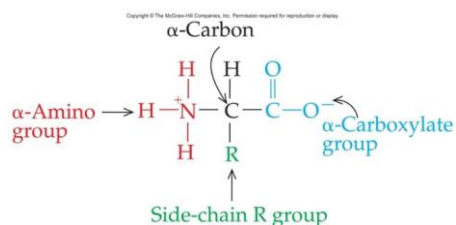


# BIO312 note

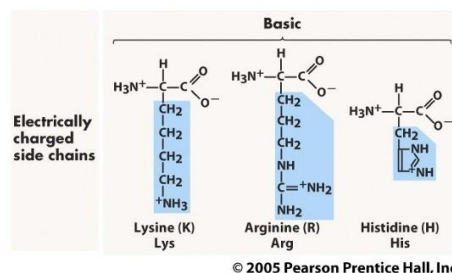
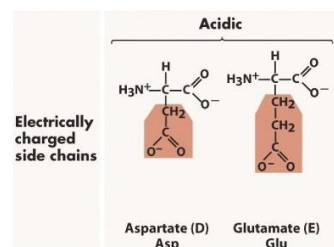
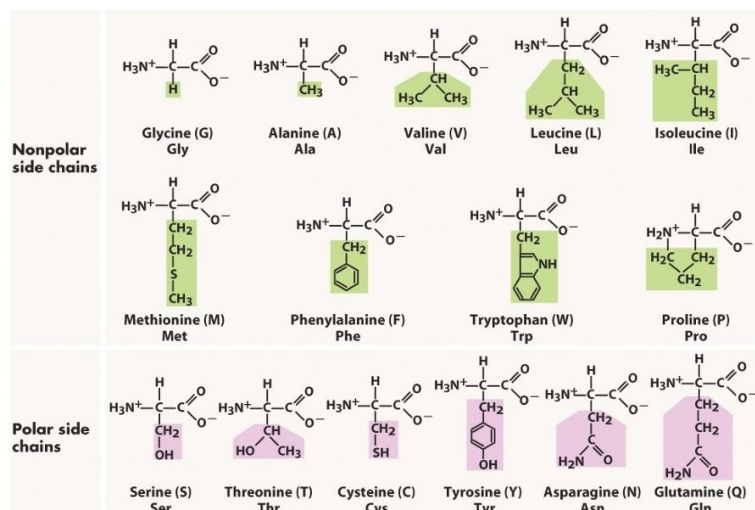
(Protein Structure and Function)

#	Method	Assessment Type	Learning outcomes assessed	Duration	Week	% of Final Mark	Resit (Y/N)
1	Assignment	CW	1-2		Available in LMO on April 9 <sup>th</sup> . Due at 5 pm on April 23 <sup>th</sup>	12.5%	N
2	Assignment	CW	3-4		Available in LMO on April 23 <sup>th</sup> . Due at 5 pm on May 7 <sup>th</sup>	12.5%	N
3	Final Exam	EXAM	ALL	3 Hours		75%	N

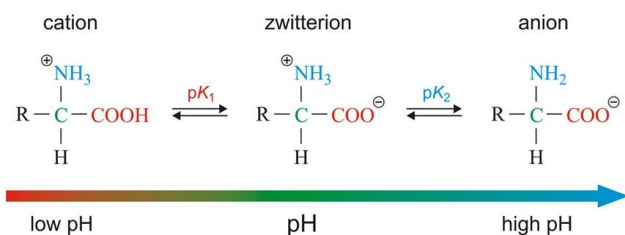
## Amino Acid Structure:



## 20 Common Amino Acids:

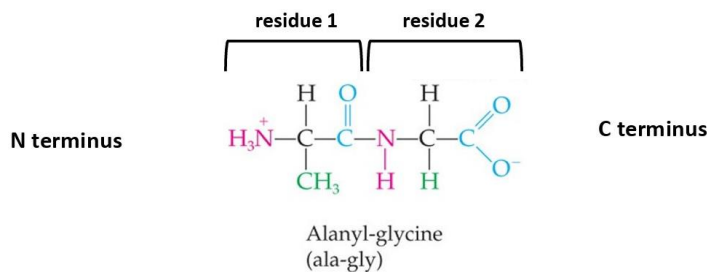


## Zwitterions (两性离子) :



- pH < pI, positively charged
- pH > pI, negatively charged

Condensing or dehydrating two amino acids produces a **dipeptide** (二肽) :



- Genetic information flows: DNA→RNA→Protein

### Primary Structure:

- Primary structure is the amino acid sequence of the polypeptide chain connected by the peptide bonds

CLUSTAL O(1.2.1) multiple sequence alignment (Partial)

```

RitR      MGKRILLLEKERNLAHFLSLELQKEQYRVDLVEEGQKALSMAIQTDDYDILLNVNLGDM
PhoB      MARRILVVEDEAPIREMVCVFLEQNGFQFVEAEDYDSAVNQLNEPWPDLILDWMLPGGS
          *:***:*. * : : : : * : : : : . * : : * : . : * : : * :
          . : * : : * : . : * : : : : : : : : * . * * * . : : * : : * :

RitR      AQDFAEKLSTRT---KPASVIMILDHWEDLQEELEVVRFAVSYYKPVLIENLVARISAI
PhoB      GIQFIKHLKRESMTDIPVVMLT-ARGEEDRVRGLETGADDYITKPFSPKELVARIKAV
          . : * : : * : . : * : : : : : : : : * . * * * . : : * : : * :

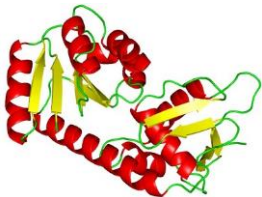
```

- Sequence alignment is a way of arranging primary sequences

### Secondary Structure:

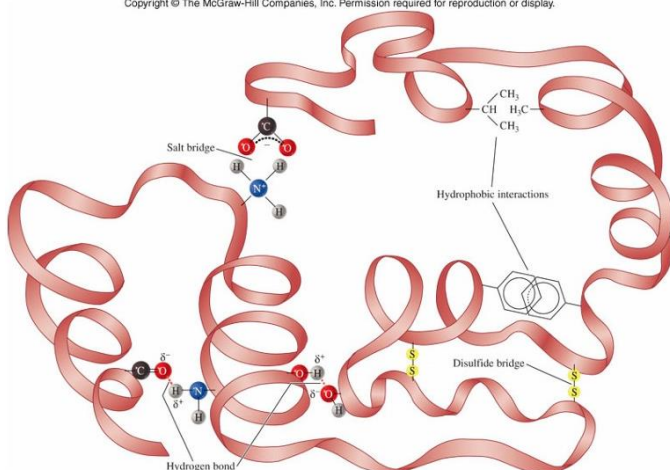
- Common secondary structures:

- $\alpha$ -helix
- $\beta$ -strain



### Tertiary Structure:

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



- Weaker interactions in proteins are dynamic

### Quaternary Structure:

- Quaternary structure is maintained by the same forces which are active in maintaining tertiary structure

### Stability of Proteins:

- The unfolding of proteins is called **denaturation** (变性) .
  - Temperature
  - pH
  - Organic Solvents/Detergents
  - Heavy Metals

- Mechanical Stress
- Peptide bonds between amino acids can be broken through **hydrolysis** (水解) .
  - This changes the primary structure!
  - Strong acid or strong base
  - Certain enzymes

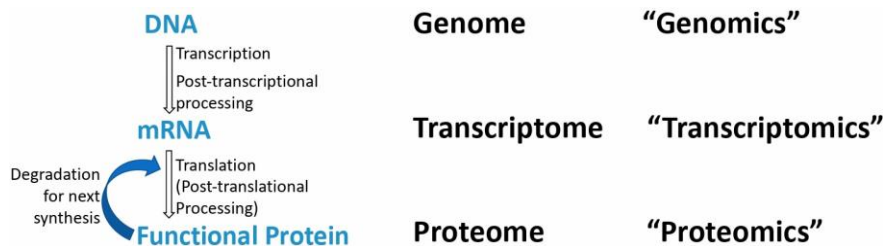
### Storage of Proteins:

- Similar conditions apply as with Stability
- Freezing with Liquid N<sub>2</sub> (and thaw) is typically OK
- Lyophilization



### Introduction of Proteomics

- Determine the functions of genes and their products, allowing them to be linked into pathways and networks, and ultimately providing a detailed understanding of how biological systems work.

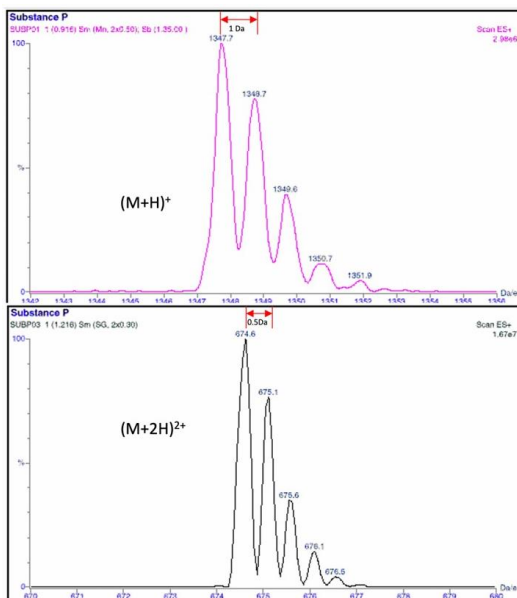


### Mass Resolution:

$$\text{Resolution} = \frac{m/z}{\Delta m/z}$$

full width at half maximum

$$m/z = \frac{m+z}{z}$$



Isotopes of singly charged ions are separated by 1 Da

Isotopes of doubly charged ions are separated by 0.5 Da

### MS plot:

1. **m/z 轴**: 横轴通常代表离子的质量/电荷比 (m/z)，即离子的质量除以其电荷数。这个轴表示了质谱中不同离子的质量相对比较。
2. **Intensity 轴**: 纵轴通常代表离子的相对丰度，也就是离子的相对数量。峰的高度代表了相应离子的丰度，高峰通常代表较丰富的离子。
3. **波峰**: 图中的每个峰代表一个离子，其位置由 m/z 值确定，其高度代表了相应离子的丰度。
4. **峰与峰之间的距离**: 两个峰之间的距离通常代表两个离子的 m/z 之间的差异。这些峰之间的距离可能对应着不同的碎片离子或者是同一分子的不同电荷状态。

5. **峰的形状**: 峰的形状可以提供关于离子的分布和离子产生的机制的信息。例如, 对称的高峰通常代表单一的离子种类, 而不对称或扁平的峰可能表示复杂的离子组合或质谱的噪音。

## Peptide:

**Peptide: A-B-C-D-E**

<u>b-ions</u>	<u>y-ions</u>	
$b_1^+$ A	BCDE	$y_4^+$
$b_2^+$ AB	CDE	$y_3^+$
$b_3^+$ ABC	DE	$y_2^+$
$b_4^+$ ABCD	E	$y_1^+$

- b-ions**: b-ions 是由肽链的 N-端部分解离形成的碎片离子。具体来说, 当肽链中的碳氮键 (peptide bond) 在 N-端的氨基基团处断裂时, 产生了一个带有 N-端残基的离子。b-ions 通常用字母 b 来表示, 后面跟着一个数字, 表示该离子中包含的氨基酸残基的数量。例如,  $b_1$  代表一个含有一个氨基酸残基的 b-ion,  $b_2$  代表一个含有两个氨基酸残基的 b-ion, 依此类推。
- y-ions**: y-ions 是由肽链的 C-端部分解离形成的碎片离子。当碳氮键在 C-端的羧基处断裂时, 产生了一个带有 C-端残基的离子。y-ions 通常用字母 y 来表示, 后面跟着一个数字, 表示该离子中包含的氨基酸残基的数量。例如,  $y_1$  代表一个含有一个氨基酸残基的 y-ion,  $y_2$  代表一个含有两个氨基酸残基的 y-ion, 依此类推。

## Mass Accuracy:

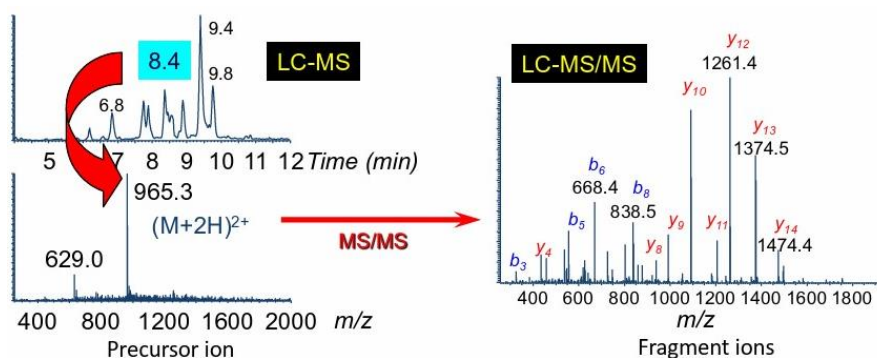
$$\frac{M_{\text{exp}} - M_{\text{true}}}{M_{\text{true}}} \times 10^6$$

- 测量质量与真实质量之间的相对百分比差(通常以 ppm 表示)。
- 数值越低质量精度越好。

## Searching with Peptide Mass Fingerprints (PMF):

搜索肽质量指纹 (Peptide Mass Fingerprints, PMF) 是一种生物信息学技术, 用于识别蛋白质样本中存在的蛋白质。在 PMF 中, 首先将蛋白质样本使用特定的酶 (例如胰蛋白酶) 进行消化, 产生一系列的肽段。接下来, 利用质谱仪将这些肽段进行质量分析, 得到每个肽段的质量。最后, 将实验获得的肽段质量与蛋白质数据库中已知的蛋白质质量进行比对, 以确定样本中存在的蛋白质。

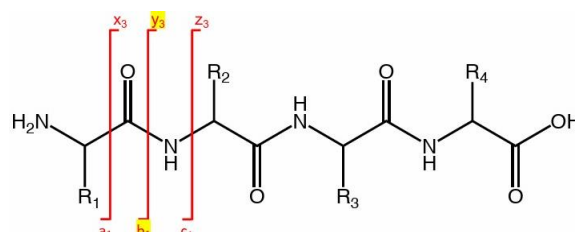
PMF 是一种常用的蛋白质鉴定技术, 特别适用于未知蛋白质样本的初步鉴定和分类。然而, 它也有一些局限性, 例如可能存在质谱图中的杂质或者未知的肽段可能导致鉴定的准确性下降。

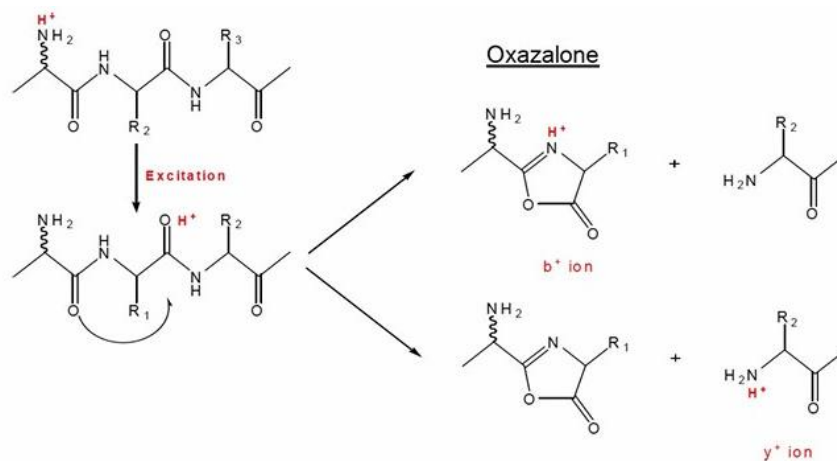


数据依赖的扫描能力可以显著提高蛋白质鉴定的容量和吞吐量。

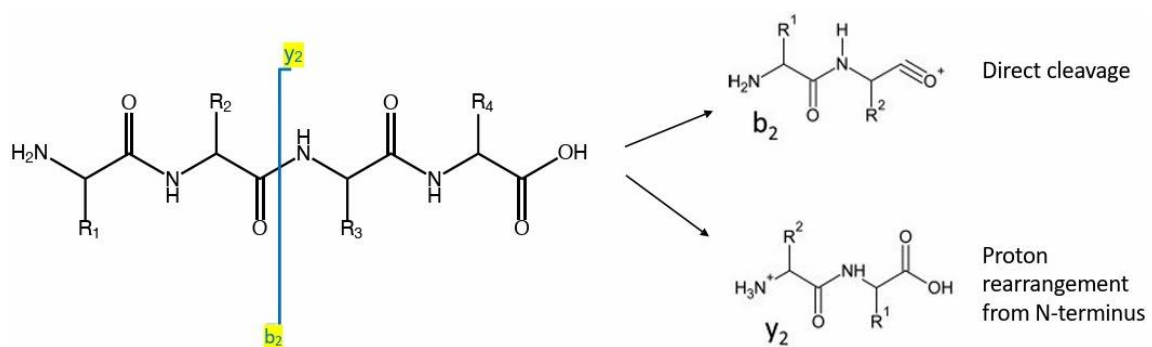
## Fragment Ion Analysis:

肽可以通过碰撞诱导解离 (collision-induced dissociation, CID) 破碎。首先进行肽的质子化, 然后再沿 CO-NH 键的切割是最常见的, 产生了 'b' 和 'y' 离子:





For a singly protonated peptide,  $\begin{matrix} \Rightarrow \\ \text{OR} \\ \Rightarrow \end{matrix}$  Singly charged N term ion ( $+H^+$ ) and neutral C-term  
Neutral N term and Singly charged C-term ion ( $+H^+$ )

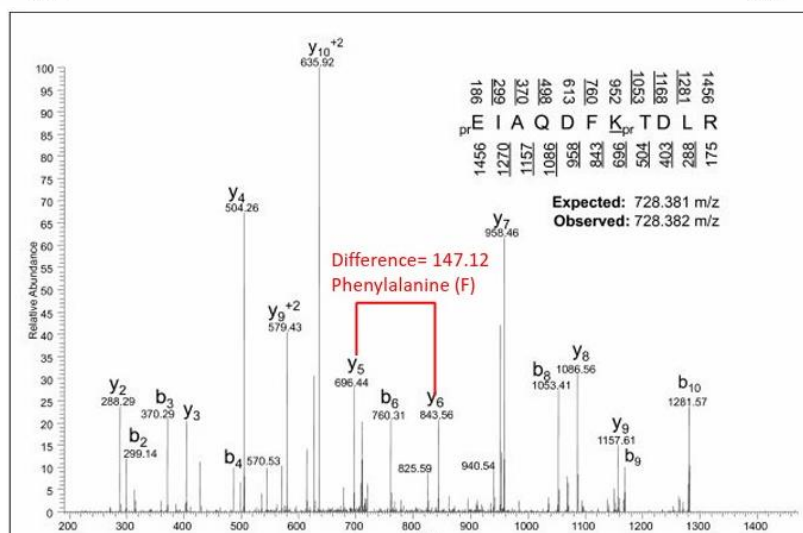
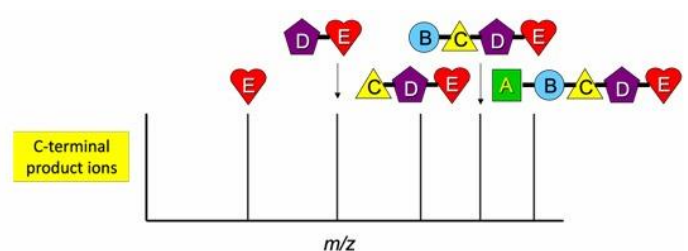
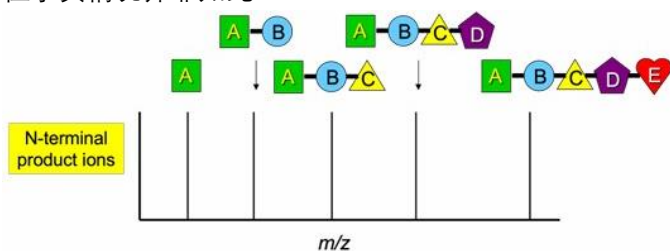


## Peptide Sequencing:

理想情况下，人们可以测量产物离子峰之间的间隔来推断序列

- 如果每个酰胺键解离的概率相等
- 如果每个分子只有一个酰胺键断裂
- 如果只形成 c 端或 n 端产物离子

但事实情况并非如此.....





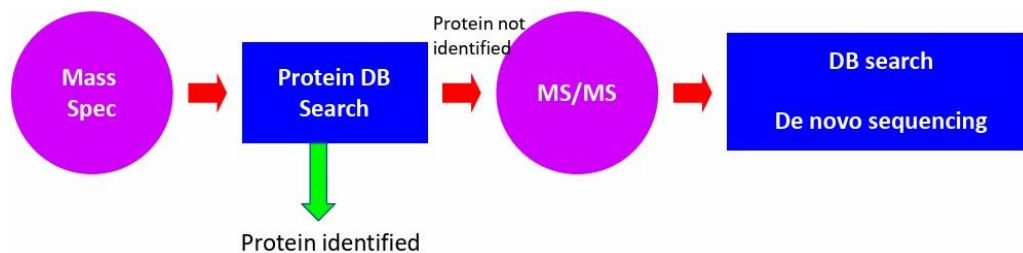
Mass of amino acid fragment ion

Name	3-letter code	1-letter code	Residue Mass	Immonium ion	Related ions	Composition
Alanine	Ala	A	71.03711	44		C <sub>3</sub> H <sub>7</sub> NO
Arginine	Arg	R	156.10111	129	59,70,73,87,100,112	C <sub>6</sub> H <sub>12</sub> N <sub>4</sub> O
Asparagine	Asn	N	114.04293	87	70	C <sub>4</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>
Aspartic Acid	Asp	D	115.02694	88	70	C <sub>4</sub> H <sub>7</sub> NO <sub>3</sub>
Cysteine	Cys	C	103.00919	76		C <sub>3</sub> H <sub>5</sub> NOS
Glutamic Acid	Glu	E	129.04259	102		C <sub>5</sub> H <sub>7</sub> NO <sub>3</sub>
Glutamine	Gln	Q	128.05858	101	56,84,129	C <sub>5</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>
Glycine	Gly	G	57.02146	30		C <sub>2</sub> H <sub>3</sub> NO
Histidine	His	H	137.05891	110	82,121,123,138,166	C <sub>6</sub> H <sub>7</sub> N <sub>3</sub> O
Isoleucine	Ile	I	113.08406	86	44,72	C <sub>6</sub> H <sub>11</sub> NO
Leucine	Leu	L	113.08406	86	44,72	C <sub>6</sub> H <sub>11</sub> NO
Lysine	Lys	K	128.09496	101	70,84,112,129	C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> O
Methionine	Met	M	131.04049	104	61	C <sub>5</sub> H <sub>9</sub> NOS
Phenylalanine	Phe	F	147.06841	120	91	C <sub>9</sub> H <sub>9</sub> NO
Proline	Pro	P	97.05276	70		C <sub>5</sub> H <sub>7</sub> NO
Serine	Ser	S	87.03203	60		C <sub>3</sub> H <sub>5</sub> NO <sub>2</sub>
Threonine	Thr	T	101.04768	74		C <sub>4</sub> H <sub>7</sub> NO <sub>2</sub>
Tryptophan	Trp	W	186.07931	159	11,117,130,132,170,100	C <sub>11</sub> H <sub>9</sub> N <sub>2</sub> O
Tyrosine	Tyr	Y	163.06333	136	91,107	C <sub>9</sub> H <sub>9</sub> NO <sub>2</sub>
Valine	Val	V	99.06841	72	44,55,69	C <sub>5</sub> H <sub>9</sub> NO

Mass of b2 ions in peptide fragmentation

	G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
G	115																	
A	129	143																
S	145	159	175															
P	155	169	185	195														
V	157	171	187	197	199													
T	159	173	189	199	201	203												
C	161	175	191	201	203	205	207											
I/L	171	185	201	211	213	215	217	227										
N	172	186	202	212	214	216	218	228	229									
D	173	187	203	213	215	217	219	229	230	231								
K/Q	186	200	216	226	228	230	232	242	243	244	257							
E	187	201	217	227	229	231	233	243	244	245	258	259						
M	189	203	219	229	231	233	235	245	246	247	260	261	263					
H	195	209	225	235	237	239	241	251	252	253	266	267	269	275				
F <sup>b</sup>	205	219	235	245	247	249	251	261	262	263	276	277	279	285	295			
R	214	228	244	254	256	258	260	270	271	272	285	286	288	294	304	313		
Y	221	235	251	261	263	265	267	277	278	279	292	293	295	301	311	320	327	
W	244	258	274	284	286	288	290	300	301	302	315	316	318	324	334	343	350	373

GG=N=114; GA=K/Q=128; GV=R=156; GE=AD=SV=W=186.



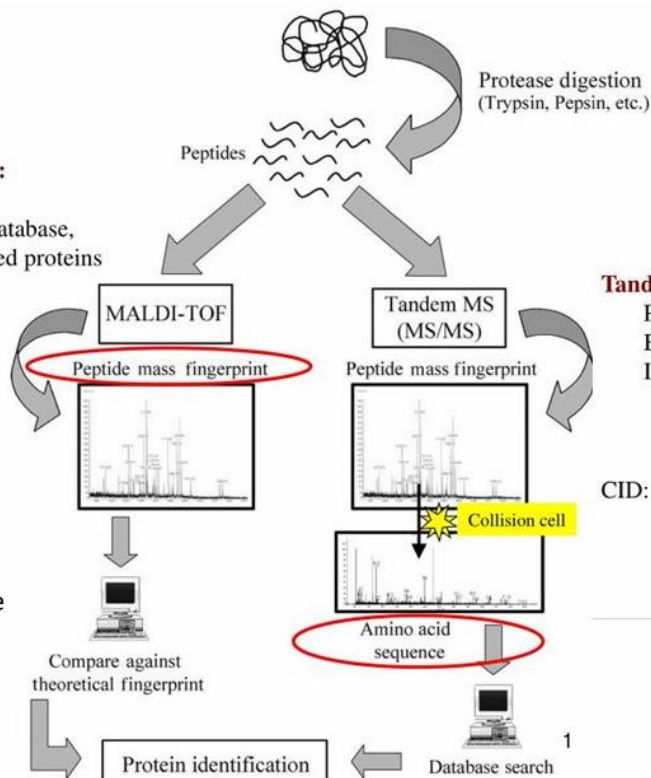
De novo 测序 (de novo sequencing) 是指通过质谱分析或 DNA/RNA 测序等技术, 在没有参考序列的情况下对未知序列进行直接鉴定和确定其组成的过程。

在蛋白质研究领域, de novo 质谱测序是指通过质谱技术直接解析未知蛋白质序列的过程。在这种情况下, 没有可用的数据库或已知的蛋白质序列信息作为参考。通过分析质谱图谱中肽段的碎片离子, 以及这些碎片的质量和相对丰度, 可以推断出原始蛋白质的氨基酸序列。这种方法通常需要使用先进的质谱仪器和计算算法, 以尽可能准确地重建未知蛋白质的序列。

### Proteomics Analysis:

#### Single MS (peptide fingerprinting):

Identifies m/z of peptide only  
Peptide id'd by comparison to database,  
of predicted m/z of trypsinized proteins



#### Tandem MS/MS (peptide sequencing):

Pulls each peptide from the first MS  
Breaks up peptide bond  
Identifies each fragment based on m/z

CID: collision induced dissociation

**MALDI-TOF** (基质辅助激光解吸/飞行时间质谱):

- MALDI-TOF 是一种质谱技术，用于分析生物大分子（如蛋白质和多肽）以及其他大分子化合物。
- 在 MALDI-TOF 中，样品通常与一种能够吸收激光光源的基质混合，然后通过激光照射样品，基质吸收能量后将其转移给样品中的分子，导致分子从样品表面脱离，并形成带电离子。
- 这些带电离子通过飞行时间质谱仪器进行分析，其原理是根据带电粒子在电场中飞行的时间来确定其质量/电荷比 ( $m/z$ ) 值。
- MALDI-TOF 主要用于蛋白质和多肽的鉴定、定量和分析。

#### **Tandem MS** (MS/MS, 串联质谱):

- Tandem MS 是一种质谱技术，它结合了两个或多个质谱过程，通常用于分析复杂样品中的分子结构。
- 在 Tandem MS 中，首先通过质谱仪器进行质谱分析，得到样品中分子的质谱图谱。然后，选定某些特定的离子进行进一步分析。
- 选定的离子通过碰撞电解质 (CID) 或其他碰撞技术与气体分子发生碰撞，导致它们解离成更小的碎片离子。
- 接下来，这些碎片离子被再次进入另一个质谱仪器进行质谱分析，以获得碎片离子的质谱图谱。
- 通过分析初始分子的质谱图谱和其碎片的质谱图谱，可以推断出分子的结构信息，包括它们的组成和连接方式。
- Tandem MS 常用于蛋白质组学、代谢组学和化合物结构鉴定等领域。