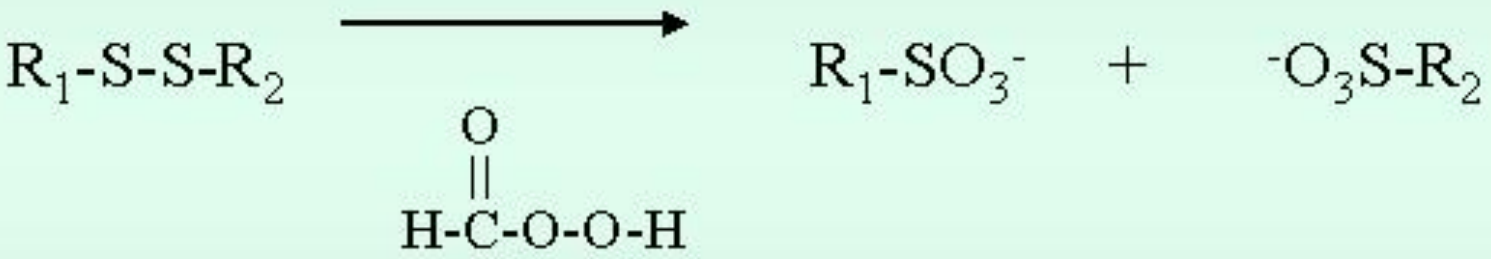
Oxidation of disulphide bridges

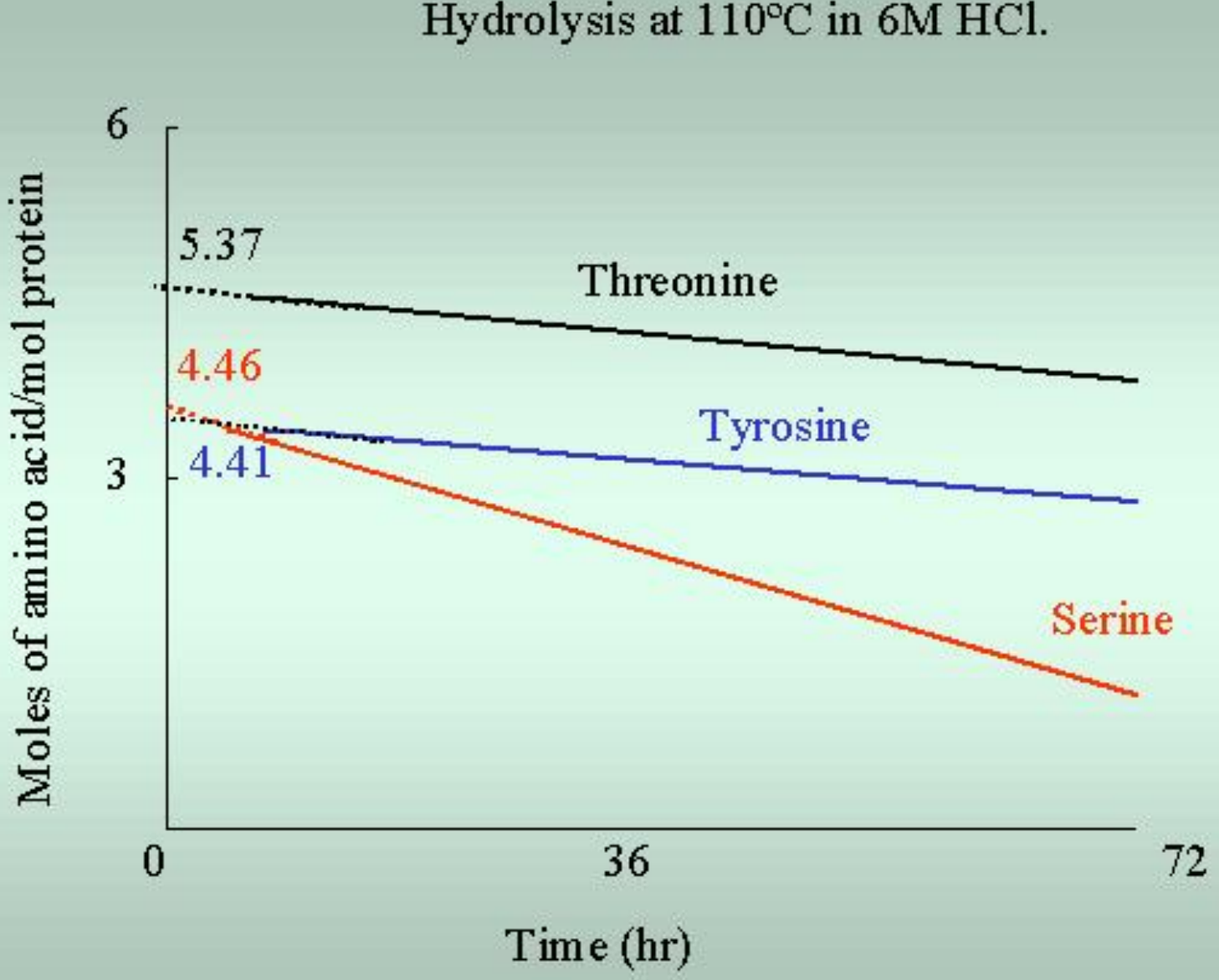
- Or …… oxidise with performic acid.



- The negatively charged -SO_3^- groups electrostatically repel each other and are incapable of forming bridge structures.

Step 3 - amino acid composition.

- To work out amino acid composition, the protein needs to be completely hydrolyzed. This is done using 6M HCl, 110°C.
  - Note that Try will be lost
  - AsN and Gln hydrolyse to Asp and Glu
  - Ser, Thr and Tyr are partly degraded
- The resulting amino acids are separated by ion exchange or reverse-phase HPLC - identities worked out by size and charge.
- The separated amino acids are usually analyzed with an **amino acid analyzer**, which can also conduct the whole separation process automatically.


Hydrolysis at 110°C in 6M HCl.
6
3
0
36
72
5.37
4.46
4.41
Threonine
Tyrosine
Serine
Moles of amino acid/mol protein
Time (hr)

Step 4 - identification of N and C terminal residues.

- Also called **end group analysis**.
- End group analysis reveals the identity of N terminal and C terminal residues.
- Also reveals how many ends there are, and so how many polypeptide chains there are.

Steps 5 and 6 -fragmentation of polypeptide chain.

- Can be done with enzymes, or by chemical means.
- Can be specific or non-specific.
- Specificity: some enzymes only cleave at particular peptide bonds -> immediately gives information about the product fragments.
- The idea is to **produce peptide fragments** that are neither too big nor too small for Edman degradation and end group analysis.

Step 7 - reconstruction of overall amino acid sequence.

- Overlapping peptides generated from different hydrolysis procedures can be aligned to reveal the overall amino acid sequence.

"Diagonal" analysis

- Electrophoresis of peptide fragments in 2 dimensions with identical conditions for both dimensions leads to a "**diagonal" pattern** of peptide fragments.
- If a peptide is modified in some way before the 2nd dimension is run, the modified peptide will have a different mobility on the 2nd dimension and fall outside the diagonal pattern.

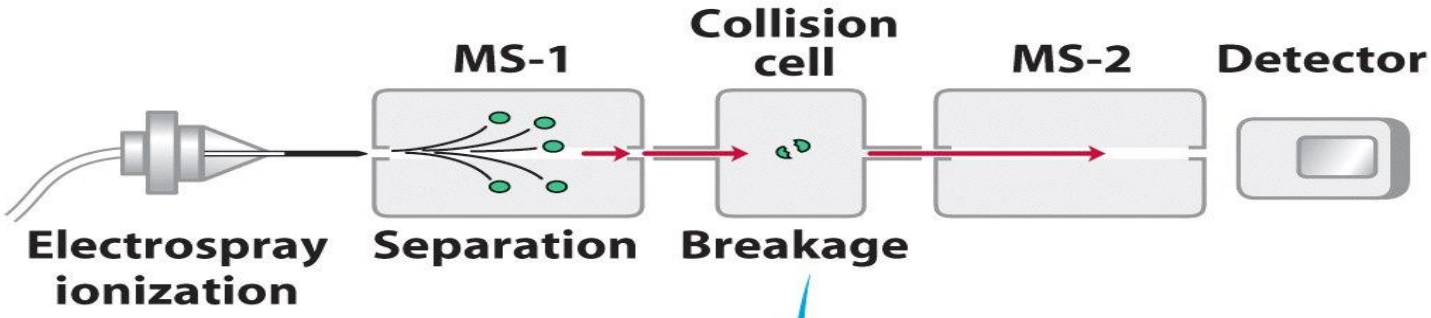
Step 8 - location of disulphide bridges.

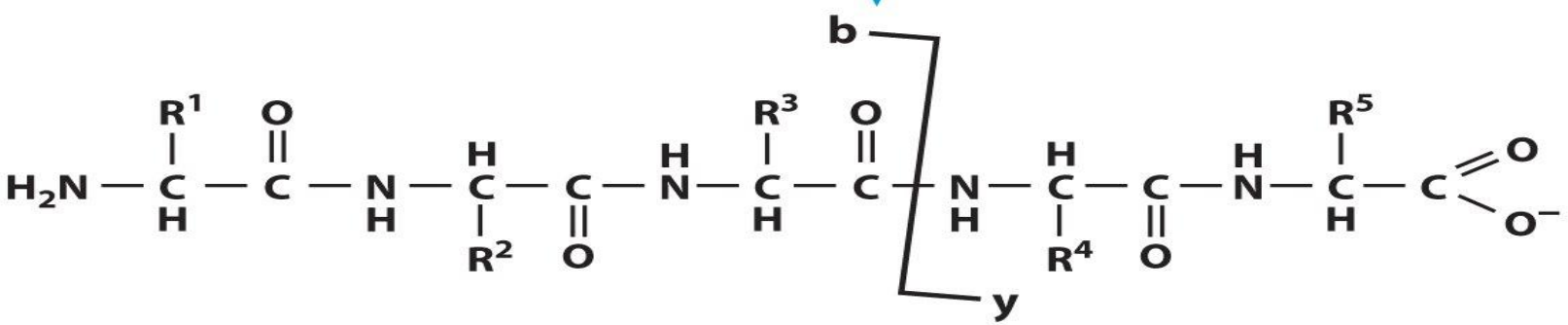
- To identify locations of -S-S- bridges, the protein is cleaved (e.g. with trypsin) into peptides without prior disruption of disulphide bridges.
- 2-D electrophoresis is then carried out, but before the 2nd dimension is run, -S-S- bonds are disrupted (usually by oxidation to $-SO_3^-$)
- Peptides originally joined together with -S-S- bridges will now separate and move away from the diagonal pattern.

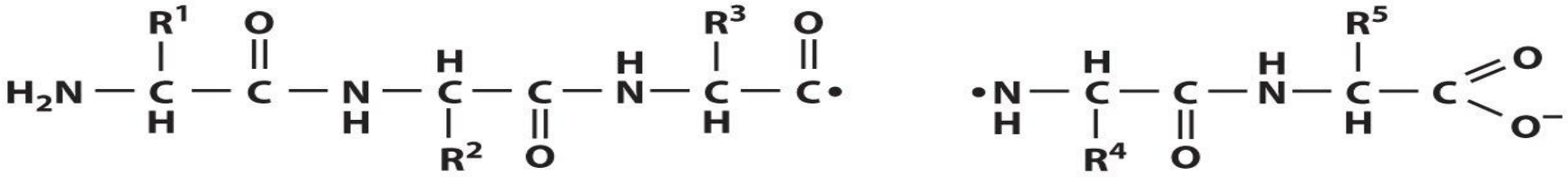
Protein sequencing by mass spectrometry

- One problem with this method is that you can only sequence peptides of 16 amino acids or less.
- However, sizes of tryptic fragments of a protein can sometimes be used to identify that protein without actually having to sequence the peptide fragments.
- This is because a database of the sizes of tryptic fragments of a variety of proteins now exists - **protein fingerprints.**

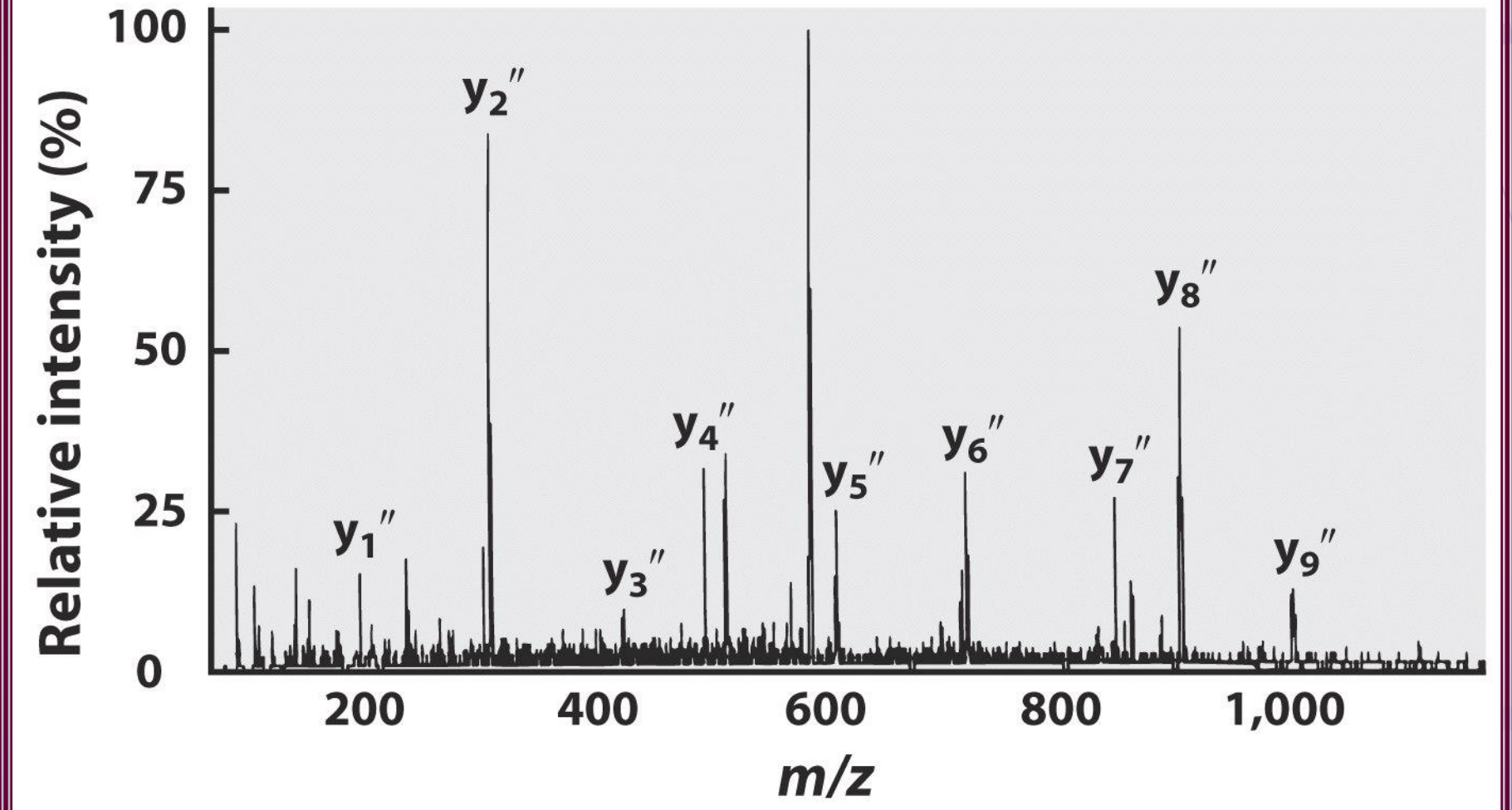
Tandem Mass Spectrometry: A New Method for Protein Sequencing







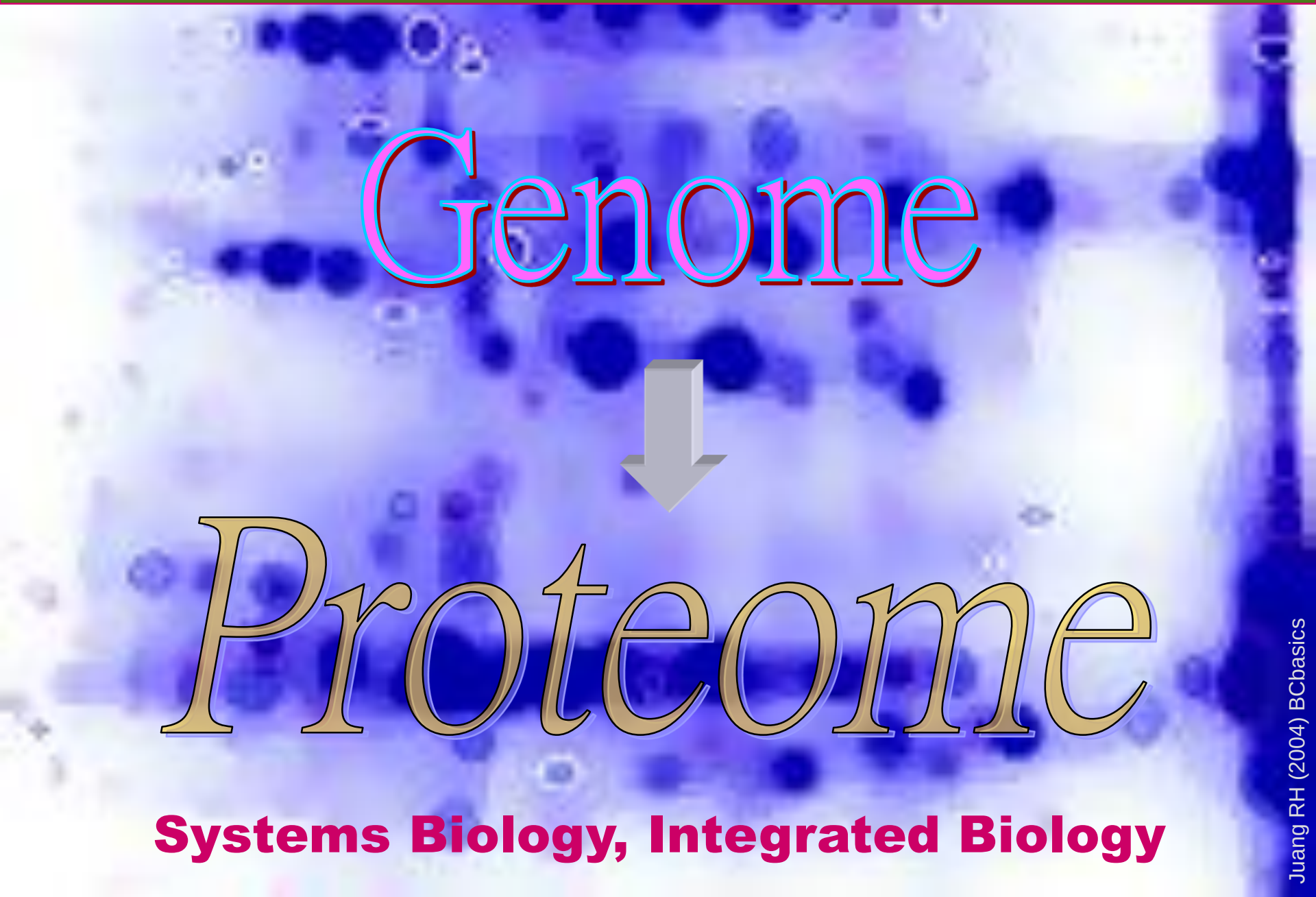
Tandem Mass Spectrometry: A New Method for Protein Sequencing



Protein sequencing by mass spectrometry

- **Machine 2.**
  - Analyzes the fragments of the peptide sent by machine 1.
  - Peptides are fragmented so that the **fragments differ in size by one amino acid.**
- **Knowing the mass of each amino acid side chain (Gly=1 to Trp=130) you can work out the sequence from the masses of the fragments of the peptide.**

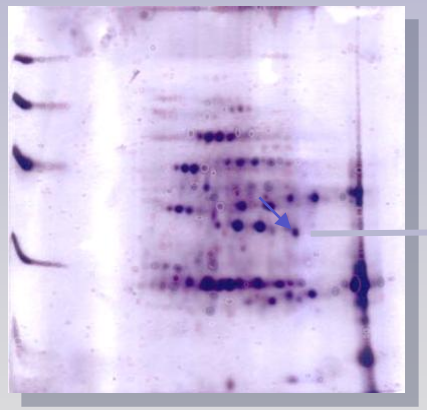
From Genome to Proteome



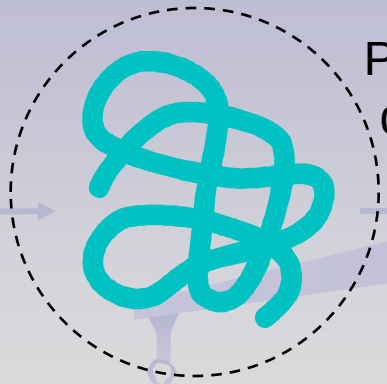
Systems Biology, Integrated Biology

Proteome Reflects Gene Expression Repertoire

Sample

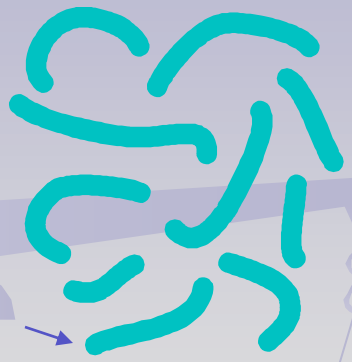


2D electrophoresis



Pure protein

Proteolytic digestion



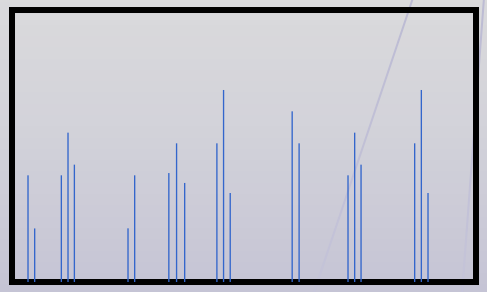
Proteolytic fragments

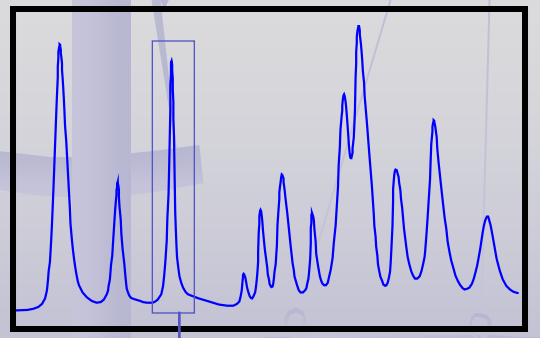
MALDI-TOF

Capillary Electrophoresis

HPLC

Mass spectrum

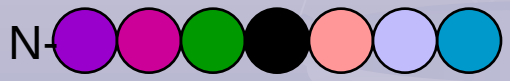




LC/MS/MS

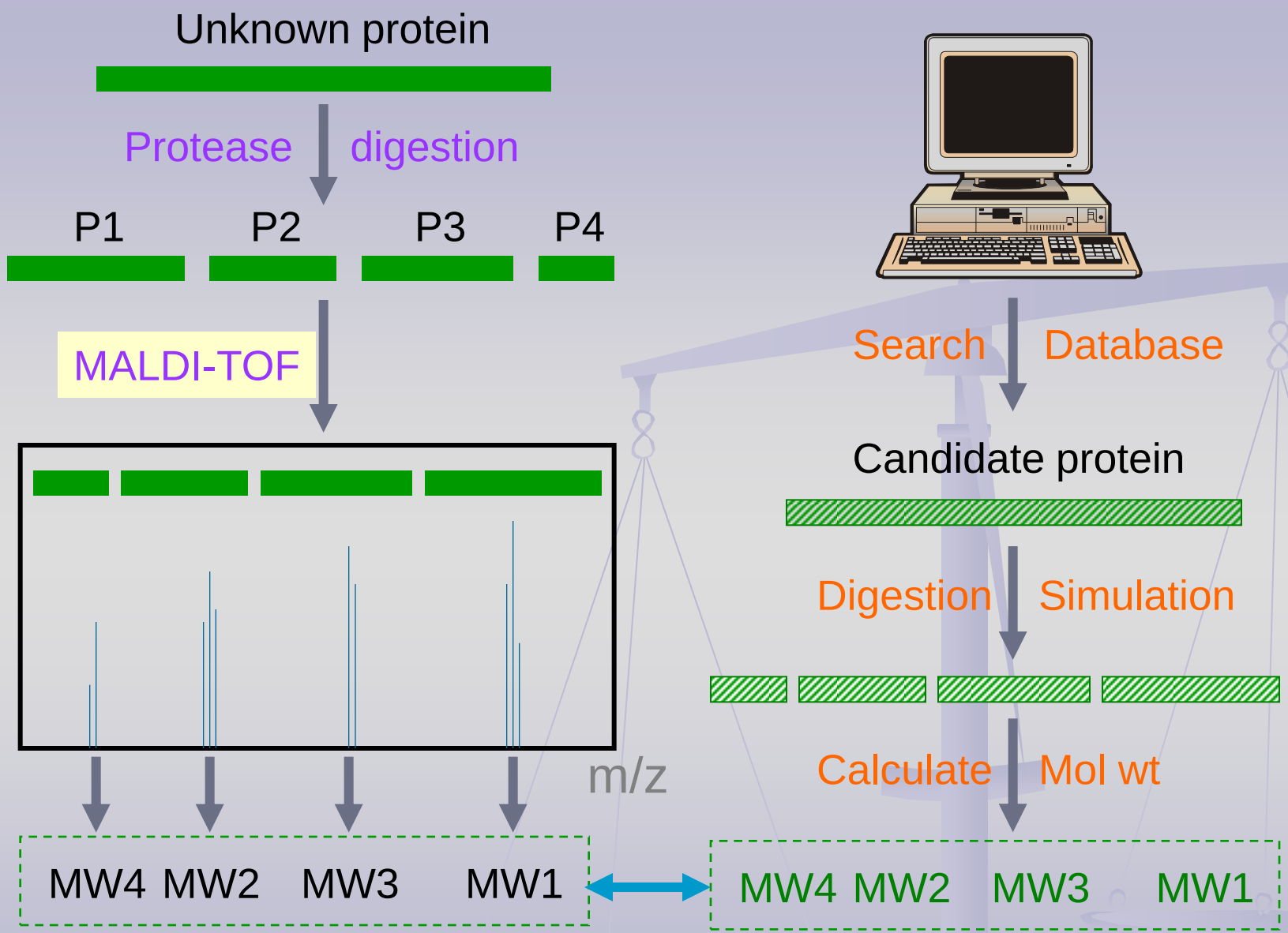
Amino acid sequencing

Database searching



Juang RH (2005) BCbasics

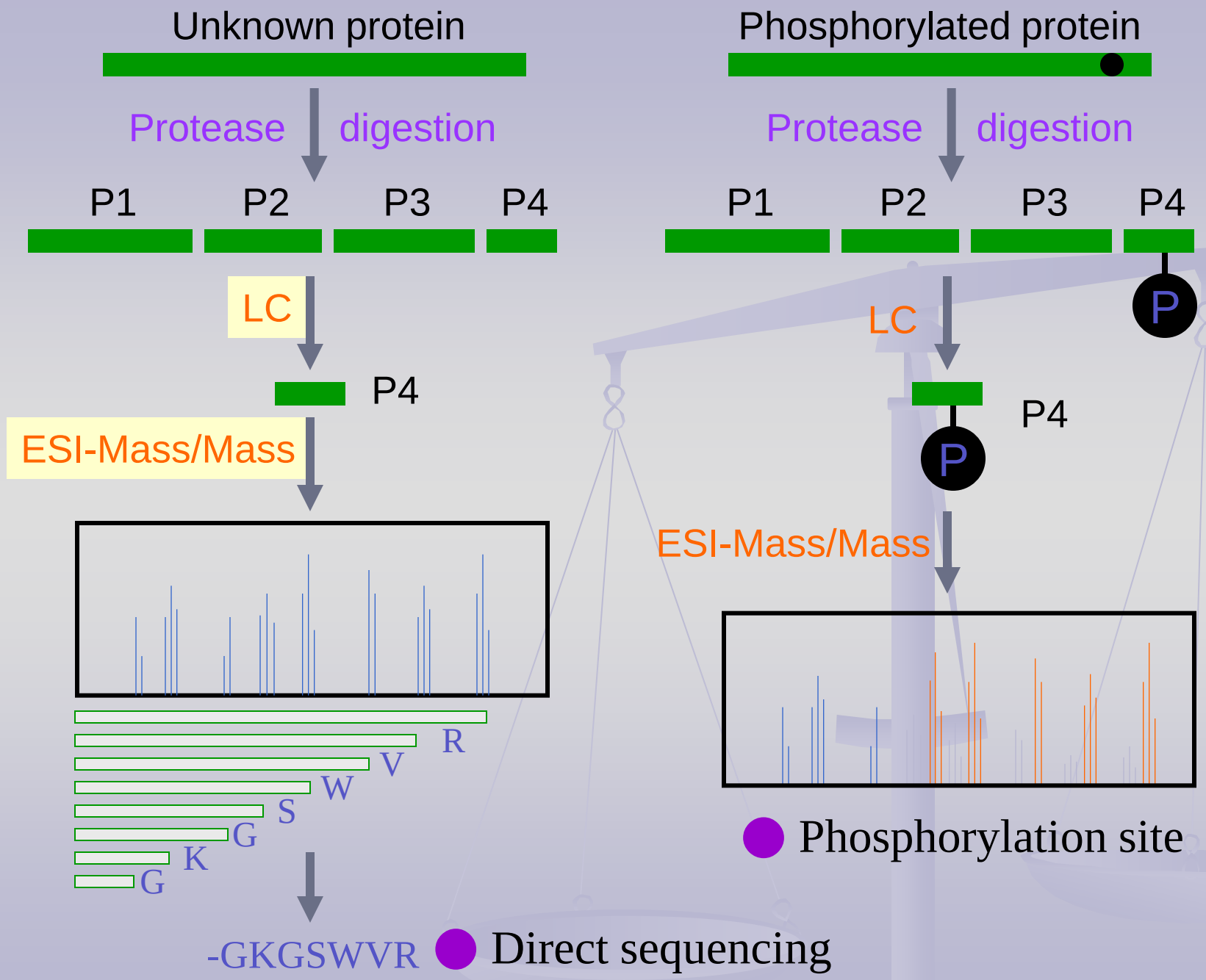
Identify Unknown Protein by Mass



- Compare peptides and identify the unknown protein

Juang RH (2004) BCbasics

Protein Sequencing by Mass



Juang RH (2004) BCbasics

Cleaving Larger Proteins Into Fragments For Sequence Analysis

Polypeptide

Procedure

hydrolyze; separate amino acids →

A	5	H	2	R	1
C	2	I	3	S	2
D	4	K	2	T	1
E	2	L	2	V	1
F	1	M	2	Y	2
G	3	P	3		

Conclusion

Polypeptide has 38 amino acid residues. Trypsin will cleave three times (at one R (Arg) and two K (Lys)) to give four fragments. Cyanogen bromide will cleave at two M (Met) to give three fragments.

react with FDNB; hydrolyze; separate amino acids →

2,4-Dinitrophenylglutamate detected

E (Glu) is amino-terminal residue.

reduce disulfide bonds (if present)

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.

establish sequence →

Amino terminus

T-2 T-1 T-4 T-3

EGAAYHDFEPIDPRGASMALIKYLIACGPMTKDCVHSD

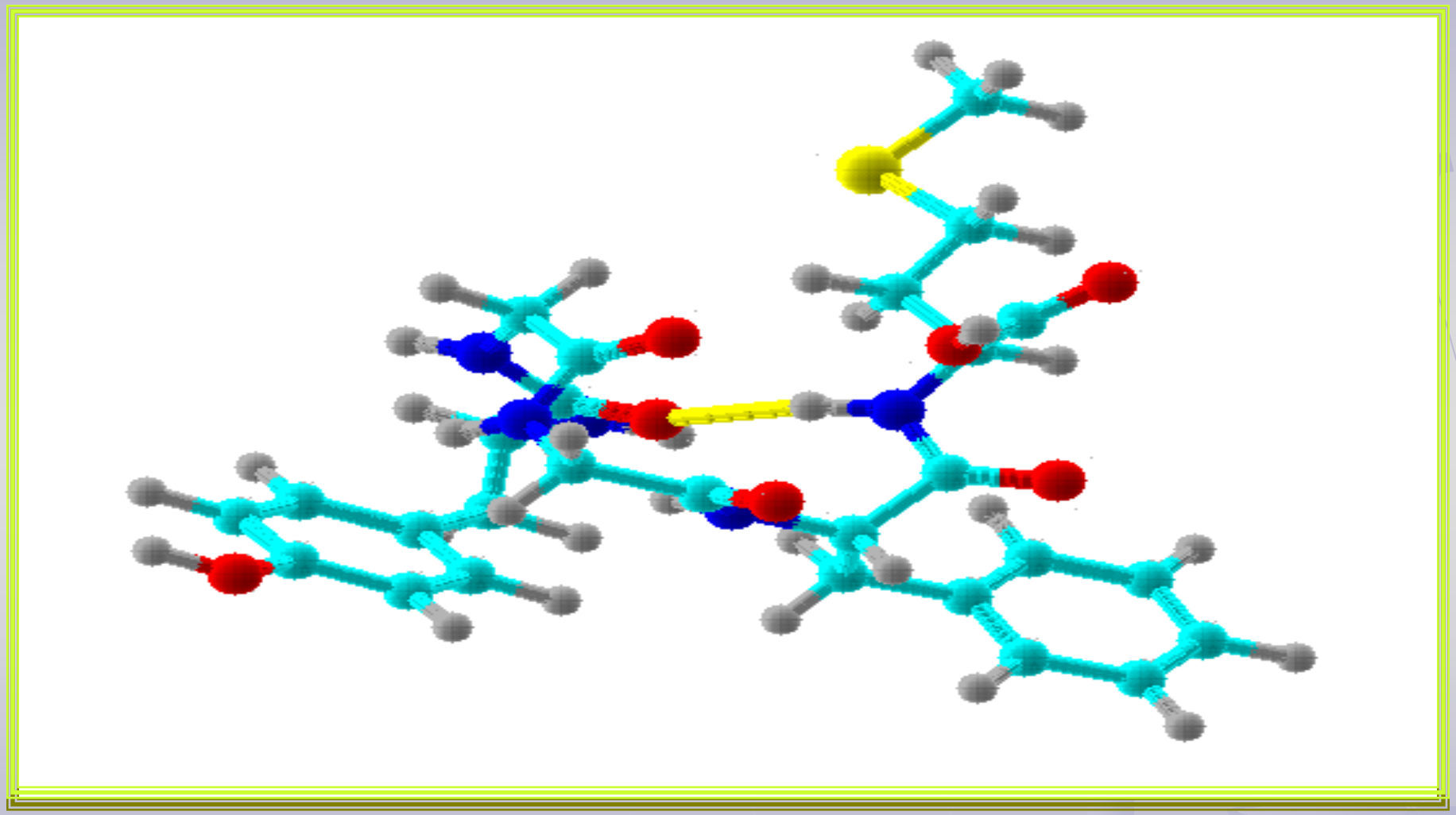
C-1 C-3 C-2

Carboxyl terminus

Enkephalins

- Wondering why the human brain should have receptor sites for alkaloids from the opium poppy led to the discovery of a family of natural painkillers, the endorphins (from ***endogenous morphines***). These substances are oligopeptides, containing from 5 to 30 amino acids.
- The two 5-peptide examples are called enkephalins, from the Greek ***kephale***, meaning "head". These have four of the five amino acids the same: Tyr-Gly-Gly-Phe, tyrosine-glycine-glycine-phenylalanine. One terminates in a leucine, and is known as ***Leu***-enkephalin; the other terminates in a methionine, and is called ***Met***- enkephalin.

Met- enkephalin



Leu-enkephalin

Solid phase peptide synthesis

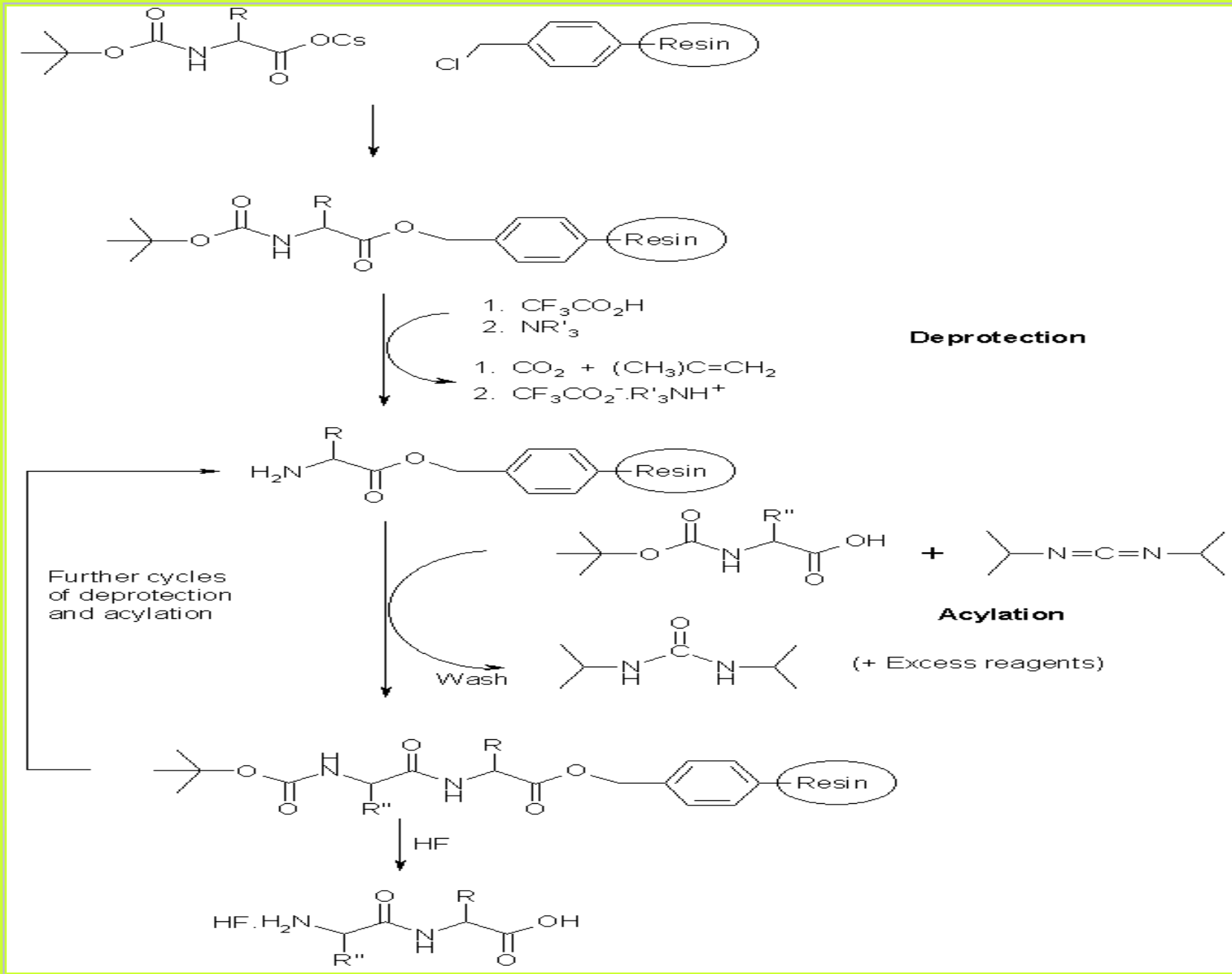
- In solid phase synthesis, the C-terminal residue is attached covalently to an insoluble resin .
- As each new residue is added, the elongation product attached to the solid resin is recovered by filtration (or centrifugation).
- This process has been automated -> **amino acid synthesizers.**
- A very high efficiency is needed for amino acid synthesis to work properly, since any mistakes made are cumulative.

Types of Solid Phase Peptide Synthesis

1. Merrifield synthesis

- This methodology is characterized by the use of *tert*-butyl based temporary α-amino protection and benzyl, or substituted benzyl, groups for permanent side chain protection.
- There are over one hundred different substituted resins suitable for peptide synthesis generally based on polystyrene and polyethylene glycol.

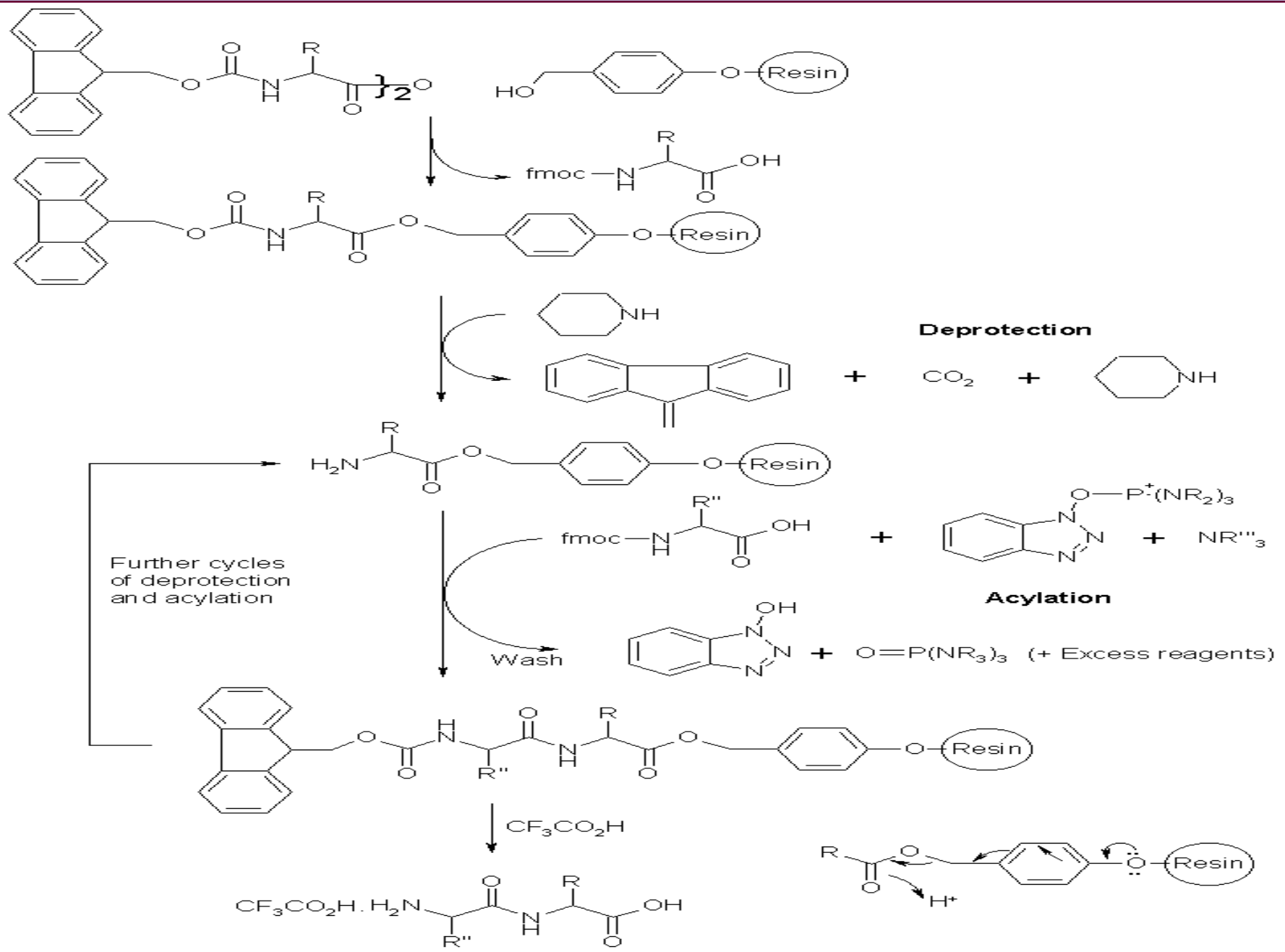
- These resins allow introduction of an amino acid through either substitution, condensation or addition reactions.
- The traditional resin used for Merrifield synthesis was a chloromethylphenyl substituted resin. The first amino acid was attached to the resin through substitution of the chloride by the caesium salt of the BOC-amino acid, generating an equivalent to a benzyl ester.


Resin
1. CF_3CO_2H
2. NR'_3
1. CO_2 + $(CH_3)C=CH_2$
2. $CF_3CO_2^-.R'_3NH^+$
Deprotection
Further cycles of deprotection and acylation
Wash
Acylation
(+ Excess reagents)
HF

- Deportation of the temporary BOC group uses a 20-50% solution of trifluoroacetic acid (TFA) in dichloromethane and has to be followed by neutralization of the resulting ammonium salt with a hindered tertiary base.
- Final cleavage from the resin as well as deprotection of benzyl based side chain protecting groups is achieved using strong acids, usually liquid hydrogen fluoride or trifluoromethane sulphonic acid.
- Such procedures require specialized apparatus and the highly acidic conditions catalyze several possible rearrangements .

2. Fmoc Polyamide Synthesis

- The fundamental differences between the Fmoc polyamide strategy when compared to the Merrifield approach are that the reactions are carried out under continuous flow and that the conditions for α-amino deprotection and cleavage from the resin are far more mild.
- This arises from the adoption of the base labile Fmoc protecting group for α-amino protection. The side chains are generally protected with *tert*-butyl based groups which, in common with the linkage to the resin, can be cleaved by TFA in the presence of scavengers.
- As discussed above, a large number of resins are available.


Resin
fmoc
OH
NH
Deprotection
CO_2
$O-P^{+}(NR_2)_3$
NR'''_3
Acylation
Further cycles
of deprotection
and acylation
Wash
$O=P(NR_3)_3$
(+ Excess reagents)
CF_3CO_2H
$CF_3CO_2H.H_2N$
H^+

- Traditionally, resins with 4-hydroxymethylphenoxy substitution were used.
- These allowed were esterified with the anhydride of the first amino acid.
- As a result of the mesomerically electron donating *para* oxygen atom stabilizing the resultant carbocation, cleavage of the peptide from the resin occurs under more mild acid conditions, typically using trifluoroacetic acid with scavengers.
- The use of continuous flow means that the reagents are passed through a reaction chamber containing the resin supported peptide.

- Having passed through this chamber, the reagents can be recirculated back into the chamber again or taken to a waste collection bottle.
- This allows the resin to be washed clean of excess reagents and unwanted reaction products, which in turn helps drive deprotection steps to completion following Le Chatelier's principle.
- In addition, the solution can be passed through a u.v. detector and monitored at a suitable wavelength for the Fmoc chromophore.
- A typical cycle consists of:

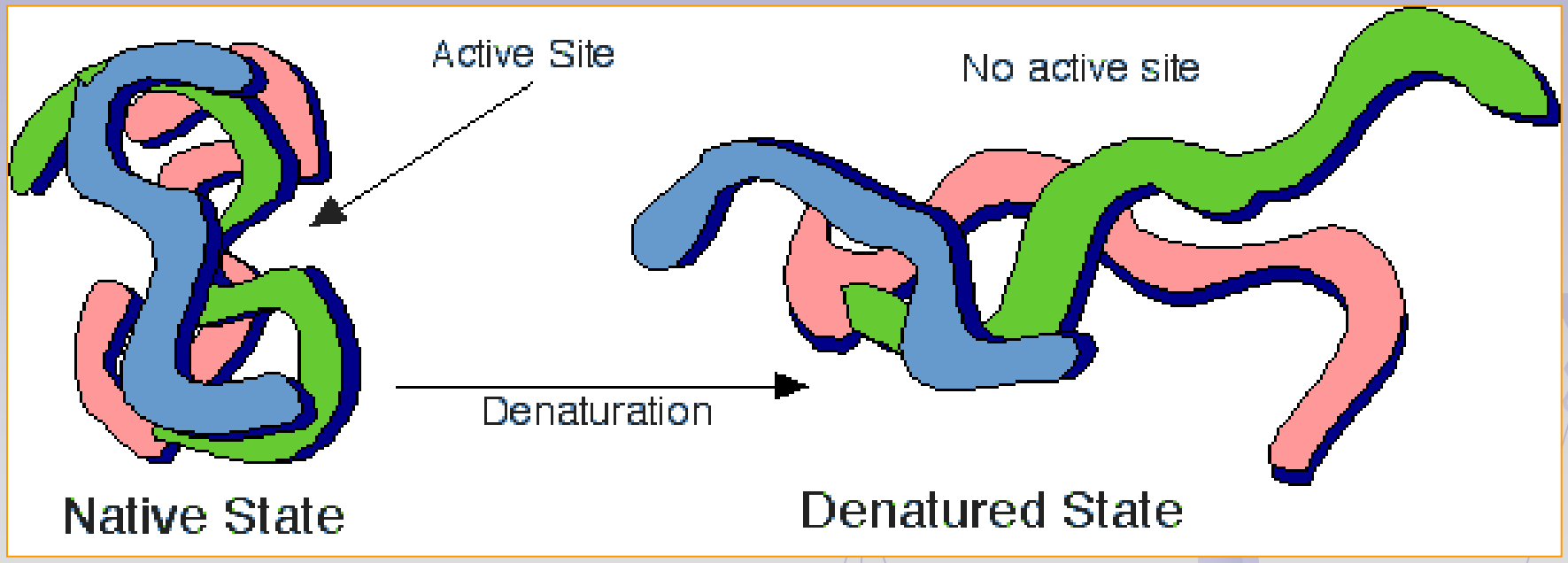
1. Deprotection of the preceding residue with piperidine.
2. Wash to remove any remaining reagents from 1

3. Acylation in a recirculatory mode.
4. Wash to remove excess reagents.

- In this way, the cycle can be qualitatively monitored.
- Such monitoring can be used in conjunction with automation.
- For example, a slow deprotection step suggests inaccessibility of the peptide amino terminal, indicating that the next coupling step may require a longer acylation time.
- If this is detected then the cycle can be interrupted for later manual intervention.

Protein :Denaturation & Renaturation

1. When proteins are heated, or exposed to acids or bases, or high salt concentrations, the variety of weak bonds holding tertiary and quaternary structure together can be disrupted so that the protein unfolds. Unfolding = **denaturation** resulting in loss of function.
2. Unfolding can proceed even to disrupt secondary structure.



3. Denaturation is sometimes **reversible** ; an unfolded protein can be restored to correct folding and regain biological activity. This is called **renaturation** .
4. Denaturation can also occur **irreversibly** (as when egg white protein, albumin , is denatured by boiling to congeal as egg white). Renaturation is then no longer possible.

GOOD LUCK