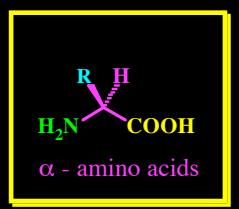
Amino Acids - The Building Blocks of Proteins

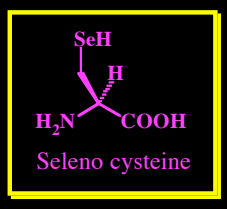
Manickam Sugumaran
University of Massachusetts Boston
Boston, MA 02125

Amino acids

- There are 20 common amino acids (R group is variable) that are genetically coded in proteins.
- Two uncommon amino acids that are present in some proteins are selenocysteine (1986) and pyrrolysine (2002). They have their own codons.
- Sec – UGA (opal stop codon); pyrrolysine –UAG (amber stop codon).
- All other amino acids arise by the modification of these amino acids.
- Ultimately over 200 different amino acids are found in proteins.



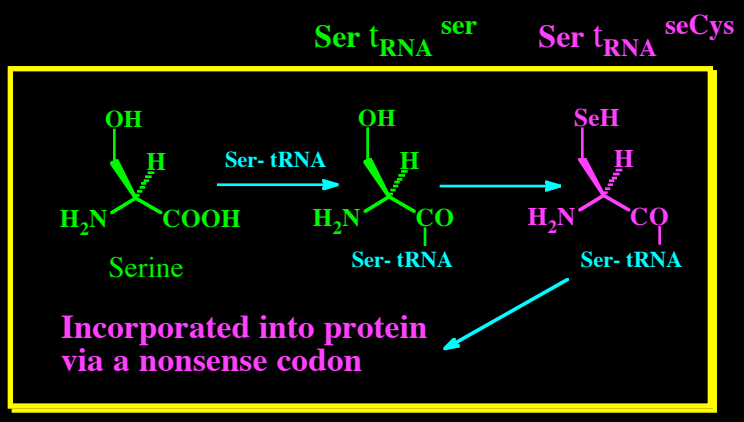
α - amino acids



Seleno cysteine

Selenocysteine is made on Ser - tRNA

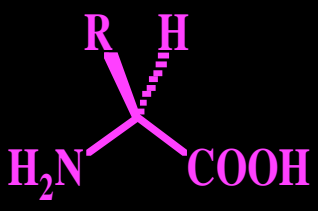
The codon for selenocysteine is in frame UGA opal
Examples of enzymes that have selenocysteine:
Formate dehydrogenase and glutathione reductase



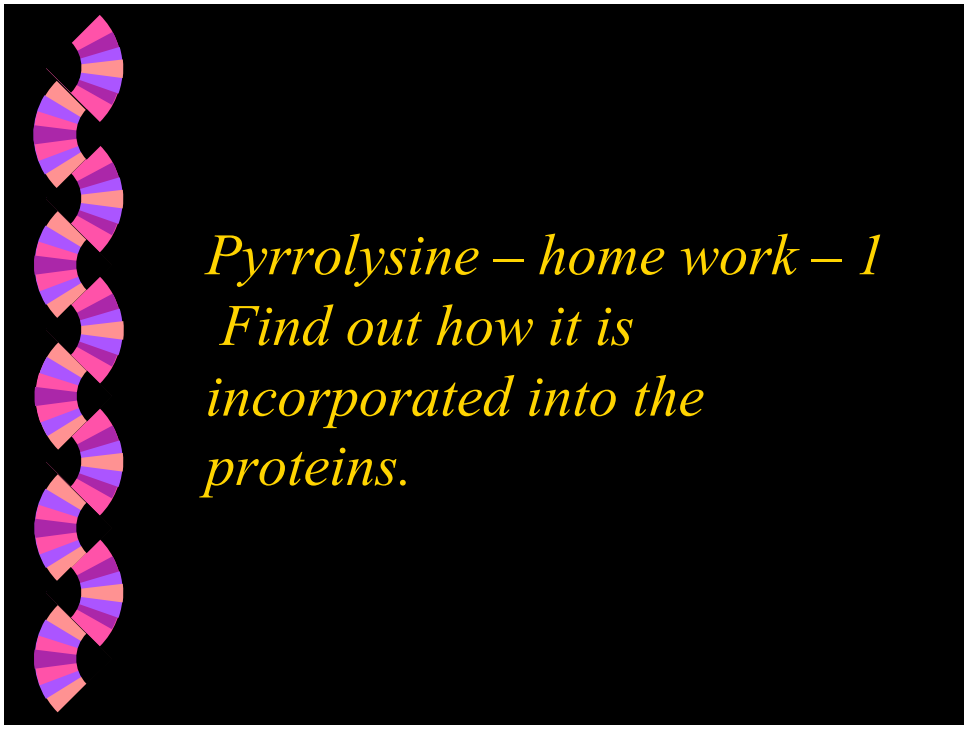
Leinfelder et al., Nature 331, 723 - 725 (1988).

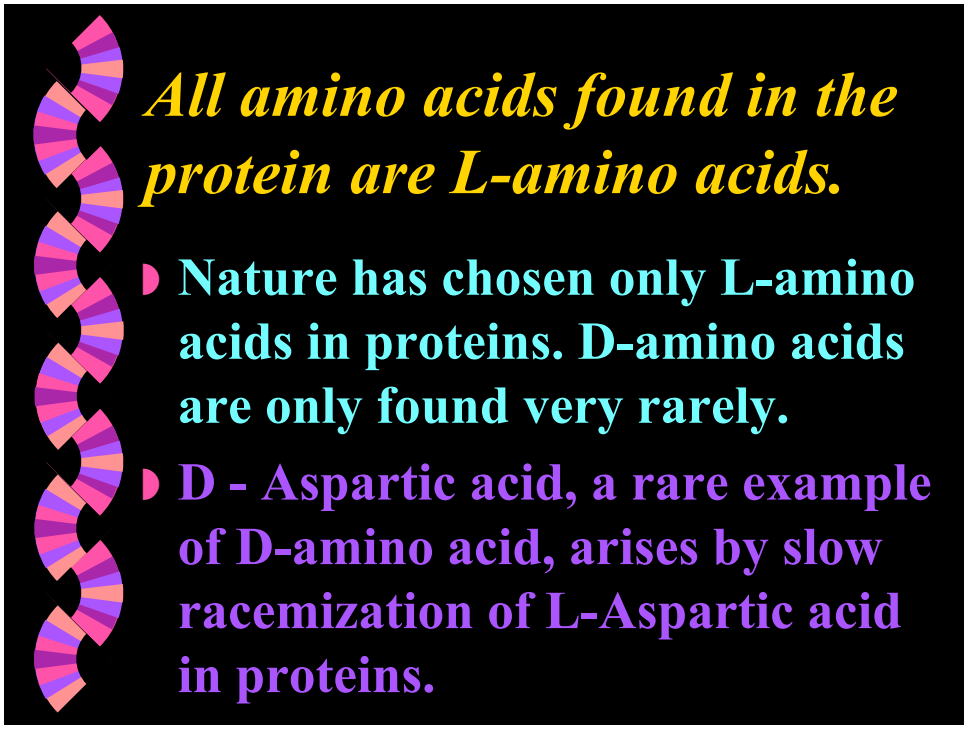
How to build the structure of some amino acids

R =	Name
- H	Glycine
- CH_3	Alanine
CH_3-CH-CH_3	Valine
– CH_2OH	Serine
-CH_2COOH	Aspartic acid
Benzyl (-CH_2Ph)	Phenylalanine



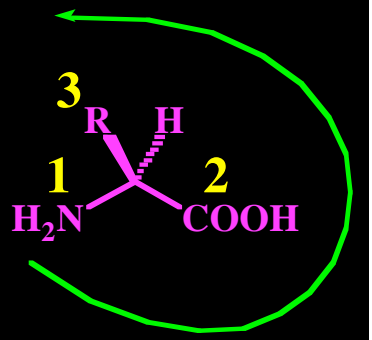
Write always from N to C
Keep the smallest group (H) away from you. R should be in the front (facing you)


Pyrrolysine – home work – 1
Find out how it is incorporated into the proteins.


All amino acids found in the protein are L-amino acids.
Nature has chosen only L-amino acids in proteins. D-amino acids are only found very rarely.
D - Aspartic acid, a rare example of D-amino acid, arises by slow racemization of L-Aspartic acid in proteins.

L-amino acids are S-chiral molecules

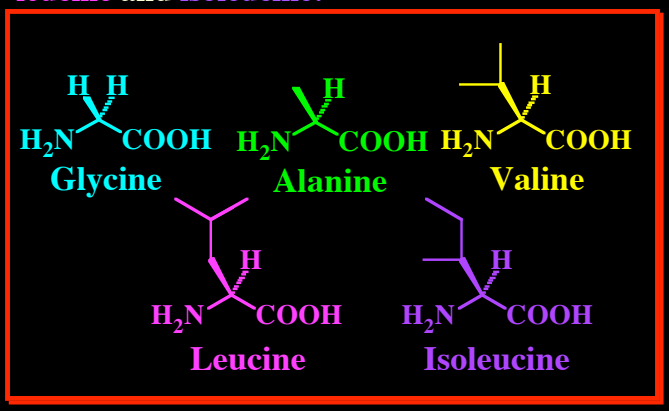
By keeping the smallest group away from you, you prioritize the substituents based on their atomic number.
N gets first (1)
Followed by COOH (2)
Followed by R (3).



In L -amino acid going 1-2-3 takes anti-clock wise direction.
So L-amino acids are all S- chiral molecules.

Simple amino acids with aliphatic hydrophobic side chain

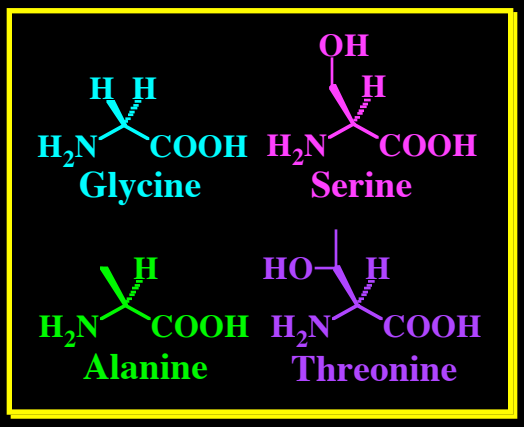
On the α- carbon atom of glycine, if you substitute
A) a methyl group for a hydrogen, you get alanine.
B) Isopropyl group, you get valine
C) Isobutyl group in two different way, you get leucine and isoleucine.



Hydroxyl group containing aliphatic amino acids -

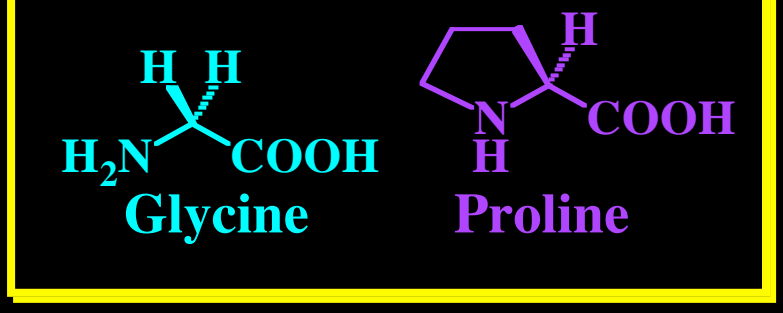
Substitute a - CH_2OH group for H in glycine to get serine.

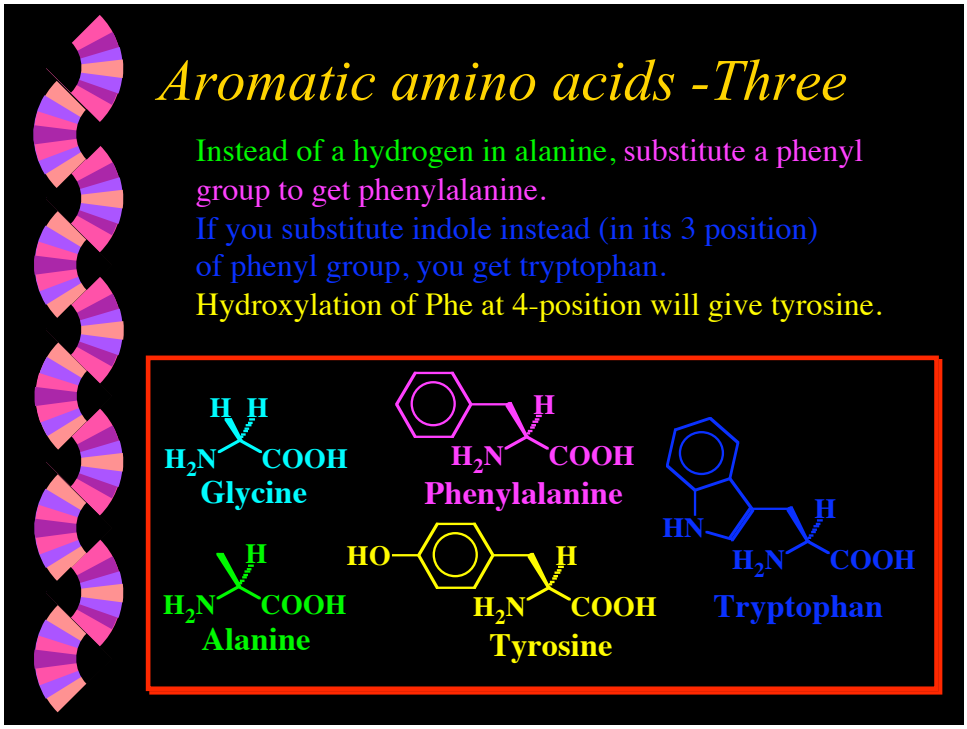
Substitution of - CHOH CH_3 group gives threonine.

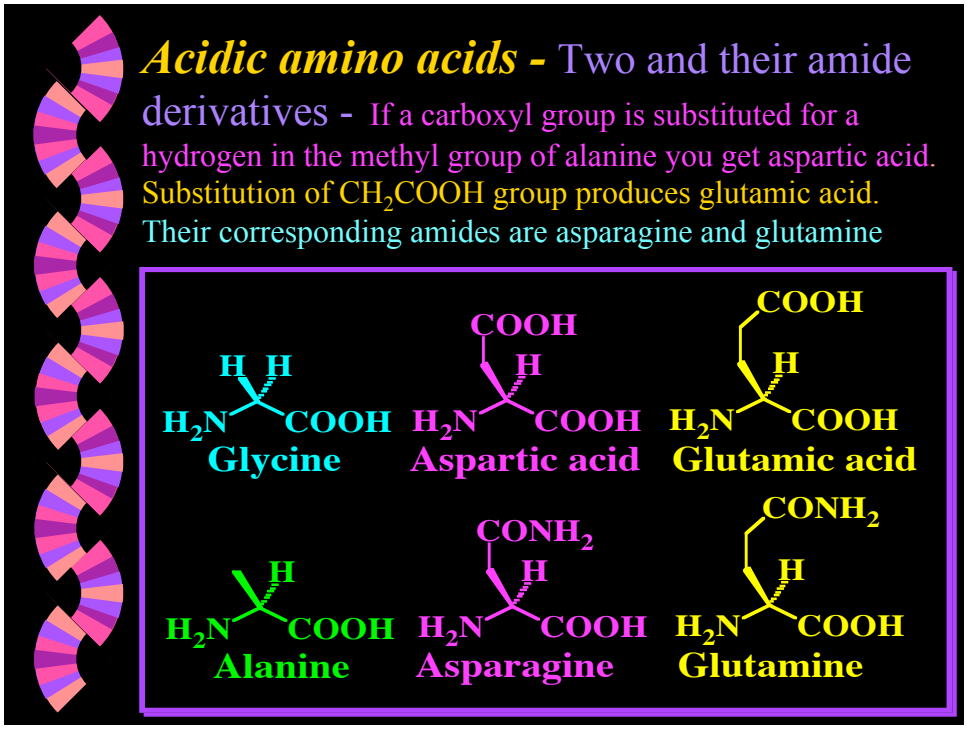


Proline is the only imino acid.

Note that it is not an α-amino acid. The amine group is cyclized with the R side chain *via* three CH_2 groups. It is a secondary amine. It is called an imino acid.

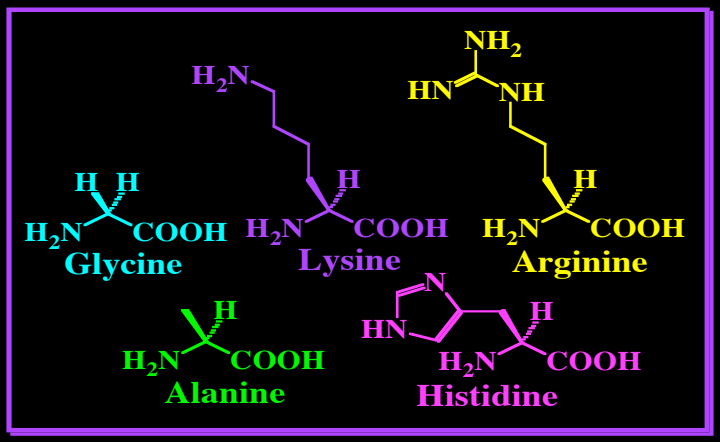



Aromatic amino acids -Three
Instead of a hydrogen in alanine, substitute a phenyl group to get phenylalanine.
If you substitute indole instead (in its 3 position) of phenyl group, you get tryptophan.
Hydroxylation of Phe at 4-position will give tyrosine.
Glycine
Phenylalanine
Alanine
Tyrosine
Tryptophan


Acidic amino acids - Two and their amide derivatives - If a carboxyl group is substituted for a hydrogen in the methyl group of alanine you get aspartic acid.
Substitution of CH2COOH group produces glutamic acid.
Their corresponding amides are asparagine and glutamine
Glycine
Aspartic acid
Glutamic acid
Alanine
Asparagine
Glutamine

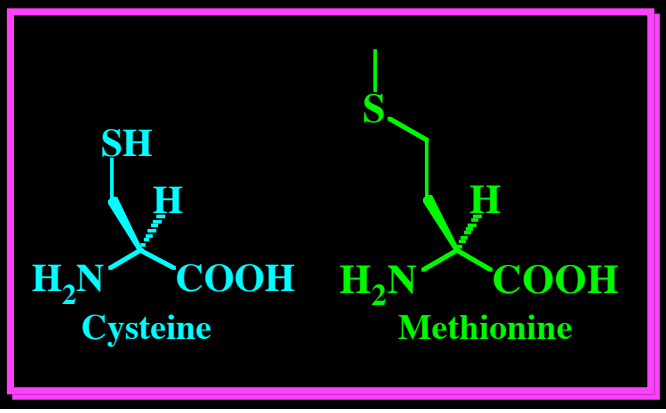
Basic amino acids - lysine, arginine and histidine.

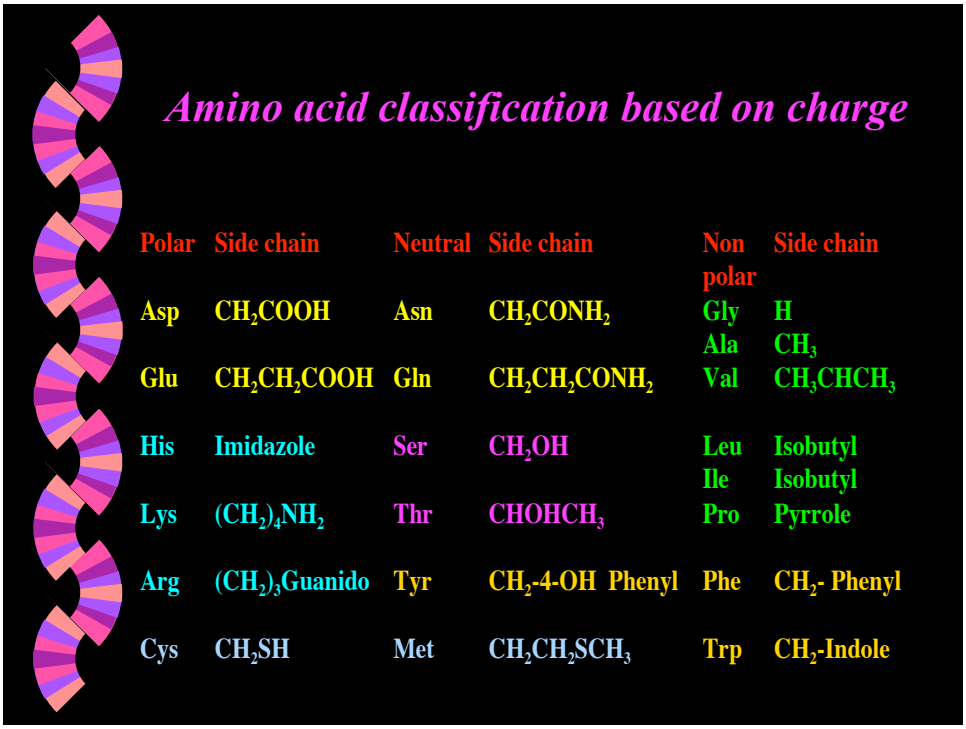
Lysine has -$(CH_2)_4$ NH_2 instead of H (in glycine)
Arginine has - $(CH_2)_3$-guanido group and
Histidine has -CH_2-imidazole group.

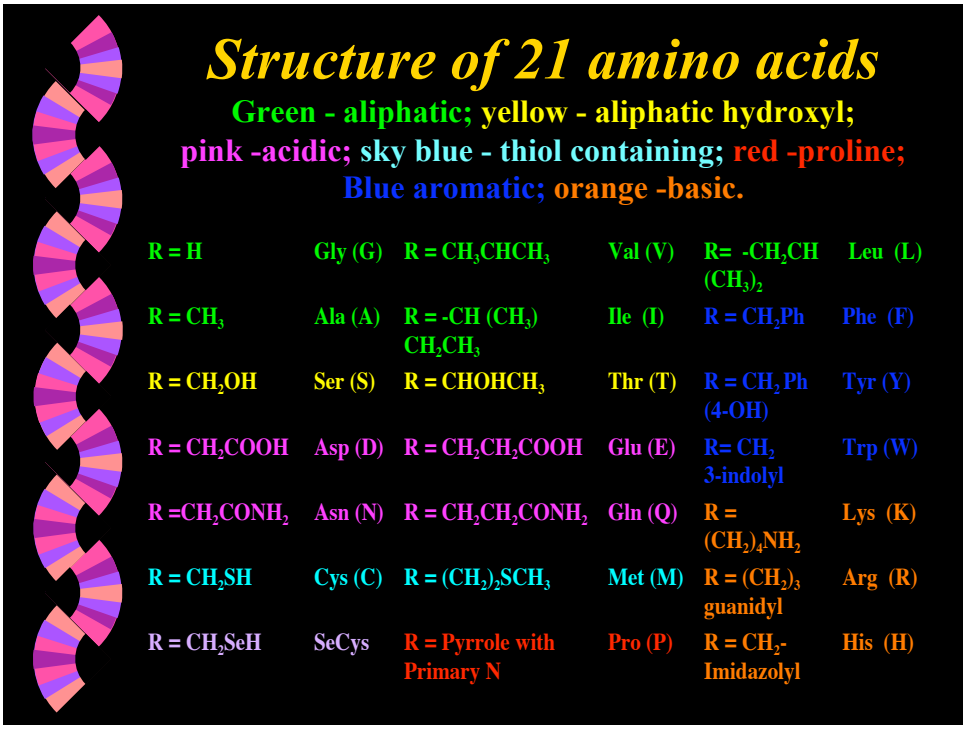


Sulfur containing amino acids - Cysteine and Methionine

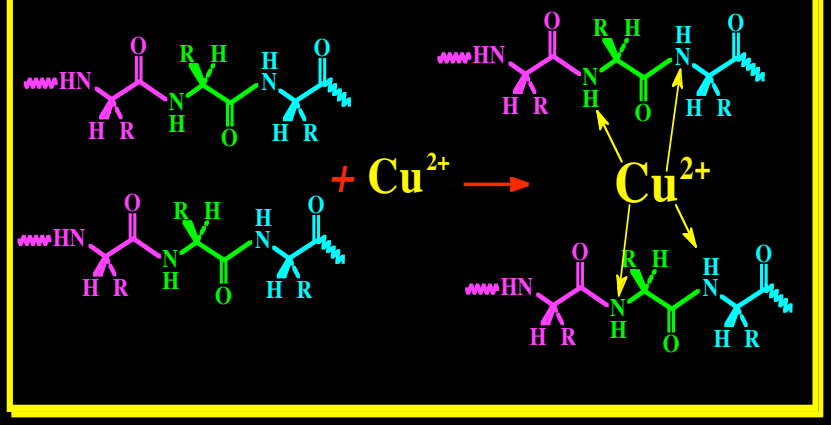
Cysteine has CH_2SH and methionine has
- $(CH_2)_2SCH_3$ instead of H (in glycine)



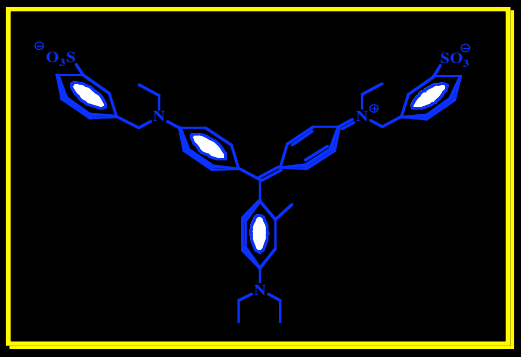

Amino acid classification based on charge
Polar Side chain
Asp CH2COOH
Glu CH2CH2COOH
His Imidazole
Lys (CH2)4NH2
Arg (CH2)3Guanido
Cys CH2SH
Neutral Side chain
Asn CH2CONH2
Gln CH2CH2CONH2
Ser CH2OH
Thr CHOHCH3
Tyr CH2-4-OH Phenyl
Met CH2CH2SCH3
Non polar Side chain
Gly H
Ala CH3
Val CH3CHCH3
Leu Isobutyl
Ile Isobutyl
Pro Pyrrole
Phe CH2- Phenyl
Trp CH2-Indole


Structure of 21 amino acids
Green - aliphatic; yellow - aliphatic hydroxyl;
pink -acidic; sky blue - thiol containing; red -proline;
Blue aromatic; orange -basic.
R = H Gly (G)
R = CH3 Ala (A)
R = CH2OH Ser (S)
R = CH2COOH Asp (D)
R =CH2CONH2 Asn (N)
R = CH2SH Cys (C)
R = CH2SeH SeCys
R = CH3CHCH3 Val (V)
R = -CH (CH3) CH2CH3 Ile (I)
R = CHOHCH3 Thr (T)
R = CH2CH2COOH Glu (E)
R = CH2CH2CONH2 Gln (Q)
R = (CH2)2SCH3 Met (M)
R = Pyrrole with Primary N Pro (P)
R= -CH2CH (CH3)2 Leu (L)
R = CH2Ph Phe (F)
R = CH2 Ph (4-OH) Tyr (Y)
R= CH2 3-indolyl Trp (W)
R = (CH2)4NH2 Lys (K)
R = (CH2)3 guanidyl Arg (R)
R = CH2- Imidazolyl His (H)

Quantification of Proteins: 1. Biuret reagent



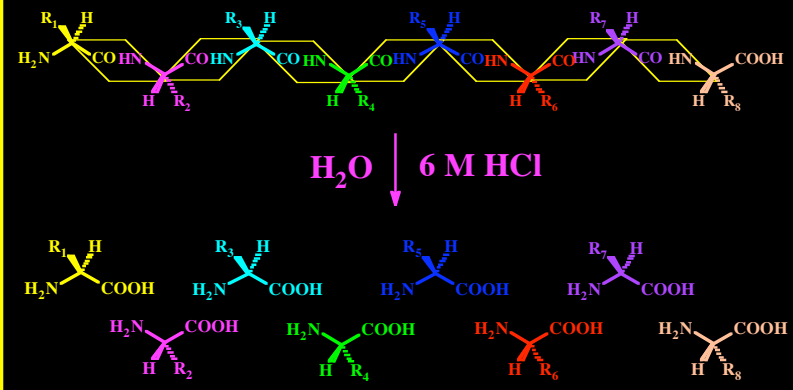
Quantification of Proteins: 2. Coomassie Brilliant Blue R-250



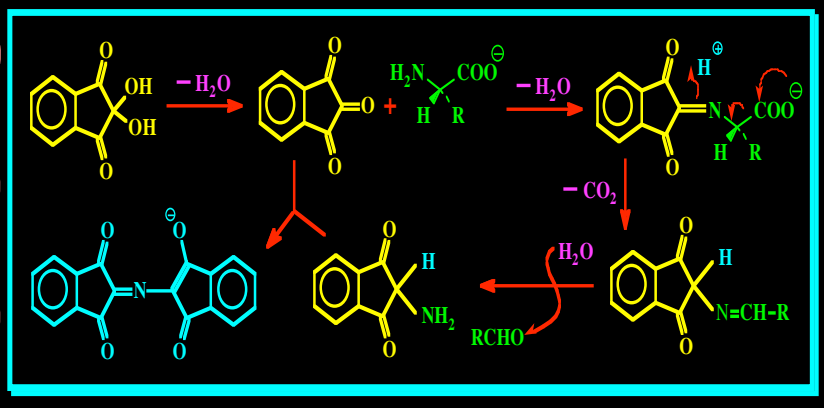
The dye coomassie blue is red in color. Upon binding to protein, it changes its color to blue. This change is used to quantify proteins. (Bradford Protein assay)

Peptide bond is susceptible to hydrolysis.

Peptide bonds can be hydrolyzed to generate individual amino acids

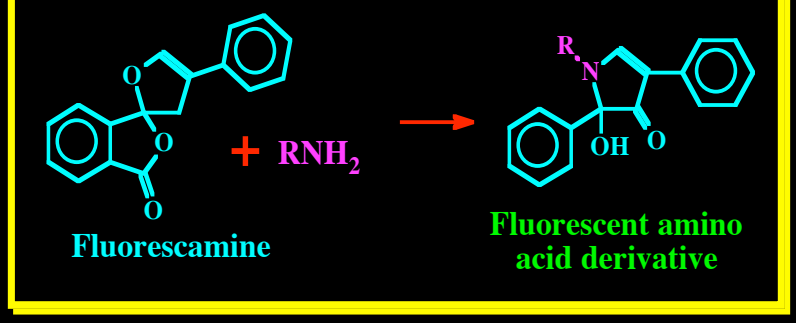


Ninhydrin reaction with amino acids.
Amino acids are usually quantified by ninhydrin reaction.



Fluorescamine reaction.

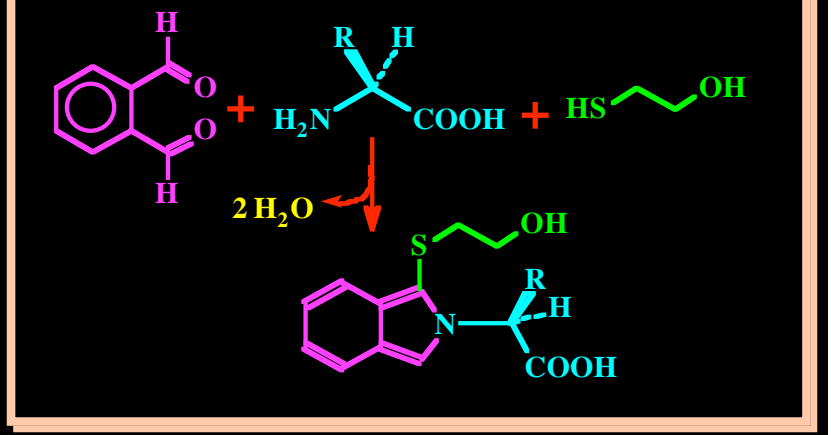
Fluorescamine is another reagent used for quantification of amino acid. It is more sensitive than ninhydrin

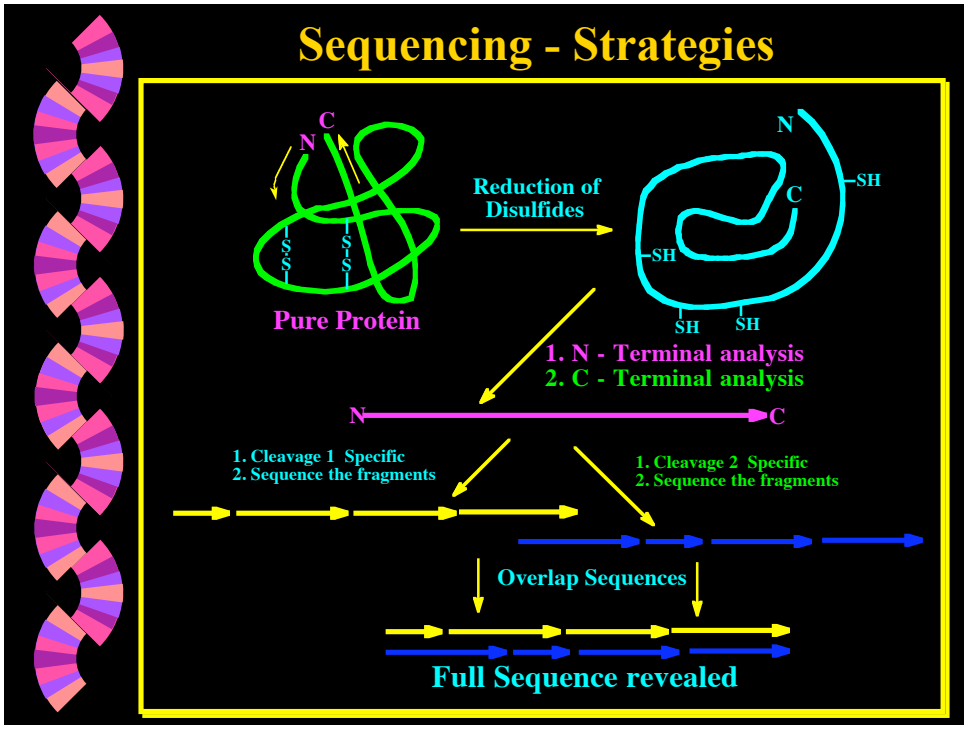


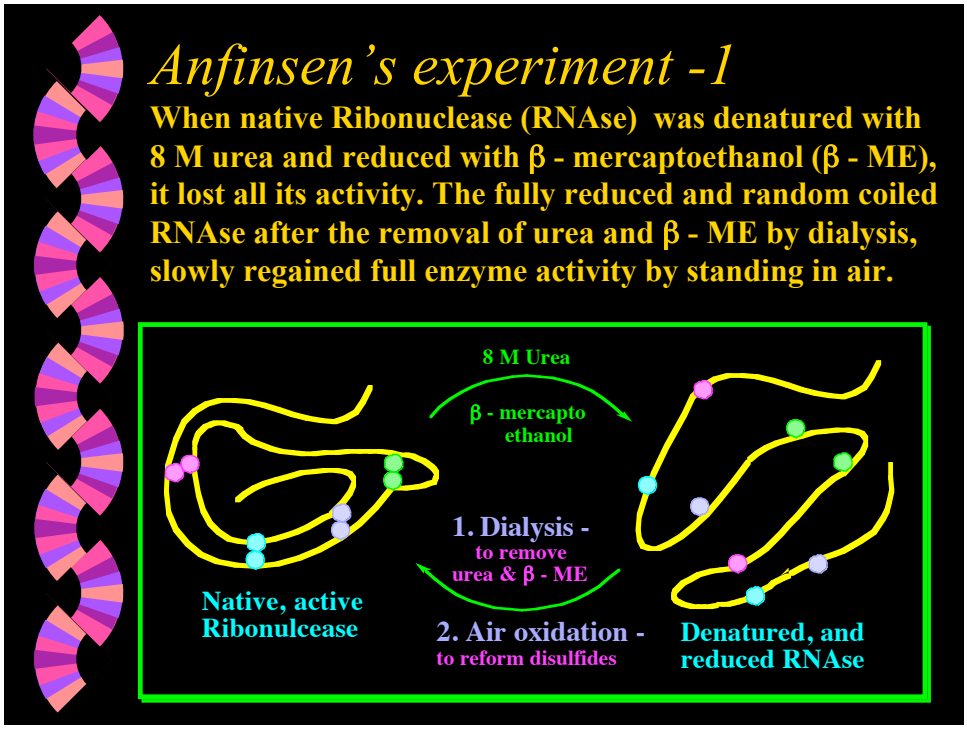
Fluorescamine

Fluorescent amino acid derivative

o-Phthalaldehyde yields a fluorescent derivatives by reaction with amino acid and β-mercaptoethanol for quantification of amino acids

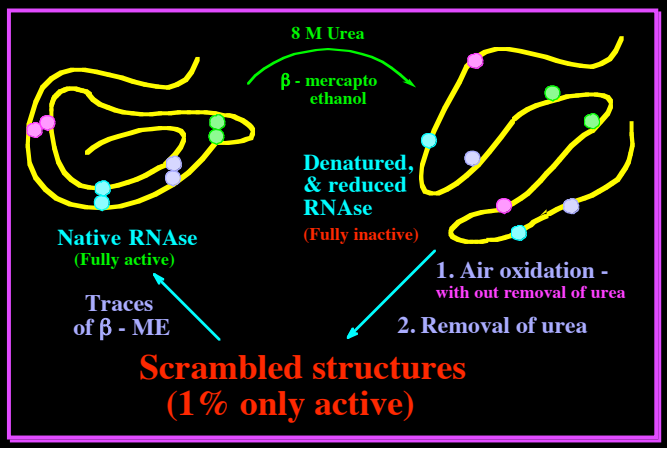



Sequencing - Strategies
C
N
S-S
S-S
Pure Protein
Reduction of Disulfides
N
C
SH
SH
SH
SH
1. N - Terminal analysis
2. C - Terminal analysis
N
C
1. Cleavage 1 Specific
2. Sequence the fragments
1. Cleavage 2 Specific
2. Sequence the fragments
Overlap Sequences
Full Sequence revealed


Anfinsen's experiment -1
When native Ribonuclease (RNAse) was denatured with 8 M urea and reduced with β - mercaptoethanol (β - ME), it lost all its activity. The fully reduced and random coiled RNAse after the removal of urea and β - ME by dialysis, slowly regained full enzyme activity by standing in air.
8 M Urea
β - mercapto ethanol
1. Dialysis - to remove urea & β - ME
2. Air oxidation - to reform disulfides
Native, active Ribonulcease
Denatured, and reduced RNAse

Anfinsen's second experiment -

After generating denatured and reduced RNAse, if the disulfide bridges were formed without removing urea and then the protein was made urea free, the resultant protein had scrambled structures. It exhibited only 1% of the total activity. However, when traces of β - mercaptoethanol was added to scrambled protein, slowly fully active protein was reformed.



Explanation for the second experiment

The denatured and reduced RNAse when allowed to form disulfide bridges in presence of urea, generated scrambled structures with different combination of disulfide bridges. Only one of them (the correct structure) exhibited biological activity. The rest did not. These structures were allowed to refold into correct structure by traces of β - mercaptoethanol (by thiol - disulfide exchange reaction).

The first SH has 7 possible SH to form disulfide; Once, it is forms a disulfide, the third SH has 5 SH to form disulfide bridge. The fifth has 3 SH to form disulfide and the seventh has the last SH.

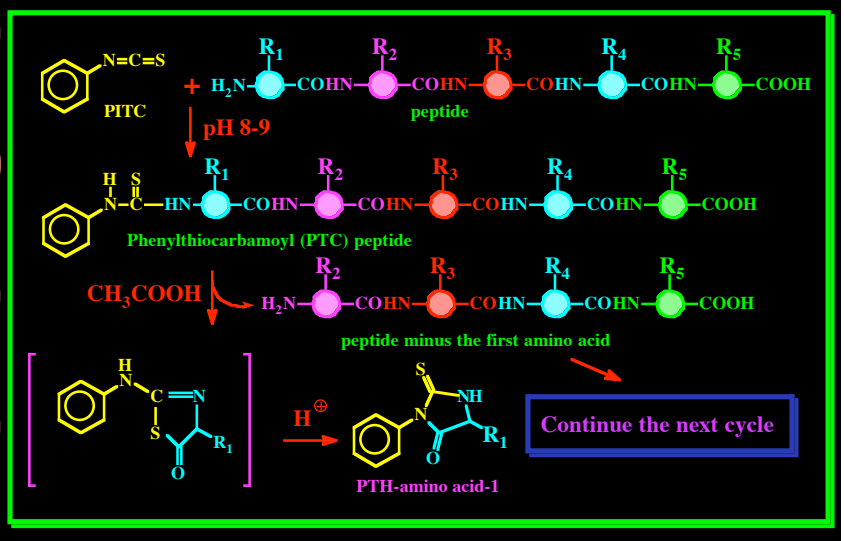
7 x 5 x 3 x 1 = 105

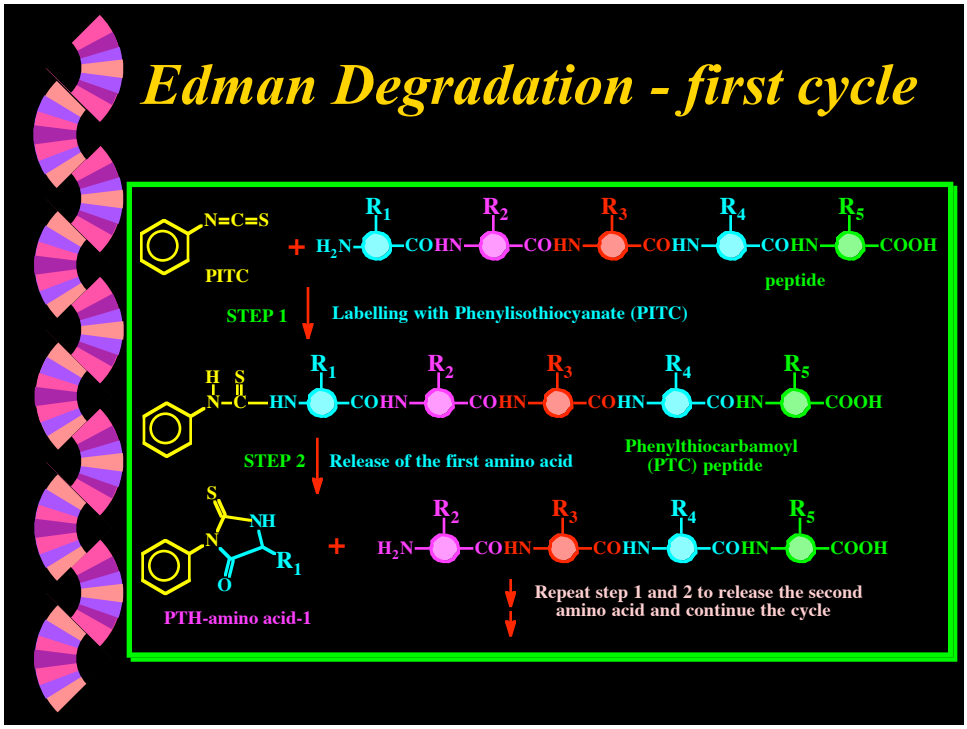
(Of the 105 possible structures, only one will be the correct structure and hence only 1% activity is observed)

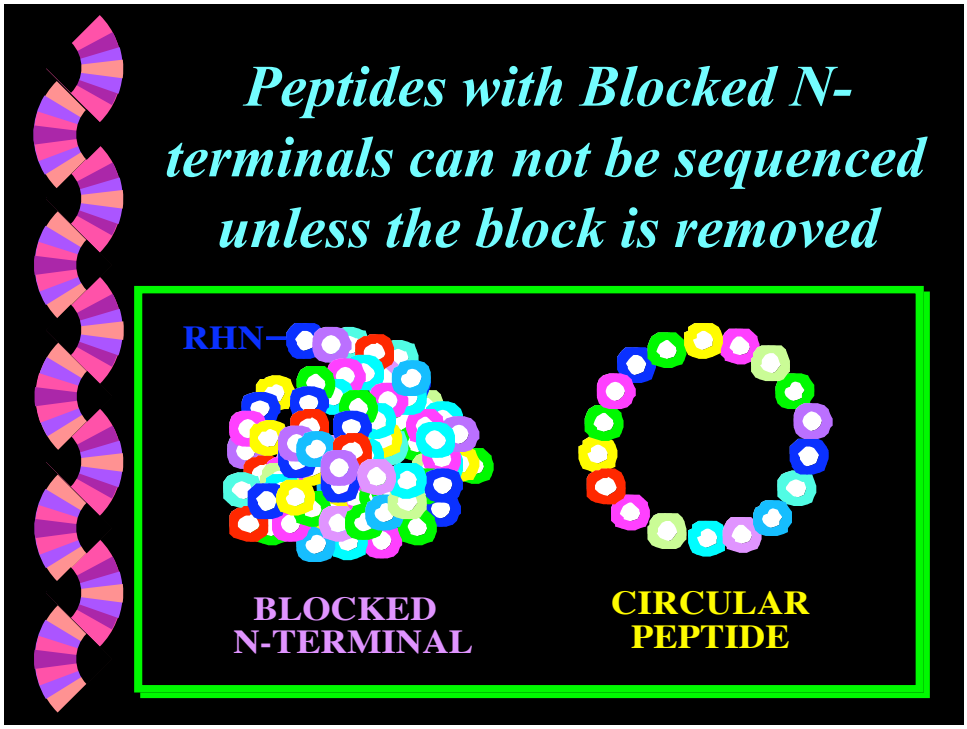
Protein terminal analysis - End group analyzing strategies

Protocol	Site	Specificity
Edman Degradation	C side of N terminus	Nonspecific
Sanger's reagent	N- terminal analysis	Nonspecific
Carboxypeptidase A	N side of C terminus	$R_n \neq$ Arg, Lys, Pro $R_{n-1} \neq$ Pro
Carboxypeptidase B	N side of C terminus	$R_n =$ Arg, Lys, AECys $R_{n-1} \neq$ Pro
Hydrazinolysis	C terminal analysis	Only C – terminal comes out free

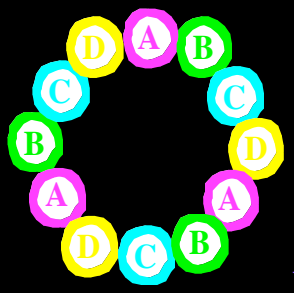
Edman Degradation - Overall reaction

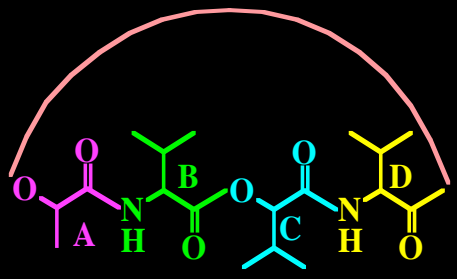



Edman Degradation - first cycle
N=C=S
PITC
+ H2N–COHN–COHN–COHN–COHN–COOH
peptide
STEP 1
Labelling with Phenylisothiocyanate (PITC)
Phenylthiocarbamoyl (PTC) peptide
STEP 2
Release of the first amino acid
PTH-amino acid-1
Repeat step 1 and 2 to release the second amino acid and continue the cycle


Peptides with Blocked N-terminals can not be sequenced unless the block is removed
RHN–
BLOCKED N-TERMINAL
CIRCULAR PEPTIDE

Cyclic peptides (valinomycin for example) cannot be sequenced as they do not have a free N terminus



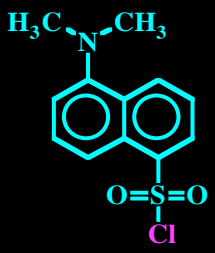


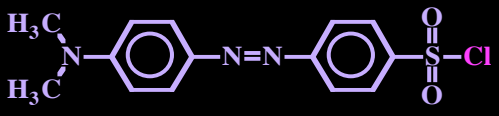
VALINOMYSIN

L-LACTATE - L-VALINE- D-HYDROXYISOVALERATE- D-VALINE

Reagents used for N-terminal analysis

Fluorodinitrobenzene

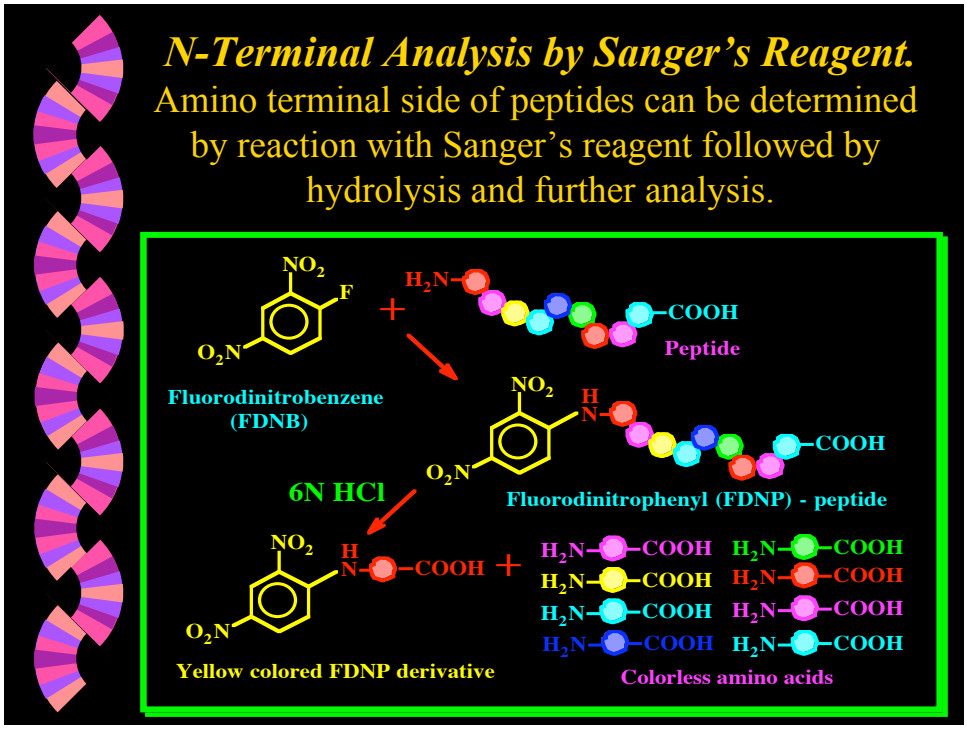


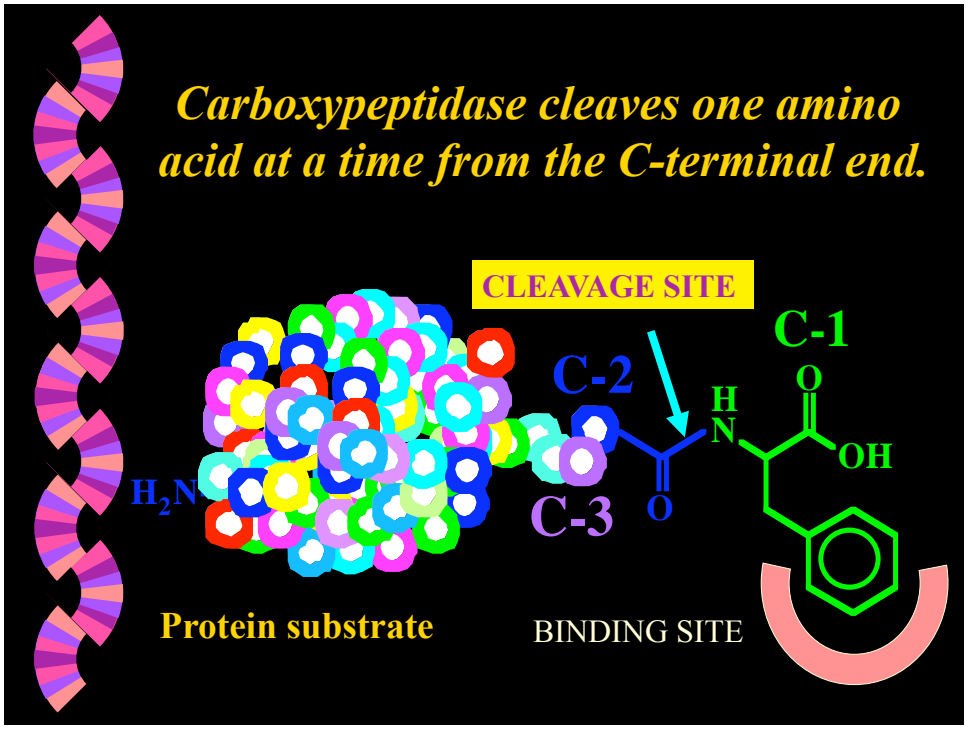


Dabsyl chloride

Dansyl chloride

The reactions of fluorodinitrobenzene (FDNB) is shown in the next slide. Other two reagents also react similarly.

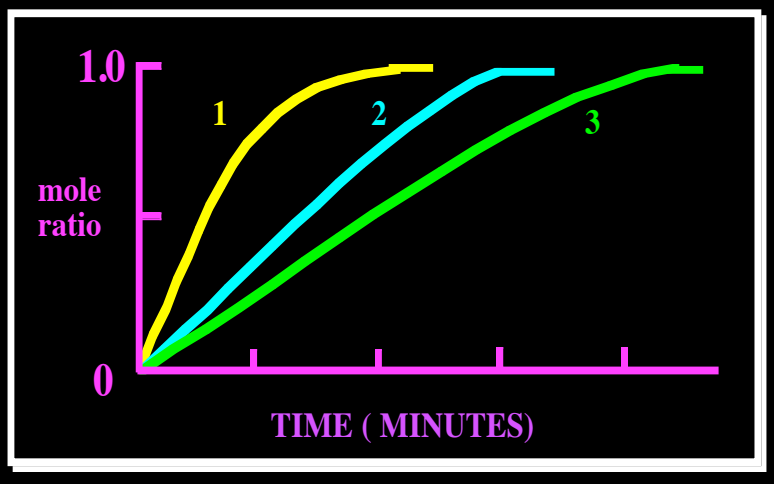

N-Terminal Analysis by Sanger's Reagent.
Amino terminal side of peptides can be determined by reaction with Sanger's reagent followed by hydrolysis and further analysis.
Fluorodinitrobenzene (FDNB)
Peptide
Fluorodinitrophenyl (FDNP) - peptide
6N HCl
Yellow colored FDNP derivative
Colorless amino acids

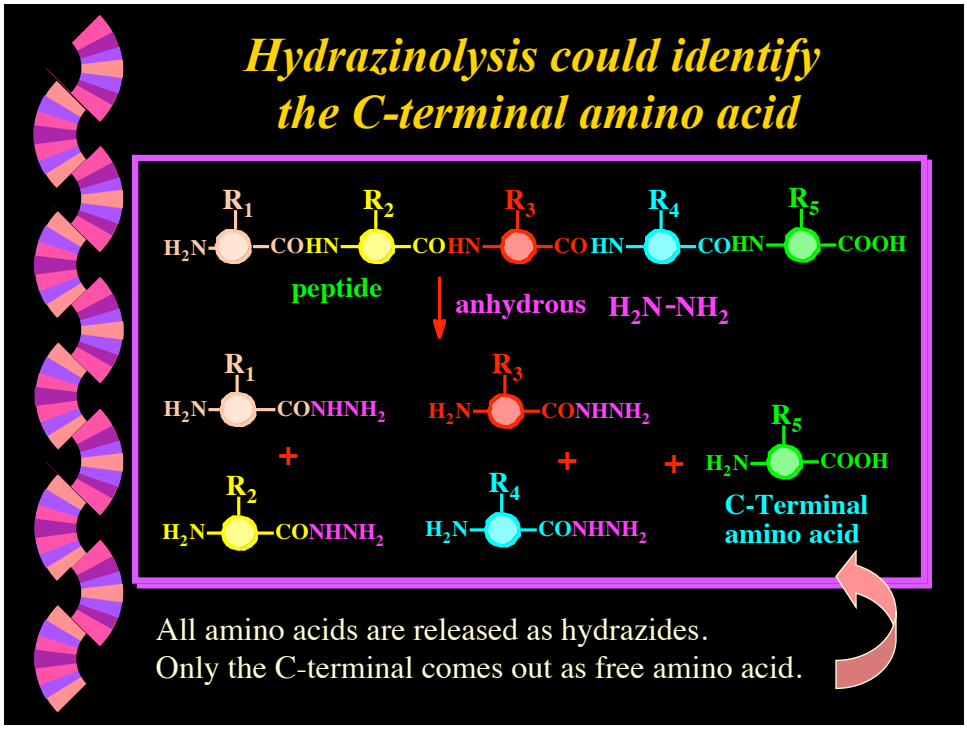

Carboxypeptidase cleaves one amino acid at a time from the C-terminal end.
CLEAVAGE SITE
C-1
C-2
C-3
Protein substrate
BINDING SITE

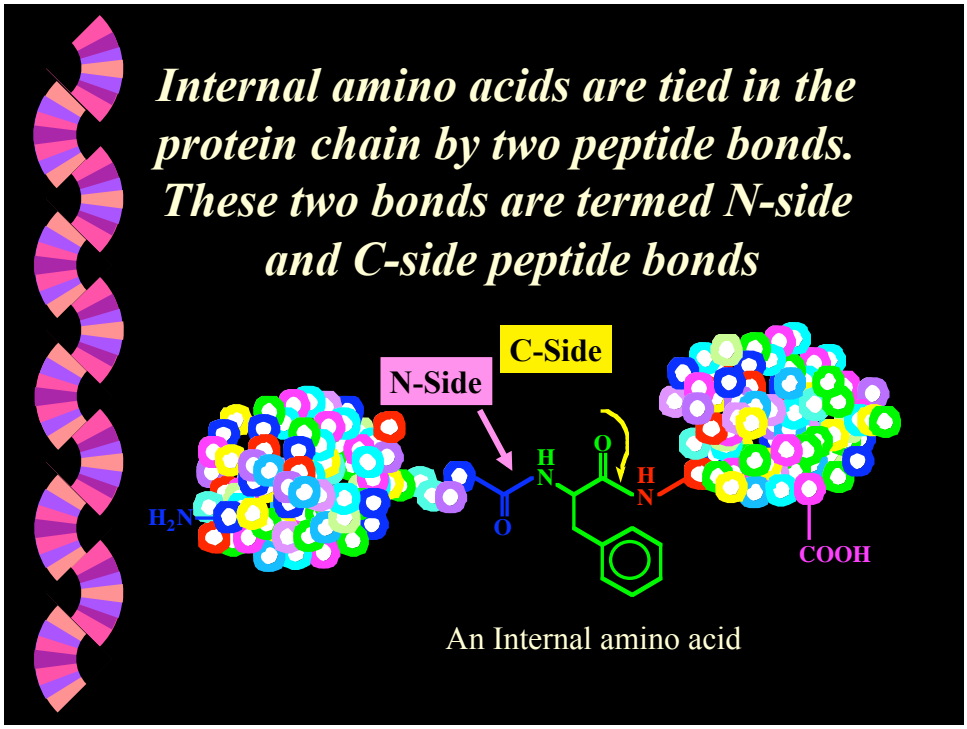
Carboxypeptidase does not stop after cleaving the first C-terminal amino acid.

- **Carboxypeptidase does not stop after cleaving the first C-terminal amino acid. It continues to cleave the next amino acid and then the next. Therefore, to assess which amino acid is released first, scientist quantify the amino acids liberated at different time intervals and plot the data. From the data one can read the C-terminal amino acid sequence.**

C-terminal sequencing by carboxypeptidases



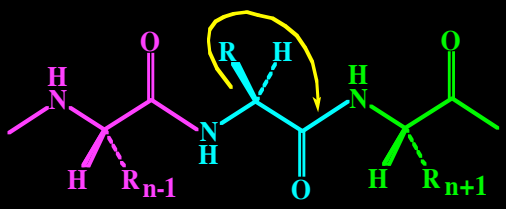

Hydrazinolysis could identify the C-terminal amino acid
peptide
anhydrous H_2N-NH_2
C-Terminal amino acid
All amino acids are released as hydrazides.
Only the C-terminal comes out as free amino acid.


Internal amino acids are tied in the protein chain by two peptide bonds. These two bonds are termed N-side and C-side peptide bonds
C-Side
N-Side
An Internal amino acid

Reagents used for protein cleavage and their specificity

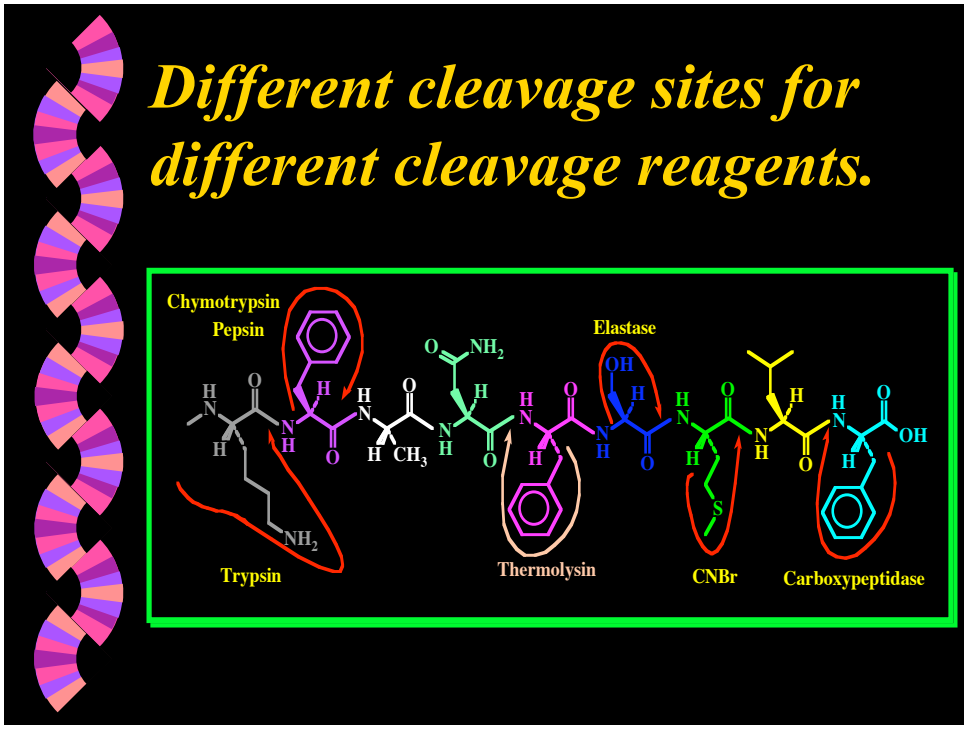
Reagent	Site	Specificity	Comment
CNBr	C side of R_n	R_n = Met	Highly specific
Anhydrous CNBr		R_n = Met, Trp	
Trypsin	C side of R_n	R_n = Lys, Arg, AECys; $R_{n+1} \neq$ Pro	Highly specific
Chymotrypsin	C side of R_n	R_n = Phe, Trp, Tyr, Leu; $R_{n+1} \neq$ Pro	Met & Asn some times
Thermolysin	N side of R_n	R_n = Leu, Ile, Phe, Trp Tyr, Val; $R_{n-1} \neq$ Pro	Some times Ala
Pepsin	N side of R_n	R_n = Leu, Asp, Glu, Phe,Tyr,Trp; $R_{n-1} \neq$ Pro	Non-specific
Acid Proteases (Phosphatase?)	C side of R_n	R_n = Asp, Glu; $R_{n-1} \neq$ Pro	Specific

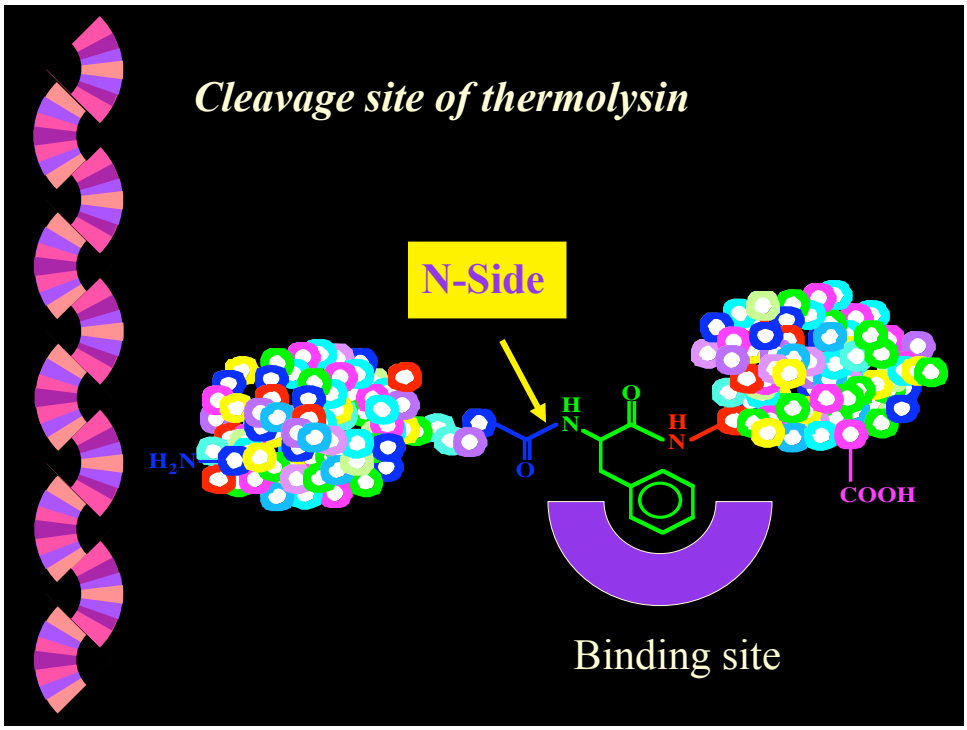
Cleavage sites for some enzymes

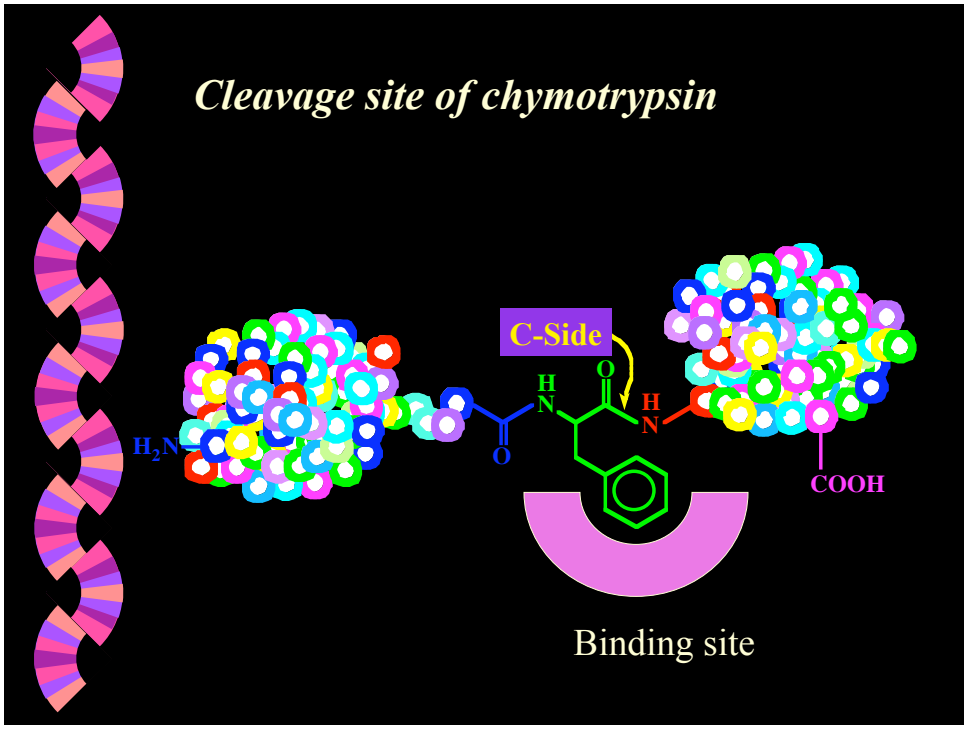


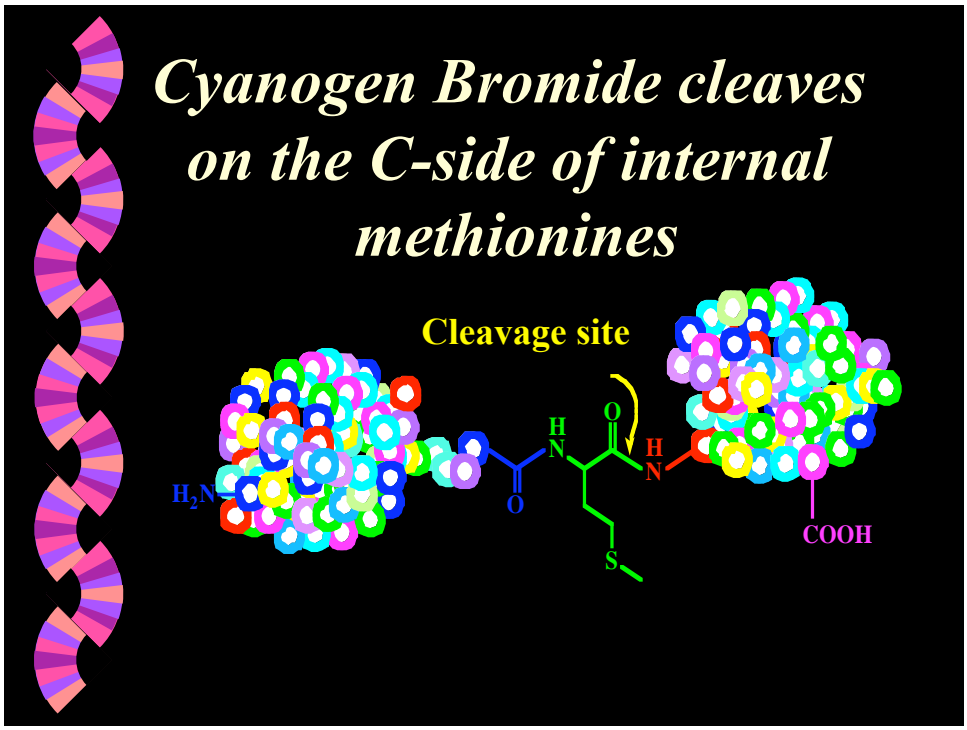
ENZYME	CLEAVAGE SITES (R =)
Trypsin	Lys, Arg
Thrombin	Arg
Papain	Arg, Lys, Phe
Pepsin	Phe, Leu and several others
Bromelain	Lys, Ala, Tyr, Gly
Chymotrypsin	Phe, Tyr, Trp, Leu, Ile, Val
Subtilisin	nonspecific

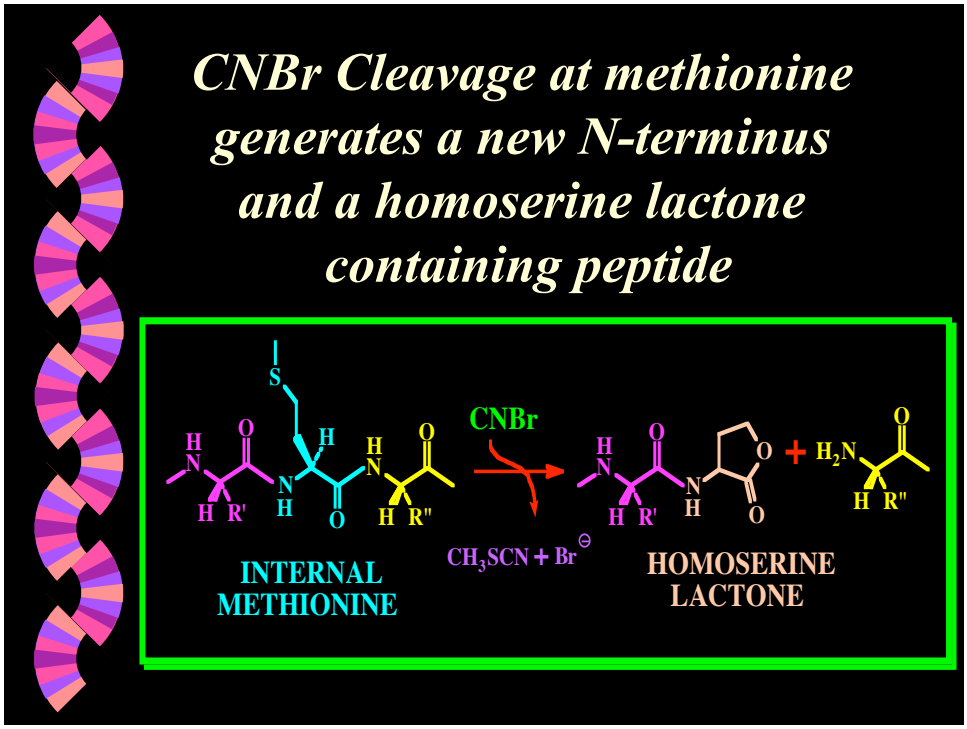
Thermolysin
same as chymotrypsin but on the N-side

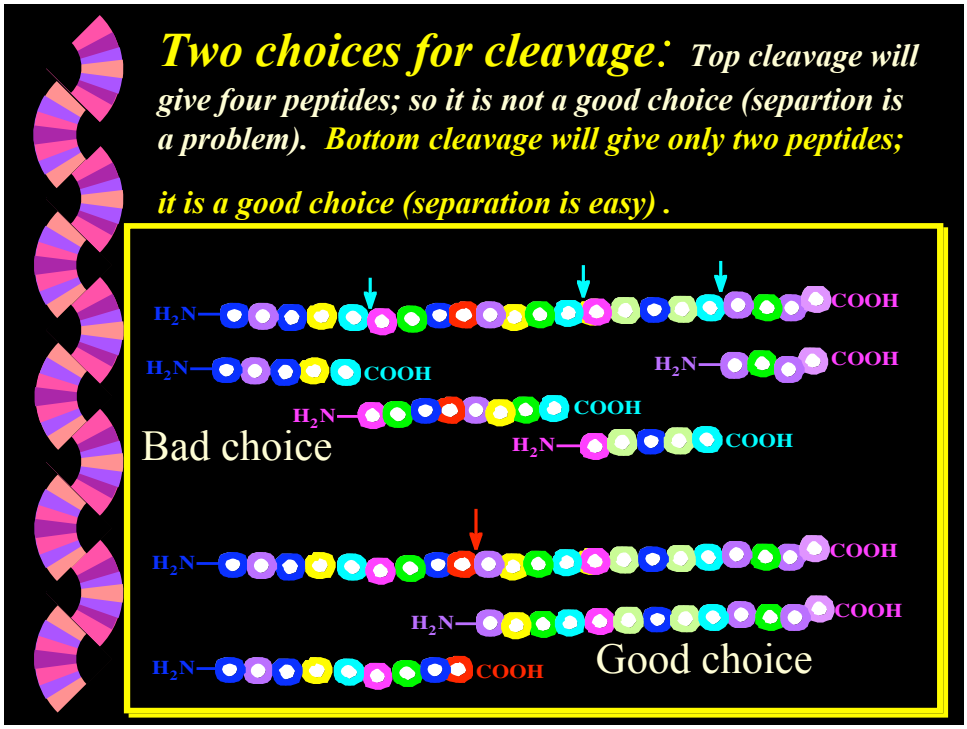

Different cleavage sites for different cleavage reagents.
Chymotrypsin
Pepsin
Elastase
Trypsin
Thermolysin
CNBr
Carboxypeptidase

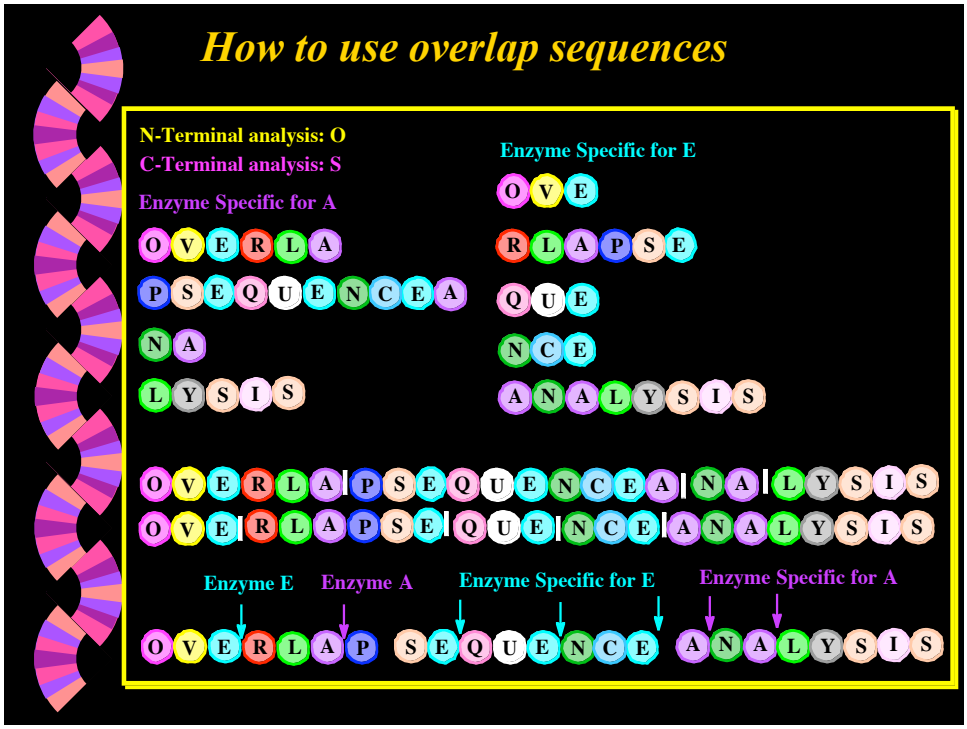

Cleavage site of thermolysin
N-Side
COOH
Binding site

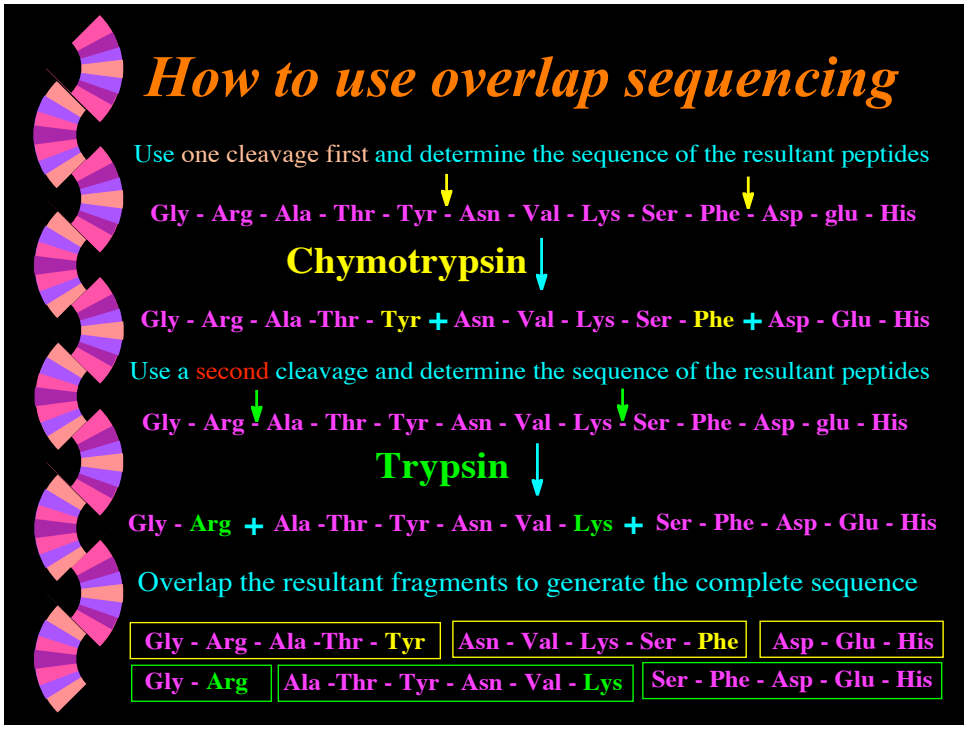

Cleavage site of chymotrypsin
C-Side
H2N
COOH
Binding site


Cyanogen Bromide cleaves on the C-side of internal methionines
Cleavage site
H2N
COOH

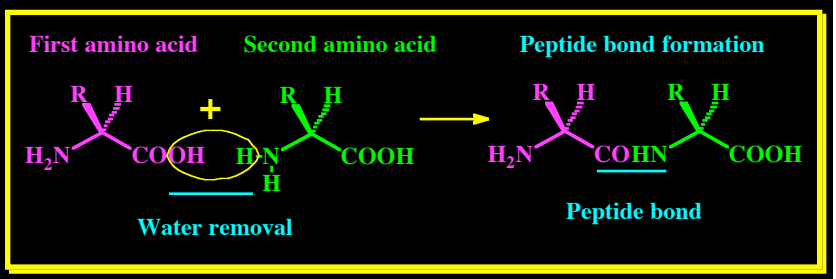

CNBr Cleavage at methionine generates a new N-terminus and a homoserine lactone containing peptide
CNBr
CH3SCN + Br⊖
INTERNAL METHIONINE
HOMOSERINE LACTONE


Two choices for cleavage: Top cleavage will give four peptides; so it is not a good choice (separtion is a problem). Bottom cleavage will give only two peptides; it is a good choice (separation is easy) .
Bad choice
Good choice

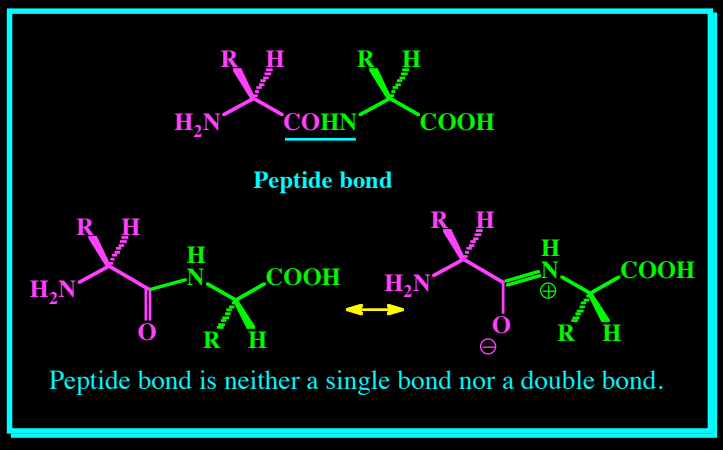

How to use overlap sequences
N-Terminal analysis: O
C-Terminal analysis: S
Enzyme Specific for A
OVERLA
PSEQUENCEA
NA
LYSIS
Enzyme Specific for E
OVE
RLAPSE
QUE
NCE
ANALYSIS
OVERLAPSEQUENCEANALYSIS
OVERLAPSEQUENCEANALYSIS
Enzyme E
Enzyme A
Enzyme Specific for E
Enzyme Specific for A
OVERLAP
SEQUENCE
ANALYSIS


How to use overlap sequencing
Use one cleavage first and determine the sequence of the resultant peptides
Gly - Arg - Ala - Thr - Tyr - Asn - Val - Lys - Ser - Phe - Asp - glu - His
Chymotrypsin
Gly - Arg - Ala -Thr - Tyr + Asn - Val - Lys - Ser - Phe + Asp - Glu - His
Use a second cleavage and determine the sequence of the resultant peptides
Gly - Arg - Ala - Thr - Tyr - Asn - Val - Lys - Ser - Phe - Asp - glu - His
Trypsin
Gly - Arg + Ala -Thr - Tyr - Asn - Val - Lys + Ser - Phe - Asp - Glu - His
Overlap the resultant fragments to generate the complete sequence
Gly - Arg - Ala -Thr - Tyr
Asn - Val - Lys - Ser - Phe
Asp - Glu - His
Gly - Arg
Ala -Thr - Tyr - Asn - Val - Lys
Ser - Phe - Asp - Glu - His

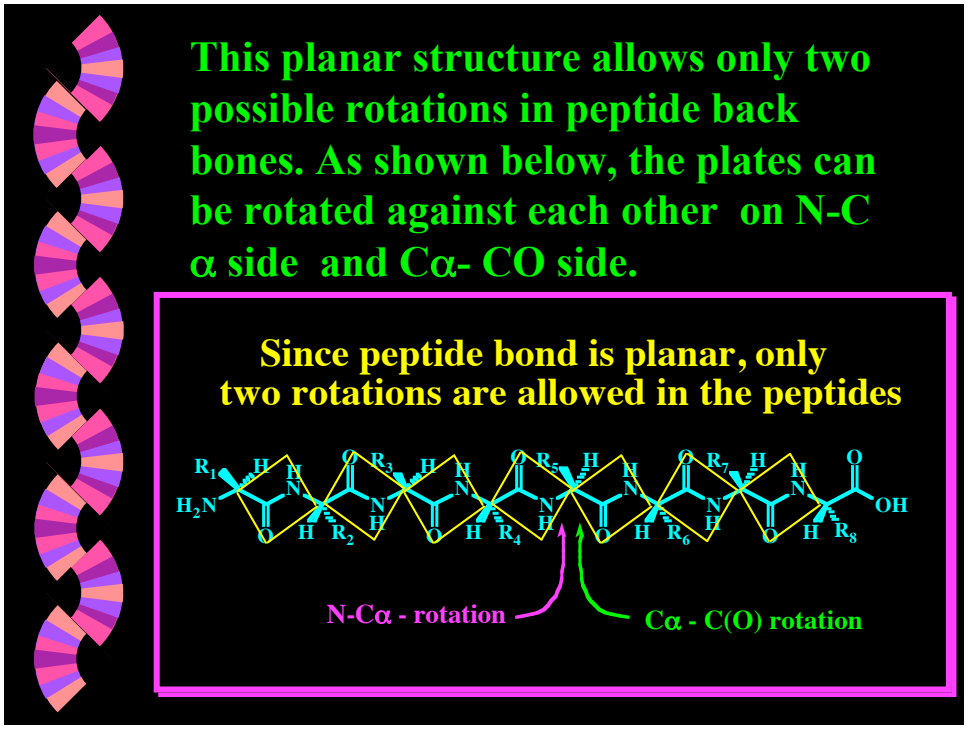
Peptide bond formation involves the removal of water between the carboxyl group of first amino acid and the amino group of second amino acid.

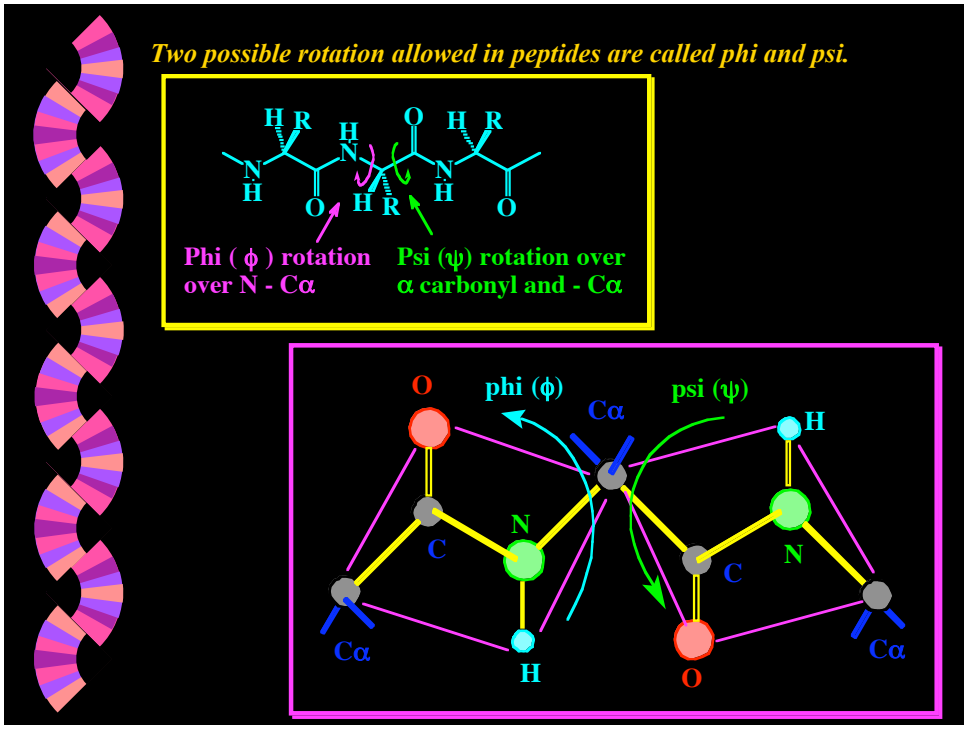


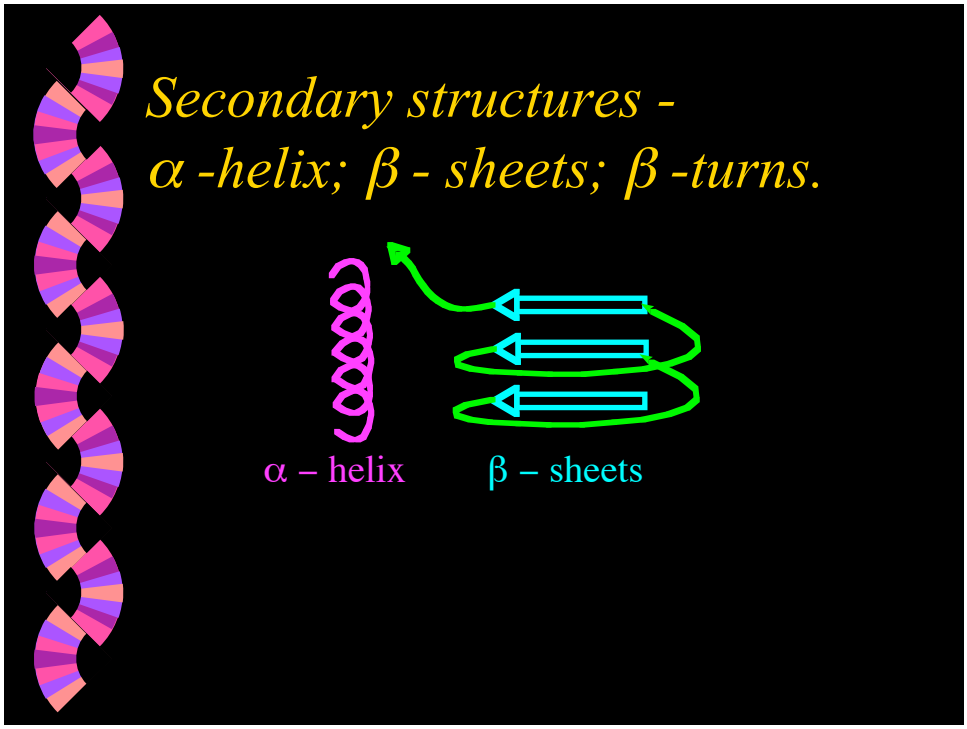
Peptide bond is planar

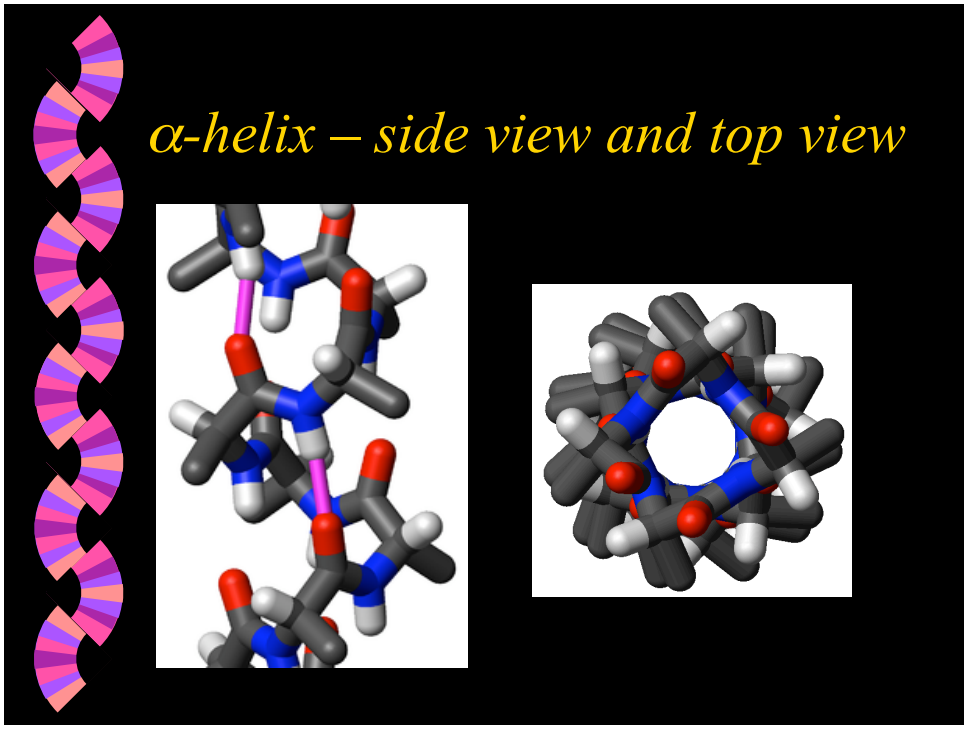


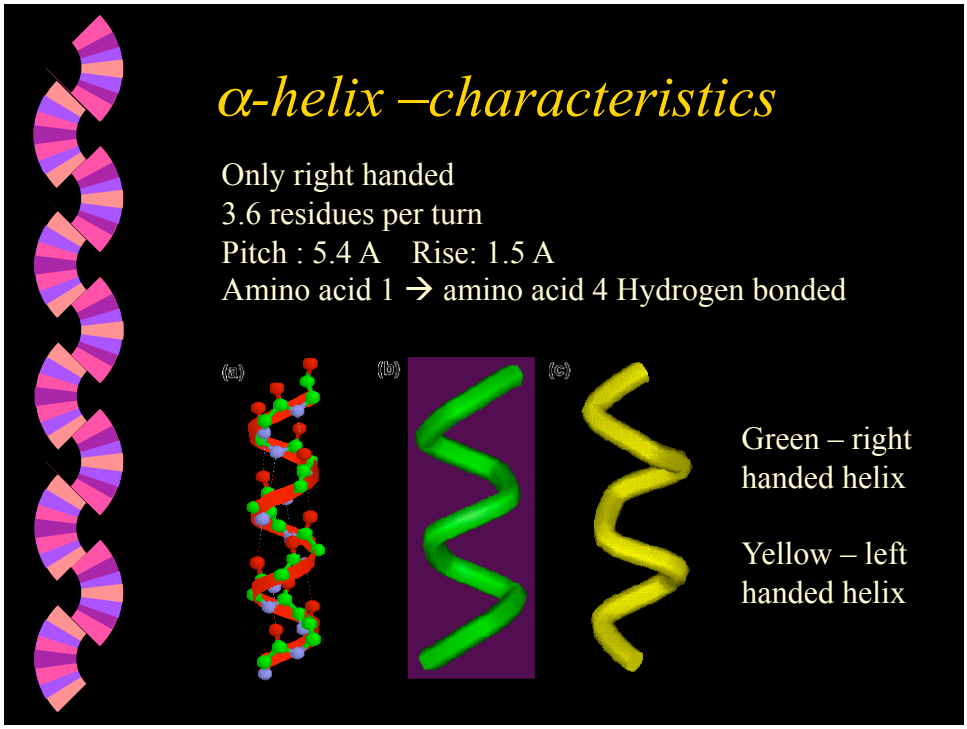
This makes the CO and NH to be in the same plane. So peptide bond is planar. In the plane, Cα carbon atom of first amino acid, CO group of the first amino acid, NH group of second amino acid and the Cα atom of the second amino acid (all together six atoms) are present.

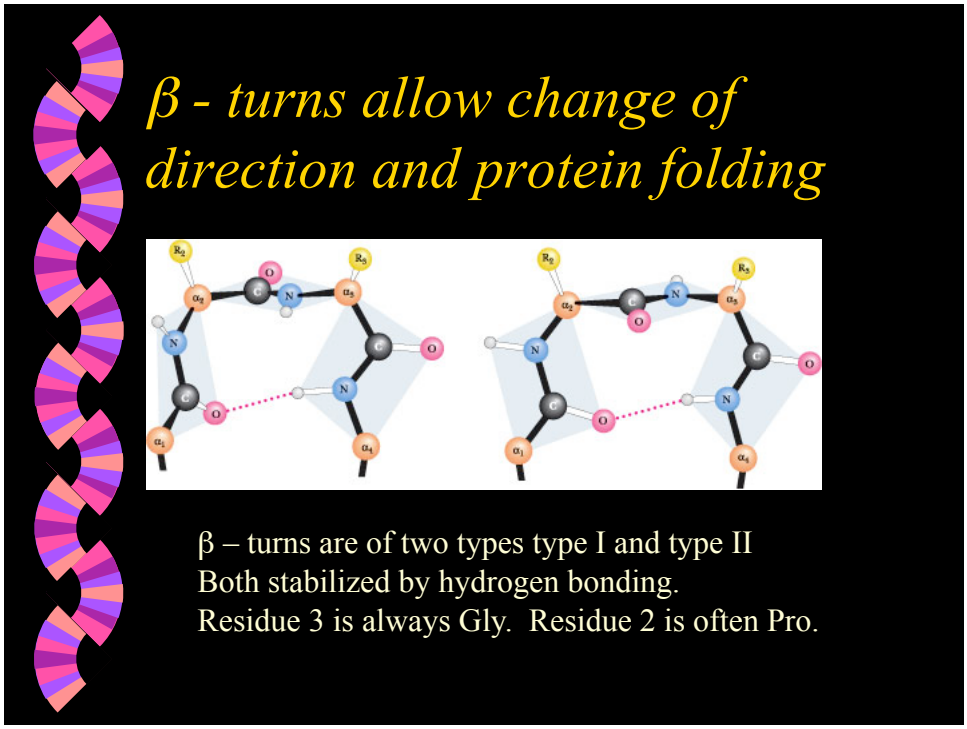

This planar structure allows only two possible rotations in peptide back bones. As shown below, the plates can be rotated against each other on N-C α side and Cα- CO side.
Since peptide bond is planar, only two rotations are allowed in the peptides
N-Cα - rotation
Cα - C(O) rotation

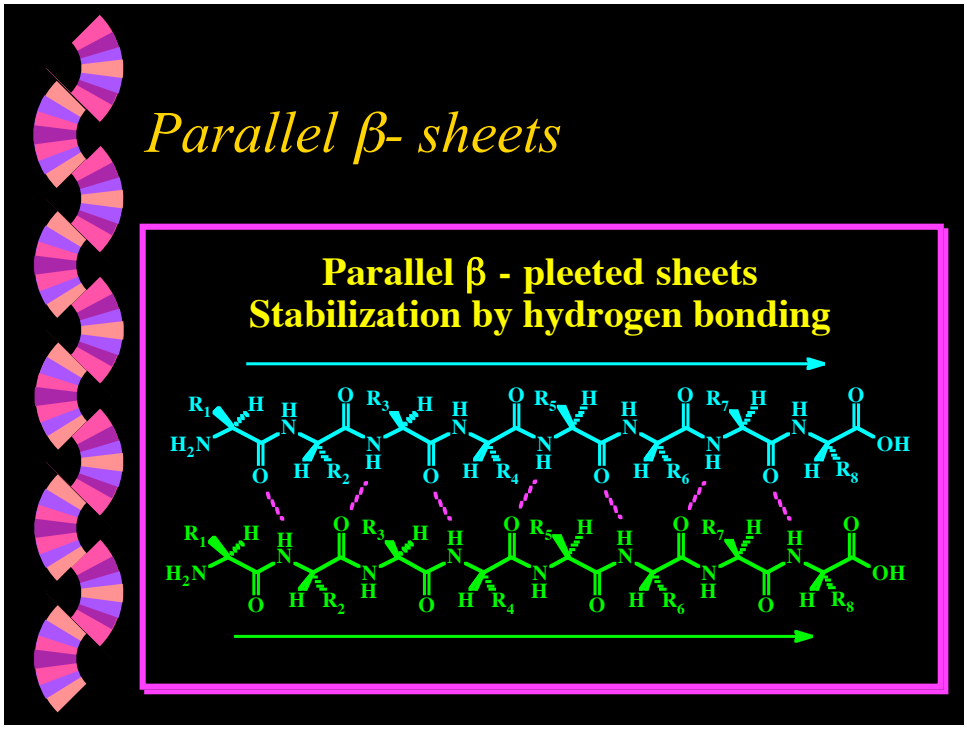

Two possible rotation allowed in peptides are called phi and psi.
Phi (ϕ) rotation over N - Cα
Psi (ψ) rotation over α carbonyl and - Cα
phi (ϕ)
psi (ψ)
Cα
N
C
H
O

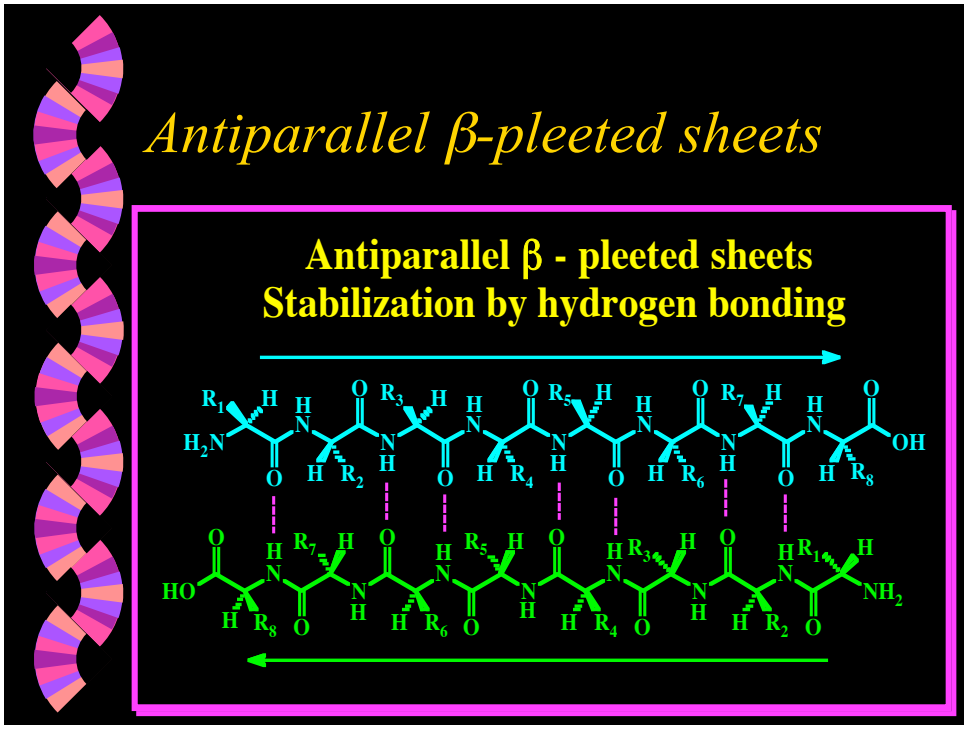

Secondary structures -
α -helix; β - sheets; β -turns.
α – helix
β – sheets

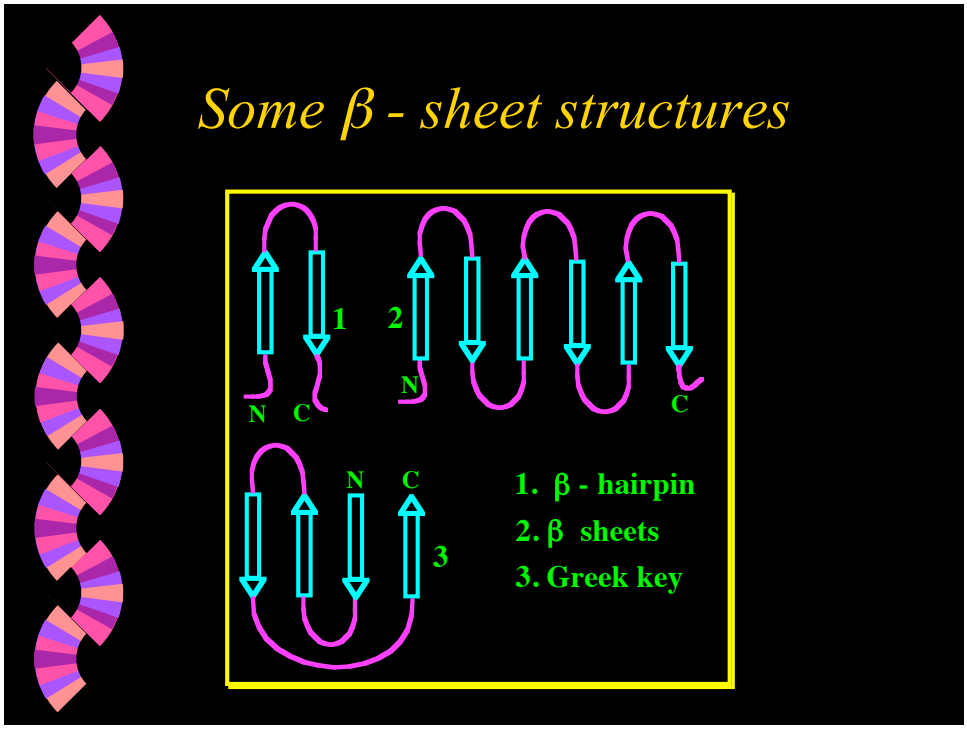

α-helix – side view and top view

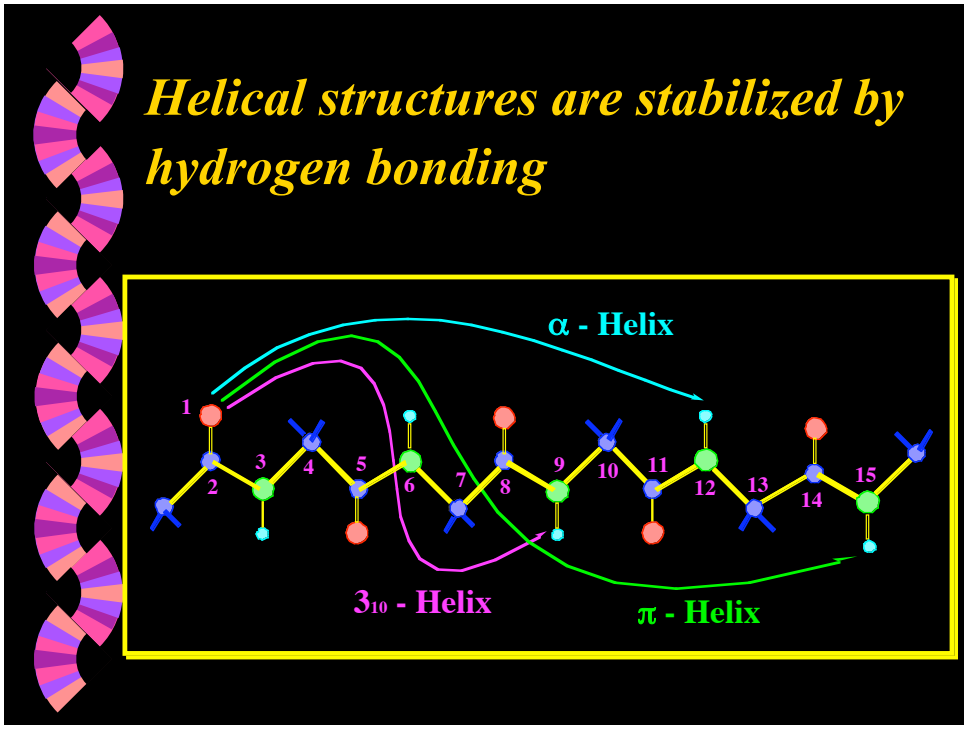

α-helix –characteristics
Only right handed
3.6 residues per turn
Pitch : 5.4 A Rise: 1.5 A
Amino acid 1 → amino acid 4 Hydrogen bonded
(a)
(b)
(c)
Green – right handed helix
Yellow – left handed helix

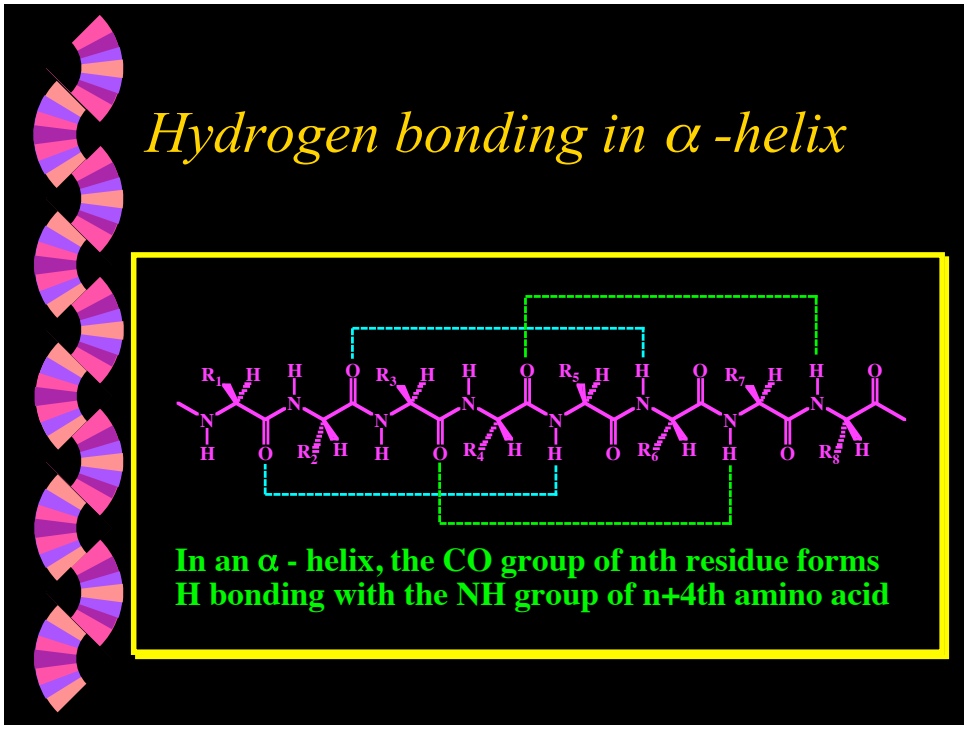

β - turns allow change of direction and protein folding
β – turns are of two types type I and type II
Both stabilized by hydrogen bonding.
Residue 3 is always Gly. Residue 2 is often Pro.

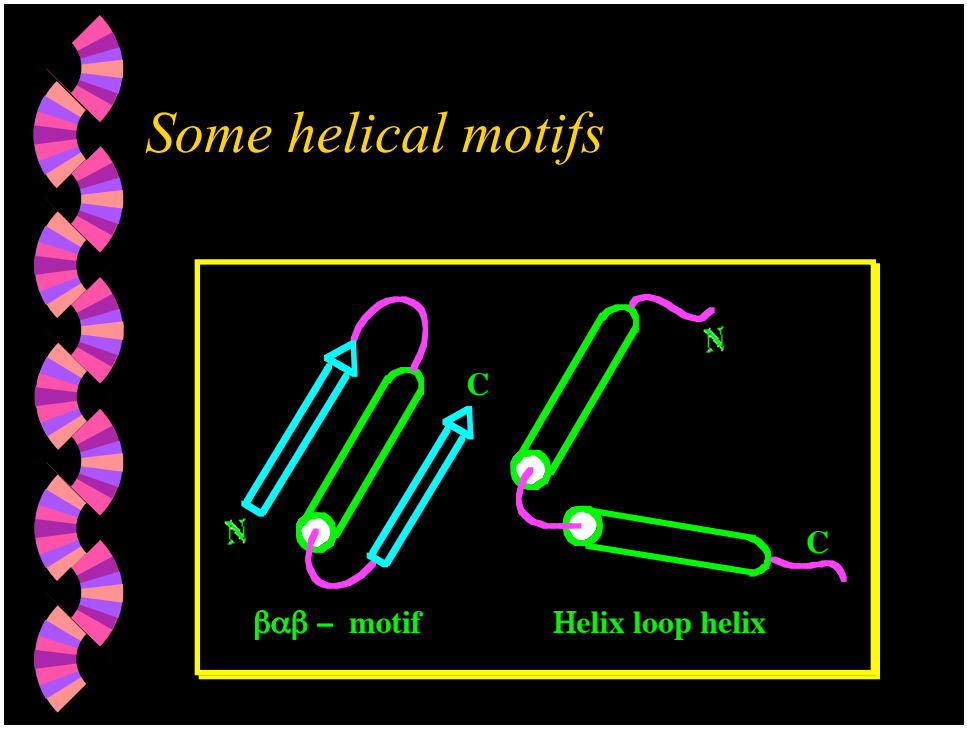

Parallel β- sheets
Parallel β - pleeted sheets
Stabilization by hydrogen bonding


Antiparallel β-pleeted sheets
Antiparallel β - pleeted sheets
Stabilization by hydrogen bonding


Some β - sheet structures
1
2
N
C
N
C
N
C
3
1. β - hairpin
2. β sheets
3. Greek key

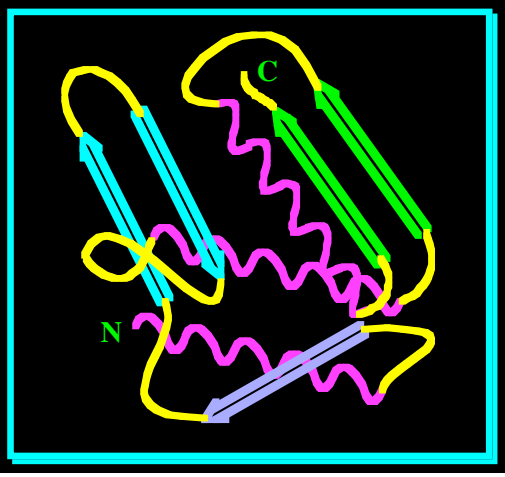

Helical structures are stabilized by hydrogen bonding
α - Helix
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
3₁₀ - Helix
π - Helix


Hydrogen bonding in α -helix
In an α - helix, the CO group of nth residue forms H bonding with the NH group of n+4th amino acid

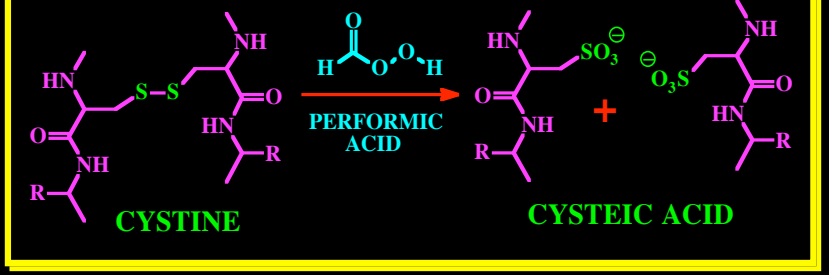

Some helical motifs
C
N
βαβ – motif
N
C
Helix loop helix

Protein three dimensional structure is made up of a combination of secondary structures. In the following structure note the presence of α -helices, β - sheets and random coils.

Helical structures in pink. Arrows are β - sheets Random coil is in yellow



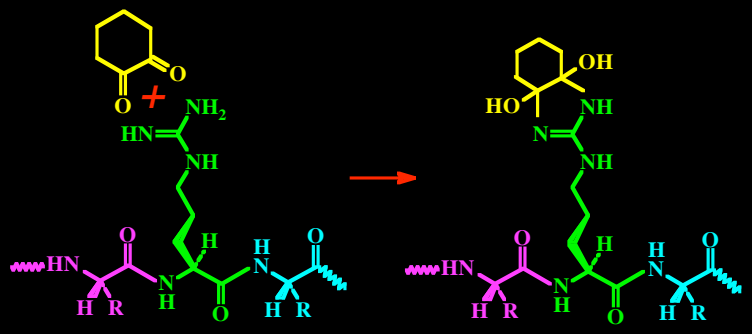
Cystine is estimated as cysteic acid after performic acid oxidation



Oxidation of disulfide linked protein with performic acid generates cysteic acid. After protein hydrolysis, cysteic acid can be quantified using an amino acid analyzer.

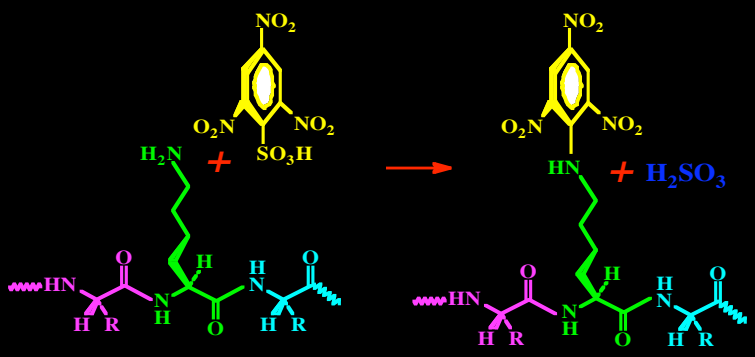
Chemical modification of Arginine

Reversible blocking of Arginine by cyclohexadione.
E. L. Smith. *Methods in Enzymology* 47, 156-161 (1977).



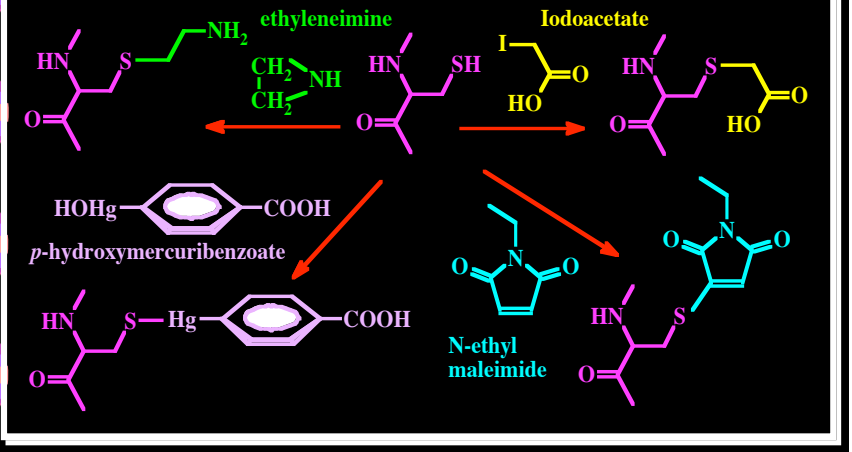
Arginine is modified by 1,2-cyclohexadione (or phenylglyoxal). The vicinyl dihydroxy adduct can be stabilized by borate.

Chemical modification of Lysine

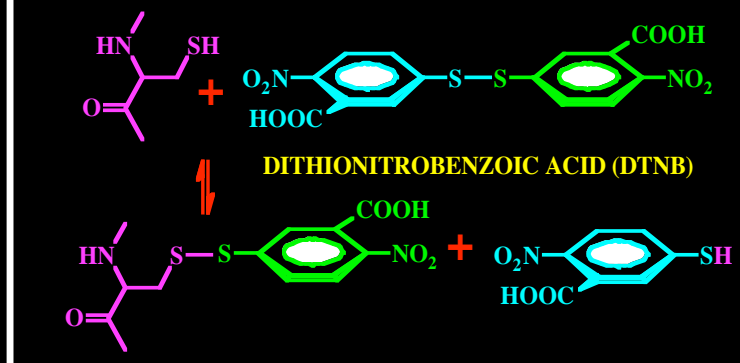


TNBS (2,4,6-trinitrobenzenesulfonic acid arylates lysines. The product can be quantified at 367 nm.

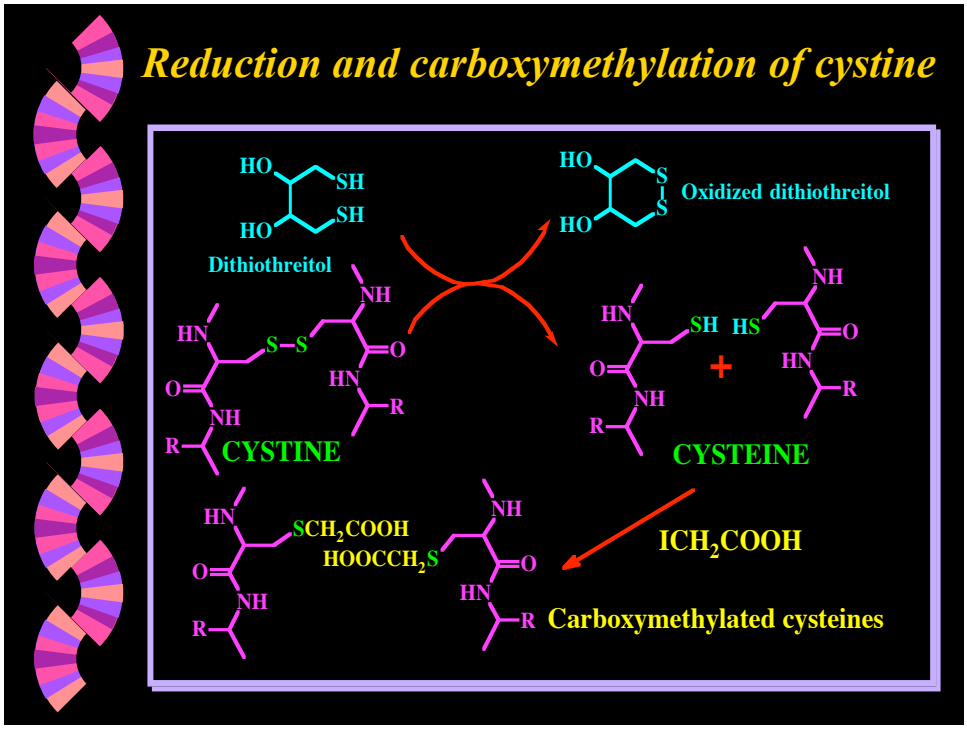
Chemical modification of Cysteine

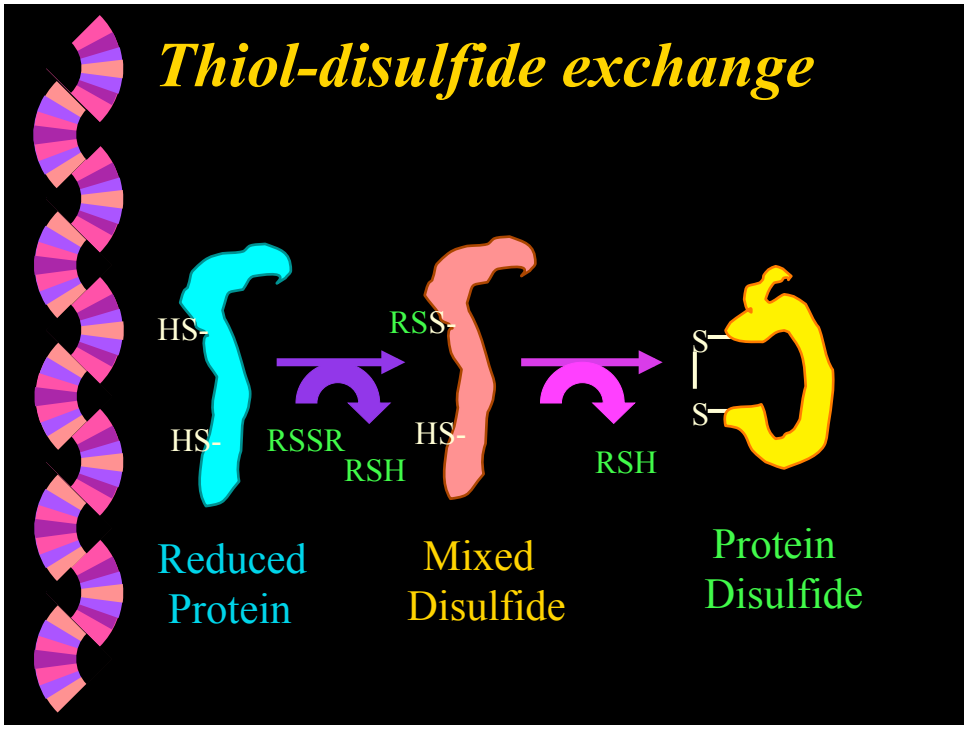


Quantitation of Cysteine in Proteins



Addition of excess DTNB leads to quantitative derivatization of cysteine and stoichiometric liberation of the yellow colored aromatic nitrothiol.


Reduction and carboxymethylation of cystine
Dithiothreitol
Oxidized dithiothreitol
CYSTINE
CYSTEINE
ICH2COOH
Carboxymethylated cysteines


Thiol-disulfide exchange
HS-
HS-
RSSR
RSH
RSS-
HS-
RSH
Reduced Protein
Mixed Disulfide
Protein Disulfide