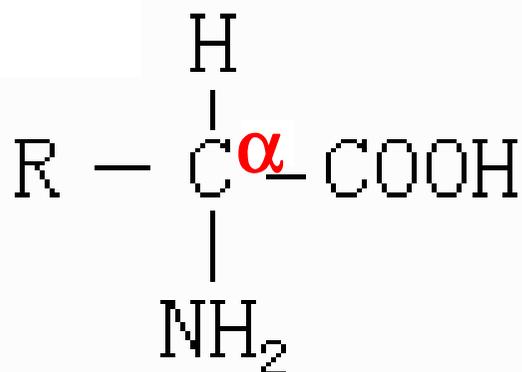
- ① 蛋白质主要元素组成为**C、H、O、N、S**。
蛋白质平均含**N**量为**16%**,这是凯氏定氮法测蛋白质含量的理论依据:
$$\text{蛋白质含量} = \text{蛋白质含N量} \times 6.25$$
- ② 蛋白质通过**酸(HCl或H₂SO₄)**水解、**碱(NaOH)**水解或酶水解可降解为多肽,最终完全水解为其组成**氨基酸**。
- ③ 常见氨基酸有**20**种,其中**8(9)**种是人体必须氨基酸。

1.1.2. 氨基酸的结构通式

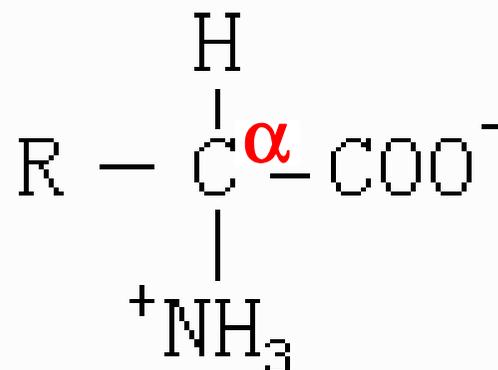


- ① 氨基连接在 α 碳上，因此称 α -氨基酸
- ② 蛋白质中的氨基酸都是 **L** 型的

1.1.3. 中性pH下，氨基酸是兼(两)性离子

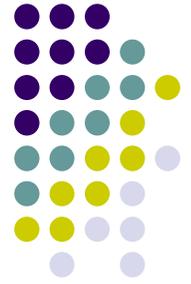


非解离形式



两性离子形式

- ① α -氨基酸都是白色晶体，熔点都在200℃以上。
- ② 每种氨基酸都有特殊的晶型，可据此鉴别不同的氨基酸。
- ③ 除酪氨酸外，一般都能溶于水。



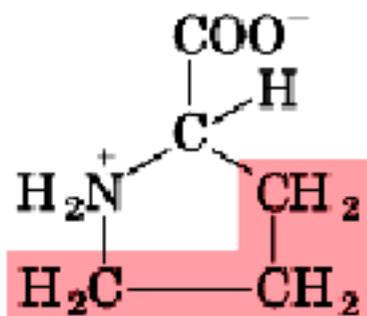
1.1.4. 生物体内氨基酸的分类

- ① 常见的蛋白质氨基酸：**20种氨基酸**
- ② 稀有的蛋白质氨基酸：从少数蛋白质中还分离出一些稀有氨基酸，它们都是相应**常见氨基酸的衍生物**。如**4-羟脯氨酸**、**5-羟赖氨酸**等。
- ③ **非蛋白质氨基酸**：生物体内呈游离或结合态的氨基酸。

1.1.5. 常见的二十种氨基酸

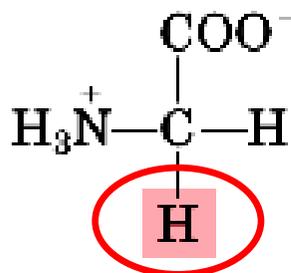


① 非极性 脂肪族 氨基酸 (7种)

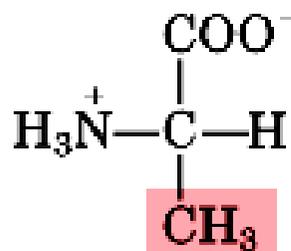


Proline
脯氨酸

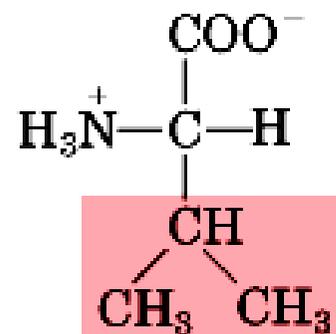
Nonpolar, aliphatic R groups



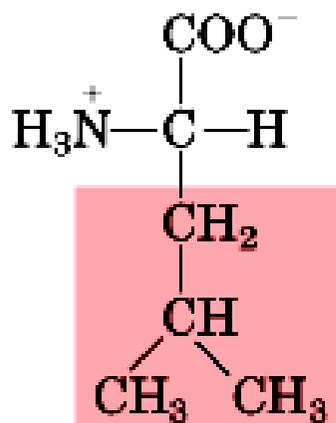
Glycine
甘氨酸



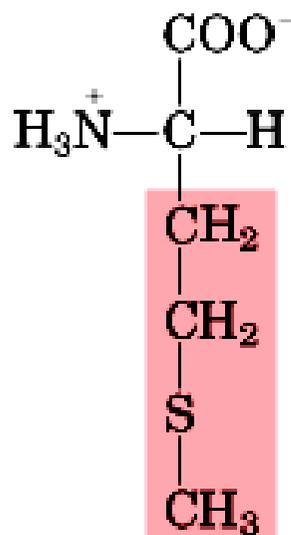
Alanine
丙氨酸



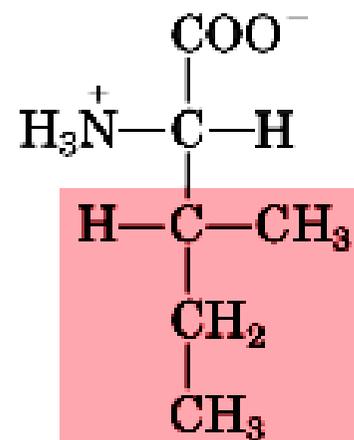
Valine
缬氨酸



Leucine
亮氨酸



Methionine
甲硫氨酸

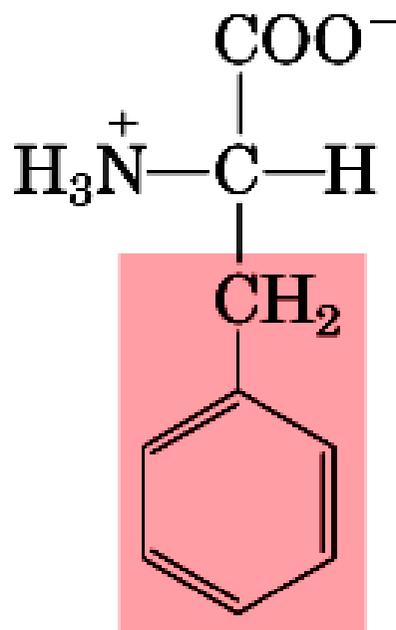


Isoleucine
异亮氨酸

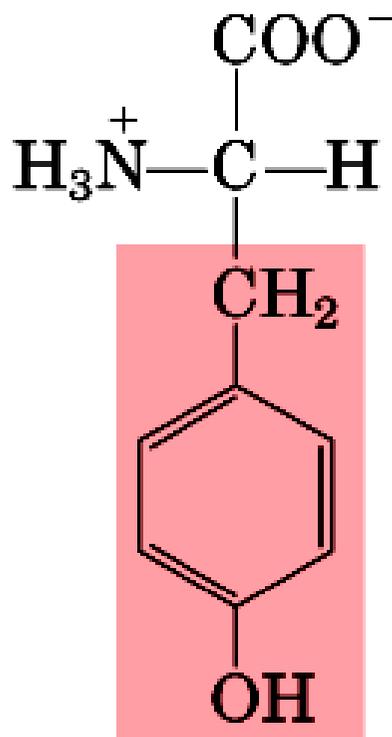
②芳香族氨基酸 (3种)



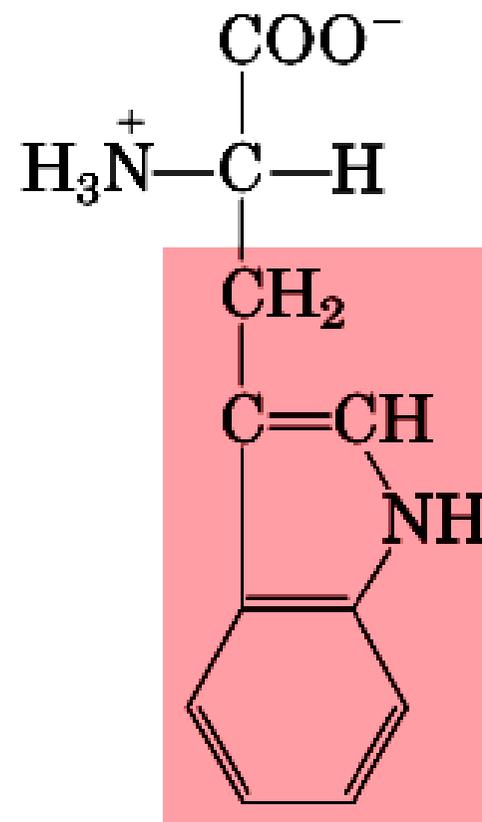
Aromatic R groups



Phenylalanine
苯丙氨酸



Tyrosine
酪氨酸

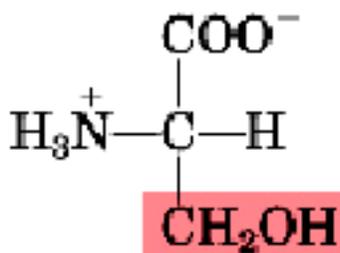


Tryptophan
色氨酸

③不带电极性氨基酸
(5种)

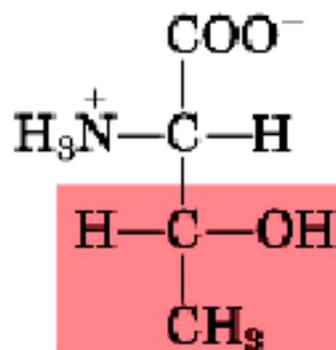
亲水性高于非极性氨基酸。

Polar, uncharged R groups



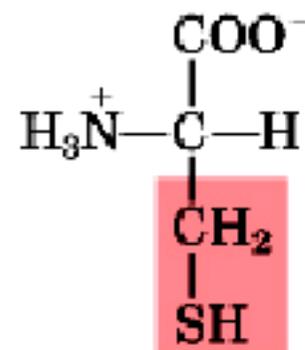
Serine

丝氨酸



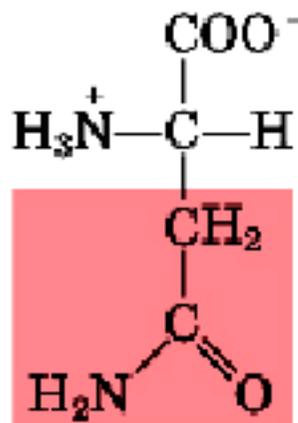
Threonine

苏氨酸

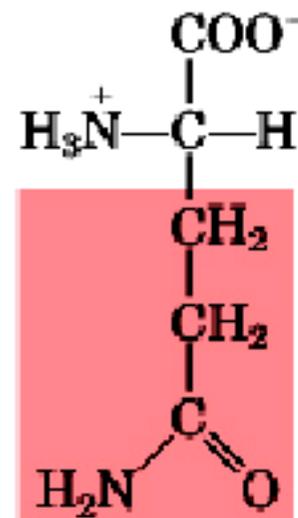


Cysteine

半胱氨酸



Asparagine
天冬酰胺



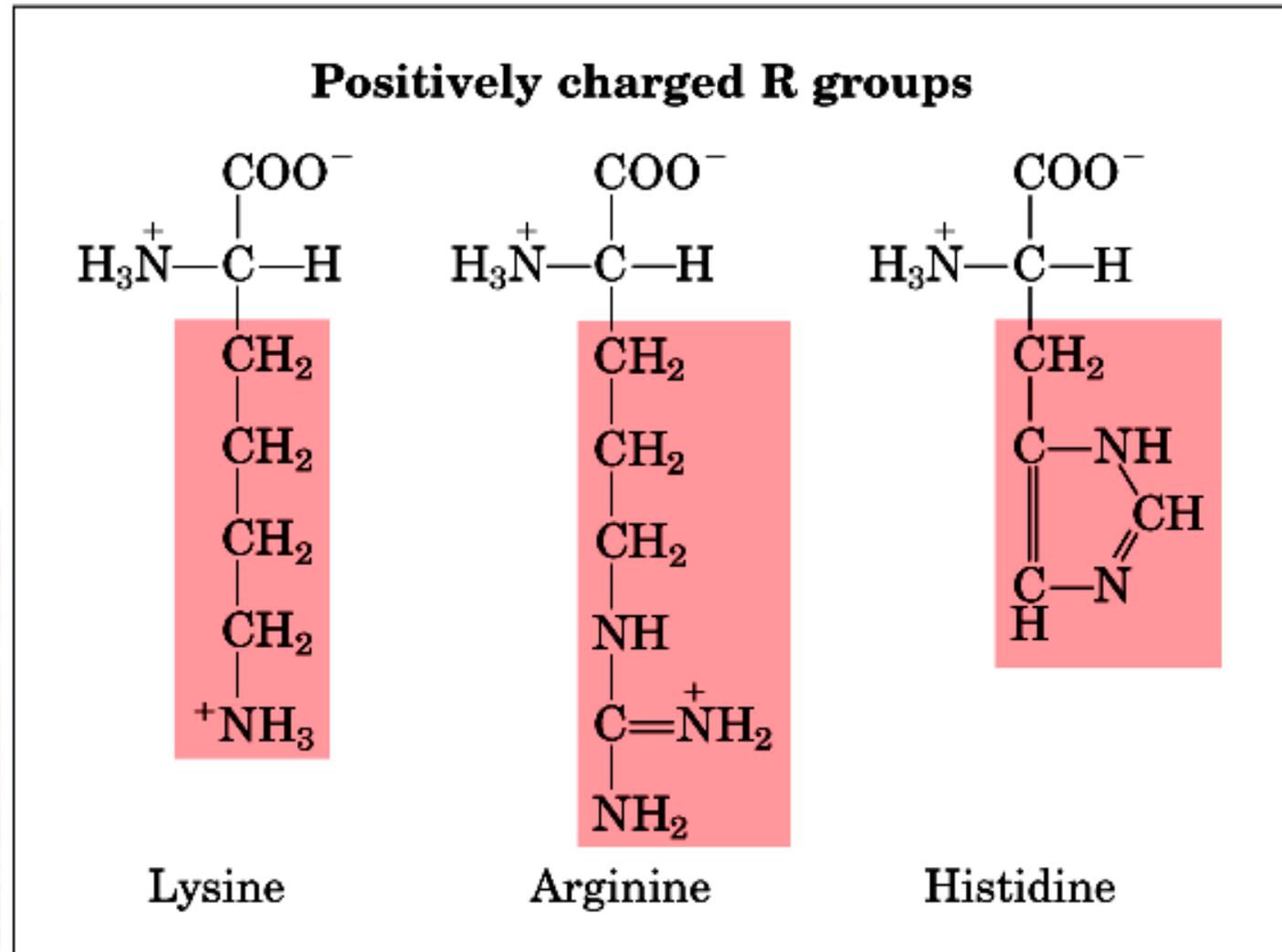
Glutamine
谷氨酰胺

④带正电（碱性）氨基酸（3种）



亲水性强

His的咪唑基其 pK_a 接近中性，是潜在质子供体/受体。



赖氨酸

精氨酸

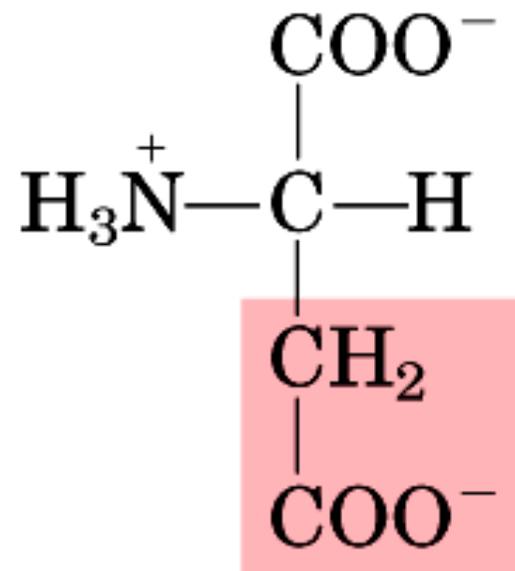
组氨酸

⑤带负电（酸性）氨基酸（2种）

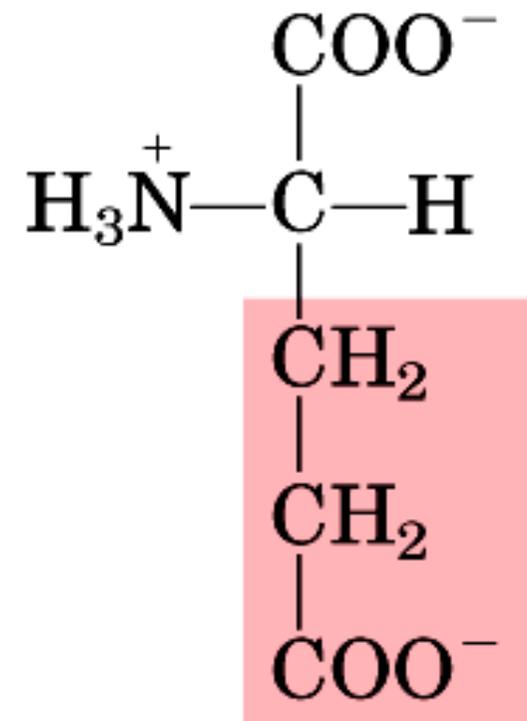


亲水性
最强

Negatively charged R groups



Aspartate
天冬氨酸



Glutamate
谷氨酸

Summery



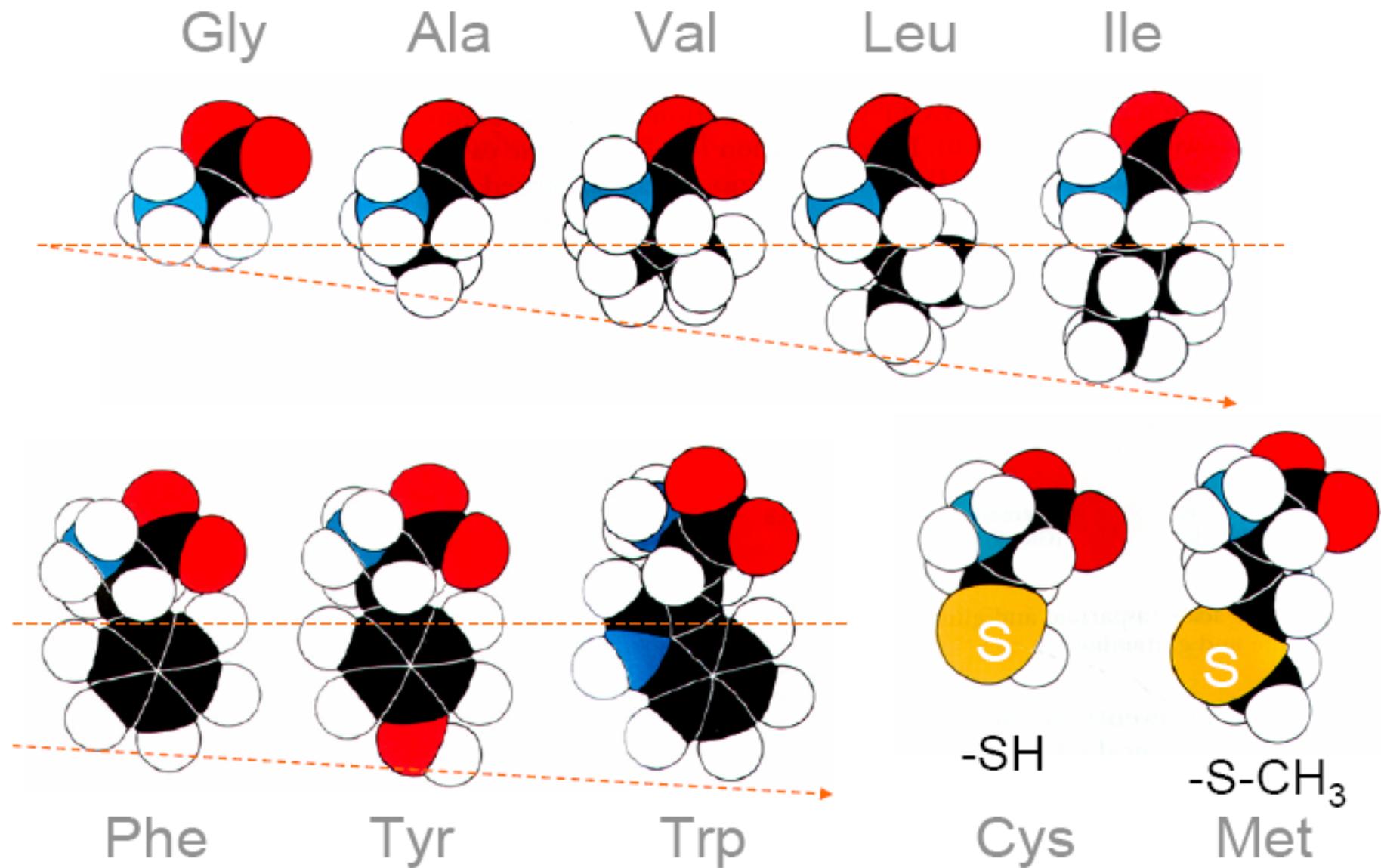
1. **脂肪族氨基酸**: Gly, Ala, Pro, Val, Leu, Ile, Met
2. **芳香族氨基酸**: Phe, Tyr, Trp
3. **极性氨基酸（不带电）**: Ser, Thr, Cys, Asn, Gln
4. **酸性氨基酸**: Asp, Glu
5. **碱性氨基酸**: Lys, Arg, His

氨基酸的侧链空间填充模型

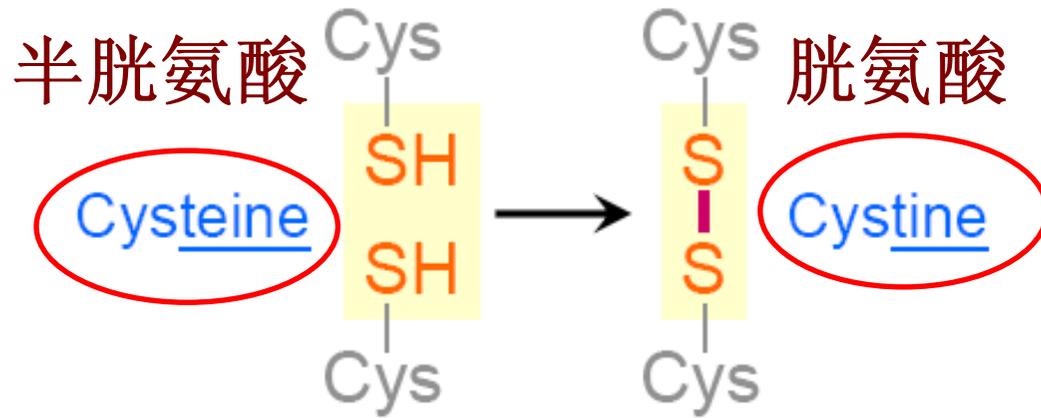
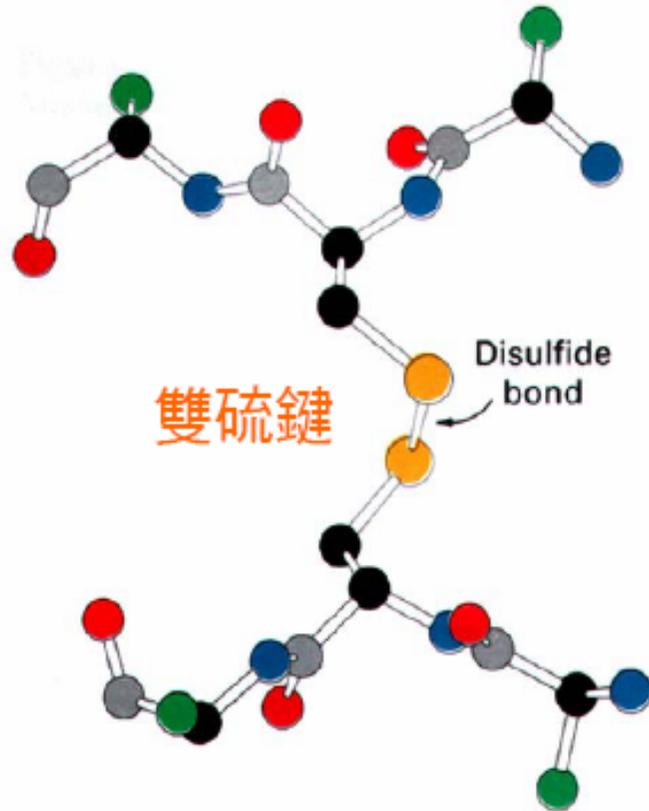


有大有小
有正有負
有極性
非極性

各式胺基酸的大小樣式齊全



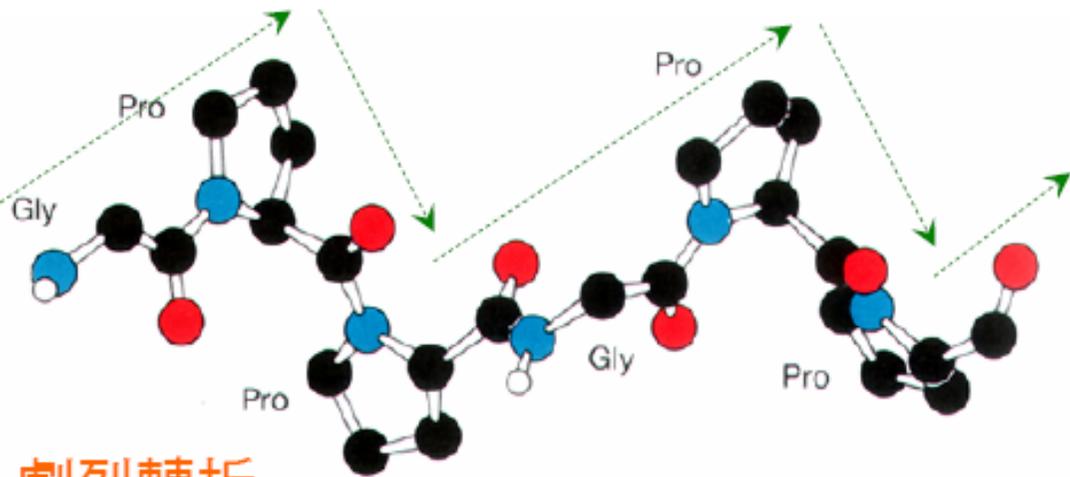
兩種胺基酸對蛋白質構造影響很大



脯氨酸



劇烈轉折



1.1.6. 常见氨基酸的三字符和单字符



中文名	英文名	三字符	单字符
甘氨酸	glycine	Gly	G
丙氨酸	alanine	Ala	A
缬氨酸*	valine	Val	V
亮氨酸*	leucine	Leu	L
异亮氨酸*	isoleucine	Ile	I
苯丙氨酸*	phenylalanine	Phe	F
脯氨酸	proline	Pro	P

1.1.6. 常见氨基酸的三字符和单字符



酪氨酸	tyrosine	Tyr	Y
色氨酸*	tryptophan	Try	W
丝氨酸	serine	Ser	S
苏氨酸*	threonine	Thr	T
天冬氨酸	aspartic acid	Asp	D
谷氨酸	glutamic acid	Glu	E

1.1.6. 常见氨基酸的三字符和单字符

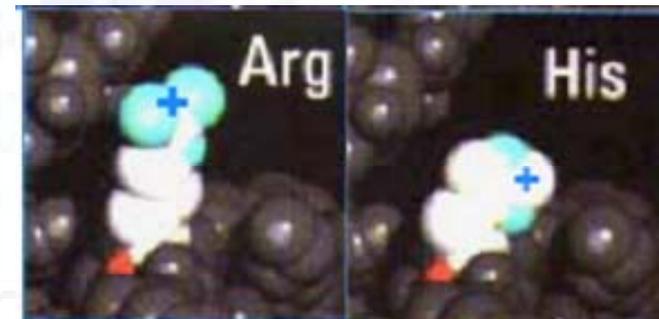


天冬酰胺	asparagine	Asn	N
谷氨酰胺	glutamine	Gln	Q
半胱氨酸	cystein	Cys	C
蛋氨酸*	methionine	Met	M
赖氨酸*	lysine	Lys	K
精氨酸	arginine	Arg	R
组氨酸	histidine	His	H

1.1.7. 八种必需氨基酸

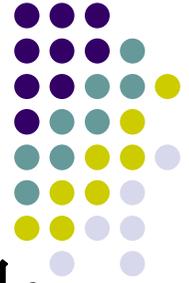


“一两色素本来淡些”（异亮氨酸、亮氨酸、色氨酸、苏氨酸、苯丙氨酸、赖氨酸、蛋氨酸、缬氨酸）。

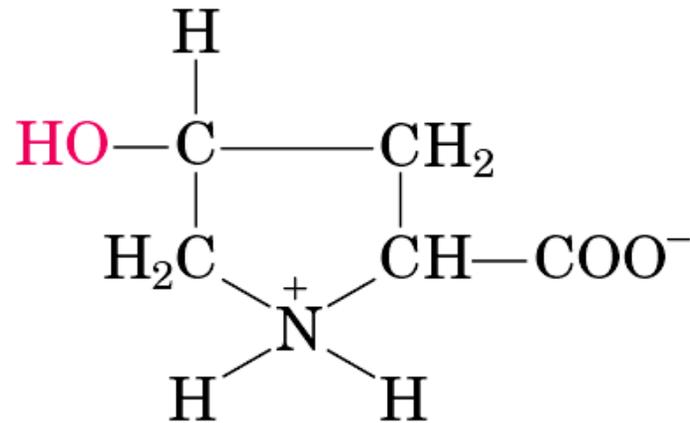


精氨酸和组氨酸

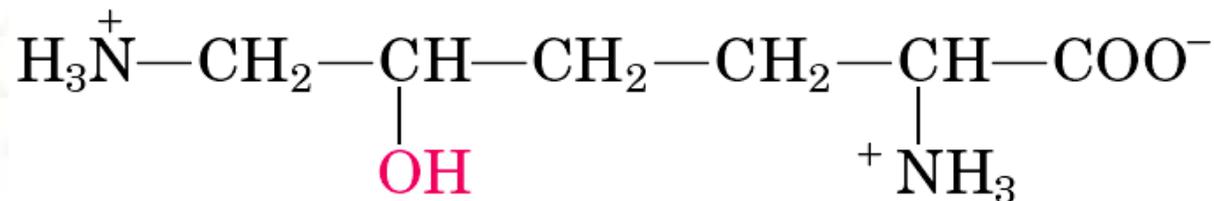
1.1.8. 稀有的蛋白质氨基酸



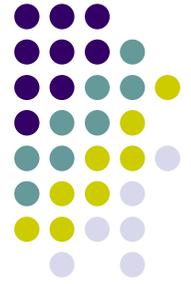
- ① **5-羟赖氨酸**和**4-羟脯氨酸**存在于**胶原蛋白**中；



4-Hydroxyproline

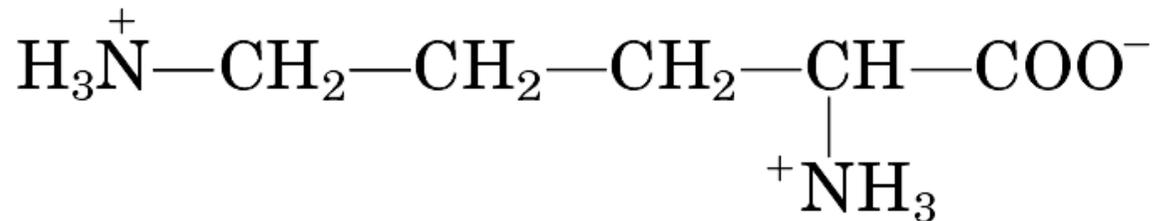


5-Hydroxylysine



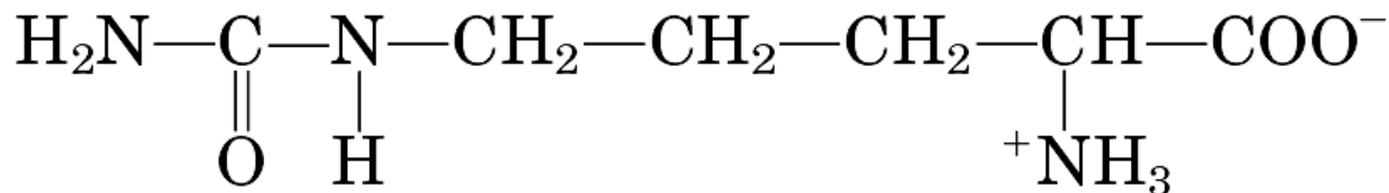
1.1.9. 非蛋白质氨基酸

- ① 在细胞中还发现了大约**300**多种不参与蛋白质组成的氨基酸，它们有广泛的功能，如**鸟氨酸**和**瓜氨酸**是尿素循环的**中间代谢物**。



Ornithine

鸟氨酸



Citrulline

瓜氨酸

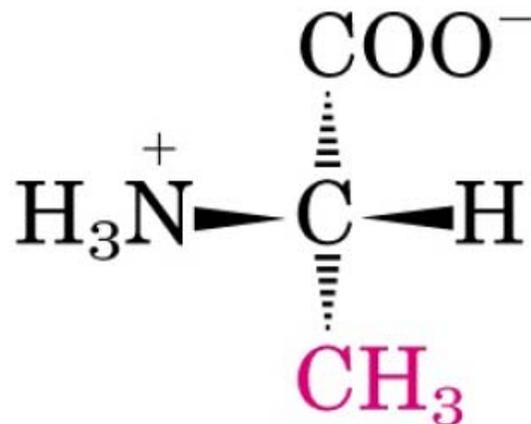
1.2. 氨基酸的基本理化性质



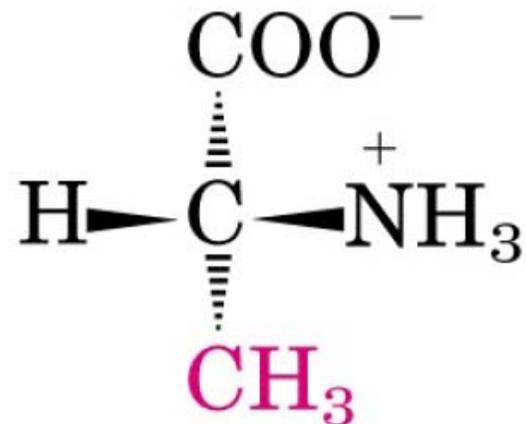
1.2.1. 氨基酸的光学性质

- ① 除甘氨酸外，所有天然 α -氨基酸都有不对称碳原子，因此所有天然氨基酸都具有旋光性。

Optical Activity



L-Alanine



D-Alanine

1.2. 氨基酸的基本理化性质



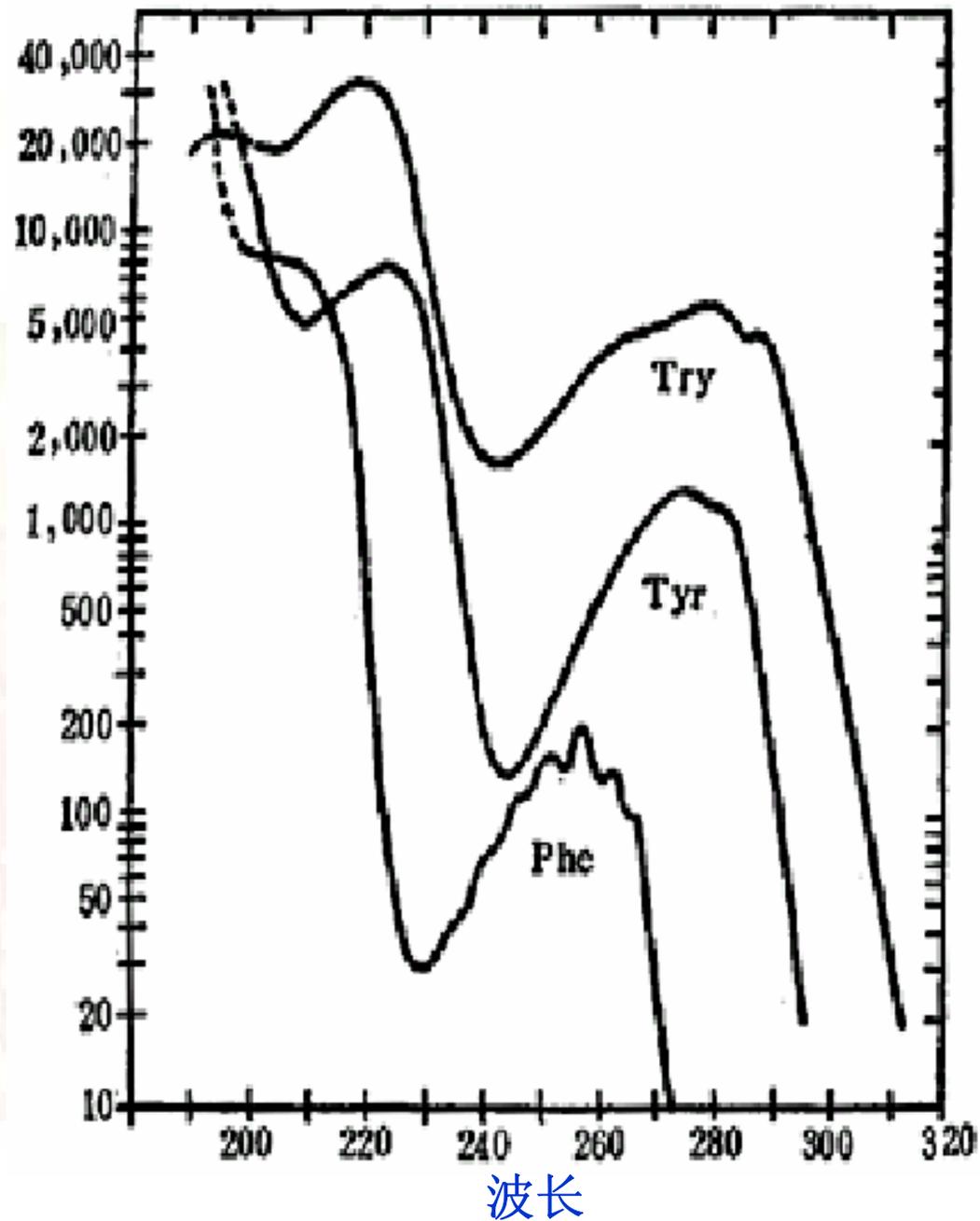
1.2.1. 氨基酸的光学性质

- ② 色氨酸(**Trp**)、酪氨酸(**Tyr**)和苯丙氨酸(**Phe**)的侧链**R**基团中含有苯环共轭双键系统，在紫外光区(**220-300nm**)有特征的吸收谱带，最大光吸收(λ_{\max})分别为**279**、**278**、和**259nm**。由于大多数蛋白质都含有这些氨基酸残基，因此用紫外分光光度法可测定蛋白质含量。

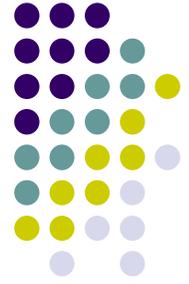
Trp
Tyr
Phe
的紫外吸收光谱

280nm

摩尔吸收系数

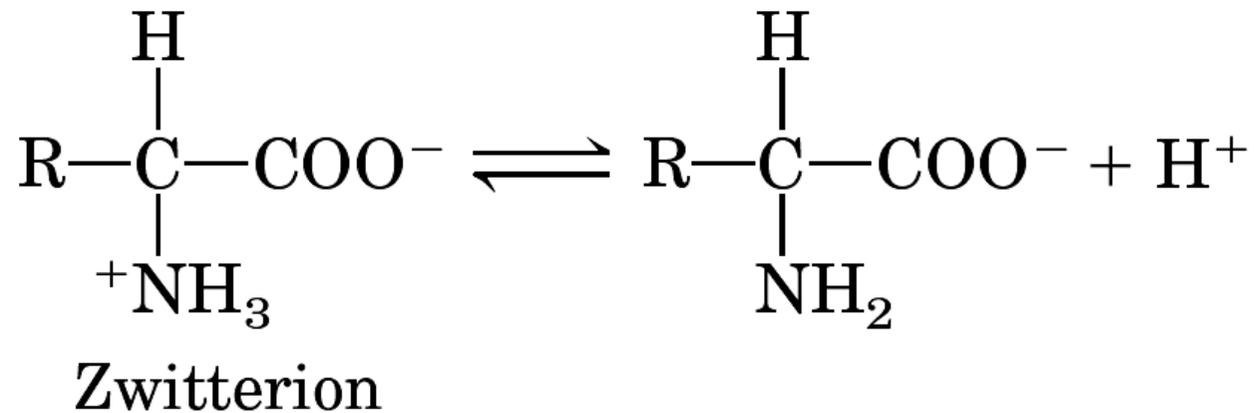


1.2.2. 氨基酸的两性解离性质及等电点

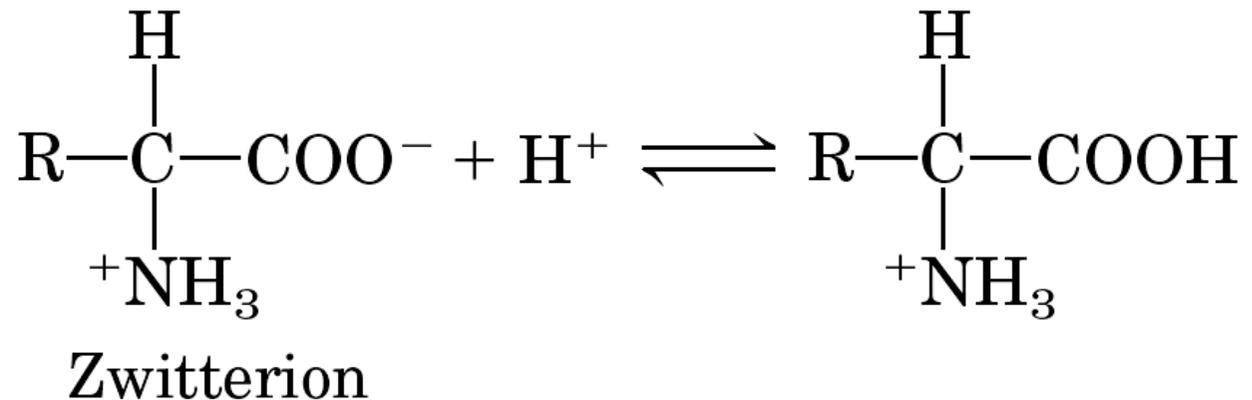


- ①氨基酸在水溶液中的**偶极离子(zwitterion)**
既是质子供体也是质子受体，它是**两性电
介质(ampholytes)**

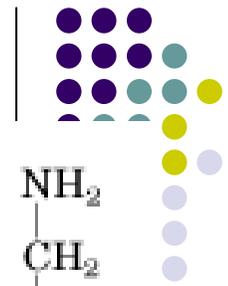
**Acid: proton
donor**



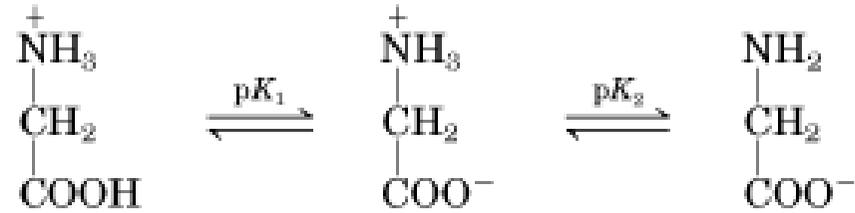
**Base: proton
acceptor**



1.2.2. 氨基酸的两性解离性质及等电点

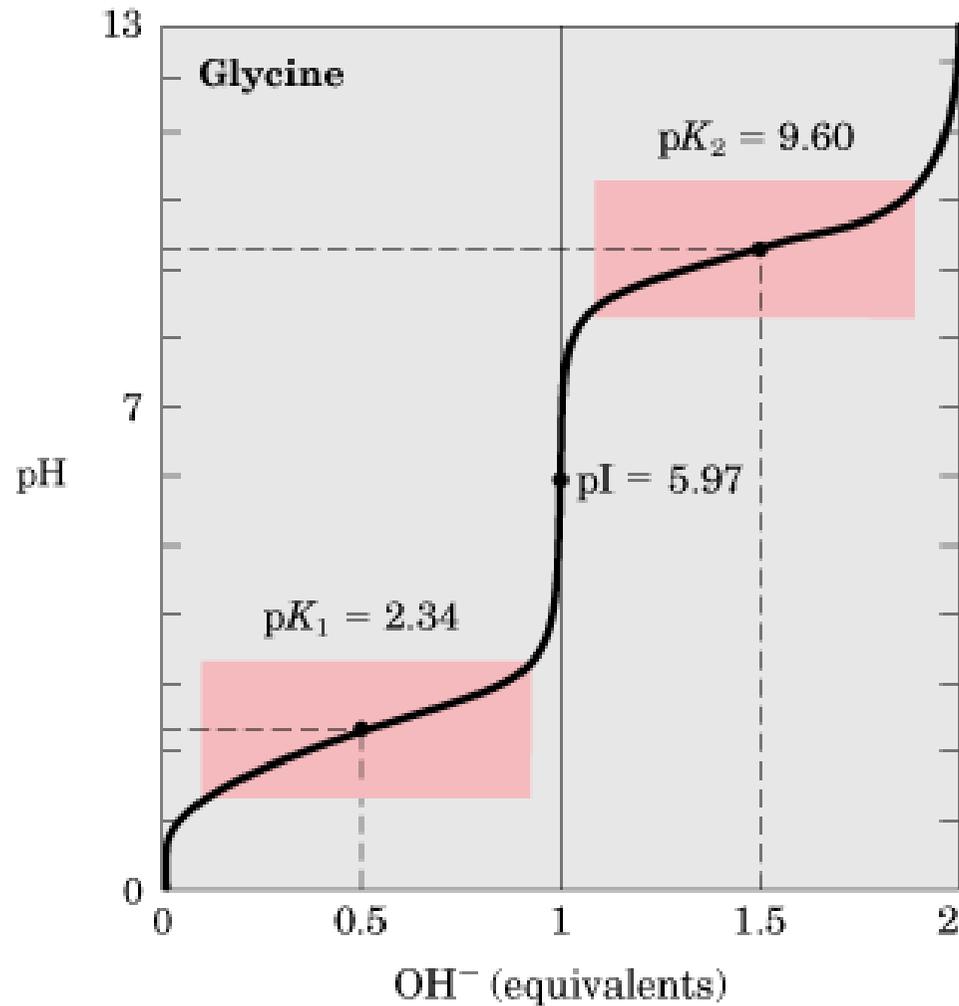


②氨基酸完全质子化时，可看成多元酸，可分部解离：

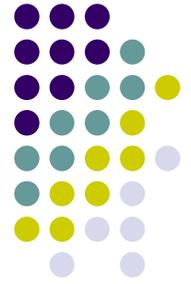


$$K_1 = \frac{[A^0][H^+]}{[A^+]}$$

$$K_2 = \frac{[A^-][H^+]}{[A^0]}$$



1.2.2. 氨基酸的两性解离性质及等电点



$$K_1 = \frac{[A^0][H^+]}{[A^+]} \quad K_2 = \frac{[A^-][H^+]}{[A^0]} \quad \longrightarrow \quad K_1 K_2 = \frac{[A^-][H^+]^2}{[A^+]}$$

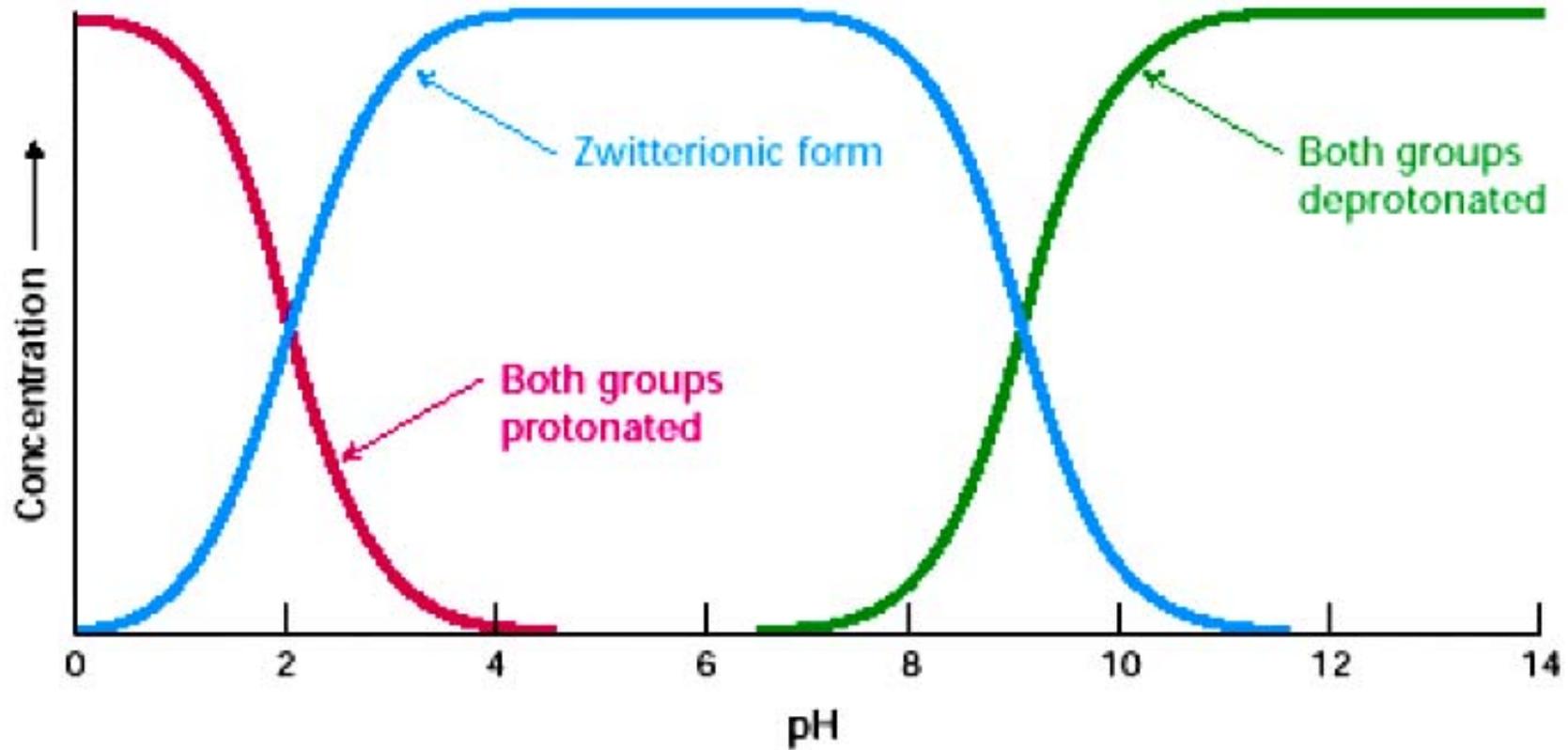
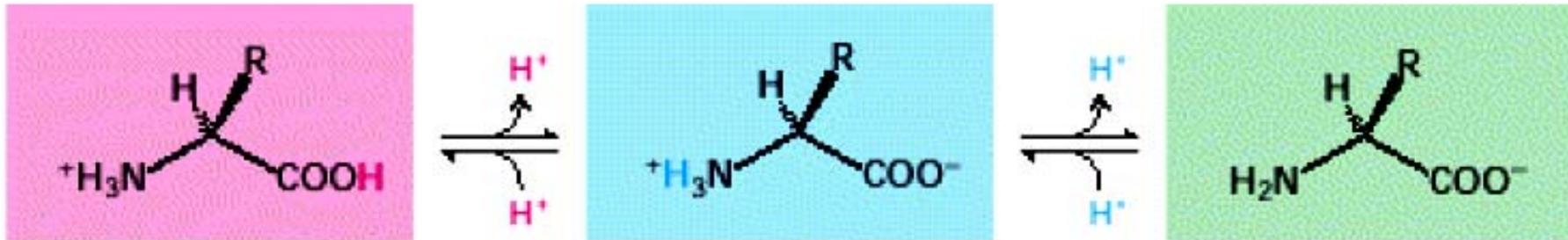
$$\longrightarrow \quad \text{pH} = \frac{\text{p}K_1 + \text{p}K_2}{2} + \log \frac{[A^-]}{[A^+]}$$

③ 对侧链不解离的中性氨基酸，当

$\text{pH} = \frac{\text{p}K_1 + \text{p}K_2}{2}$ 时，氨基酸所带静电荷为 **0**，此时的 **pH** 称为该氨基酸的 **等电点**，记为 **pI**。

即等电点(**pI**)是氨基酸静电荷为零时的**pH**

Ionization state as a function of pH

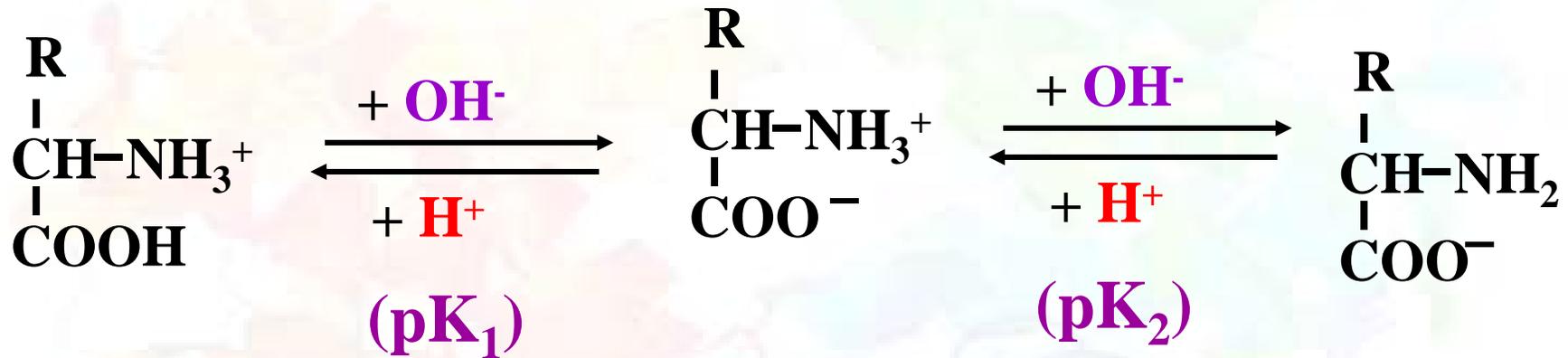


From Stryer Biochemistry

1.2.2. 氨基酸的两性解离性质及等电点



$$\text{pH} = \text{pI} + \log \frac{[\text{A}^-]}{[\text{A}^+]}$$



$\text{pH} < \text{pI}$

净电荷为正

$\text{pH} = \text{pI}$

净电荷=0

$\text{pH} > \text{pI}$

净电荷为负

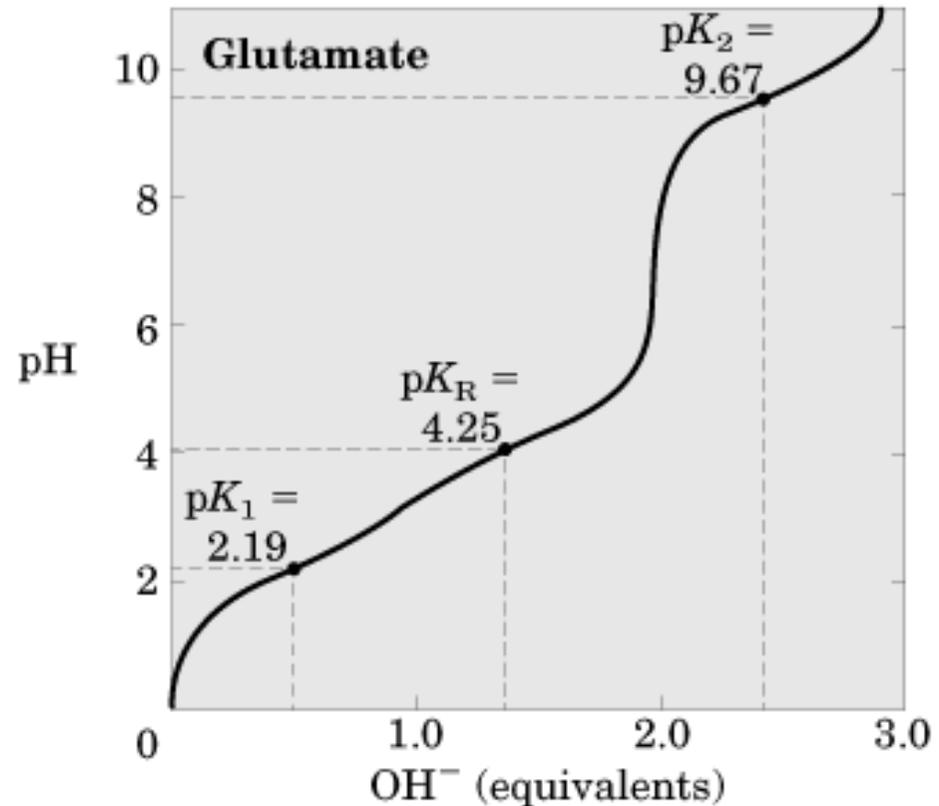
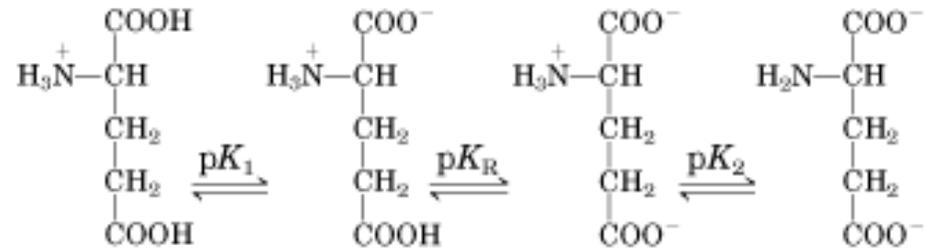
1.2.2. 氨基酸的两性解离性质及等电点



④ 氨基酸等电点的计算

一氨基二羧基AA
的等电点计算:

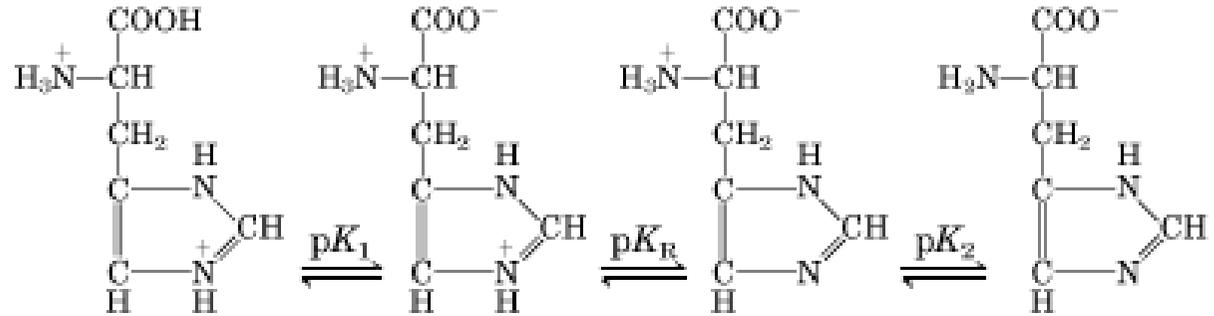
$$pI = \frac{pK'_1 + pK'_R}{2}$$



1.2.2. 氨基酸的两性解离性质及等电点

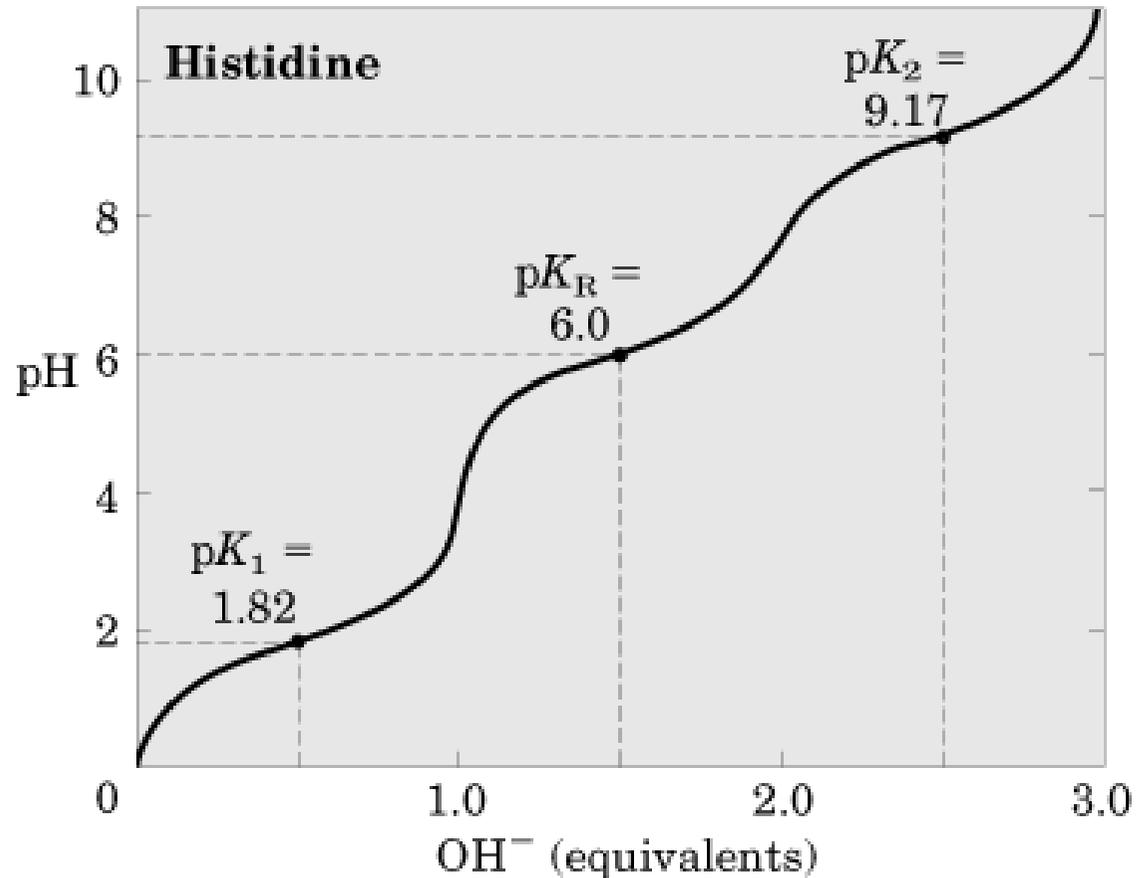


④ 氨基酸等电点的计算



二氨基一羧基AA
的等电点计算:

$$pI = \frac{pK'_2 + pK'_R}{2}$$



1.2.2. 氨基酸的两性解离性质及等电点



④ 氨基酸等电点的计算

一氨基一羧基AA的
等电点计算:

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

一氨基二羧基AA的
等电点计算:

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

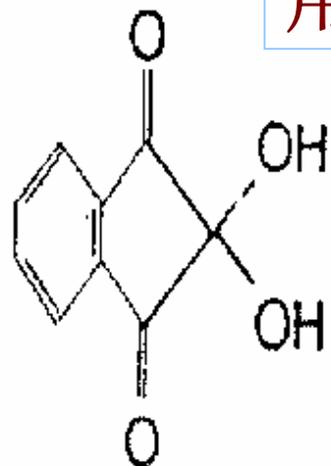
二氨基一羧基AA的
等电点计算:

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

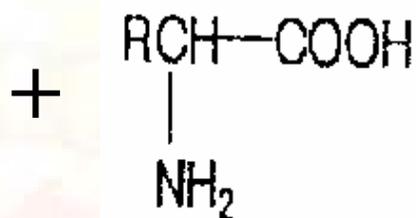
可见，氨基酸的pI值等于该氨基酸的两性离子状态两侧的基团pK值之和的二分之一。

1. 2. 3. 氨基酸与茚三酮反应

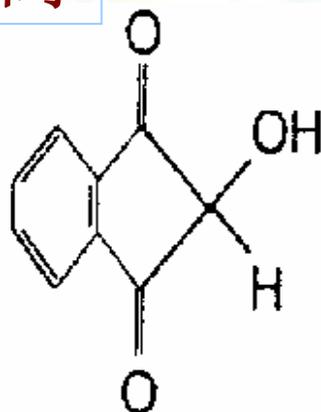
用于鉴定氨基酸，灵敏度高



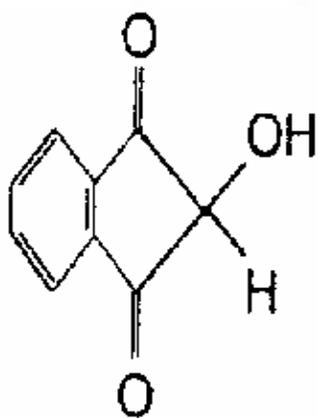
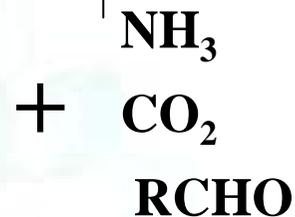
水合茚三酮(无色)



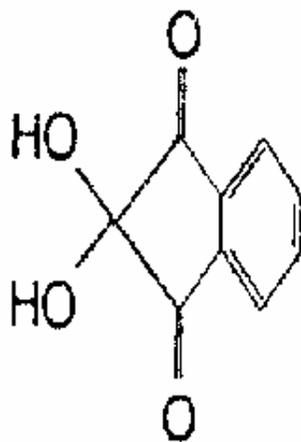
加热
(弱酸)



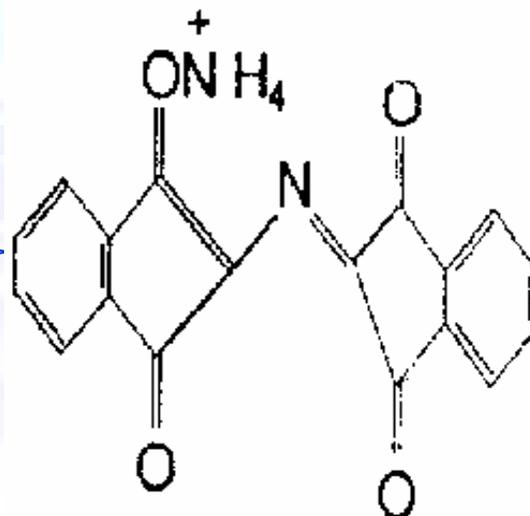
还原性茚三酮



还原性茚三酮

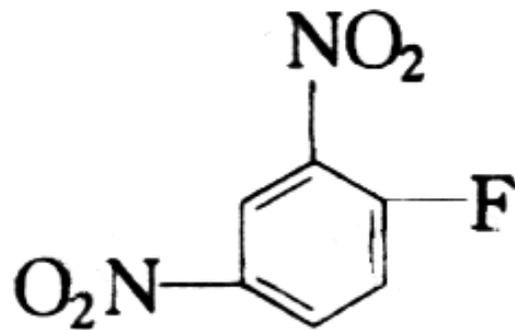


水合茚三酮



紫色化合物

1. 2. 4. 氨基酸与2, 4-二硝基氟苯 (DNFB) 的反应 (sanger反应)



DNFB
(dinitrofluorobenzene)

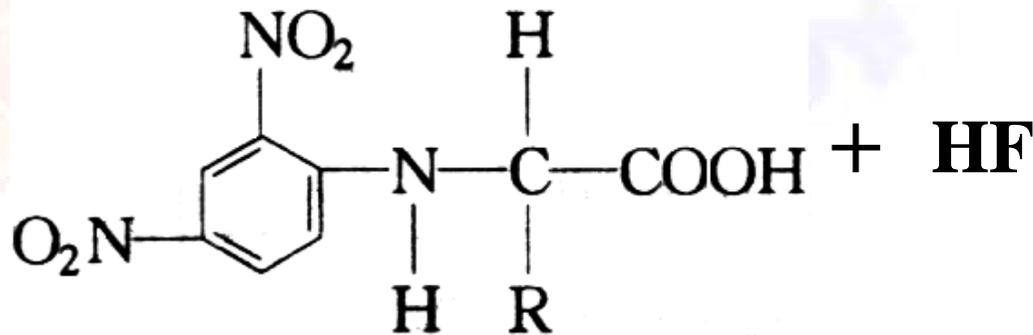
+



弱硷中



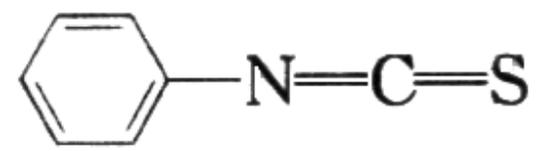
用于蛋白质
N-端测定



DNP-AA(黄色)

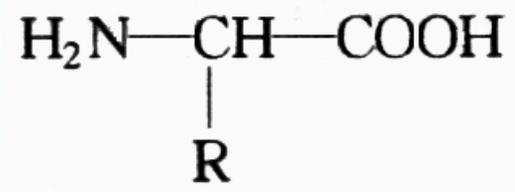


1. 2. 5. 氨基酸与苯异硫氰酯 (PITC) 的反应 (Edman反应)

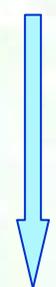


PITC (phenylisothiocyanate)

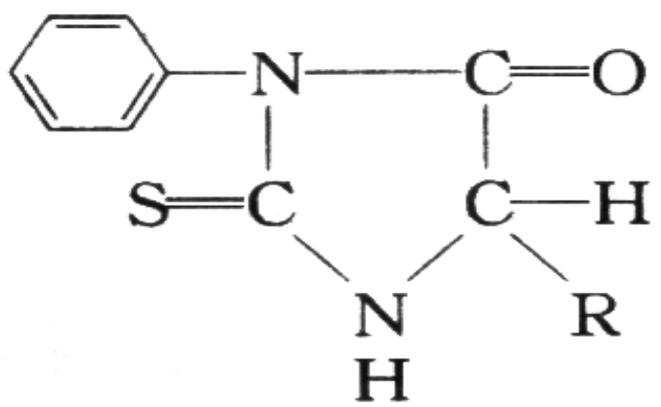
+



用于蛋白质末端测定和序列测定

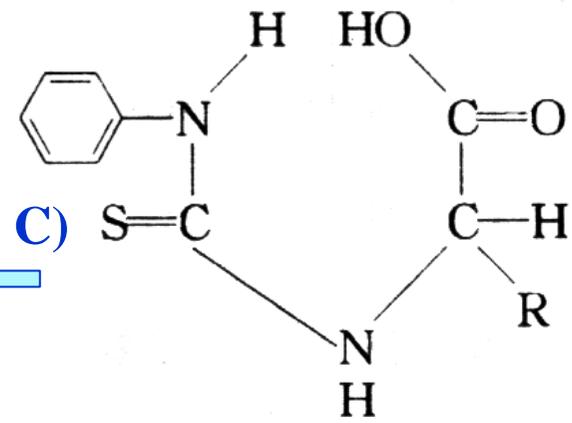


弱硷中
(40° C)



苯乙内酰硫脲衍生物(PTH-AA)
(phenylisothiohydantion-AA)

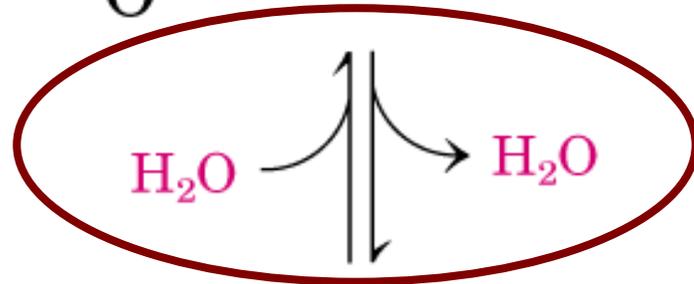
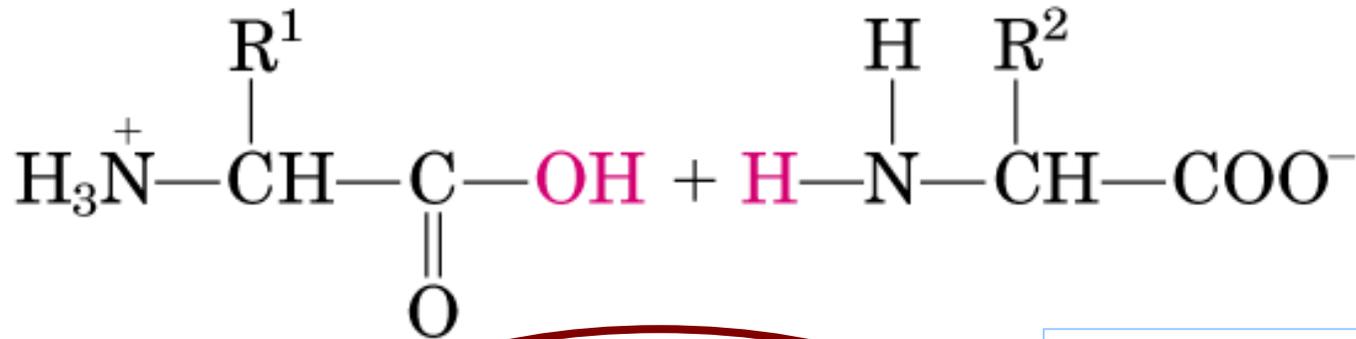
(硝基甲烷 40° C)
H⁺



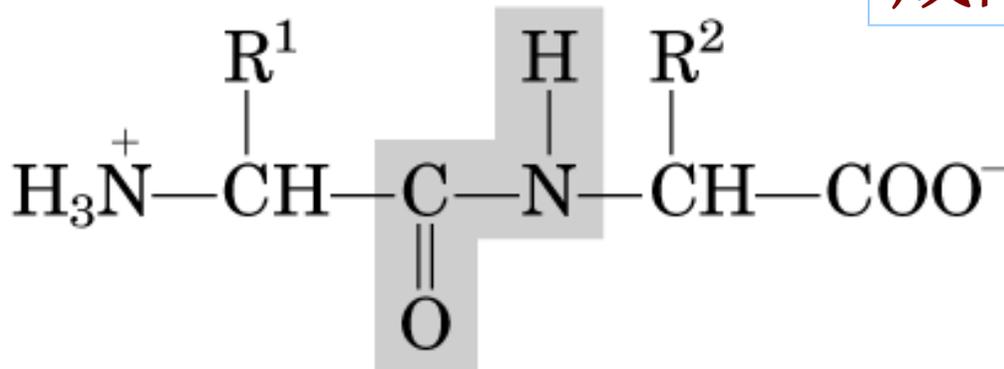
2. 肽键和多肽



2.1. 肽键（酰胺键）的形成



2分子氨基酸头尾连接，脱水缩合形成酰胺键（肽键）



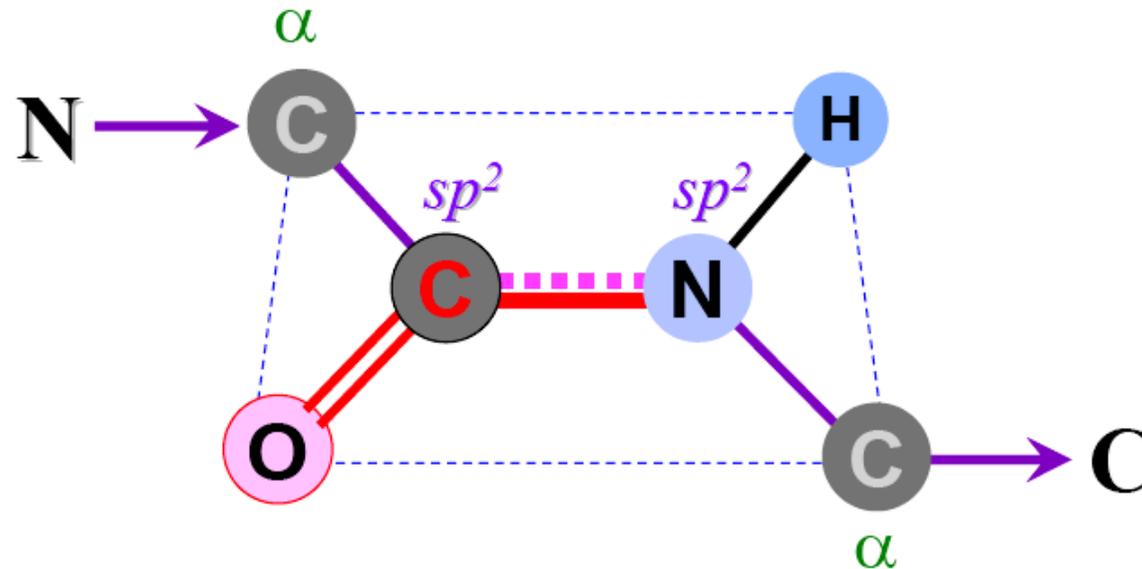
N端

C端

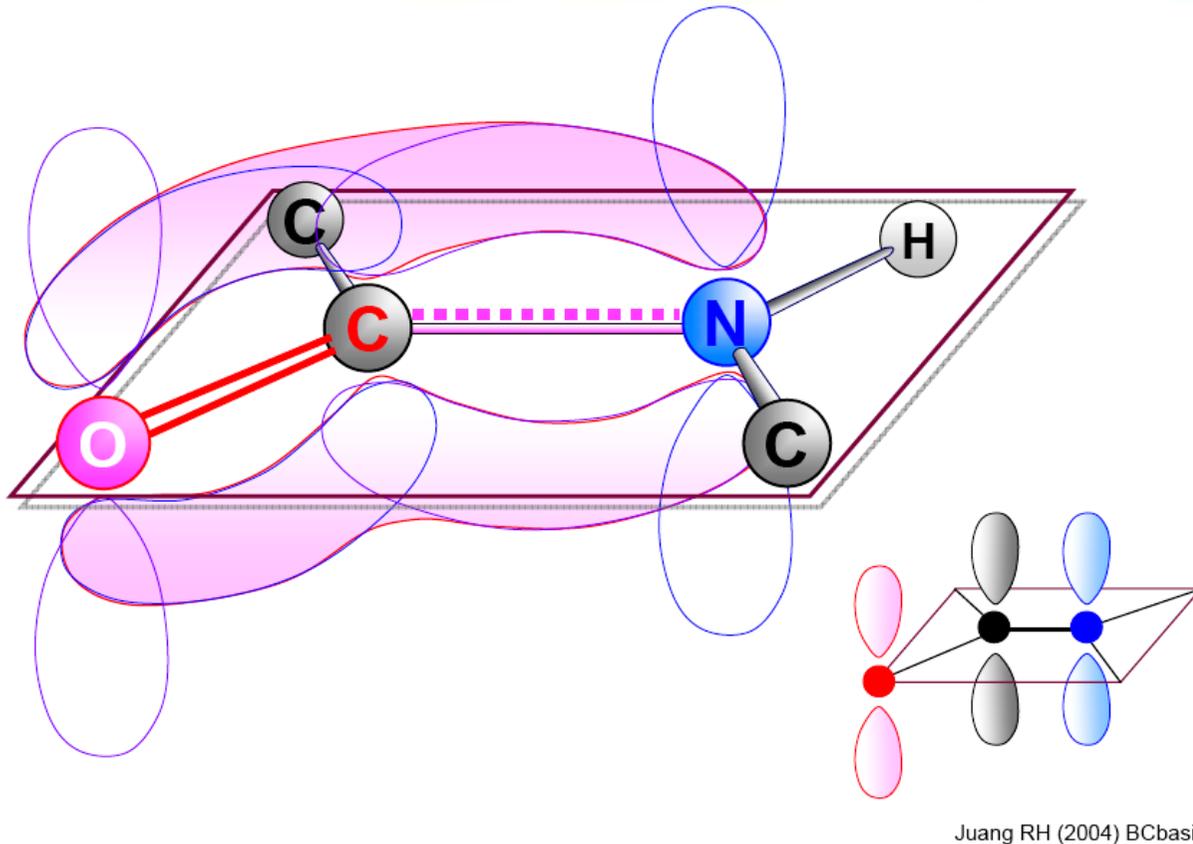


2.2. 肽键具有部分双键的性质

- ① 肽键有部分双键性质；
- ② **C—N**键不能自由旋转，周边六个原子在同一平面上，称为**肽平面**；
- ③ 前后两个氨基酸的 α -C在对角 (**trans**)。



2.2. 肽键具有部分双键的性质



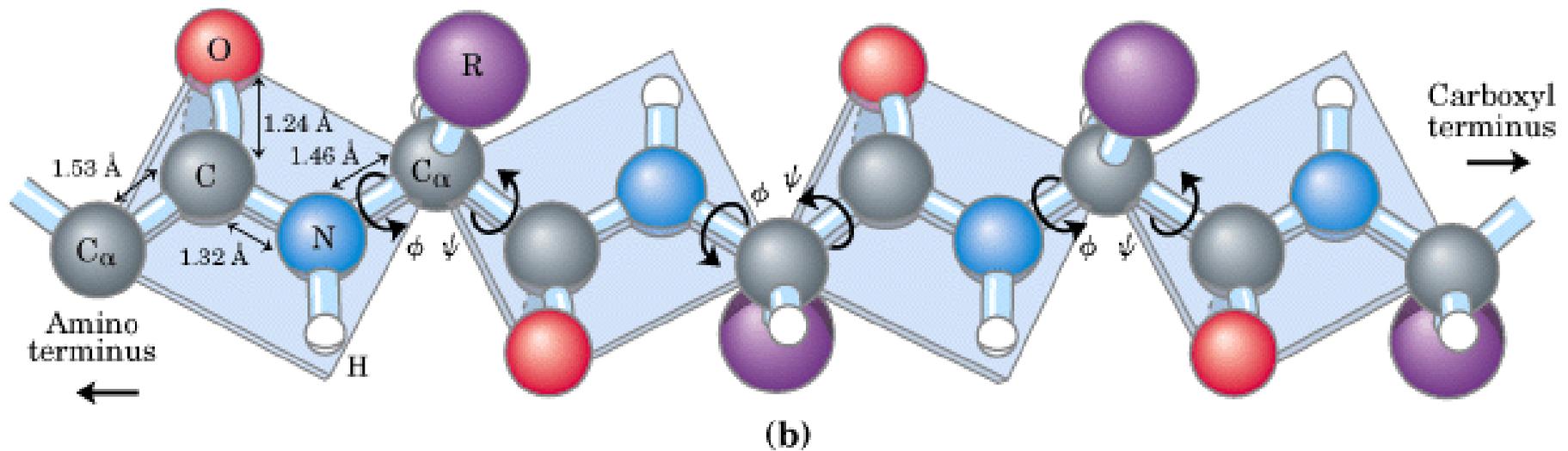
P轨道电子共振使肽键具有部分双键性质

共振杂化体
C-N键长: 0.184nm
C=O键长: 0.127nm
C-N键长: 0.132nm

C-N具40%双键性质 }
C=O具40%单键性质 } → 肽平面

2.3. 肽平面与肽单位

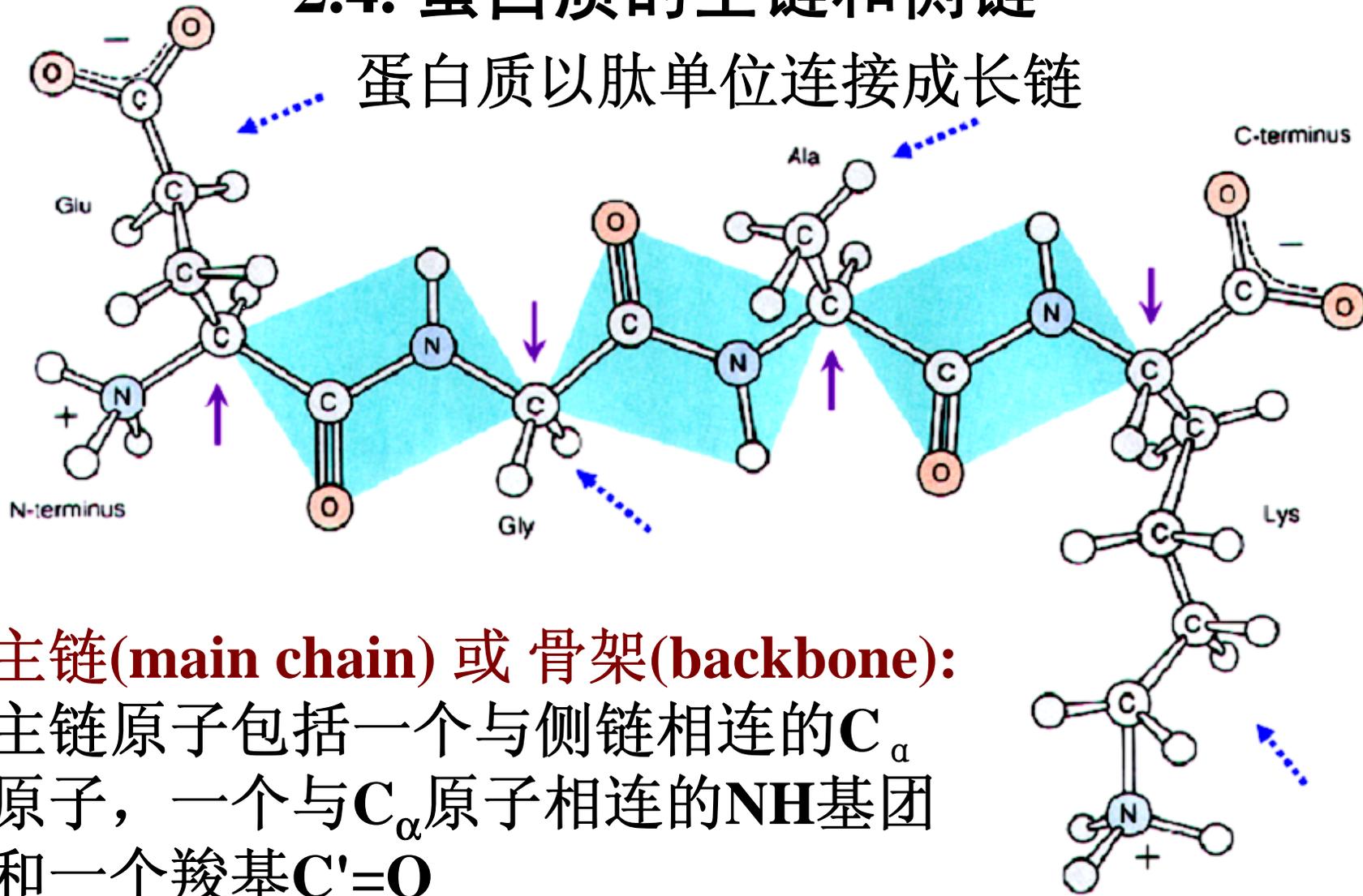
从一个 $C\alpha$ 原子到下一个 $C\alpha$ 原子之间的结构为一个肽单位



- ①两个肽平面的交接点为 $C\alpha$;
- ②前后两个 $C\alpha$ 取反式结构, $C\alpha$ 上连接侧链R基;
- ③ $C\alpha-N$ 和 $C\alpha-C'$ 是单键, 原则上可自由旋转;

2.4. 蛋白质的主链和侧链

蛋白质以肽单位连接成长链



主链(main chain) 或 骨架(backbone):
主链原子包括一个与侧链相连的C_α原子，一个与C_α原子相连的NH基团和一个羧基C'=O

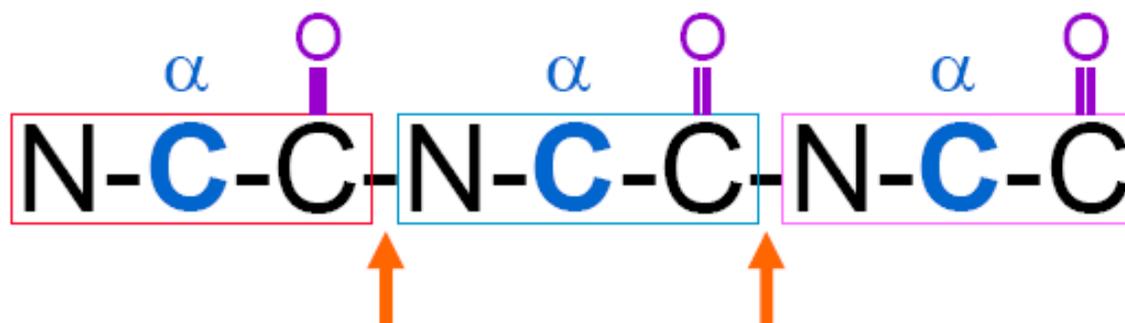
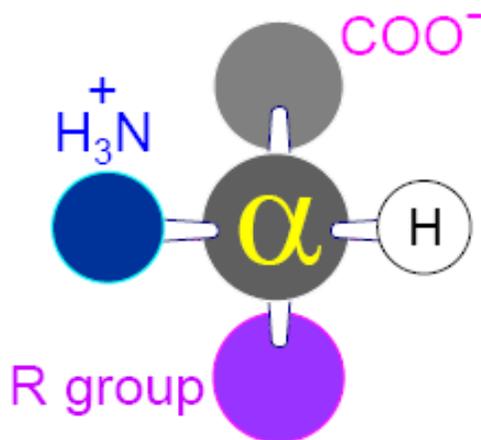
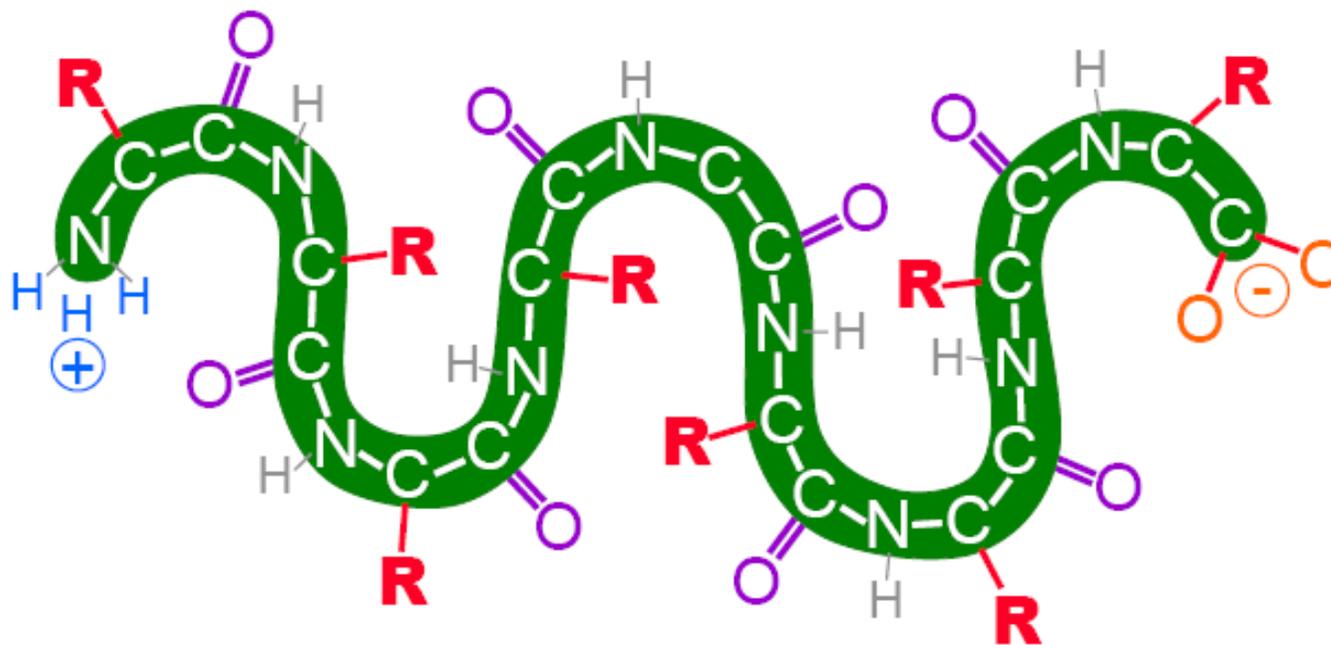
Mathews et al (2000) Biochemistry (3e) p.136

R基团为侧链，或称氨基酸残基



Constant

Variable

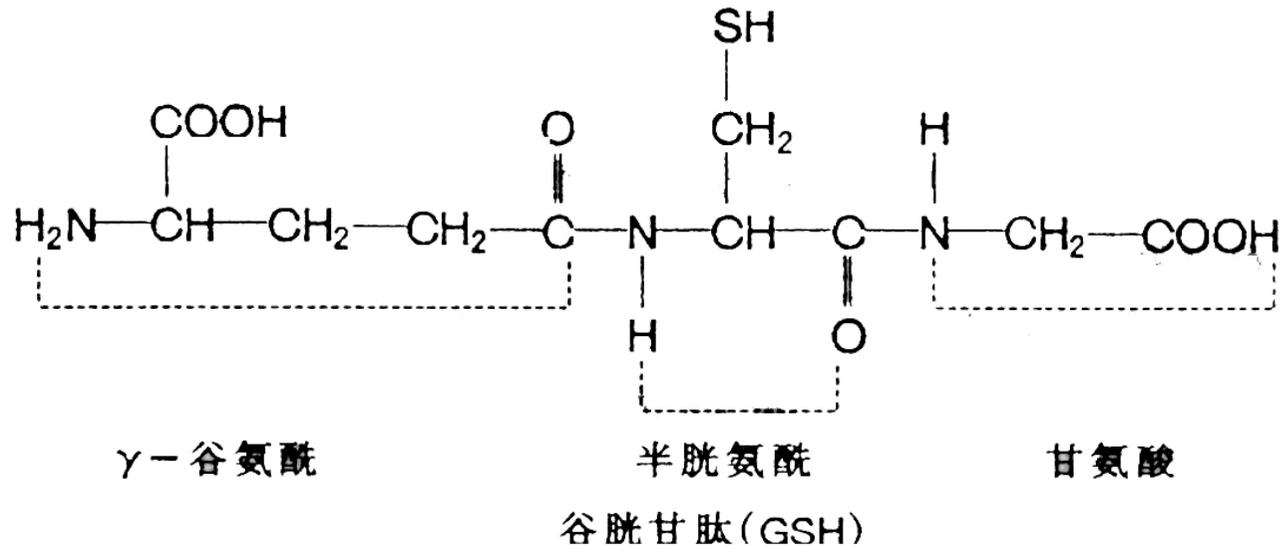


2.5. 天然存在的活性多肽

生物体内有一些活性多肽游离存在，它们有广泛的生理活性：激素、抗菌素、毒素、辅助因子等



① 谷胱甘肽



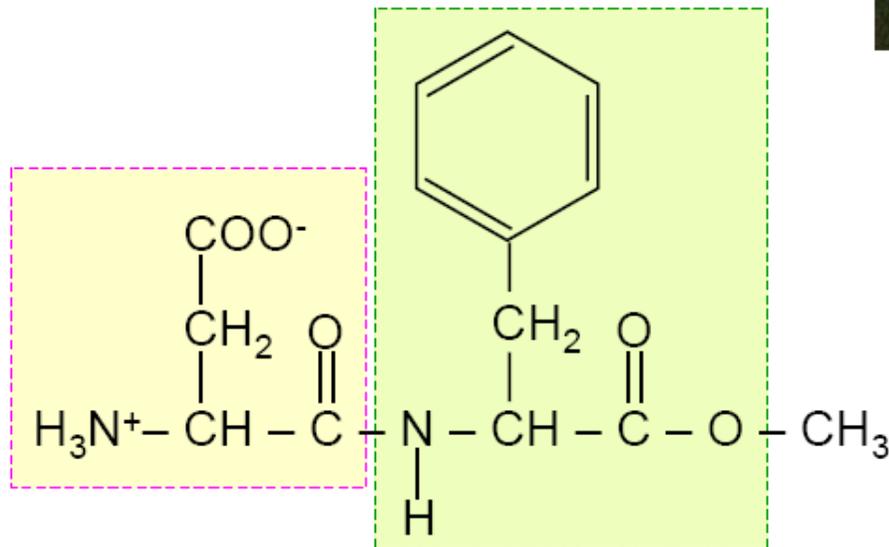
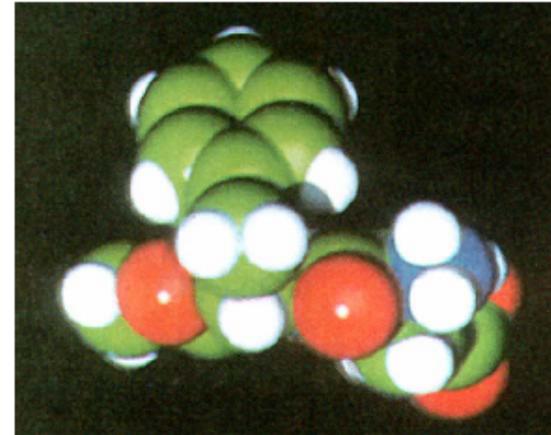
谷胱甘肽在体内参与氧化还原过程，作为某些氧化还原酶的辅因子，或保护巯基，或防止过氧化物积累。

2.5. 天然存在的活性多肽

② α -天冬氨酸-苯丙氨酸甲酯：甜味剂

阿斯巴甜

Aspartame

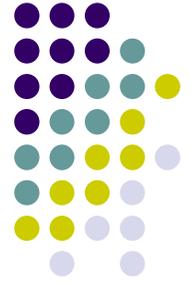


L-Aspartyl-L-phenylalanine methyl ester

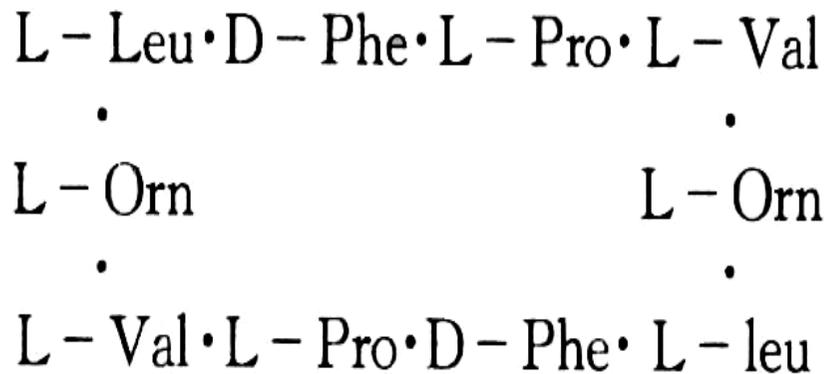


from ukhealthy.com/2007/healthyEating/20070712/15.html

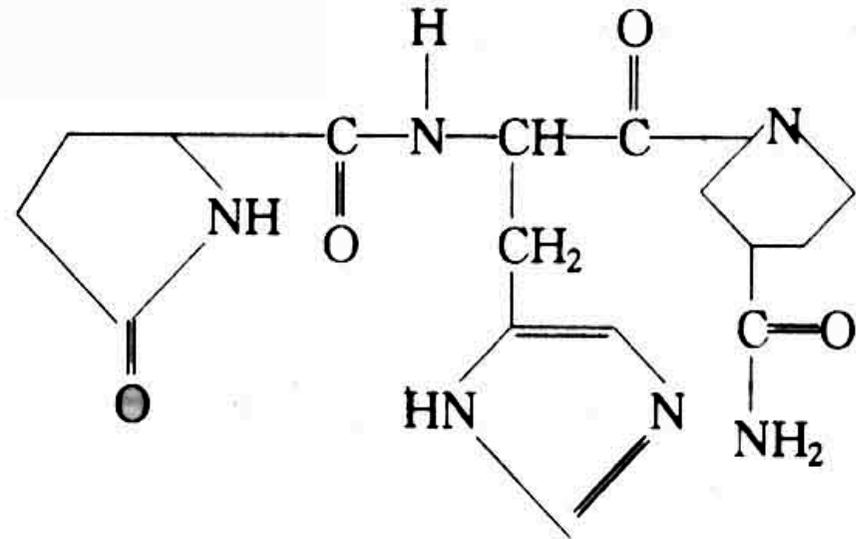
2.5. 天然存在的活性多肽



③短杆菌肽**S**：是
抗菌素，主要
作用于**G⁺**菌



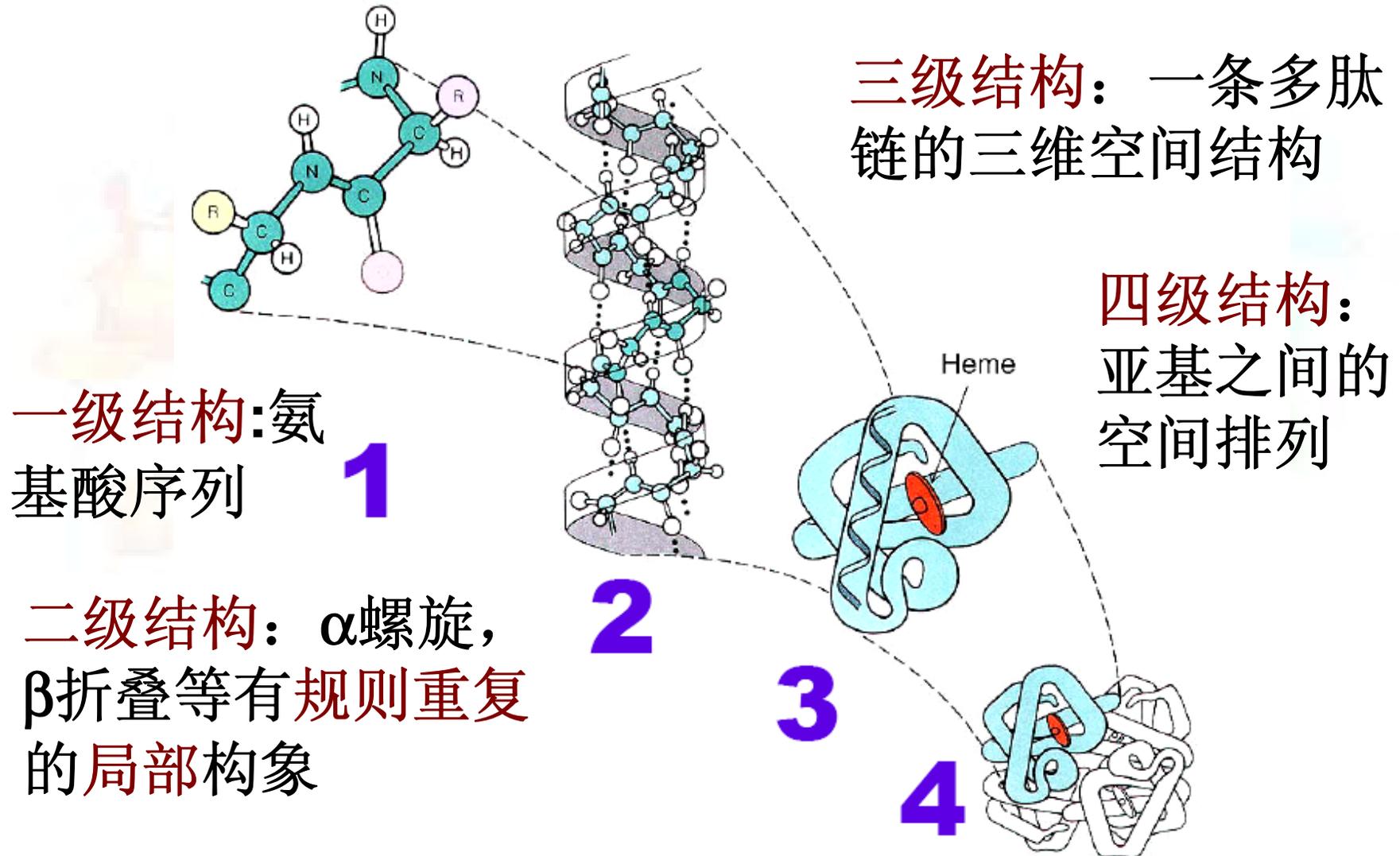
④促甲状腺素释放因子
(**TRH**)：激素



3. 蛋白质



3.1. 蛋白质分子的四级结构层次



1955年, Sanger测出了第一个蛋白质的氨基酸序列(胰岛素)

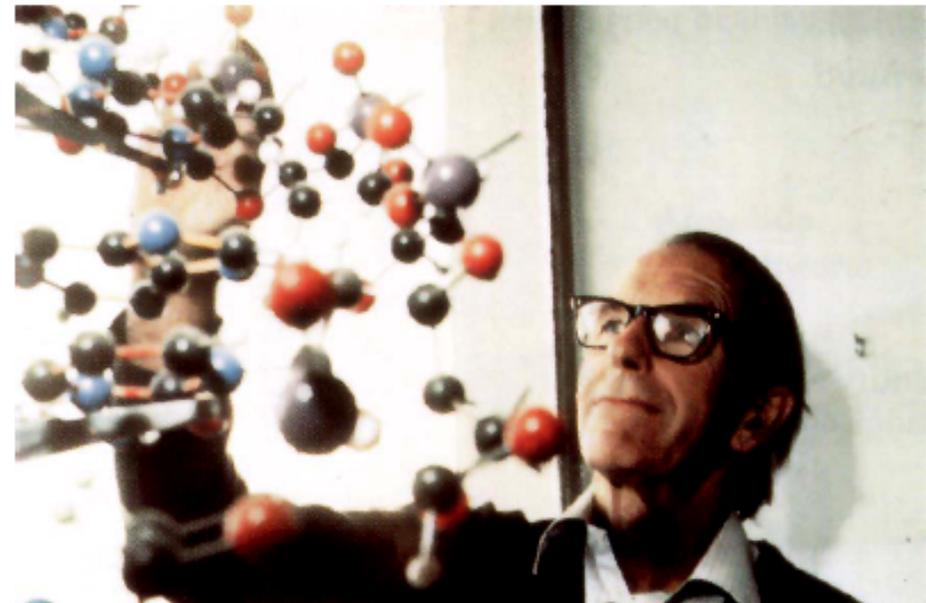
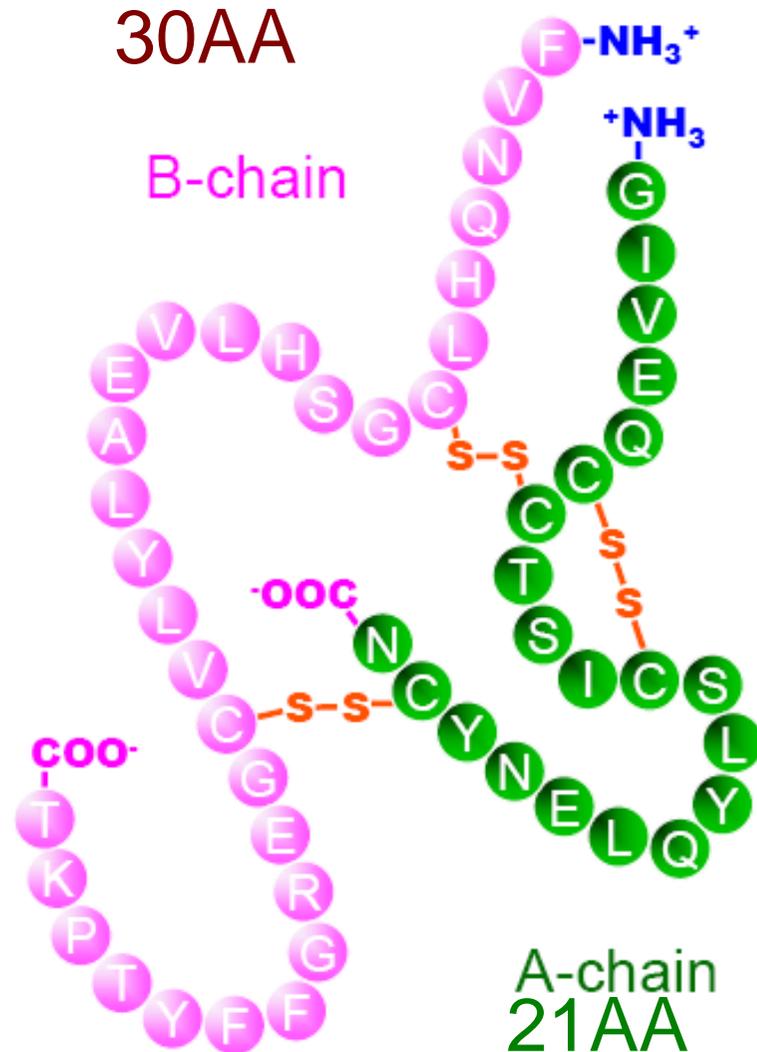


F. Sanger

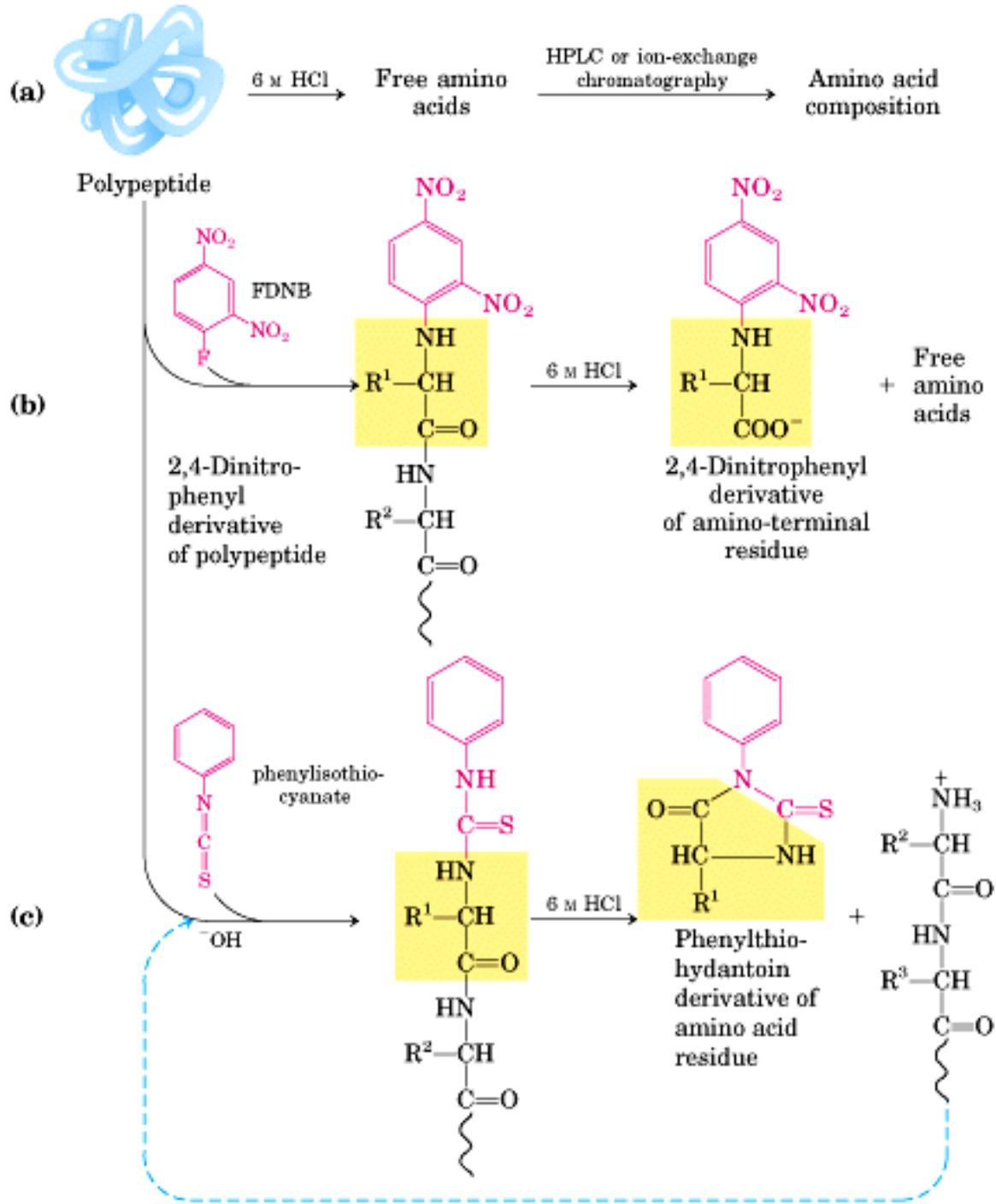
Cambridge U

Insulin 胰岛素 (A, B chains)

1945年—1955年



3.2.1. 一级结构测定 - Edman 降解法



Determine types and amounts of amino acids in polypeptide.

Sanger 法

Identify amino-terminal residue of polypeptide.

Edman 降解

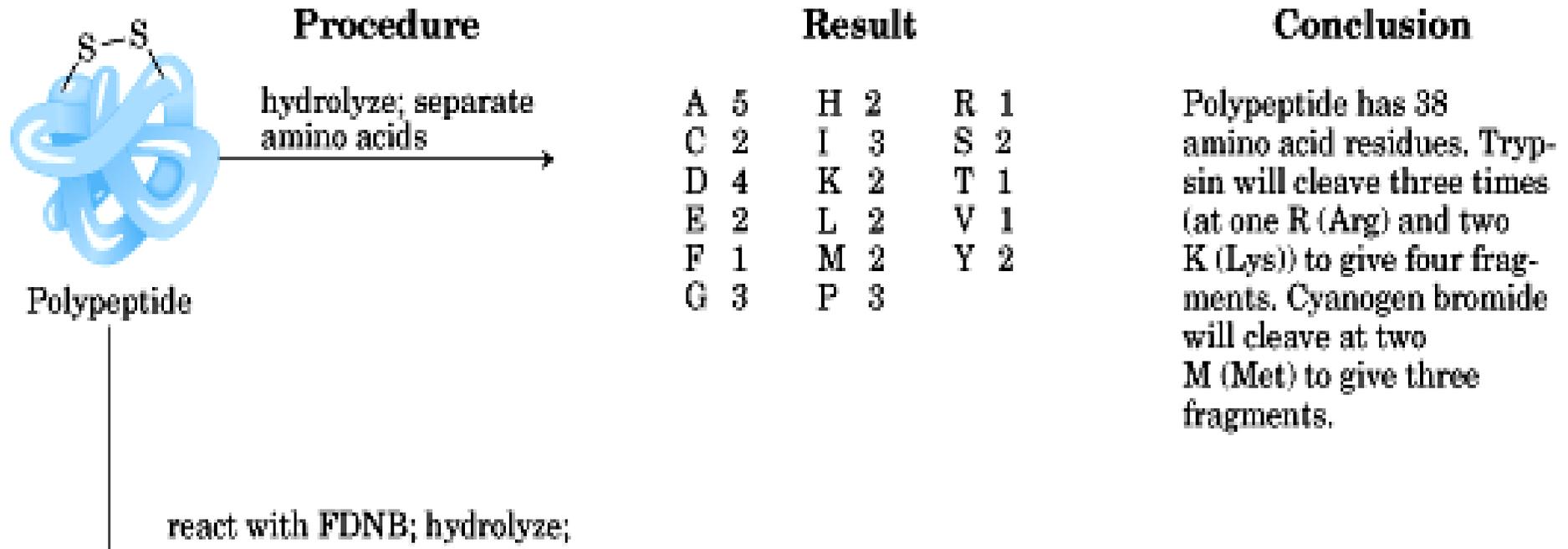
Identify amino-terminal residue; purify and recycle remaining peptide fragment through Edman process.

3. 2. 1. 一级结构测定

①蛋白质序列测定实例



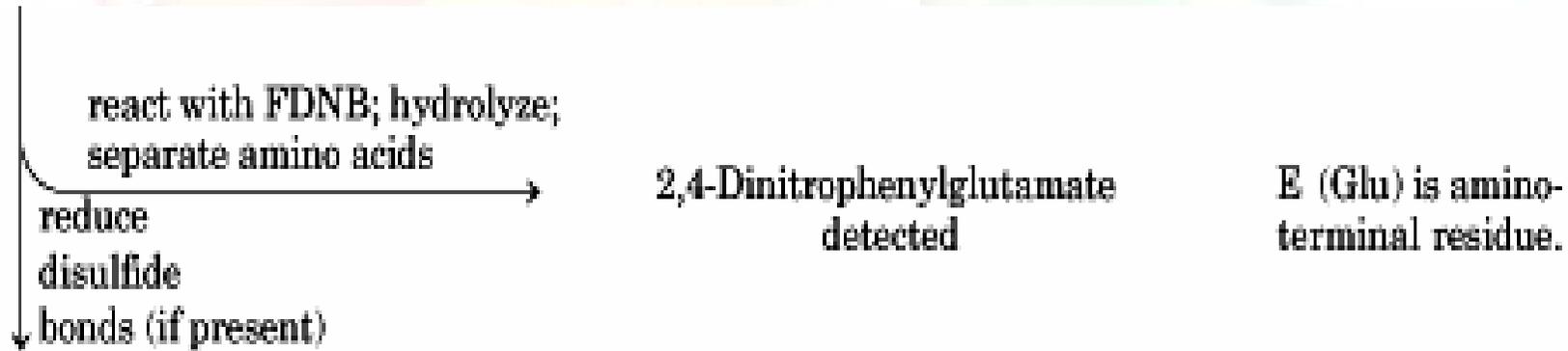
1. 断裂二硫键，彻底水解， 获得其氨基酸组成





①蛋白质序列测定实例

2. 测定N-末端氨基酸



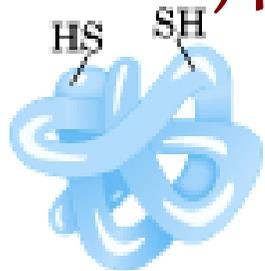
FDNB法

DNS-Cl法

①蛋白质序列测定实例



3.两种以上酶水解成小片段，分别用Edman降解法自动测序



cleave with **trypsin**
separate fragments; sequence
by Edman degradation →

- (T-1) GASMALIK
- (T-2) EGAAYHDFEPIDPR
- (T-3) DCVHSD
- (T-4) YLIACGPMTK

(T-2) placed at amino
terminus because it
begins with E (Glu).

(T-3) placed at carboxyl
terminus because it
does not end with
R (Arg) or K (Lys).

cleave with **cyanogen
bromide**; separate fragments;
sequence by Edman degradation →

- (C-1) EGAAYHDFEPIDPRGASM
- (C-2) TKDCVHSD
- (C-3) ALIKYLIACGPM

(C-3) overlaps with
(T-1) and (T-4), allowing
them to be ordered.

专一性的蛋白水解酶

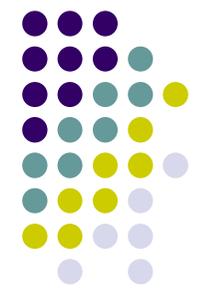


The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

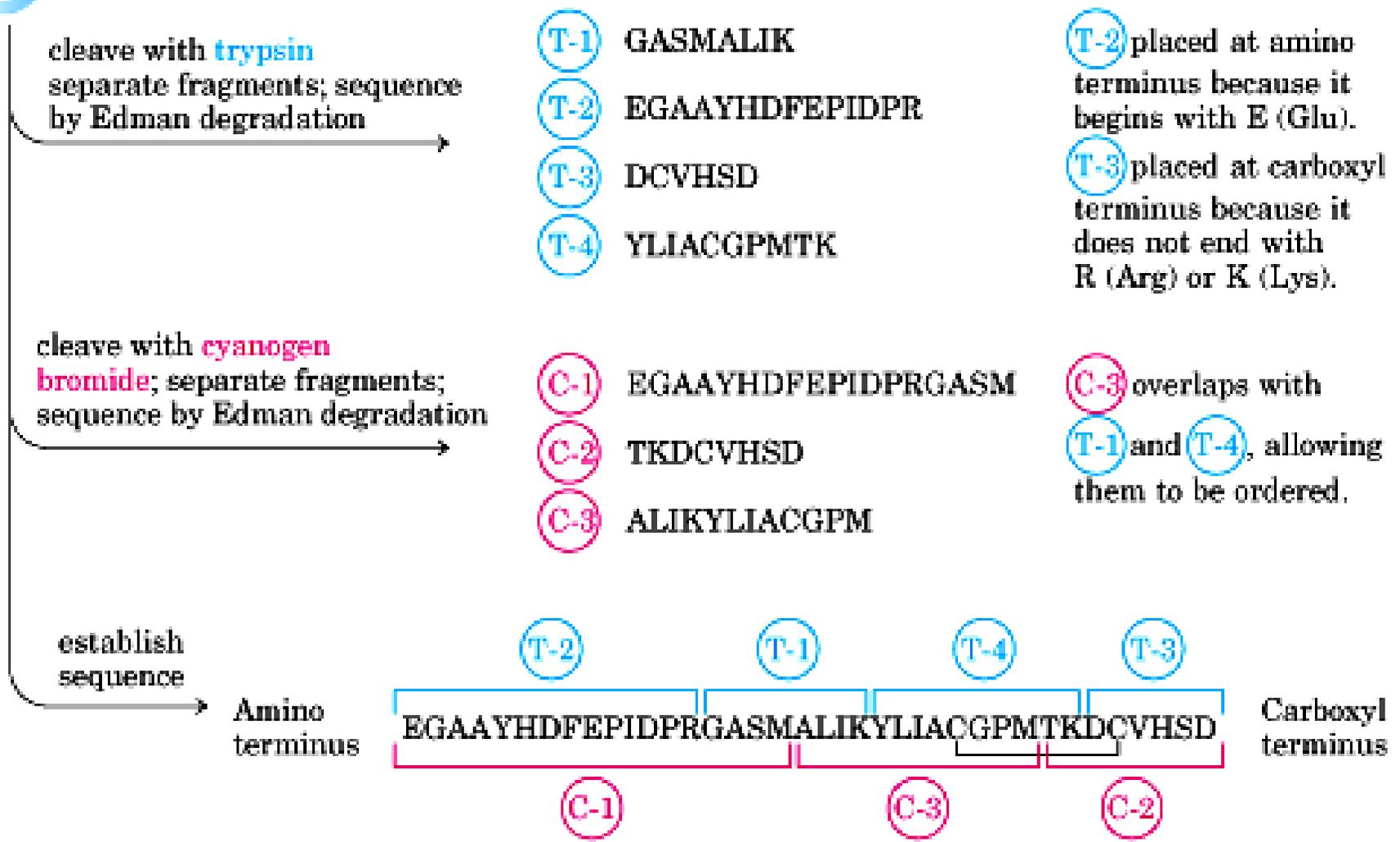
Treatment*	Cleavage points†
Trypsin	Lys, Arg (C)
<i>Submaxillarus</i> protease	Arg (C)
Chymotrypsin	Phe, Trp, Tyr (C)
<i>Staphylococcus aureus</i> V8 protease	Asp, Glu (C)
Asp-N-protease	Asp, Glu (N)
Pepsin	Phe, Trp, Tyr (N)
Endoproteinase Lys C	Lys (C)
Cyanogen bromide	Met (C)

*All except cyanogen bromide are proteases. All are available from commercial sources.

①蛋白质序列测定实例



4.将各片段拼合起来，得到完整序列



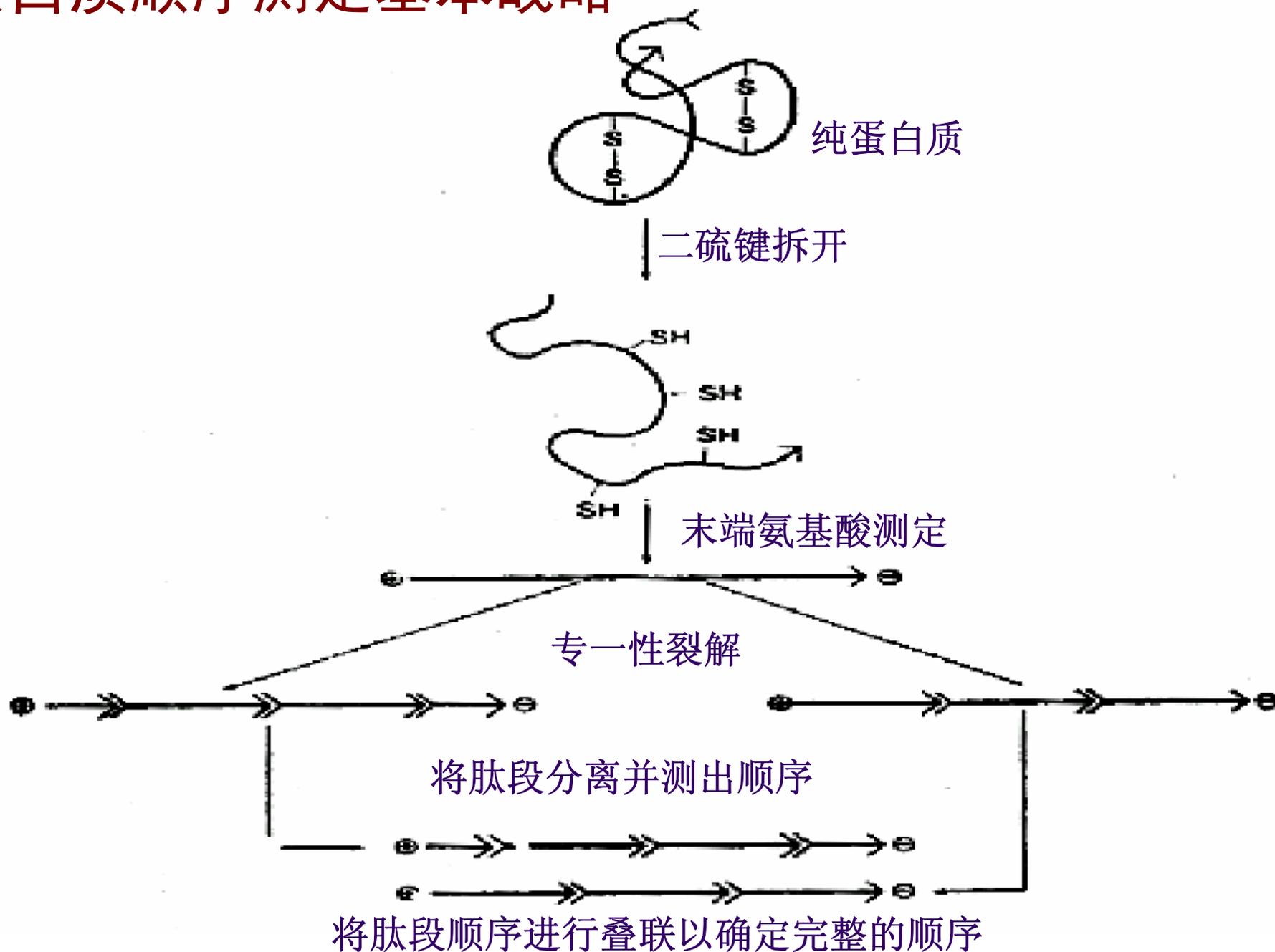
3. 2. 1. 一级结构测定



② 序列测定小结：片段重叠法+氨基酸顺序直测法
要点：

- a. 测定蛋白质的分子量及其氨基酸组分；
- b. 测定肽键的N-末端和C-末端，断开二硫键；
- c. 应用两种或两种以上的肽键内切酶或溴化氰法分别在多肽键的专一位点上断裂肽键，从而得到一系列大小不等的肽段；
- d. 分离提纯所产生的肽段，并分别测定它们的氨基酸顺序；
- e. 将这些肽段的顺序进行跨切口重叠，进行比较分析，推断出蛋白质分子的全部氨基酸序列。

蛋白质顺序测定基本战略



3.2.2. 一级结构决定三级结构

1950s, Anfinsen的核酸酶A变性复性实验

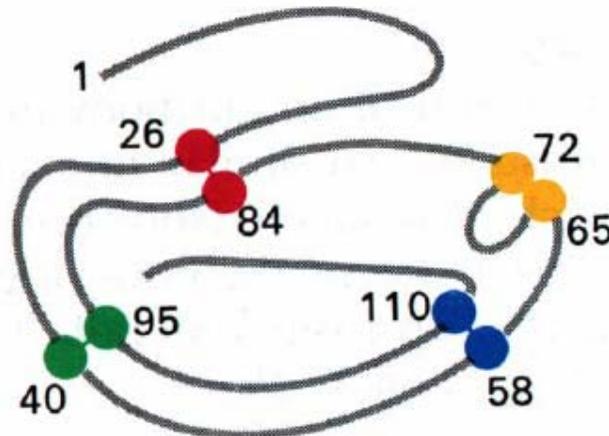


Anfinsen

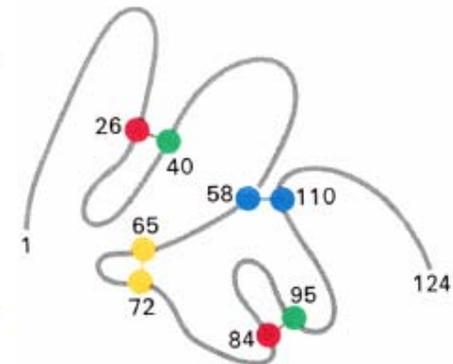
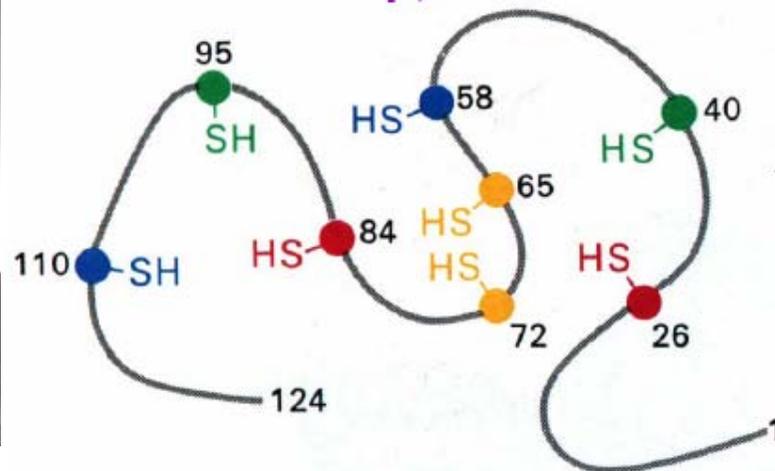


1972年Nobel

大部分可以
正确折叠



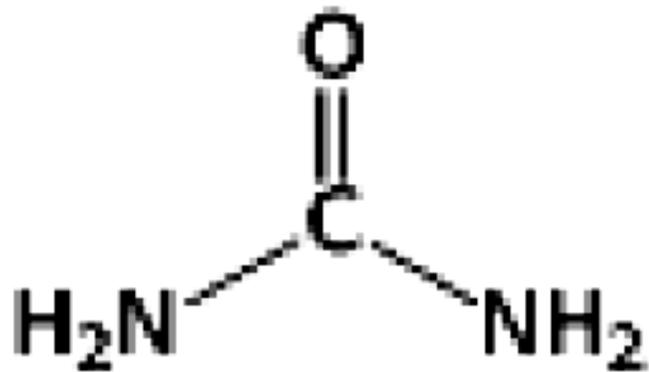
- Urea, Mercaptoethanol \rightleftharpoons + Urea, Mercaptoethanol



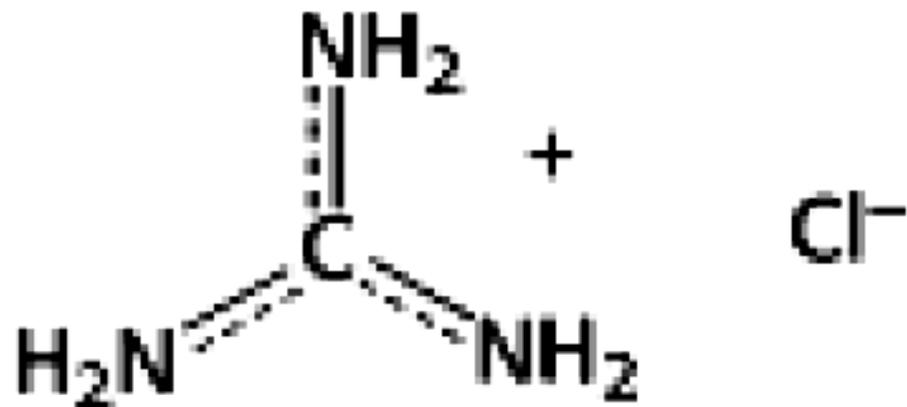
不正确摺疊

105个可能
的二硫键

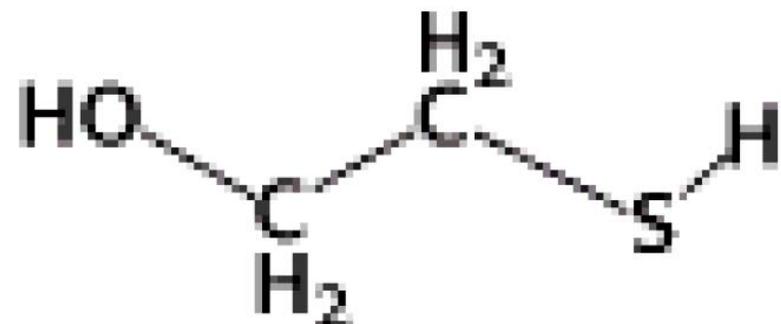




Urea



Guanidinium chloride



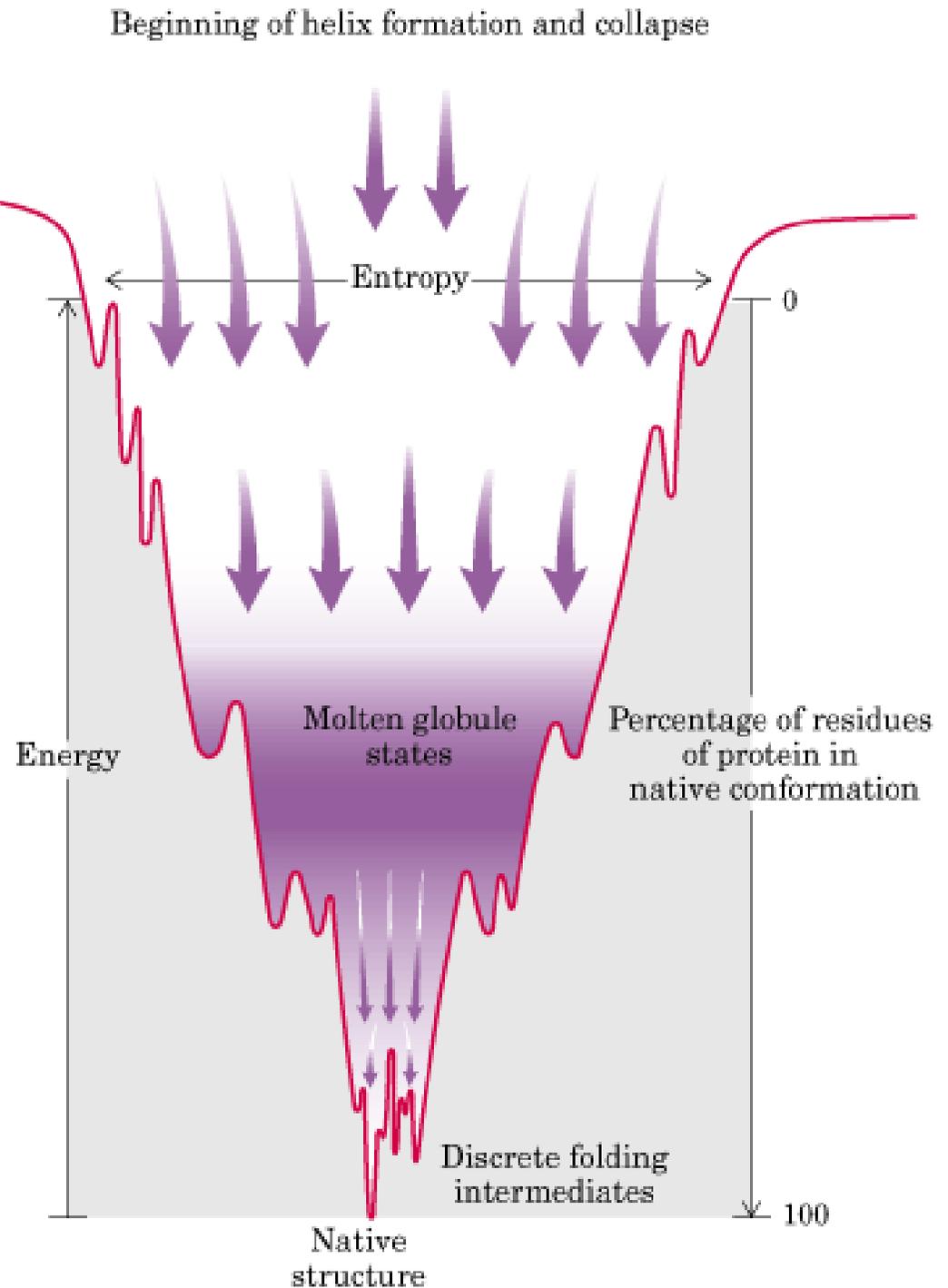
β -Mercaptoethanol



蛋白质的天然构象是能量最低的构象

热力学观点：
自由能漏斗

- 构象数目减少
- 熵值降低
- 天然构象比例增加





3.2.3. 蛋白质序列数据库

- 近年来大量生物学信息积累，形成了当前数以百计的**生物信息数据库**，并由此而诞生了**生物信息学**。
- 生物信息学是以计算机为工具对生物信息进行储存、检索和**分析**的科学。它是当今生命科学的重大**前沿领域**之一，同时也将是**21**世纪自然科学的**核心领域**之一。其研究重点主要体现在**基因组学(Genomics)**和**蛋白质组学(Proteomics)**两方面，具体说就是从核酸和蛋白质序列出发，分析序列中表达的结构与功能的生物信息。

3.2.3. 蛋白质序列数据库



- ①生物信息数据库可以分为一级数据库和二级数据库。
- ②**一级数据库**是直接来源于实验的原始数据，只经过简单的归类整理和注释；
- ③**二级数据库**是在一级数据库的基础上通过一定方式衍生而来，是对生物学知识和信息的进一步整理。
- ④国际上著名的一级核酸数据库有**Genbank数据库**、**EMBL核酸库**和**DDBJ库**等；蛋白质序列数据库有**SWISS-PROT**、**PIR**等；蛋白质结构库有**PDB**等。

3.2.3. 蛋白质序列数据库



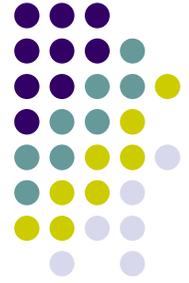
Genbank、EMBL和DDBJ每天交换数据

- **Genbank**库包含了所有已知的核酸序列和蛋白质序列，和相关的文献著作和生物学注释。它是由**美国**国立生物技术信息中心(**NCBI**)建立和维护的。
- **EMBL**是**欧洲**生物信息学研究所(**EBI**)维护的核酸序列数据库
- **日本DNA**数据仓库(**DDBJ**)也是一个全面的核酸序列数据库

截止到**2007年3月5日**所收录的**DNA**序列有**808亿bp**

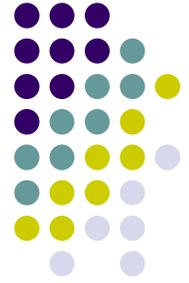
	# of Accessions	# of Basepairs/Residues
Genomic	840,657	80,808,752,887
RNA	929,109	1,632,375,659
Protein	3,438,099	1,215,085,694

3.2.3. 蛋白质序列数据库



- **PIR**是国际上最大的公共**蛋白质序列数据库**。这是一个全面的、经过注释的、**非冗余的蛋白质序列数据库**，包含超过**5,186,899**条蛋白质序列(至**07年3月**);
- **SWISS-PROT**是经过注释的蛋白质序列数据库，由欧洲生物信息学研究所(**EBI**)维护。它只接受**直接测序获得的蛋白质序列**。它与其它**30**多个数据建立了交叉引用;
- 蛋白质数据仓库(**PDB**)是国际上唯一的**生物大分子结构数据档案库**，由美国**Brookhaven**国家实验室建立。收集来源于**X光晶体衍射**和**核磁共振(NMR)**的数据。

3.2.3. 蛋白质序列与物种进化

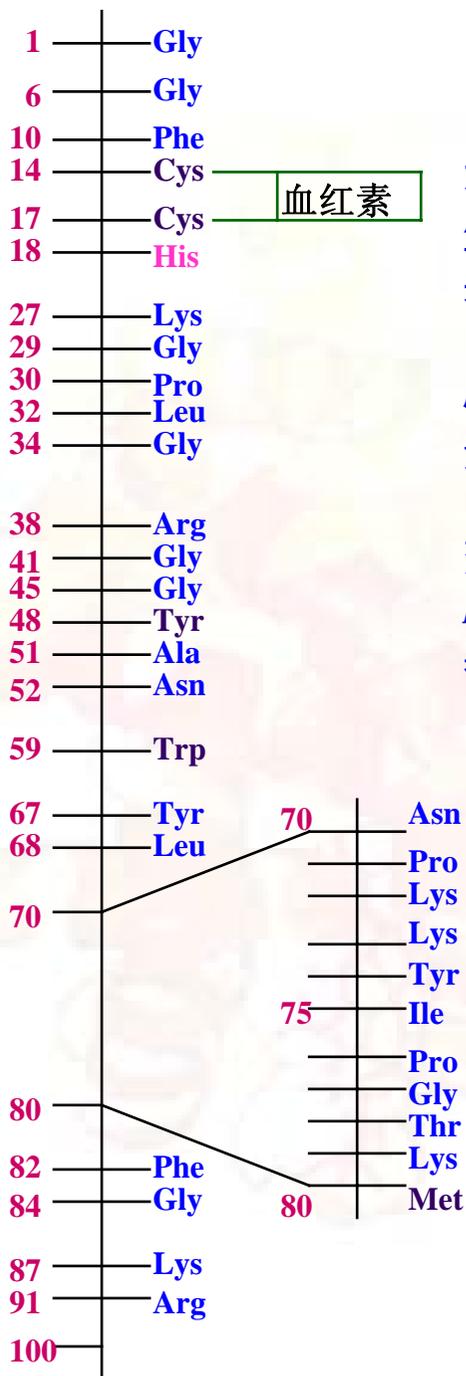


- ①**同源蛋白**：在不同物种中执行同一生物功能且氨基酸序列明显相似的蛋白质。如各种脊椎动物的血红蛋白。具有明显序列同源性的蛋白质也称同源蛋白。
- ②**保守残基(不变残基)**：同源蛋白序列中的某些位置在不同物种中都有相同的氨基酸，这样的氨基酸残基称为保守残基或不变残基。
- ③**可变残基**：同源蛋白中保守残基外的其它位置的氨基酸在不同物种中有较大变化，称可变残基。

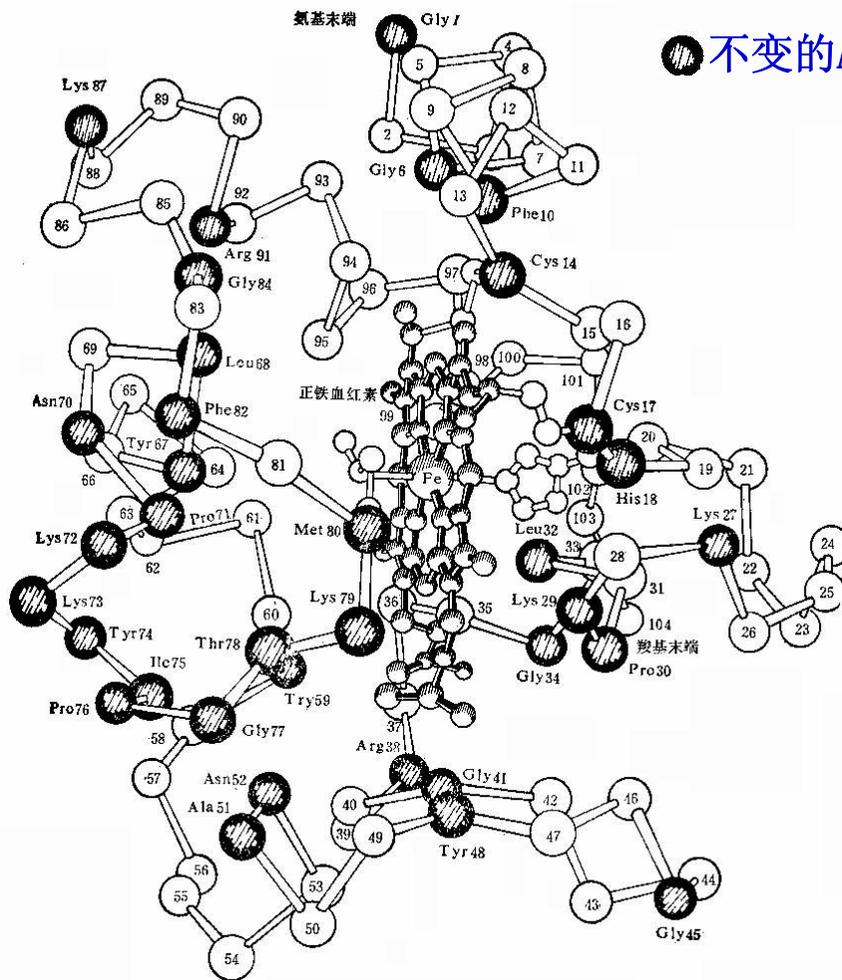
细胞色素c分子的空间结构



同源蛋白举例：细胞色素C



不同生物来源的细胞色素C中的不变的AA残基



35个不变的AA残基，是Cyt C的生物功能所不可缺少的。其中有的可能参加维持分子构象；有的可能参与电子传递；有的可能参与“识别”并结合细胞色素还原酶和氧化酶。

④ 同源蛋白的序列差异反应了物种进化关系

不同生物与人的Cyt c的AA差异数目



生物 与人不同的AA数目

黑猩猩	0
恒河猴	1
兔	9
袋鼠	10
牛、猪、羊、狗	11
马	12
鸡、火鸡	13
响尾蛇	14
海龟	15
金枪鱼	21
角蛟	23
小蝇	25
蛾	31
小麦	35
粗早链孢霉	43
酵母	44

系统树
(进化树):
反应物种
进化关系
的树状图
解。

