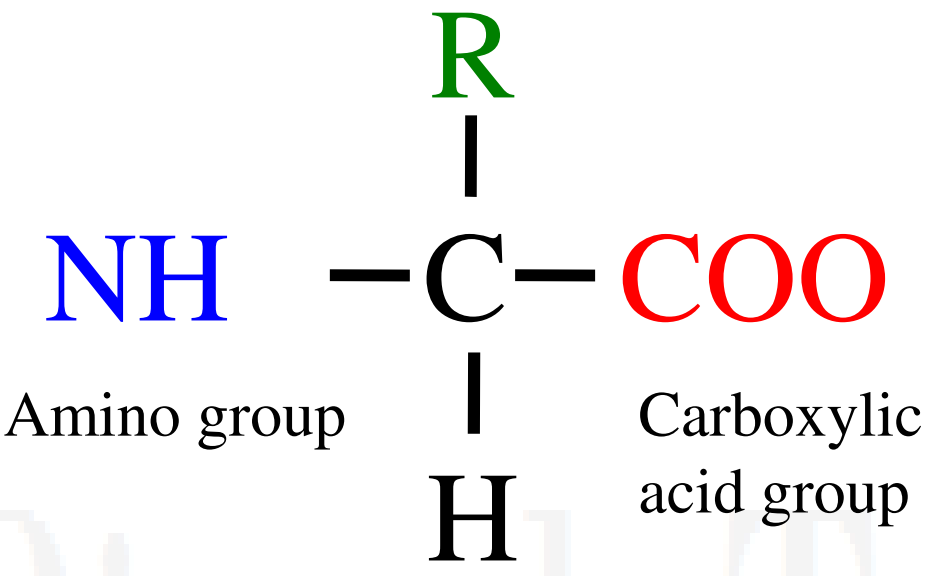
PROTEIN CHEMISTRY

Dr PIYUSH B. TAILOR
Associate Professor
Depart. Of Biochemistry
Govt. Medical College
Surat.

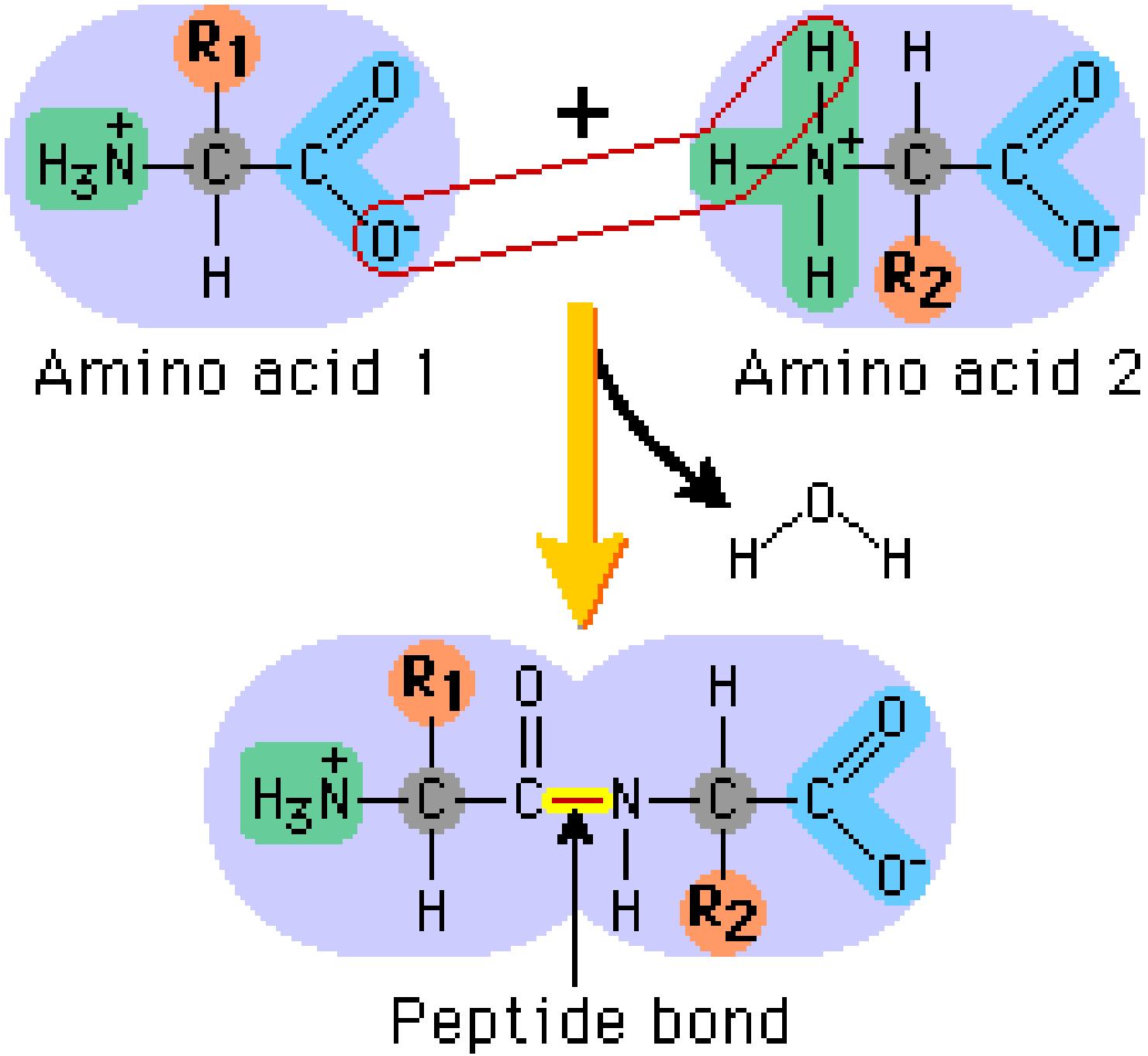
INTRODUCTION

- Abnormality in protein structure will lead to major diseases with profound alterations in metabolic functions.
- Polymers of amino acids
- Linked by peptid bond

Amino acid: Basic unit of protein



Different side chains, R, determin the properties of 20 amino acids.


R1
H3N+
C
C
O
O-
H
+
H
H
H
N+
H
C
C
O
O-
H
R2
Amino acid 1
Amino acid 2
O
H
H
R1
O
H
H3N+
C
C
N
C
C
O
O-
H
H
R2
Peptide bond

- Two amino acids = Dipeptide
- Three amino acids = Tripeptide
- Four amino acids = Tetrapeptide.
- < 10 amino acids together = Oligopeptide
- 10 -50 amino acids = Polypeptide.
- Even though there are 20 amino acids, by changing the sequence of combination of these amino acids, nature produces enormous number of markedly different proteins.

Examples on Peptides:

1- Dipeptide

Example: **Aspartame** acts as sweetening agent
= aspartic acid + phenyl alanine.

2- Tripeptide

Example: **GSH** – **Reduce Glutathione**
= Glutamic acid + Cysteine + Glycine.
= Helps in absorption of amino acids
= Protects against hemolysis of RBC

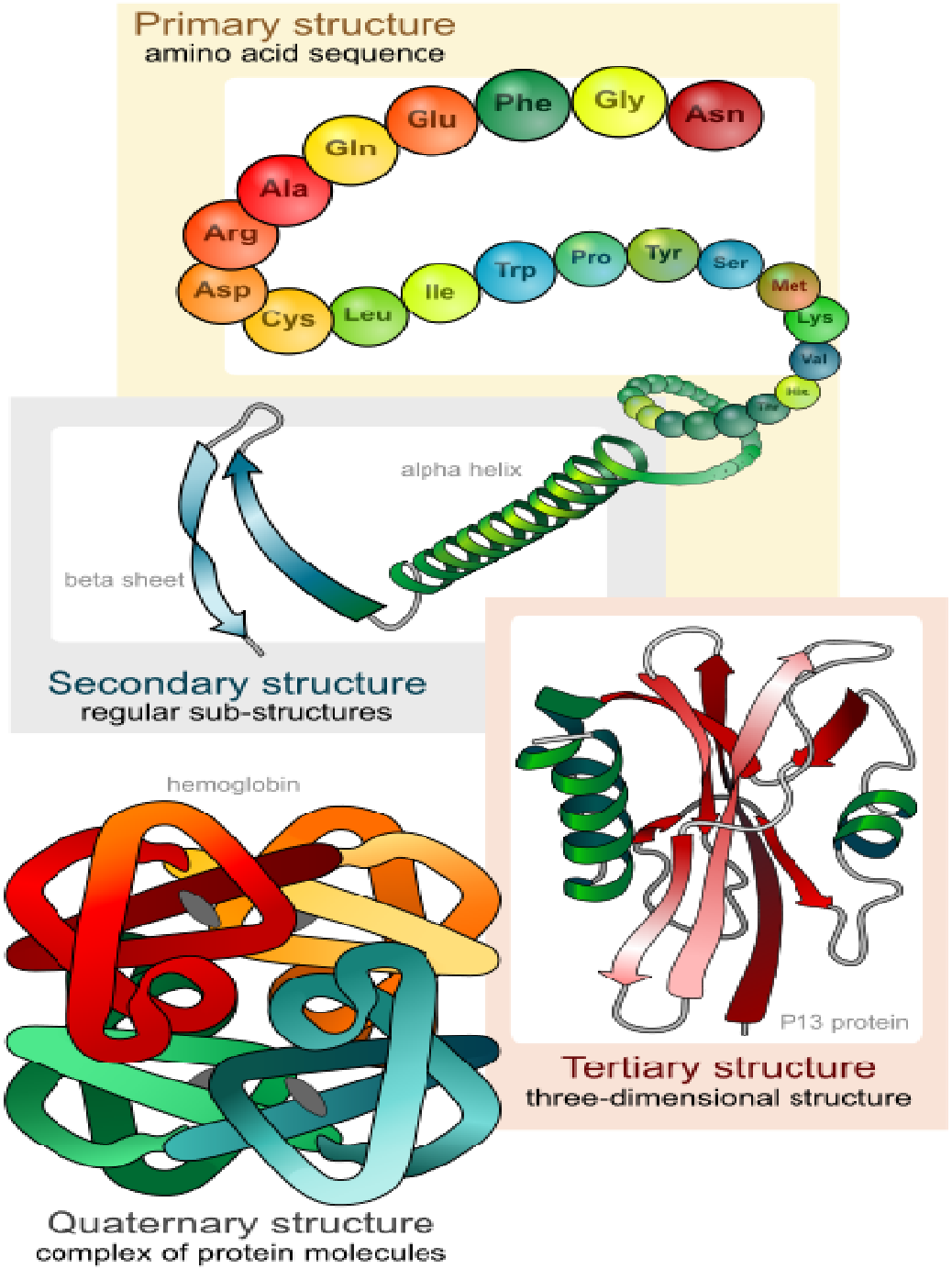
3- octapeptides: (8 amino acids)

Examples: Oxytocine and Vasopressin (ADH).

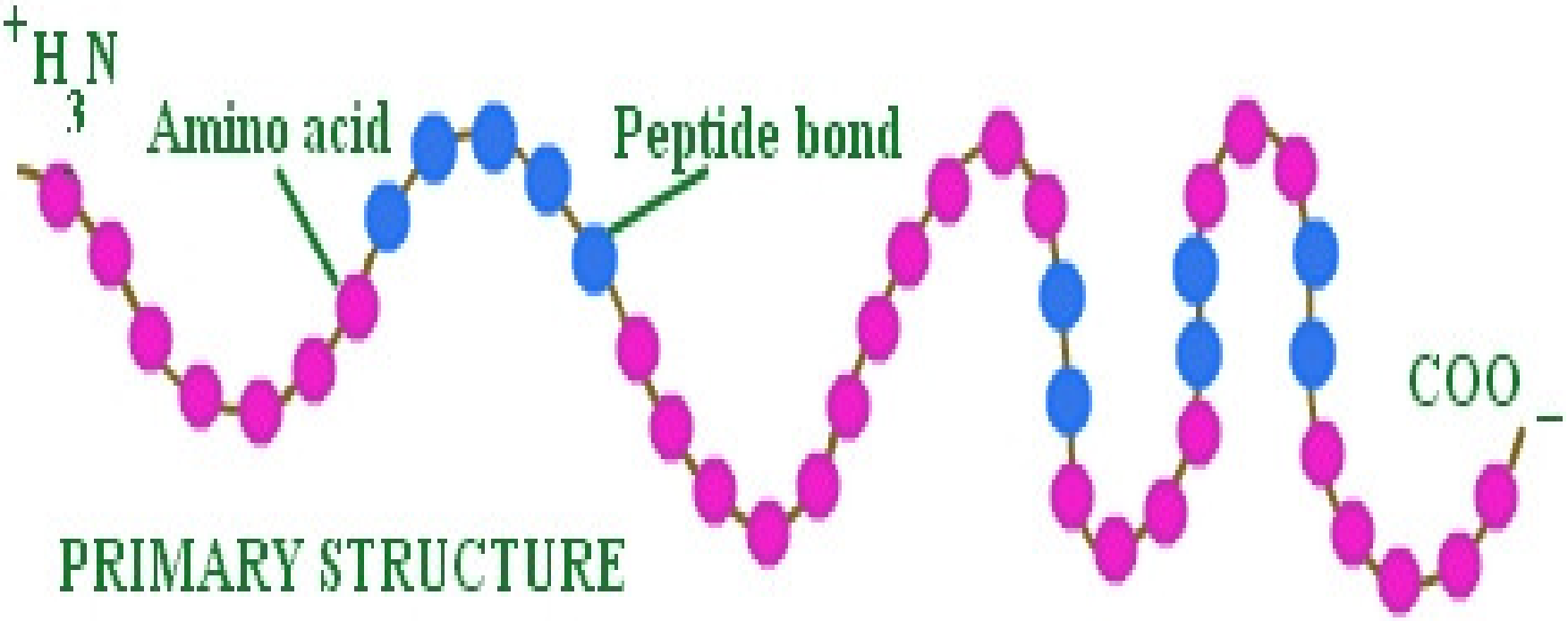
4- polypeptides: more than 10 amino acids: e.g. Insulin hormone

Type of Protein Structure

1. Primary
2. Secondary
3. Tertiary
4. Quaternary


Primary structure
amino acid sequence
Phe
Gly
Asn
Glu
Gln
Ala
Arg
Asp
Cys
Leu
Ile
Trp
Pro
Tyr
Ser
Met
Lys
Val
alpha helix
beta sheet
Secondary structure
regular sub-structures
hemoglobin
Quaternary structure
complex of protein molecules
P13 protein
Tertiary structure
three-dimensional structure

Primary Structure

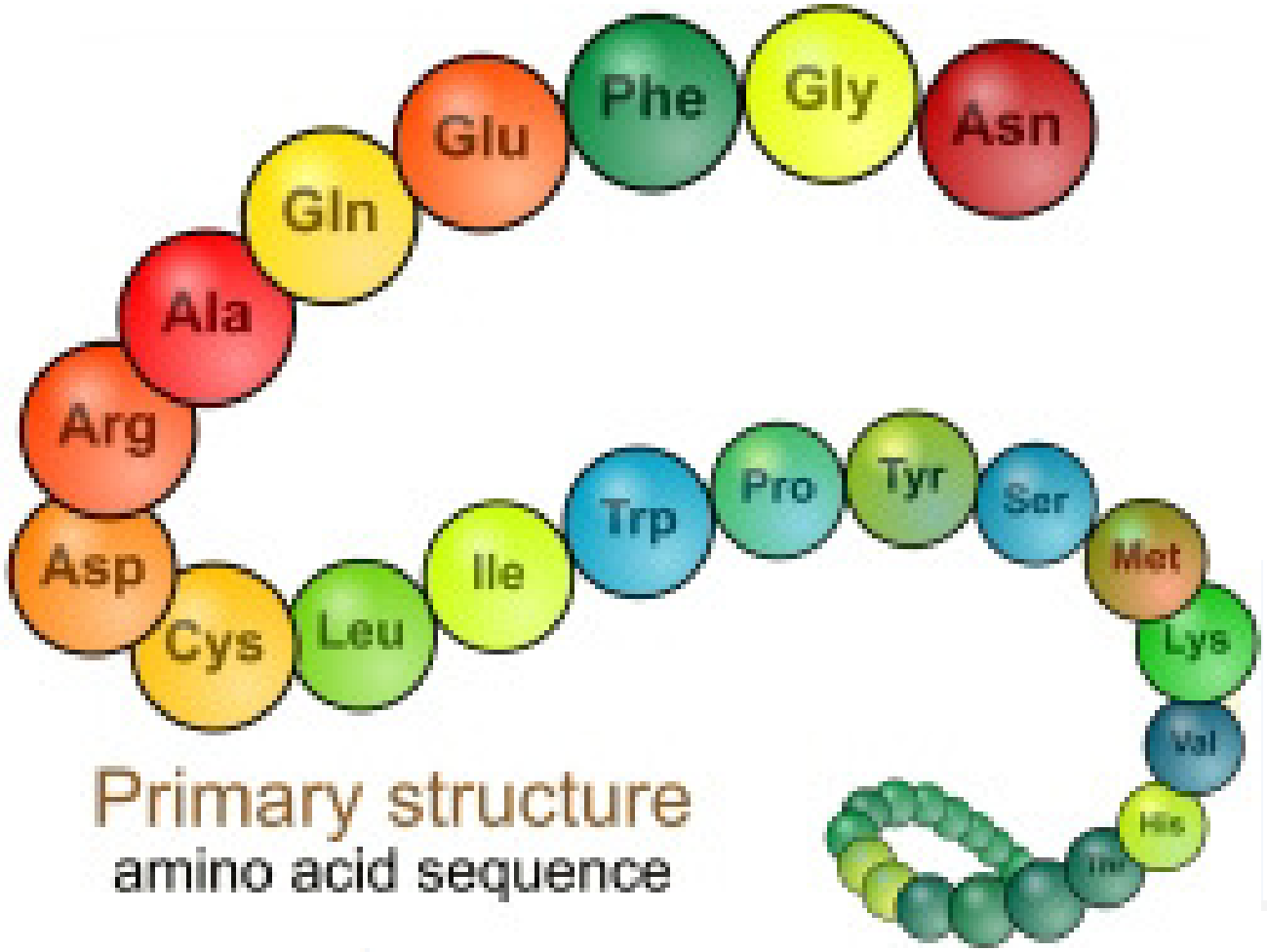


To write Amino acid Sequence of polypeptide chain

- Amino terminal on Left
- Carboxyl terminal on Right

Primary Structure

.



Primary Structure & Functional Relationship

- Example

1. Sickle Cell Disease
2. Pre-pro-insulin to Pro-insulin to Insulin
3. Thallasemia

Sickle Cell Disease

Gene to Protein Sequence

HBB Sequence in Normal Adult Hemoglobin (Hb A):

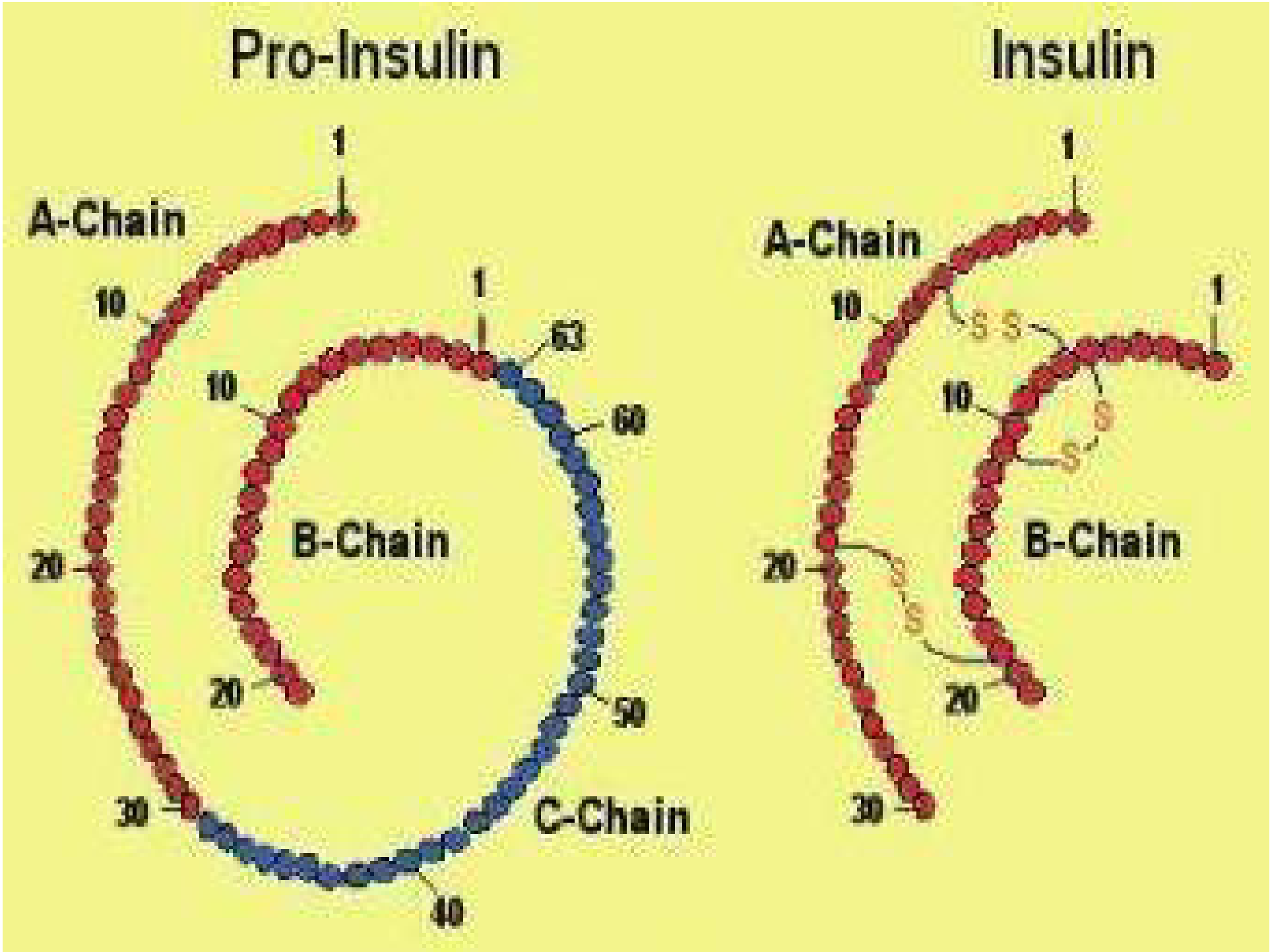
Nucleotide	CTG	ACT	CCT	GAG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Glu	Glu	Lys	Ser
	3			6			9

HBB Sequence in Mutant Adult Hemoglobin (Hb S):

Nucleotide	CTG	ACT	CCT	GTG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Val	Glu	Lys	Ser
	3			6			9

Sickle Cell Disease

- Mutation In Gene for Beta chain of Haemoglobin
  - Adenine (A) is replace by Thymine (T)
  - In genetic codon = GAG converted to GTG
- During protein synthesis (transcription)
  - GAG represent Glutamic acid
  - GTG represent Valine
- Because of mutation, on 6^{th} position in beta chain of haemoglobin, glutamic acid is replace by valine
- Normal Haemoglbin (HbA) is converted to abnormal sickle haemoglobin (HbS) – function get affected
- This is explain that
  - " Sequence of A.A. get change protein function get change"


Pro-Insulin
A-Chain
1
10
20
30
40
B-Chain
1
10
20
C-Chain
1
30
50
60
63
Insulin
A-Chain
1
10
20
B-Chain
1
10
20
S S
S S
S S

Pre-pro-insulin to Pro-insulin to Insulin

- Pre-pro-insulin & pro-Insulin ,both are inactive
- Proinsulin
  - single chain with 84 amino acids
  - Inactive form
- Insulin
  - double chain (A & B chain) with 51 amino acids
  - Active form
- When Proinsulin is converted to insulin
  - Sequence & number of amino acid, both get change
  - So Function of protein get change
  - Protein (insulin) became non-functional to functional

Thalassemia

- Normal HbA = Alpha (141 A.A.) & Beta (146 A.A.) Chain
- Mutation
  - Non sense type of mutation (one of the etiology)
  - Premature termination of haemoglobin chain synthesis
  - Alpha or Beta either absent or short
  - Alpha or Beta thalassemia occur
- Number of amino acid get changed
- Functional protein (HbA) become non-functional

Primary structural Functional Relation

- **"Whenever Suquence or / and number of amino acid get change , protein function get change"**

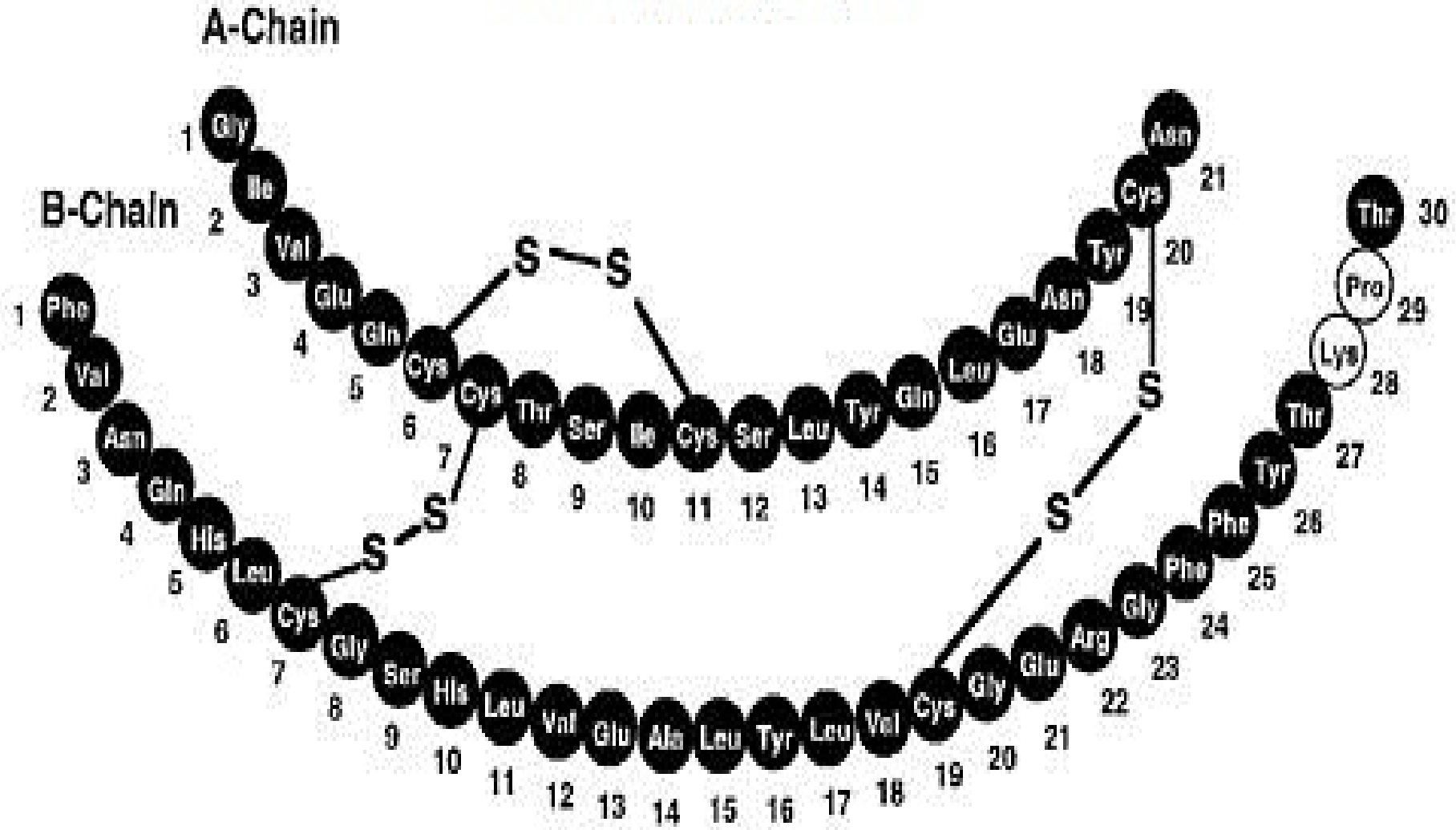
Either

- **Protein became Functional to Non-functional**
  - HbA to Sickle cell disease (sequence of A.A. changed)
  - HbA to Thalassemia (Number of A.A. changed)

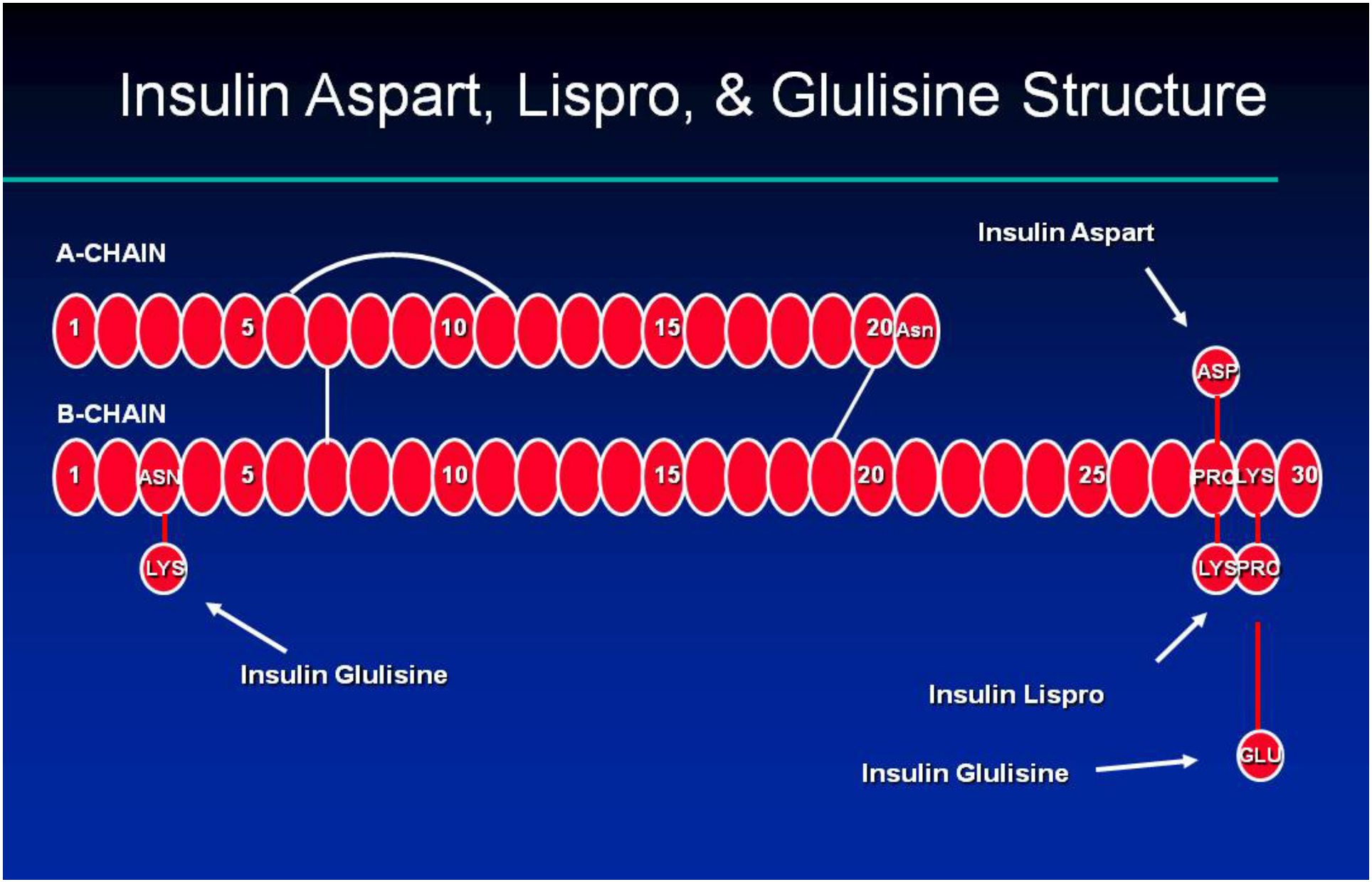
Or

- **Protein became non-functional to functional**
  - Pro-insulin to Insulin (sequence & number of A.A. both changed)

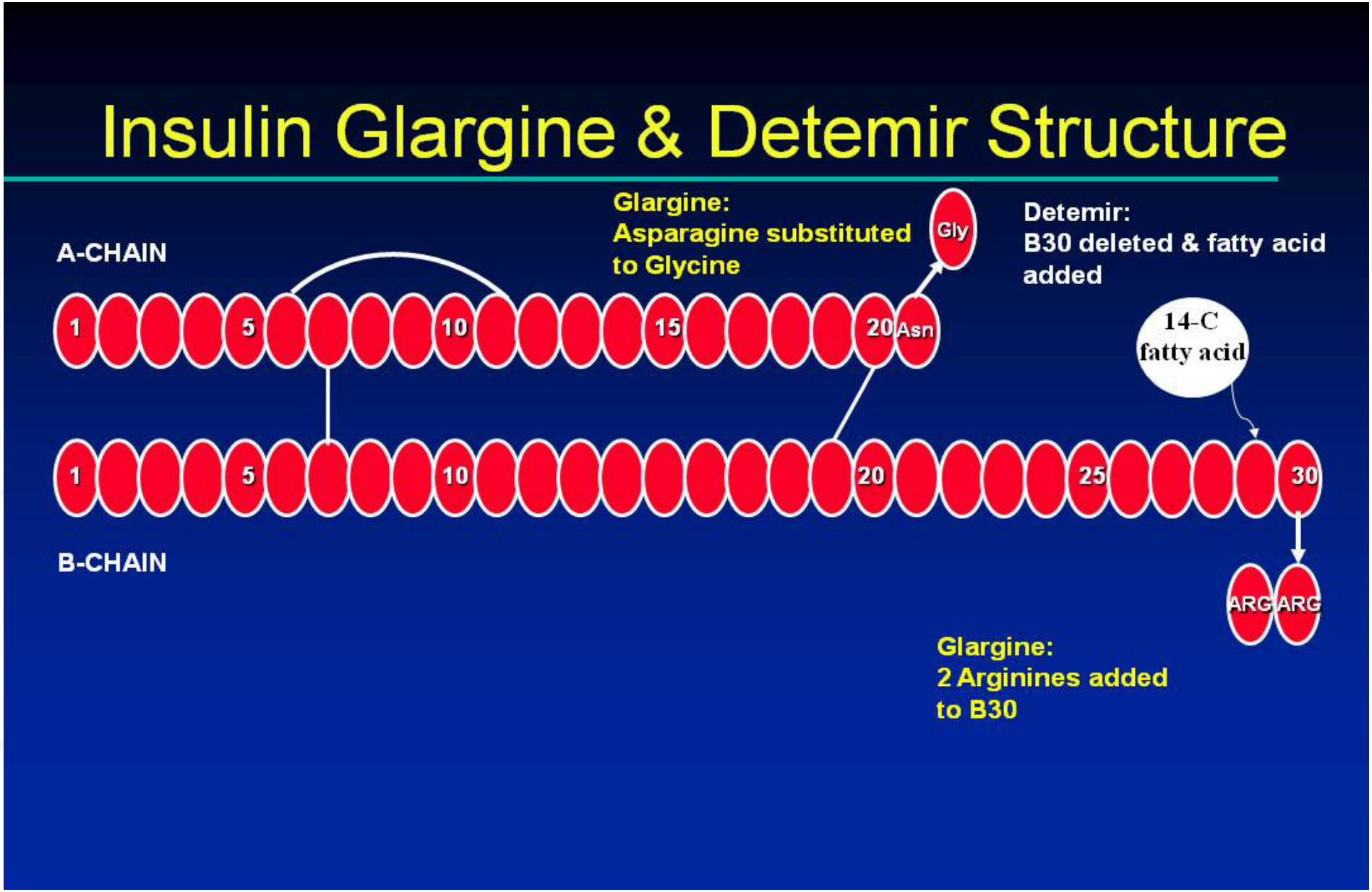
Human Insulin

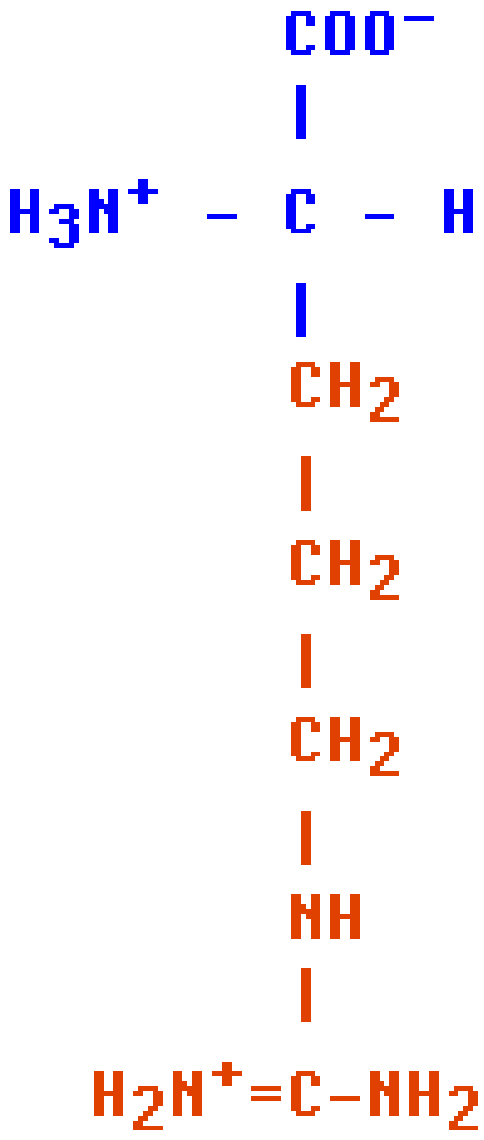


Recombinant Insulin



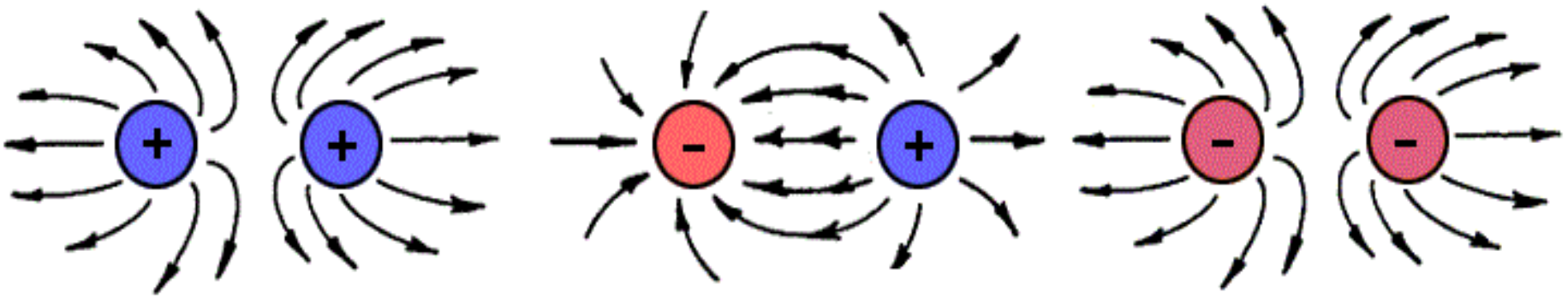
Recombinant Insulin

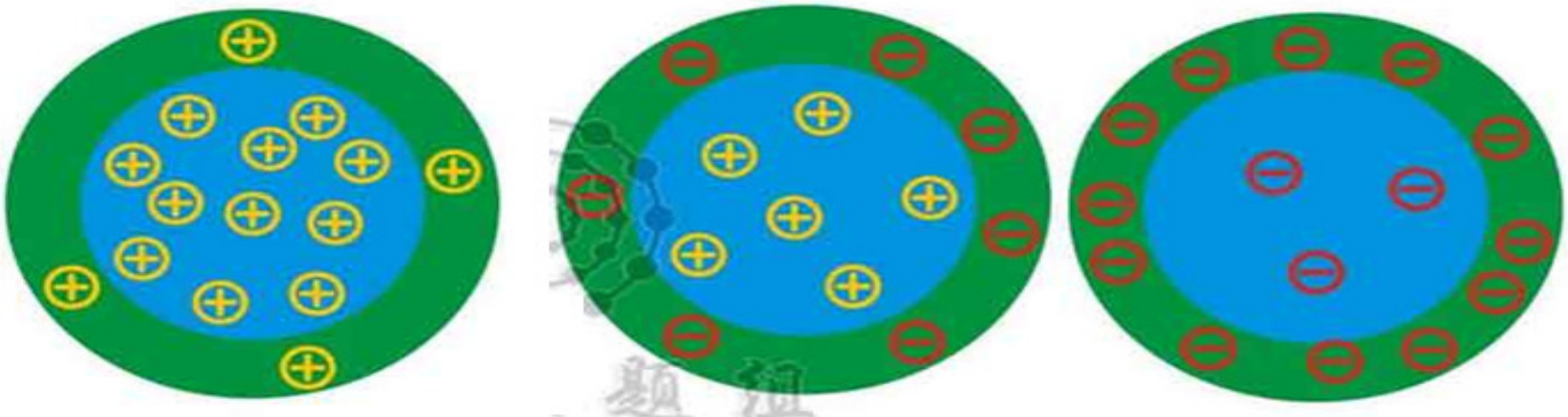




Arginine (Arg)[R]

At What pH
This ion molecule has least solubility?





pH<5 pH=7 pH>10

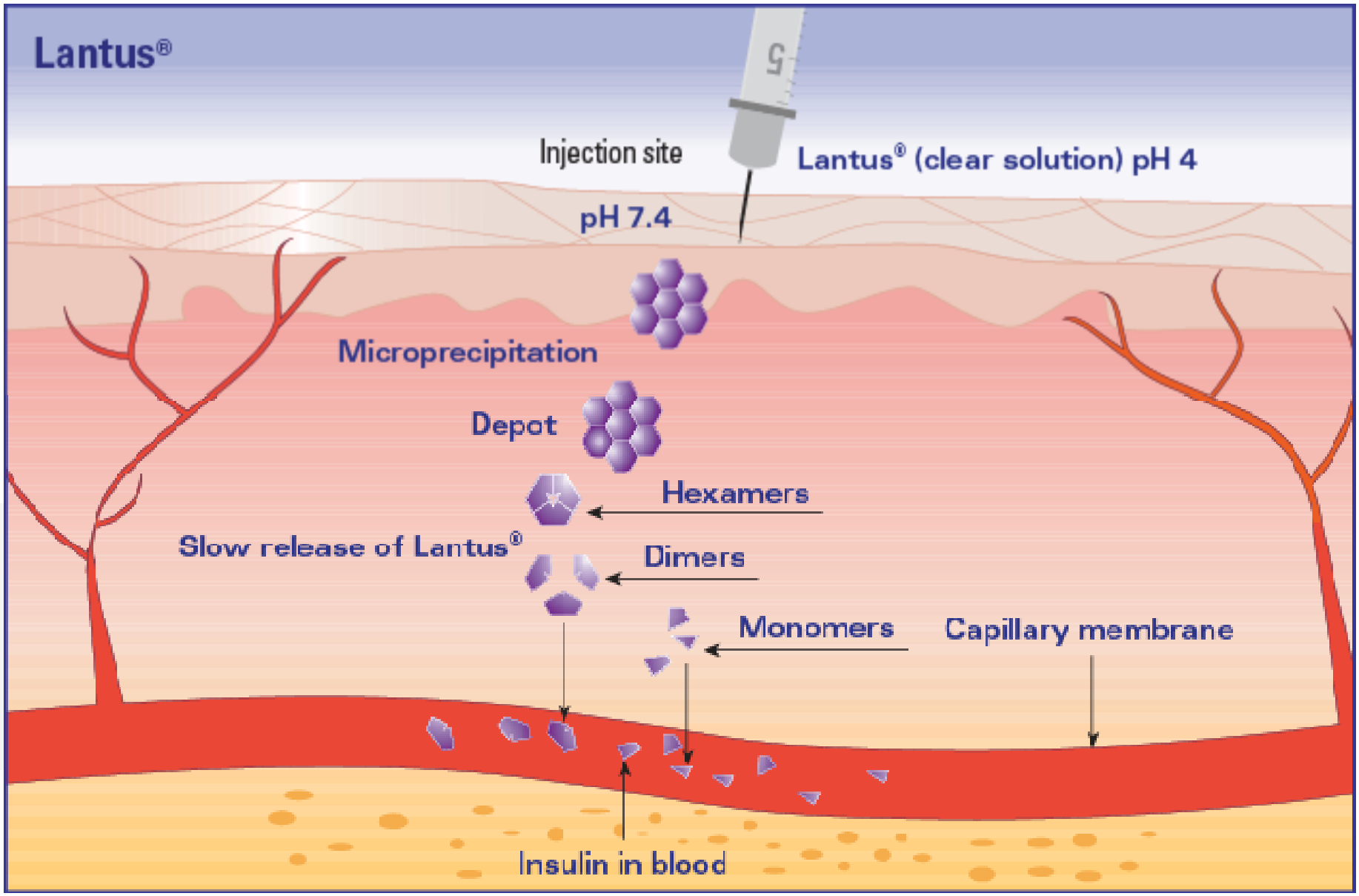
Extra 2 Arginine….Added in Human Insulin

- Is it make any change in charge?
  - More positive
  - More Negative
  - No any change
- What shell be added more to make it neutral (Net Charge Zero = Zwitter Ion)?
  - Positive ions (H+)
  - Negative ions (HCO3-)
- Can it change pI of insulin ?
  - Decrease
  - Increase
  - Unchange

Extra 2 Arginine….Added in Human Insulin

- Human insulin has isoelectric pH of 5.4. what can be pI of this newly for recombinant insulin?
  - More than 5.4
  - Less than 5.4
- In human body, this new recomb-insulin is going to expose pH of _____ .
- Recombinant insulin pI (6.7) is nearer to physiological pH (7.4) than human insulin pI (5.4). So what will be effect on it's solubility?
  - More soluble
  - Less soluble (precipitation)

Glargine Insulin (Lantus)



Primary Structure

It is representing

- **Numbers of Amino acids**
- **Sequence of Amino acids**
- Unique amino acid sequence decided by the genes.
- Maintained by peptide linkages.

This two things decide higher levels of organization (secondary, tertiary & quaternary) of protein.

Naming of Polypeptide

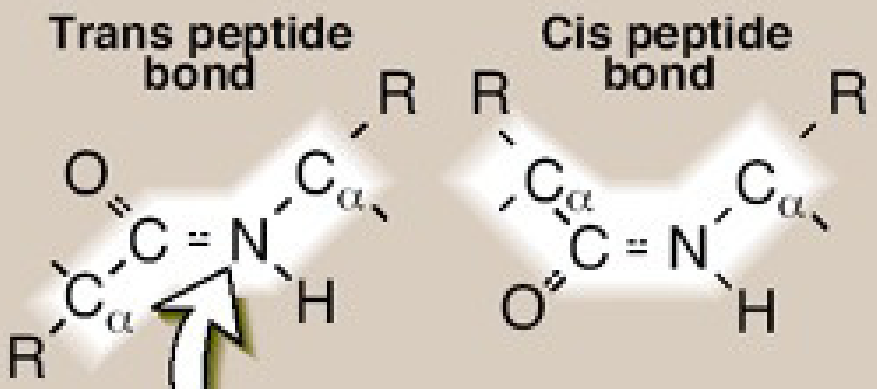
To write A.A. Sequence of polypeptide chain

- Amino terminal on Left
- Carboxyl terminal on Right
- Each component A.A. in a polypeptide is called a **"residue"**
- In a polypeptide , all A.A. residues suffixes (-ine, -an, -ic, or -ate) changed to -yl, with the exception of the C-terminal amino acid.
- E.g. Tripeptide = N-terminal valine, glycine, & C-terminal leucine
- **"valylglycylleucine"**

Charecteristics of a peptide bond

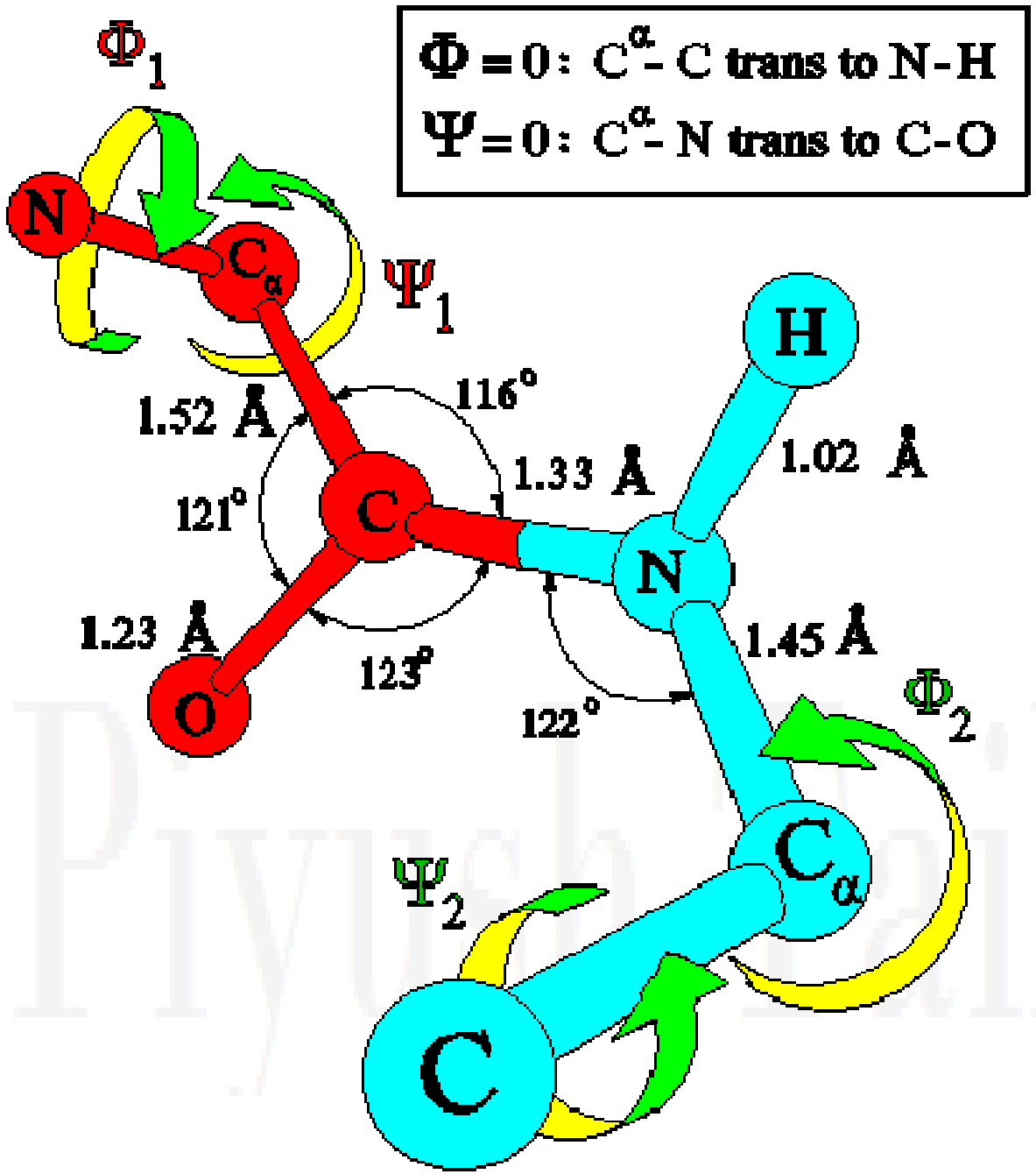
- ***It's a partial double bond***.
- 'trans' in nature
- Freedom of rotation with limitation to "R" group
- Having rotation of 180 degree
    - single bond = 360 degree
    - double bond = 0 degree
- Distance is 1.32 A°
    - single bond(1.42A°)
    - double bond(1.27A°).
- Angles of rotation known as Ramchandran angles

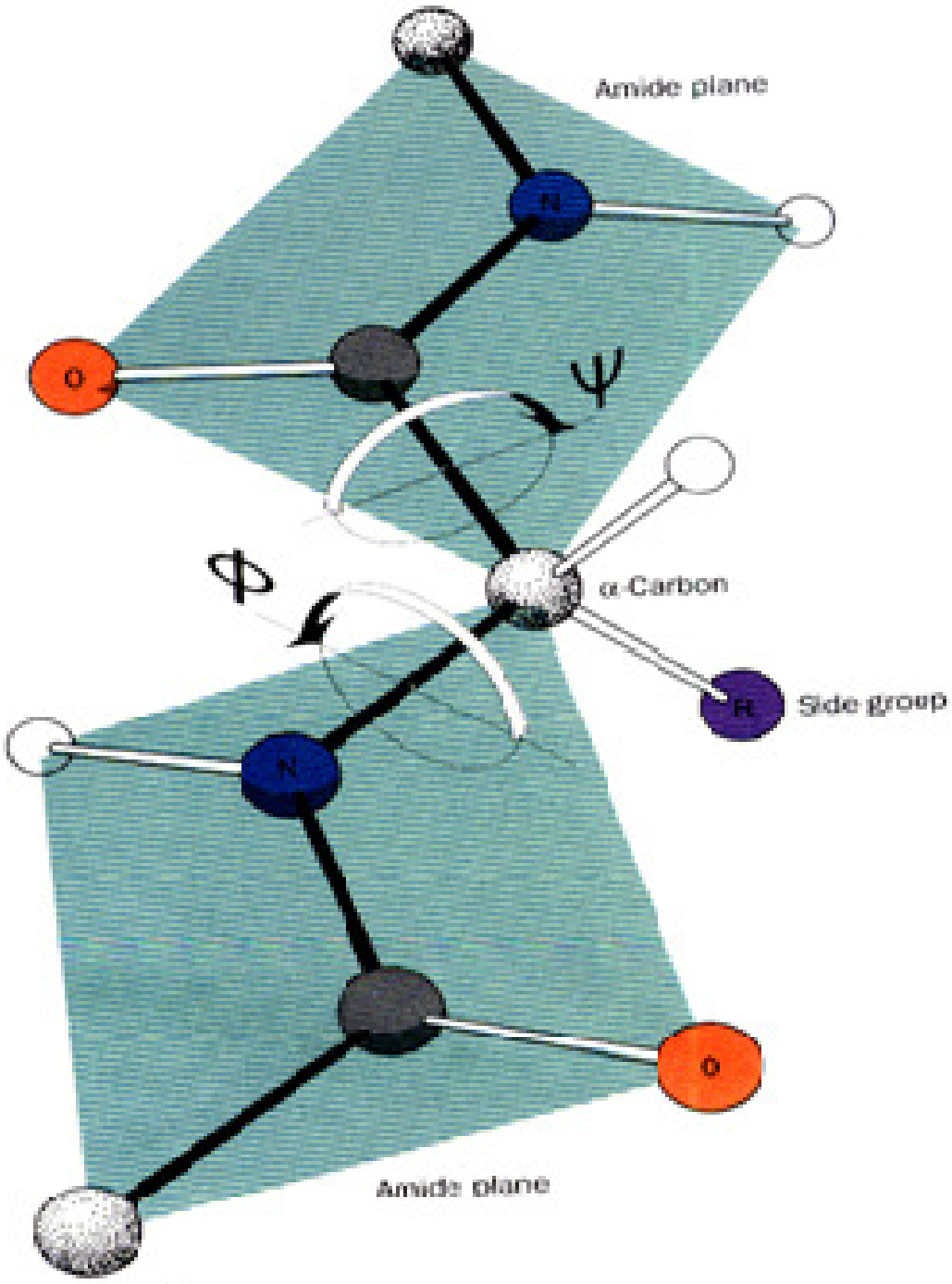
Characteristics of the peptide bond



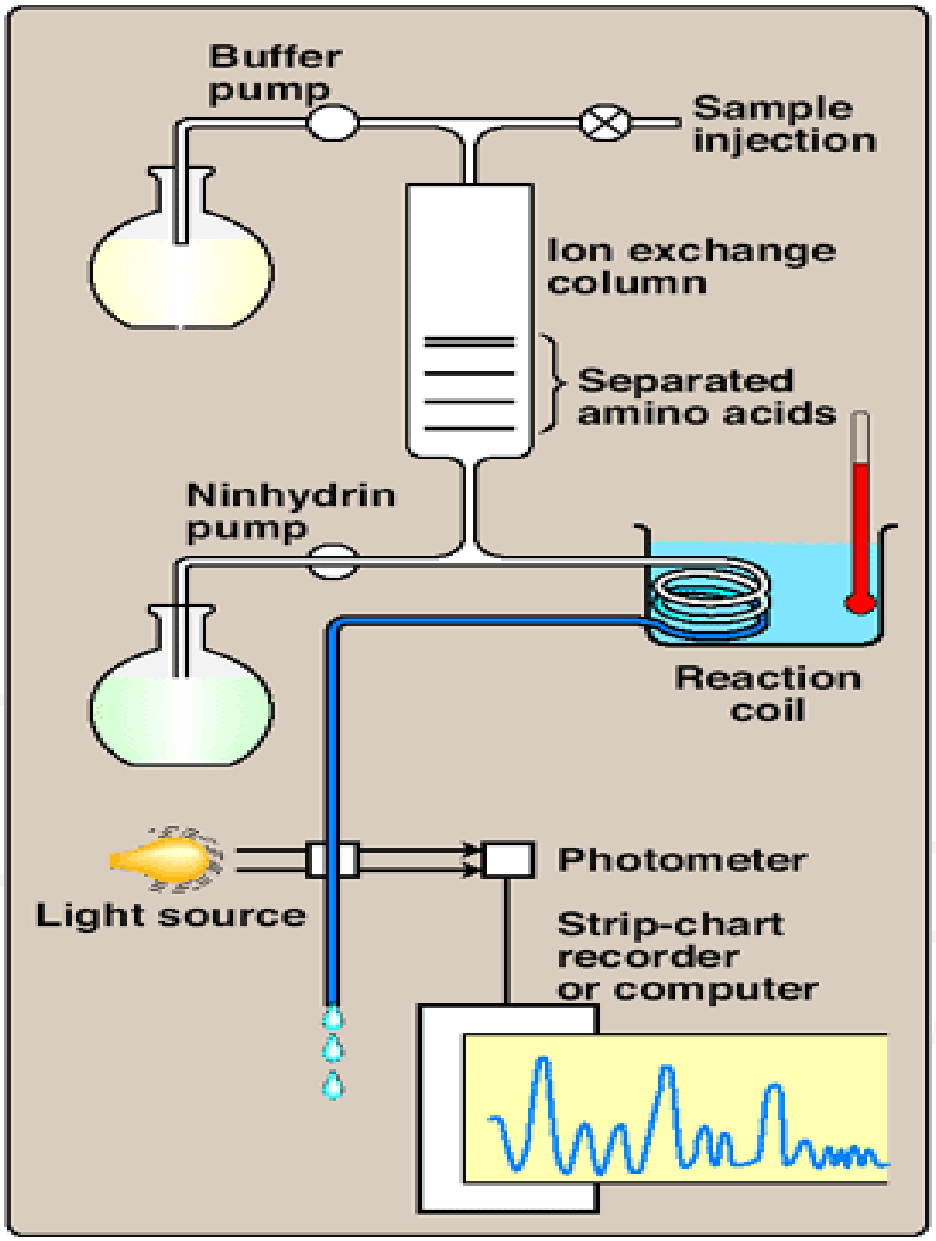
Peptide bonds in proteins

- Partial double-bond character
- Rigid and planar
- Trans configuration
- Uncharged but polar


Φ = 0 : Cα - C trans to N-H
Ψ = 0 : Cα - N trans to C-O
Φ1
Ψ1
N
Cα
C
O
H
N
Cα
C
Φ2
Ψ2
1.52 Å
1.23 Å
1.33 Å
1.02 Å
1.45 Å
116°
121°
123°
122°

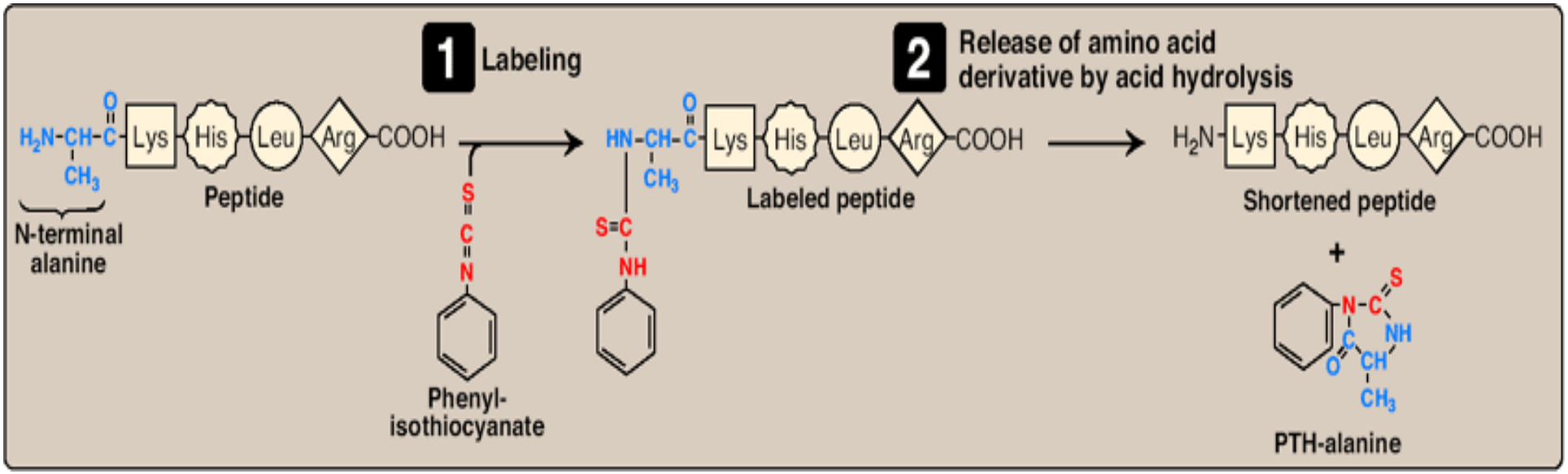

Amide plane
N
O
ψ
φ
α-Carbon
R
Side group
N
O
Amide plane

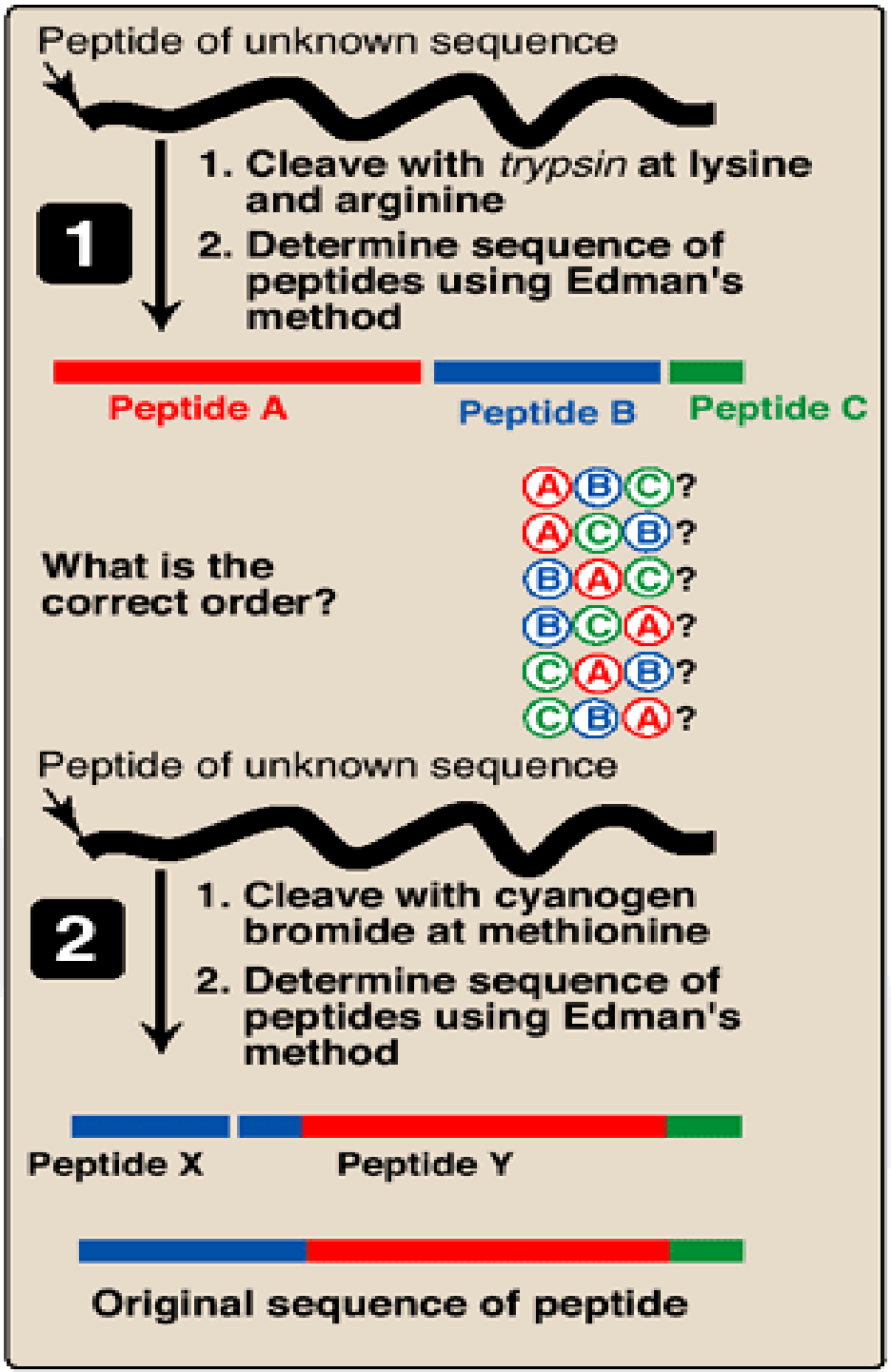
Separation of A.A. from Polypeptide



Sequencing of the peptide from its N-terminal end

- Edman reagent = Phenylisothiocyanate
- Label N-terminal A.A. residue under mildly alkaline conditions.
- Phenylthiohydantoin (PTH) create instability in the N-terminal peptide bond
- It can be selectively hydrolyzed without cleaving the other peptide bonds




Peptide of unknown sequence
1
1. Cleave with trypsin at lysine and arginine
2. Determine sequence of peptides using Edman's method
Peptide A
Peptide B
Peptide C
What is the correct order?
ABC?
ACB?
BAC?
BCA?
CAB?
CBA?
Peptide of unknown sequence
2
1. Cleave with cyanogen bromide at methionine
2. Determine sequence of peptides using Edman's method
Peptide X
Peptide Y
Original sequence of peptide

Gene to Protein Sequence

HBB Sequence in Normal Adult Hemoglobin (Hb A):

Nucleotide	CTG	ACT	CCT	GAG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Glu	Glu	Lys	Ser
	3			6			9

HBB Sequence in Mutant Adult Hemoglobin (Hb S):

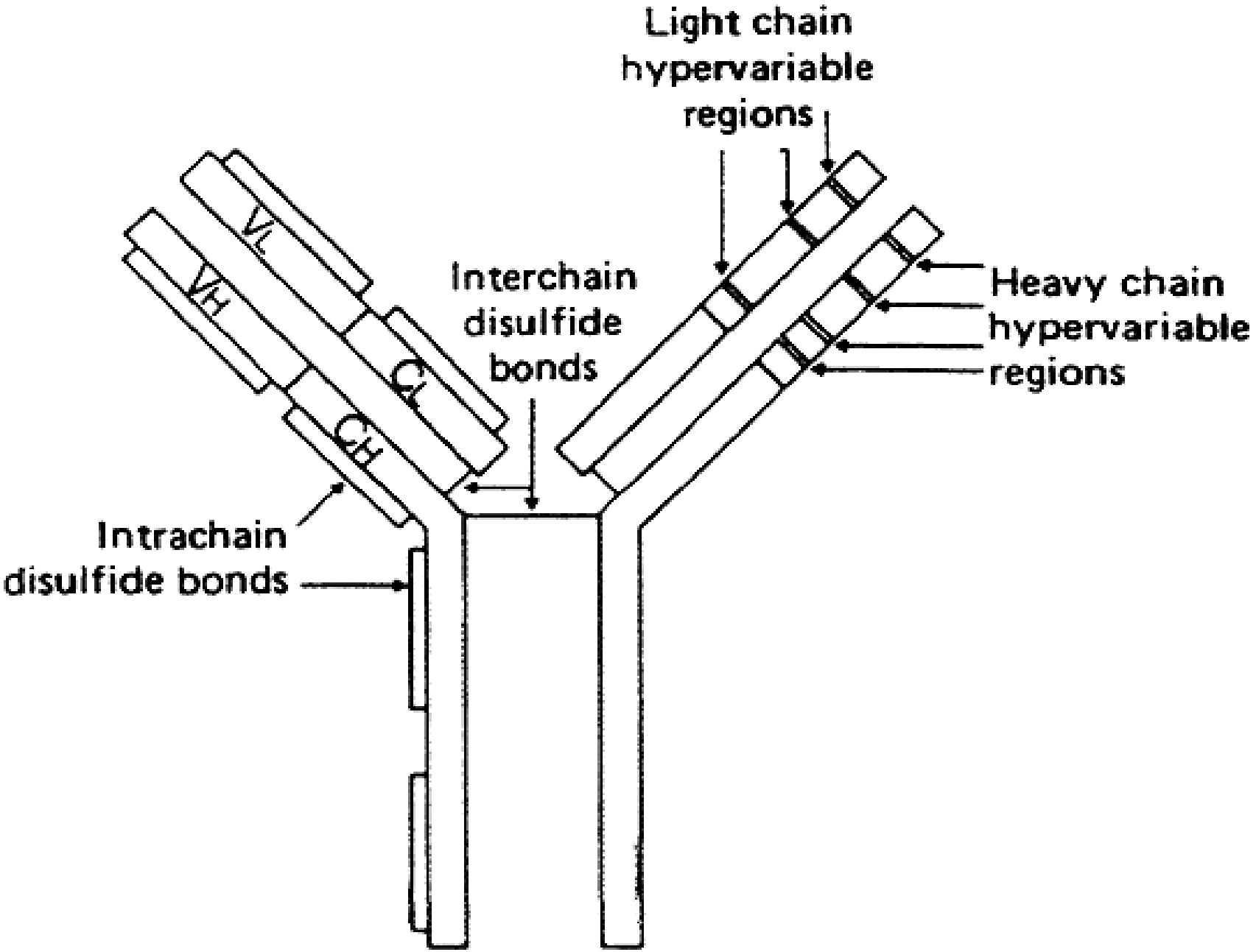
Nucleotide	CTG	ACT	CCT	GTG	GAG	AAG	TCT
Amino Acid	Leu	Thr	Pro	Val	Glu	Lys	Ser
	3			6			9

Gene to Peptide sequence

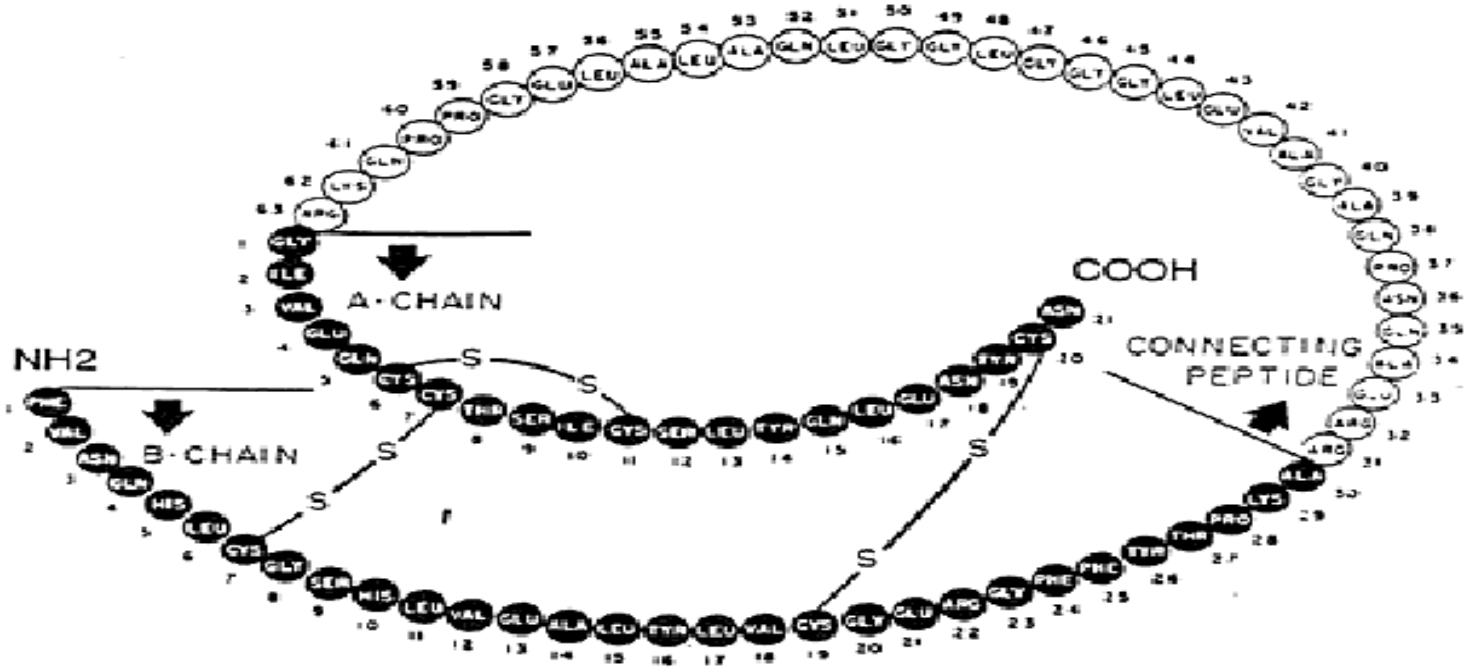
- Has limitations
  - Not able to predict positions of disulfide bonds
  - Not able to identify any posttranslational modification.
- Therefore, direct protein sequencing is an extremely important tool.

Branched and Circular proteins

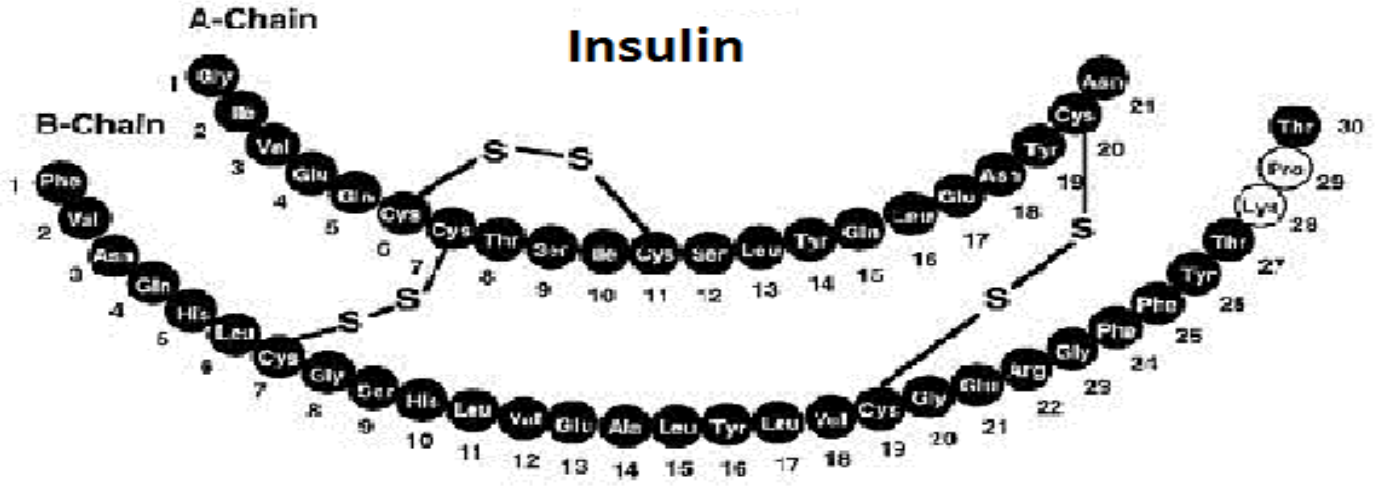
- In linear polypeptide chains, there is
    - Interchain Disulphide bridges.
- In different polypeptide chains
    - Interchain = In same protein.
    - Intrachain = Same polypeptide chain.


Light chain
hypervariable
regions
Interchain
disulfide
bonds
Heavy chain
hypervariable
regions
Intrachain
disulfide bonds
V_L
V_H
C_L
C_H

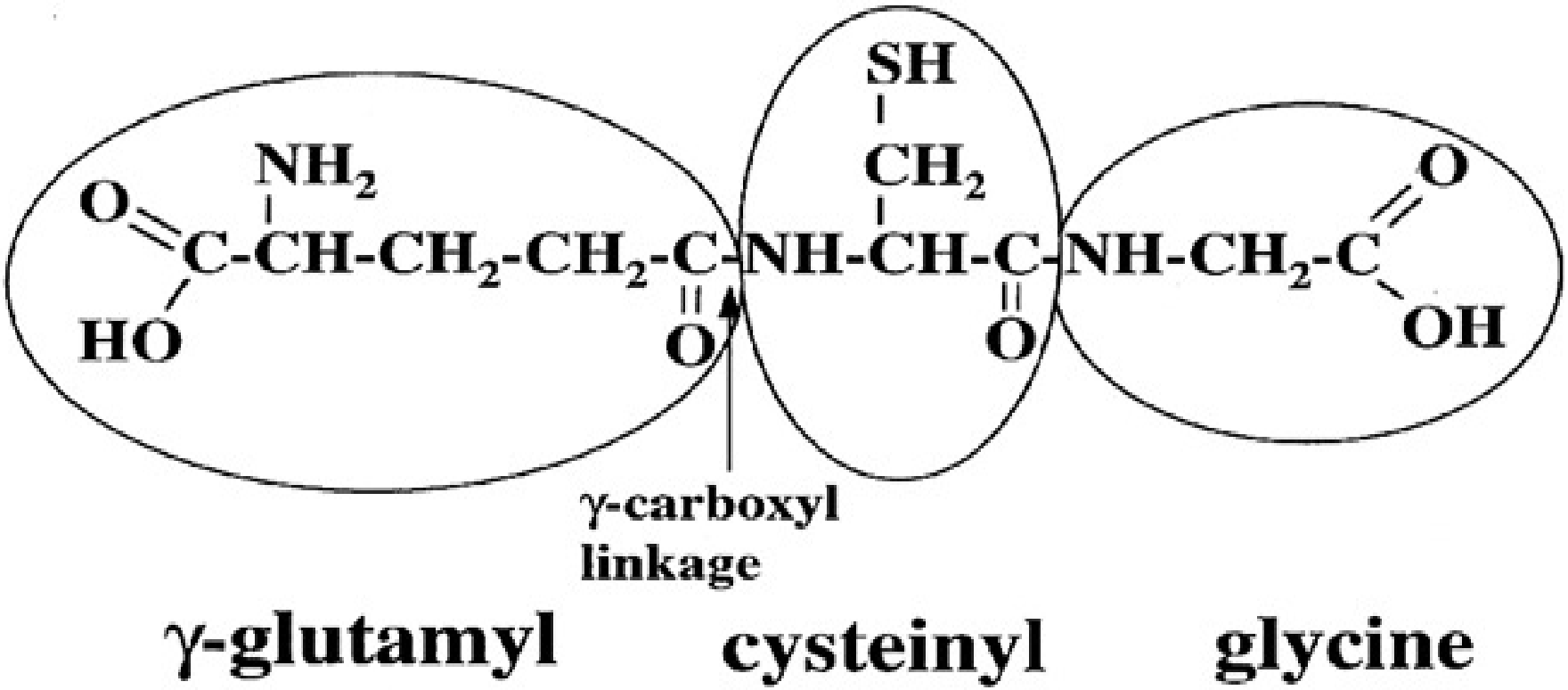
Pre-Insulin



Insulin



Glutathione = Is it protein?



Glutathione = Pseudopeptide

2. SECONDARY STRUCTURE

- Configurationally relationship between residues which are about 3-4 amino acids apart in the linear sequence.
- The structure is preserved by non-covalent forces or bonds.

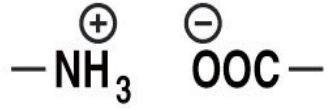
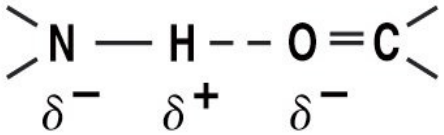
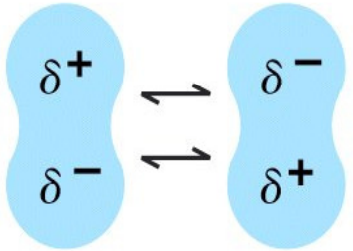
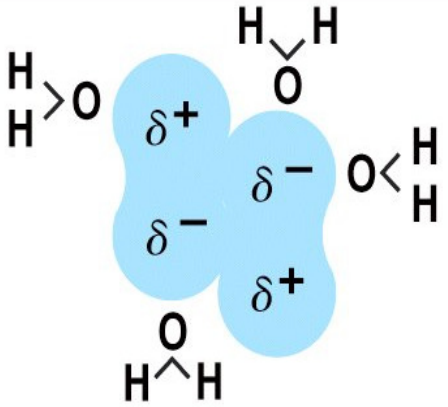
Noncovalent forces	Origin	
Electrostatic forces	Attraction between opposite charges	
Hydrogen bonds	Hydrogen shared between electronegative atoms (N,O)	
Van der Waals forces	Fluctuations in electron clouds around molecules oppositely polarize neighboring atoms	
Hydrophobic forces	Hydrophobic groups interact unfavorably with water and tend to pack together to exclude water molecules. The attraction also involves van der Waals forces	

Figure 3-9 Immunobiology, 6/e. (© Garland Science 2005)

Disulfide bonds:

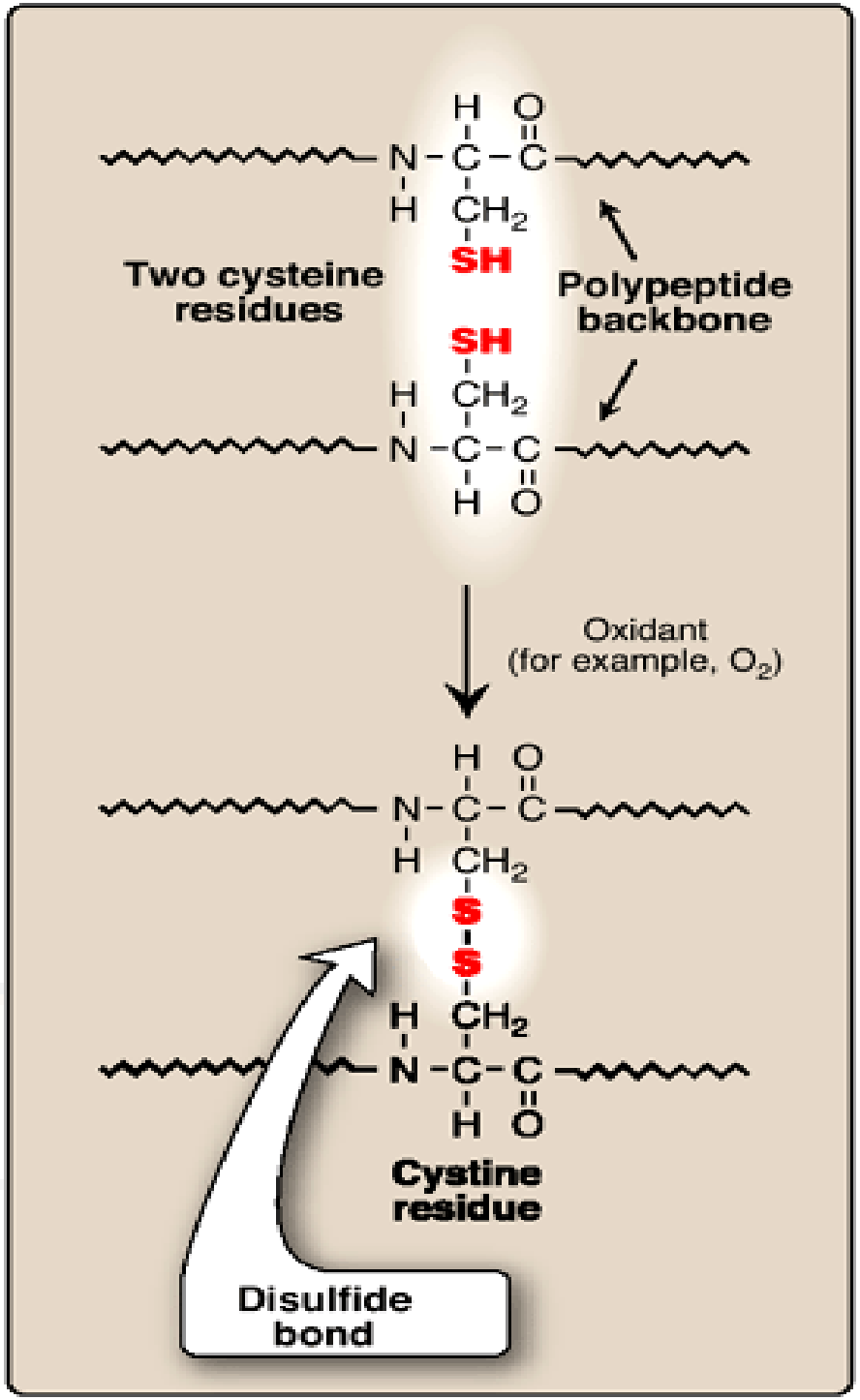
=Covalent linkage between sulfhydryl group (–SH)

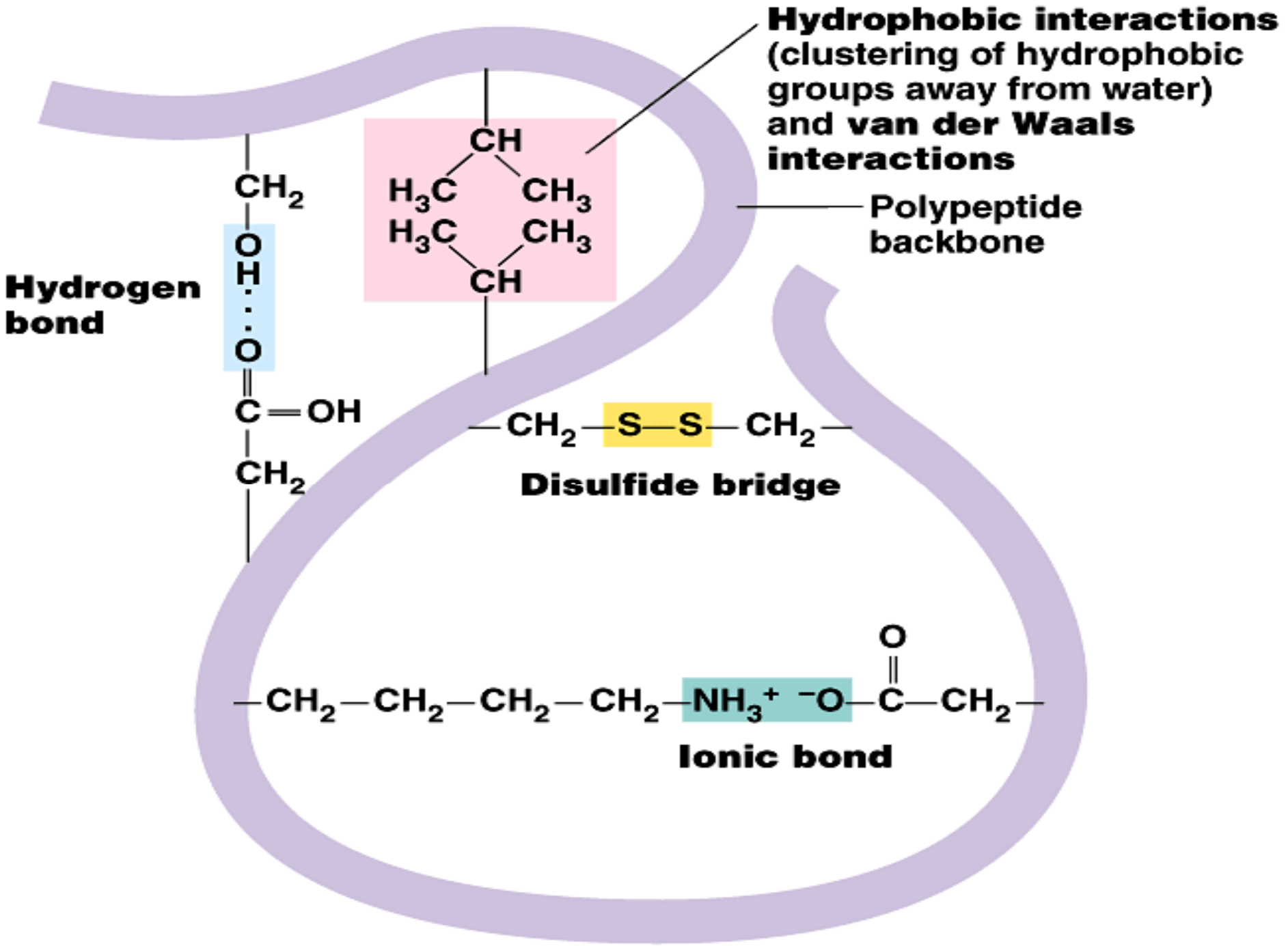
=two cysteines may be far aways in the primary sequence or may even be different polypeptide chains;

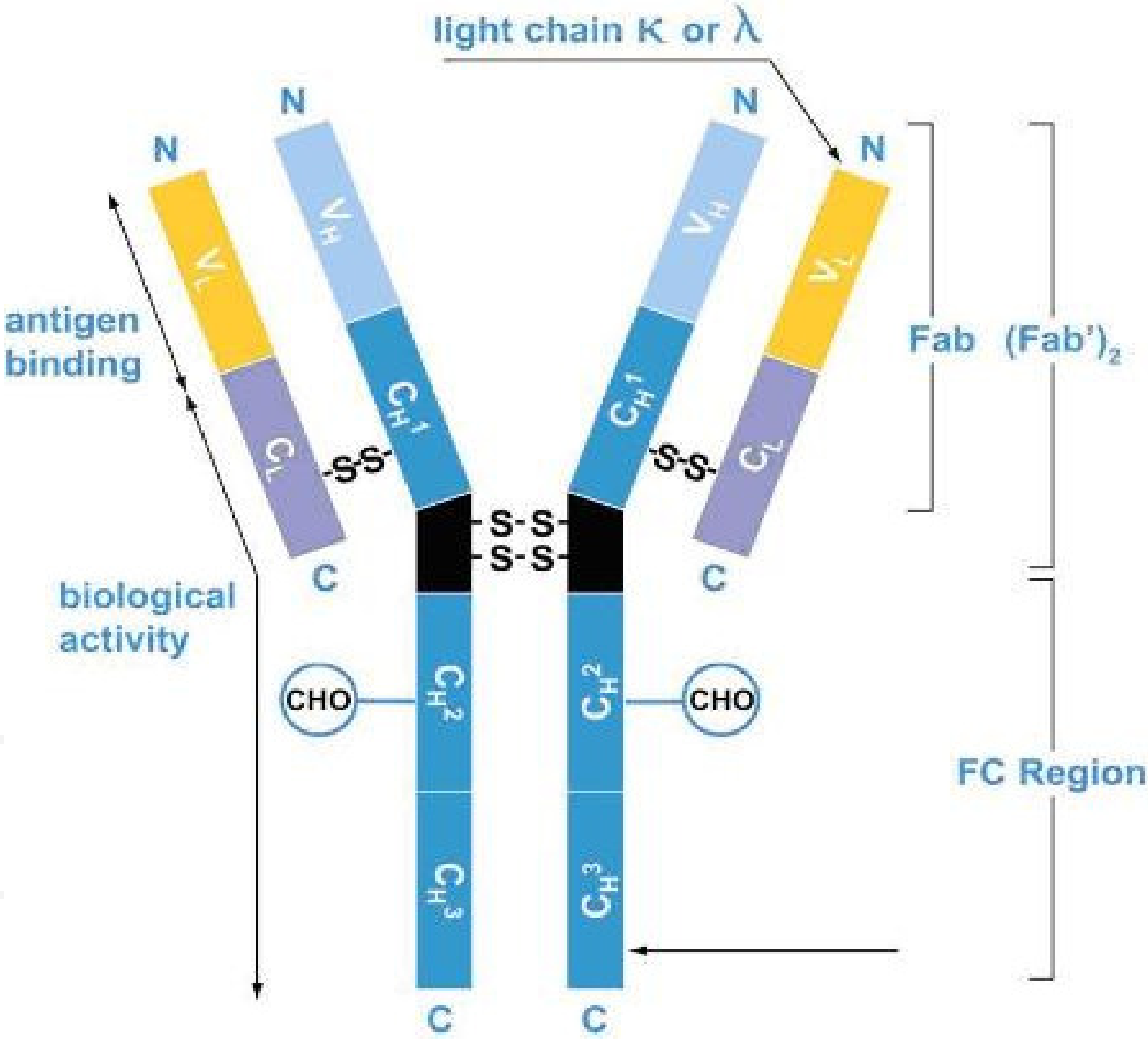
=contributes stability

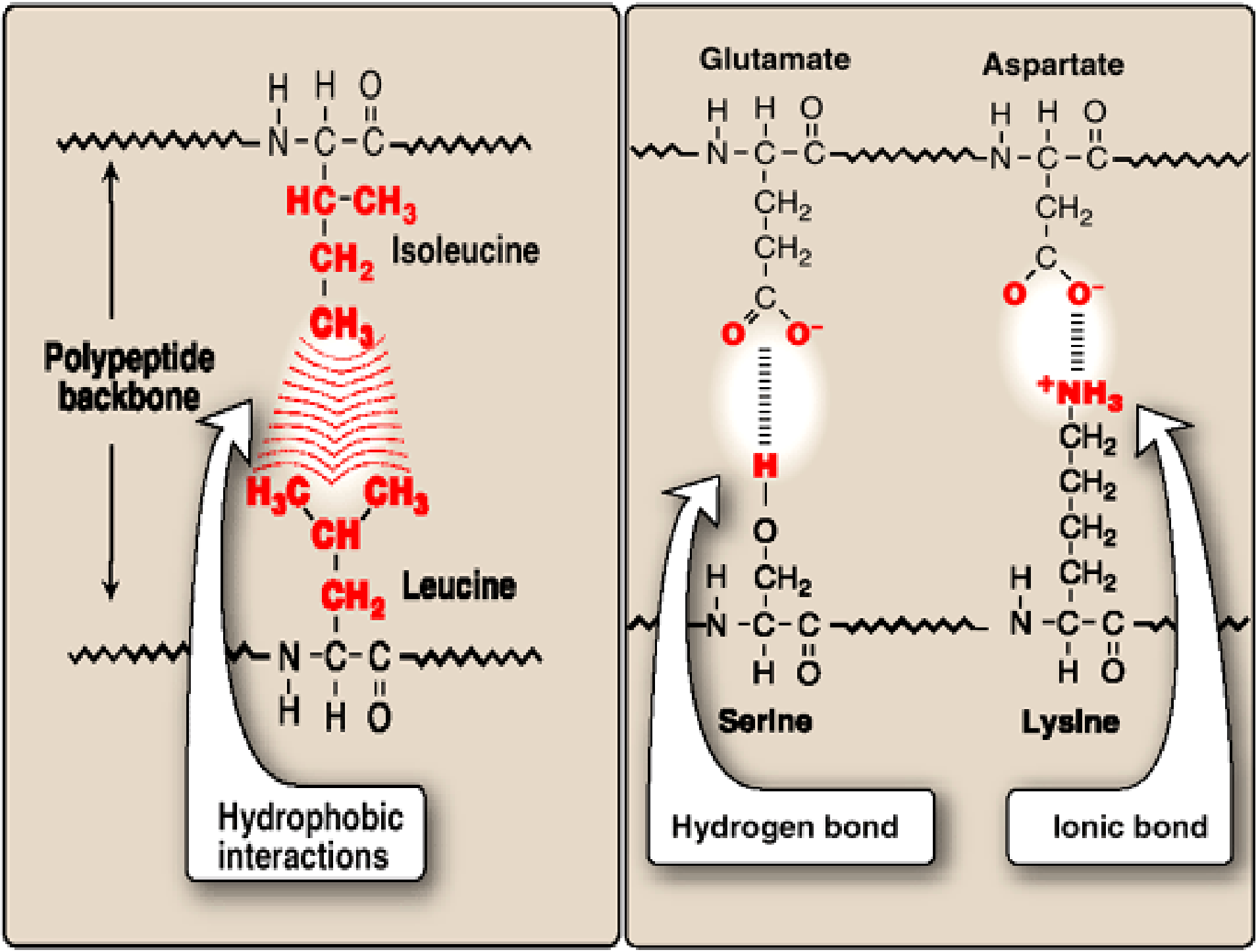
=Prevents denatured

= E.g. Immunoglobulins






light chain κ or λ
N
N
N
N
antigen binding
biological activity
V_L
V_H
C_L
C_{H1}
V_H
V_L
C_{H1}
C_L
S-S
S-S
S-S
S-S
C
C
CHO
CHO
C_{H2}
C_{H2}
C_{H3}
C_{H3}
C
C
Fab
$(Fab')_2$
FC Region


Isoleucine
Leucine
Polypeptide backbone
Hydrophobic interactions
Glutamate
Aspartate
Serine
Lysine
Hydrogen bond
Ionic bond

Hydrophobic interactions

- Non polar side chains = Remain in the interior
- Polar side chains = Remain on the surface

Hydrogen bonds

- Oxygen- or Nitrogen-bound hydrogen of side chains
- Alcohol groups of serine and threonine
- Can form hydrogen bonds with Oxygen of a carboxyl group of a peptide bond (electron-rich atoms)

Ionic interactions

- Negatively charged groups, carboxyl group ($–COO^-$) in the side chain of aspartate or glutamate.
- Positively charged groups, such as the amino group ($–NH_3^+$) in the side chain of lysine .

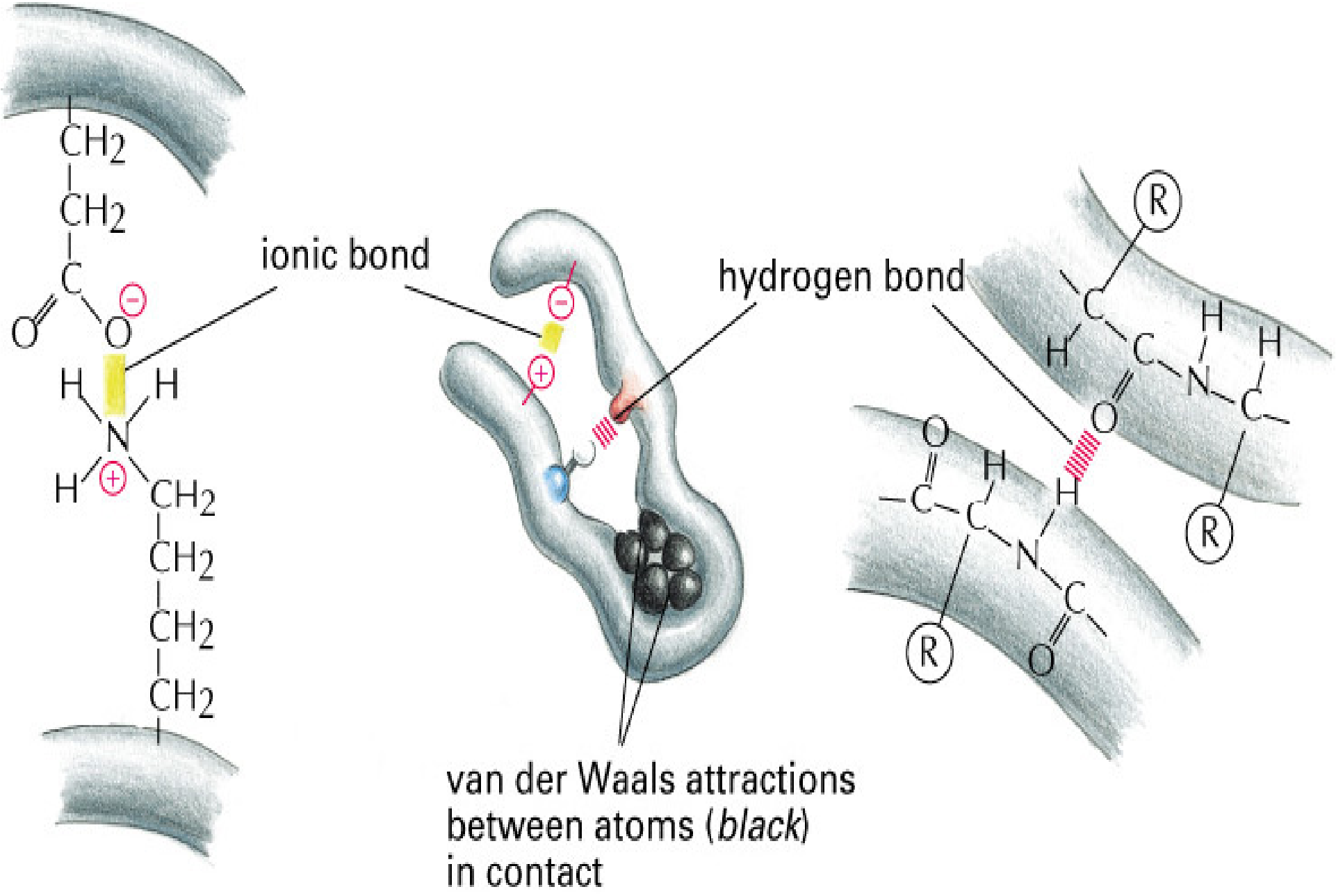


Figure 3–5. Molecular Biology of the Cell, 4th Edition.

Alpha Helix

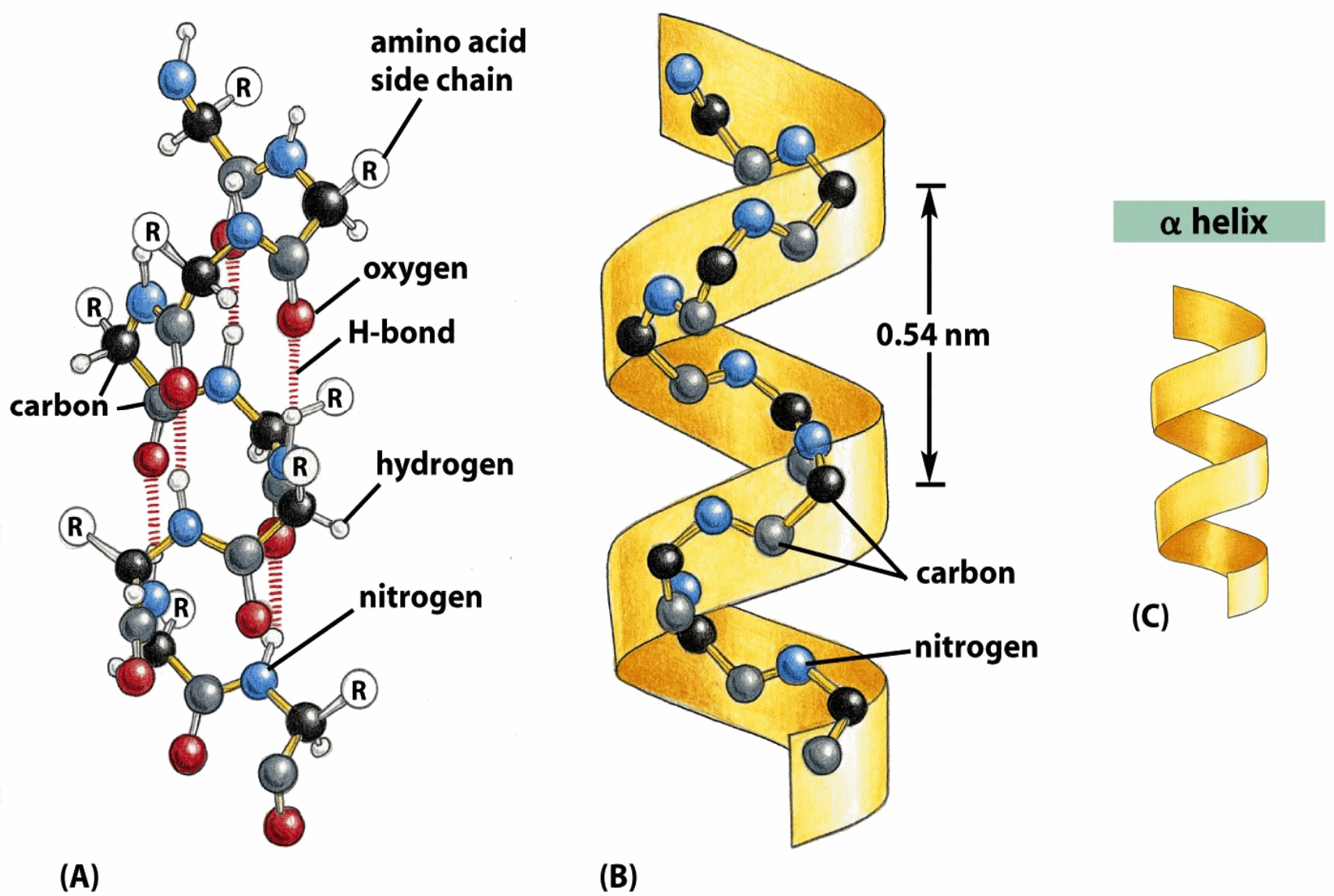
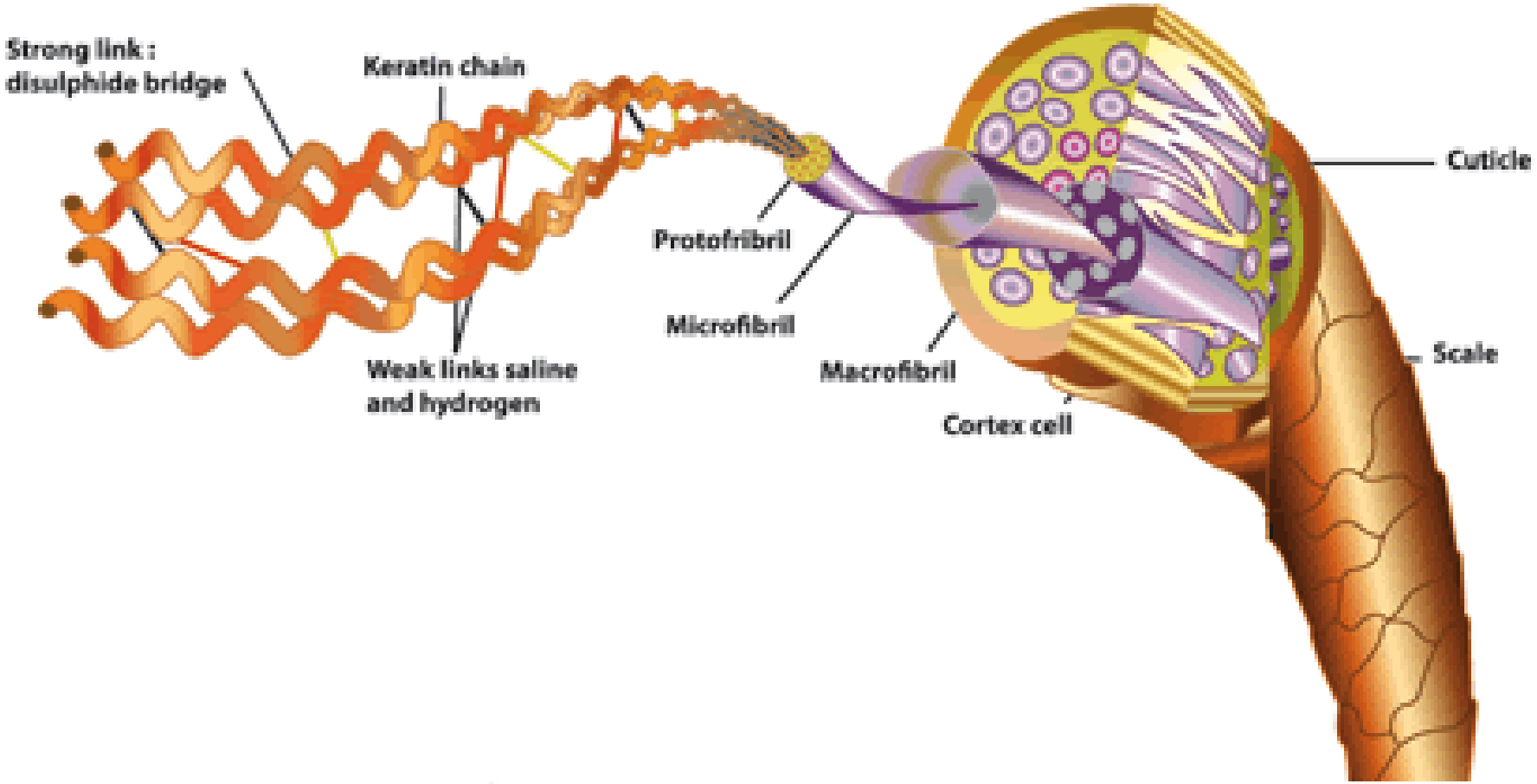


Figure 3-7a-c Molecular Biology of the Cell 5/e (© Garland Science 2008)

1. Alpha helix

- spiral structure.
- Polypeptide bonds = Backbone.
- Side chain of amino acids extend outward.
- Stabilized by hydrogen bonds
- In each turn = 3.6 A.A. residues.
- Distance between each A.A. residue is 1.5A°
- Right handed.
- As A.A. in the proteins are of L-variety.
- Proline and hydroxyproline will not allow the formation of α-helix.


Strong link :
disulphide bridge
Keratin chain
Weak links saline
and hydrogen
Protofribril
Microfibril
Macrofibril
Cortex cell
Cuticle
Scale

Example of Alpha Helix

1. Keratins = fibrous proteins in Hair & Nail

- It's rigidity is determined by the number of disulfide bonds

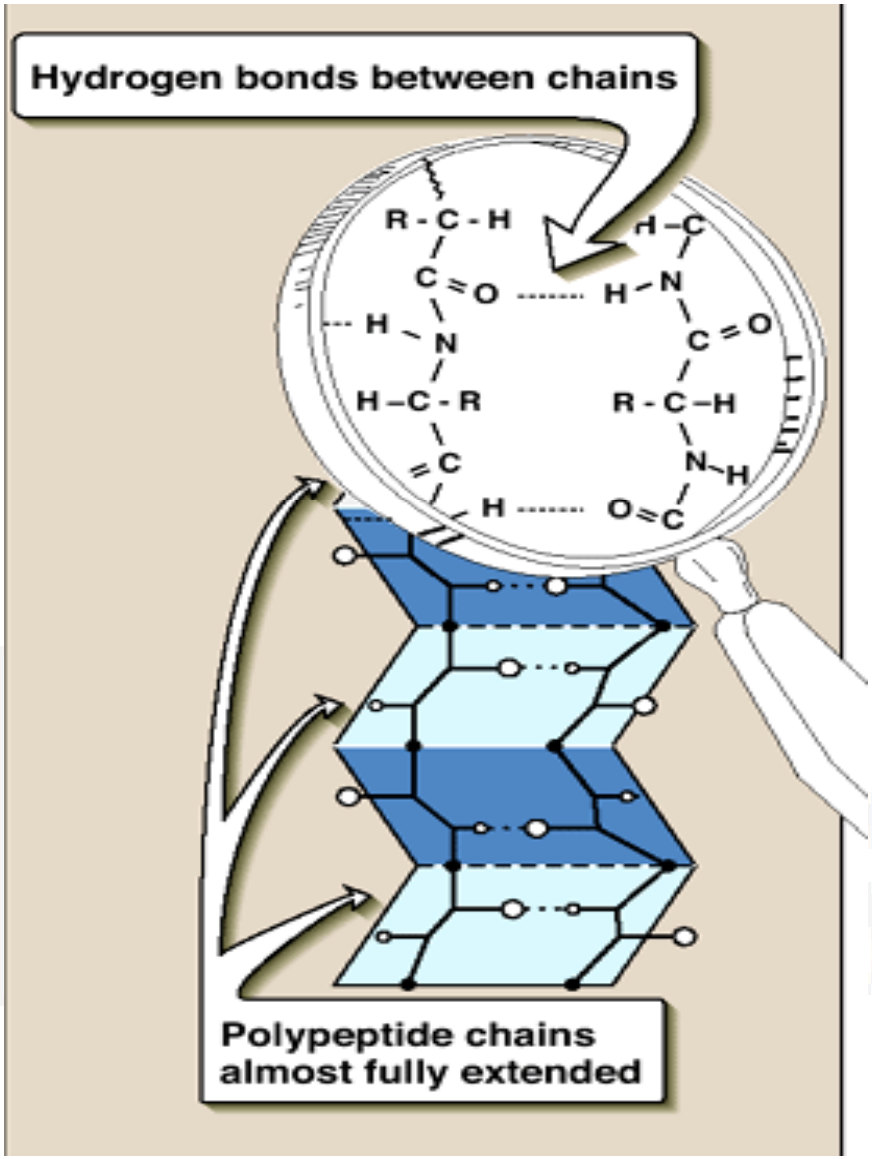
2. Myoglobin = globular & flexible molecule.

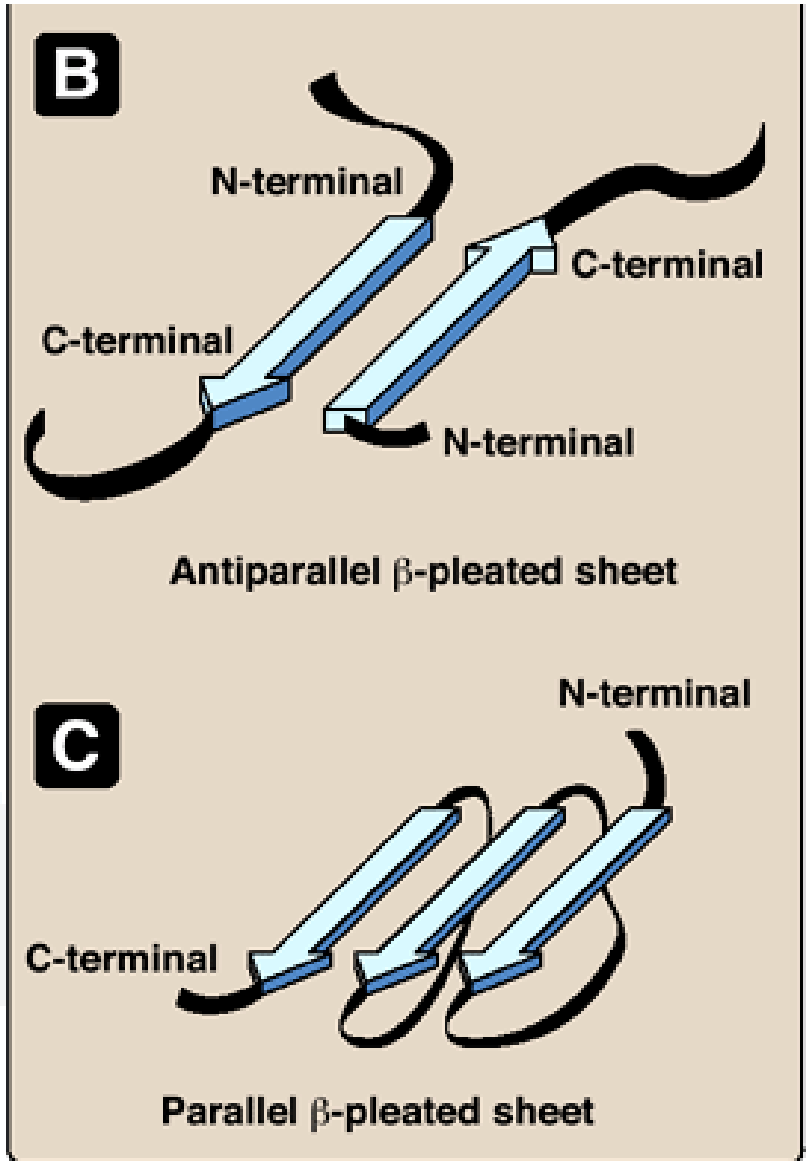
   (Contrast to Keratin)

It can disturb Alpha helix

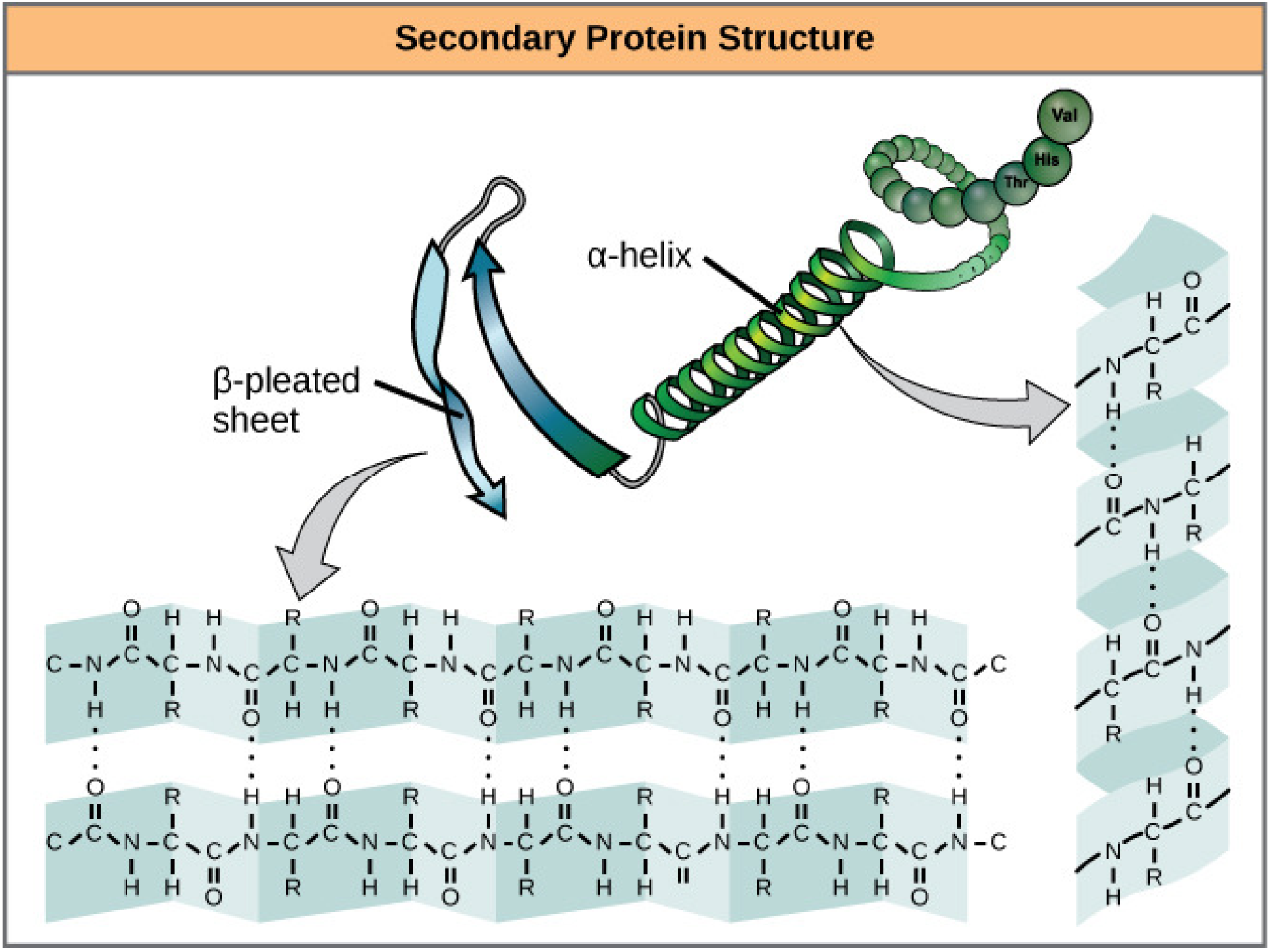
- Proline
  - Secondary amino group is not geometrically compatible
  - it inserts a kink in the chain
- Large numbers of charged amino acids
  - As it forms more ionic bonds or by electrostatically repelling each other.
  - E.g. glutamate, aspartate, histidine, lysine, or arginine
- Amino acids with bulky side chains or branched
  - Tryptophane, Valine, Isoleucine

Beta Plated Sheet





Secondary Protein Structure

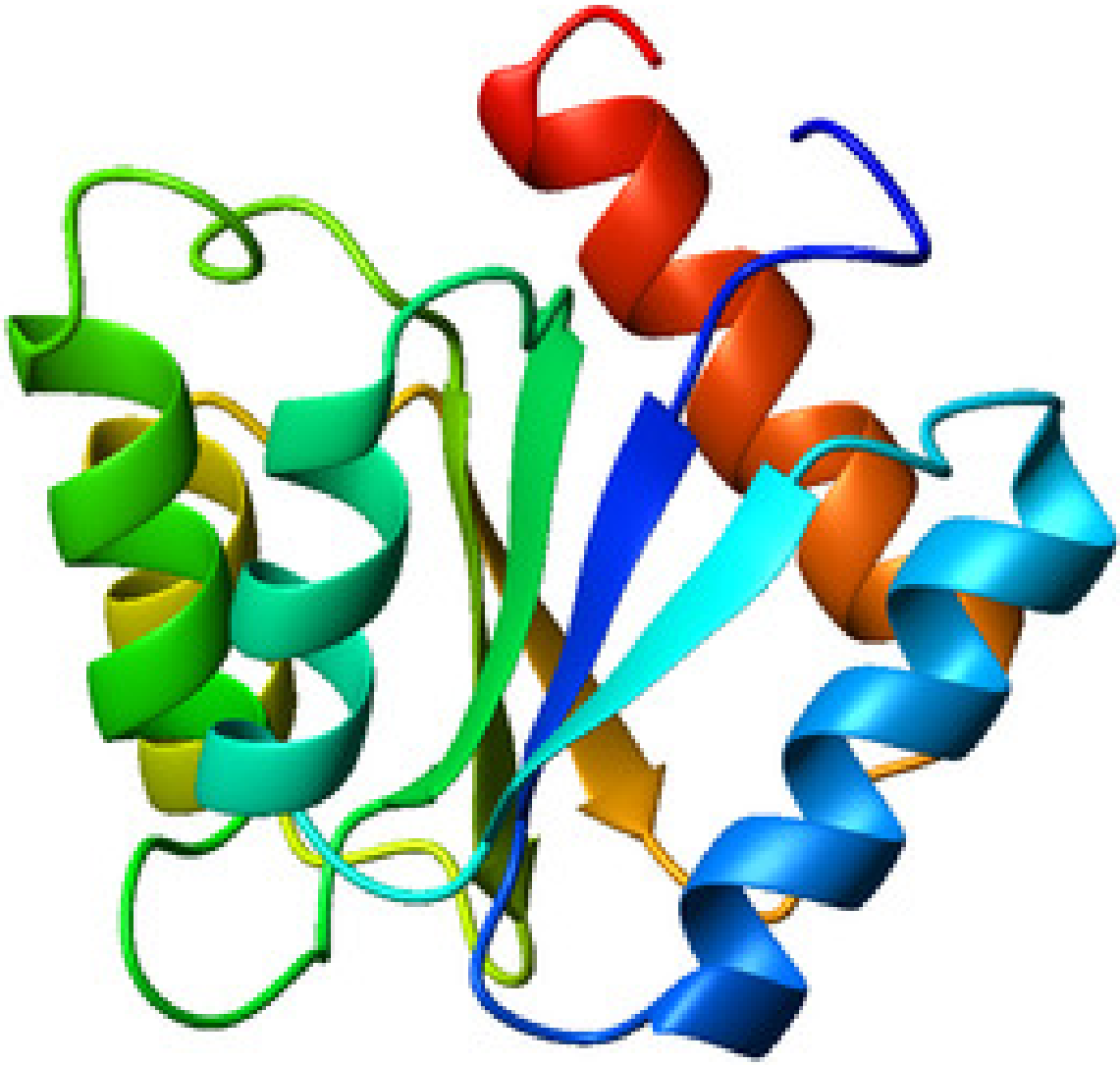


2. Beta-pleated sheet

- The distance between adjacent a.a. is 3.5A°.
- It is stabilised by hydrogen bonds.
- Adjacent strands in a sheet
    - Parallel
    - Anti parallel
- E.g. Flavodoxin , Carbonic anhydrase

  Triple helical structure in collagen

Flavodoxin



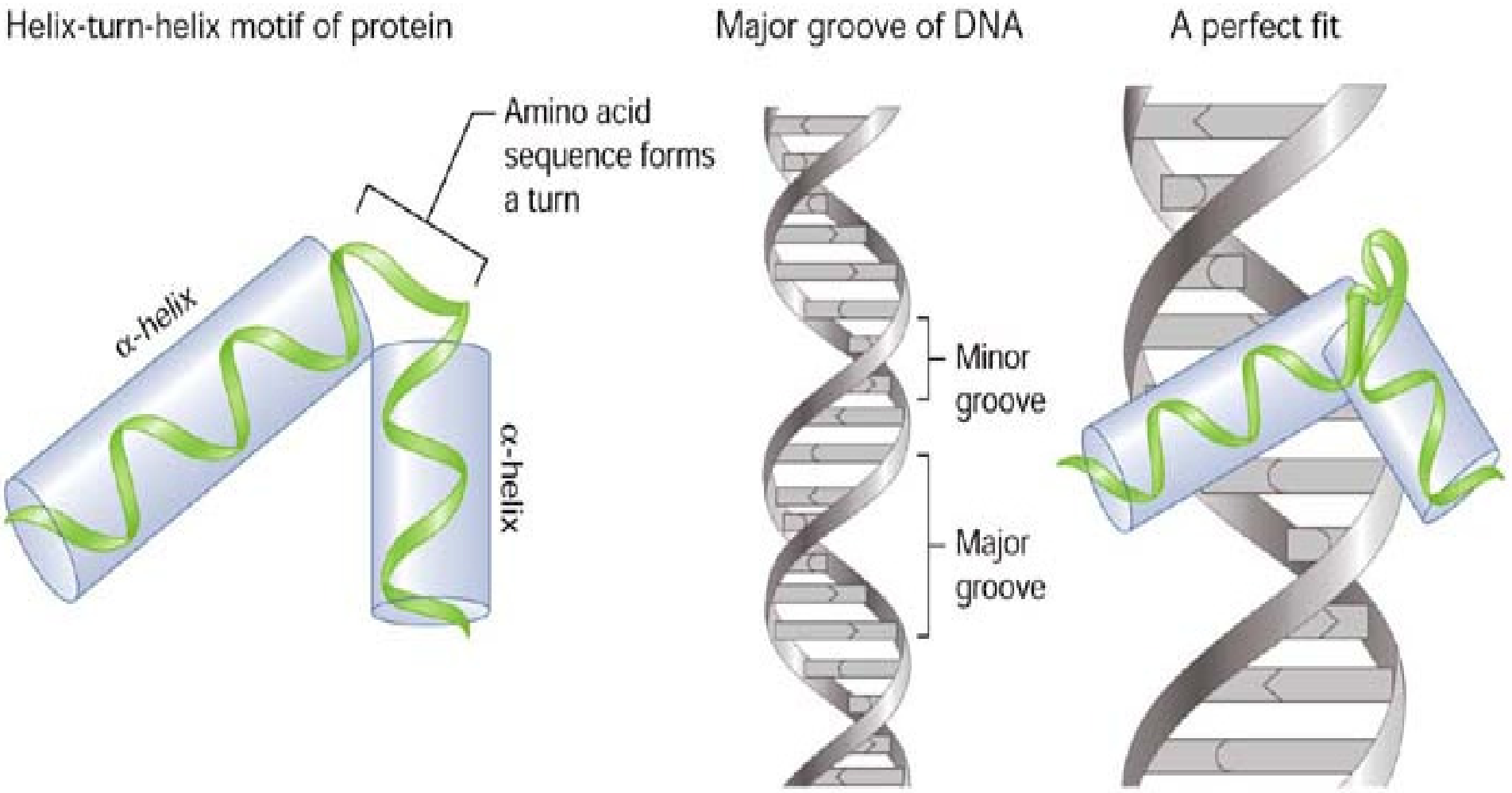
Beta Bands (beta turn)

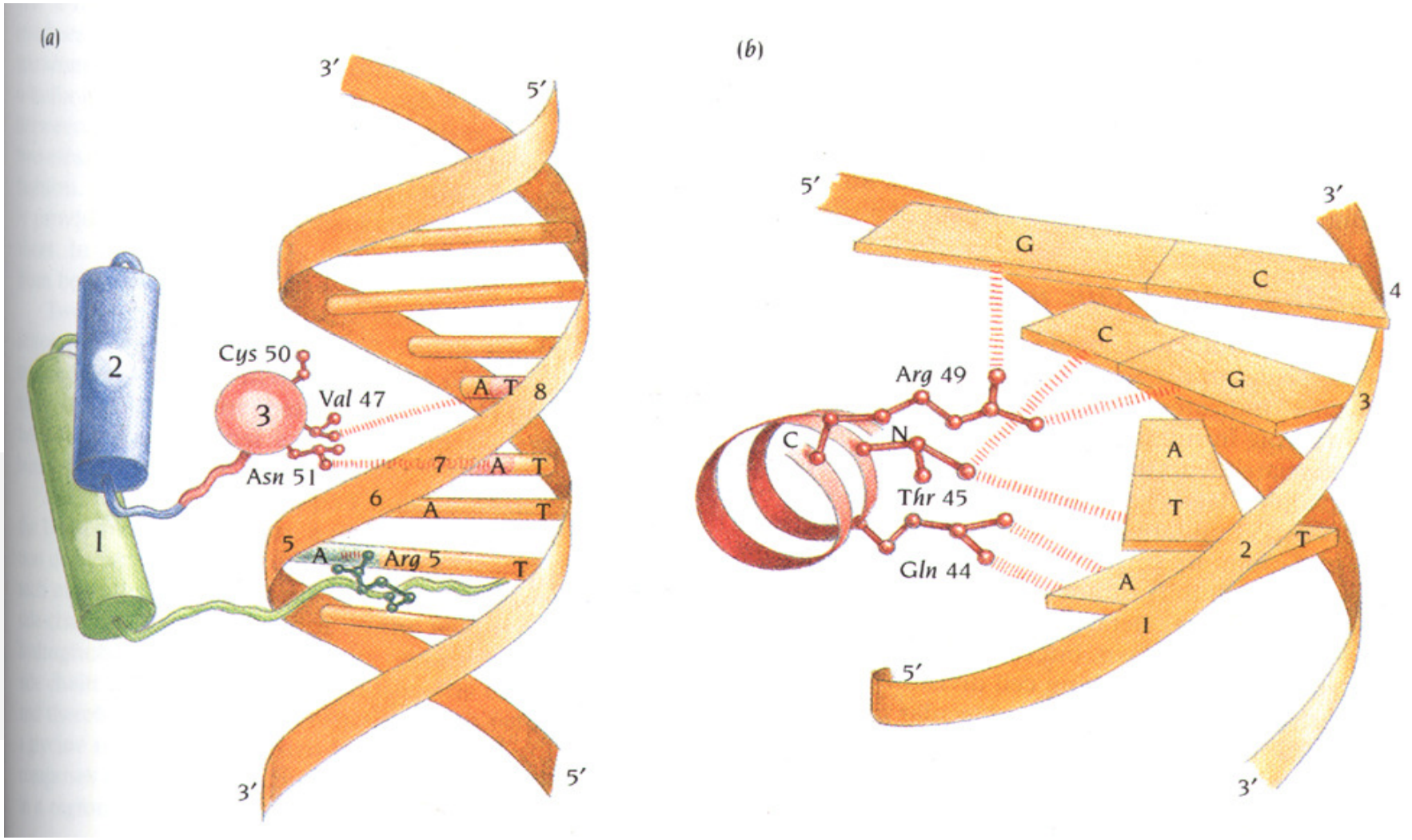
- Reverse the direction of a polypeptide chain
- Helping it form a compact, globular shape.
- Found on the surface of protein molecules

Supersecondary Structure (Motif)

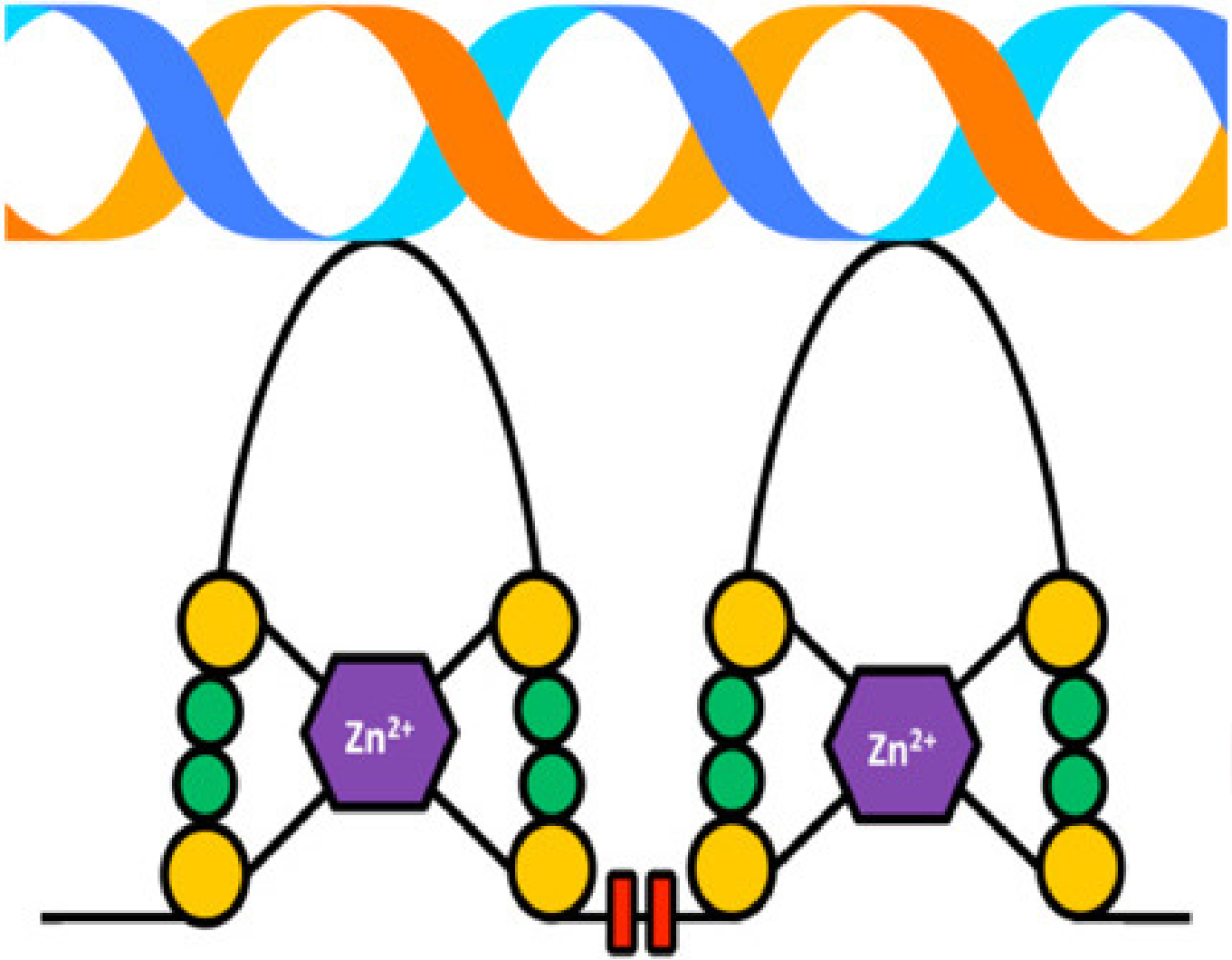
- Combining of α-helices and β-sheets
- Form primarily the core region—that is, the interior of the molecule.
- Connected by loop β-bends
- Supersecondary structures are usually produced by packing side chains from adjacent secondary structural elements close to each other.

Helix turn helix motif

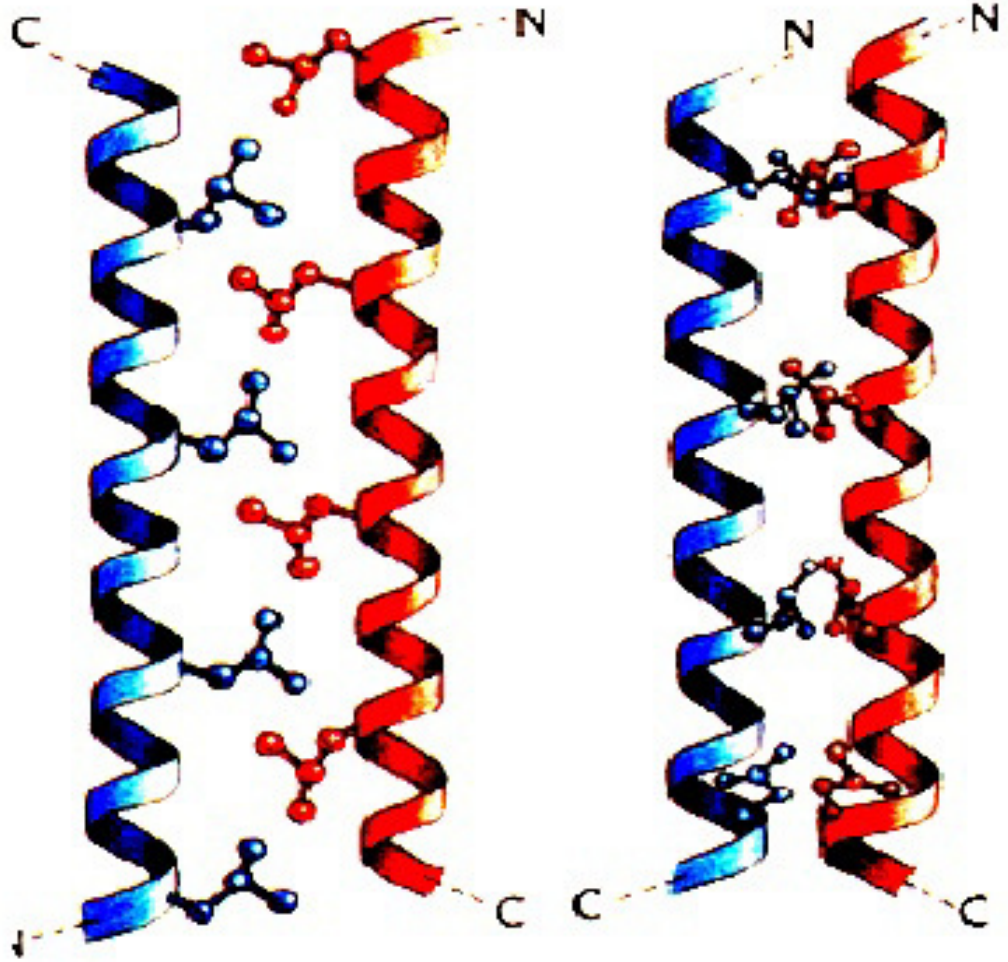



(a)
3′
5′
2
Cys 50
3
Val 47
A T 8
Asn 51
7 A T
6 A T
1
5 A Arg 5 T
3′
5′
(b)
5′
3′
G
C
4
C
Arg 49
G
3
C
N
A
Thr 45
T
2
T
Gln 44
A
1
5′
3′

Zink finger Motif



The Leucine Zipper



Original concept

More correct

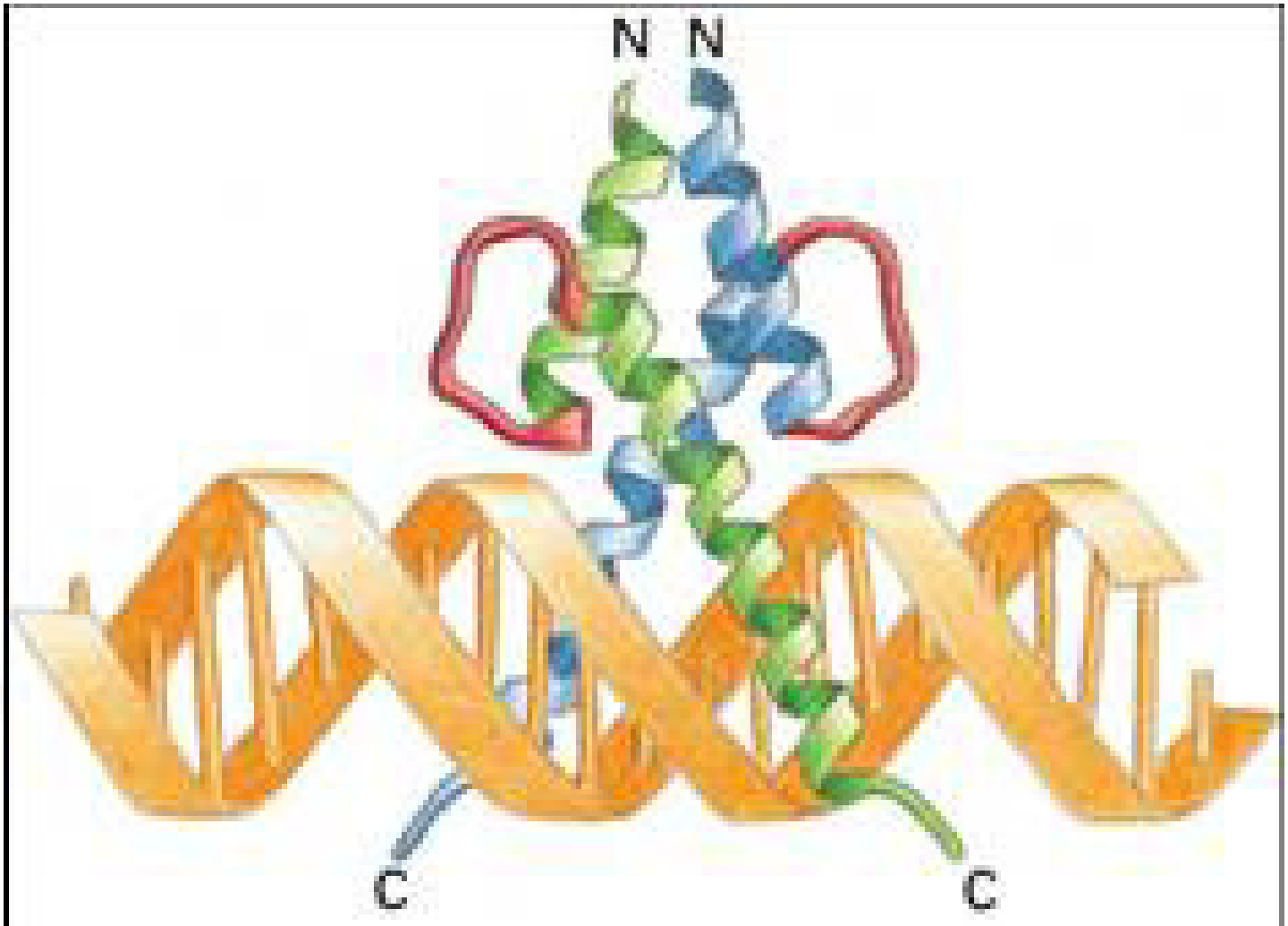
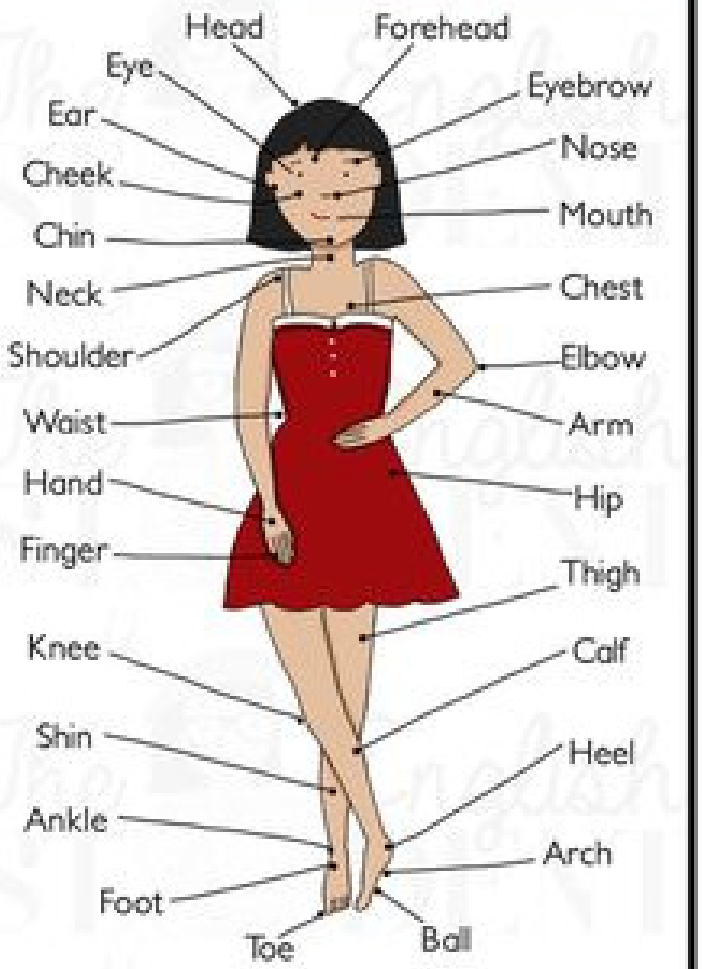


Figure 7–25. Molecular Biology of the Cell, 4th Edition.

3.TERTIARY STRUCTURE

- three dimentional structure of the whole protein.
- A.A. are far apart from each other in the linear sequence.
- But A.A. are close in the three dimention.
- It is maintained by non-covalent interactions
    - hydrophobic bonds
    - electrostatic bonds
    - Van der Waals forces.

Body Parts



head .....

ears .....

arms .....

hands .....

eyes .....

nose .....

mouth .....

teeth .....

legs .....

Domain

- **Compact Globular Functional Unit of a protein.**
- Independent region of the protein
- May be multiple = if protein has > 200 amino acids.
- Core of a domain is built from combinations of motifs.
- Folding of domain peptide chain is independently of folding in other domains.

Protein Folding

- Three-dimensional shape of the functional protein.
- Due to Interactions between A.A. side chains
  - Charges of a.a. side chains = attraction and repulsion
  - Hydrogen bonds
  - Hydrophobic interactions
  - Disulfide bonds
- Process of trial and error
- Tests many configuration
- This results in a correctly folded protein with a low-energy state

Denaturation of Protein

Secondary, Tertiary and Quaternary structures of protein molecules broken down.

- ✓ Primary structure is not altered
- ✓ Unfolding & Disorganization of Protein
- ✓ Decreases the solubility
- ✓ So protein precipitated
- ✓ Loss of biological activity.
- ✓ Denatured proteins are re-natured when the physical agent is removed. (less possible)

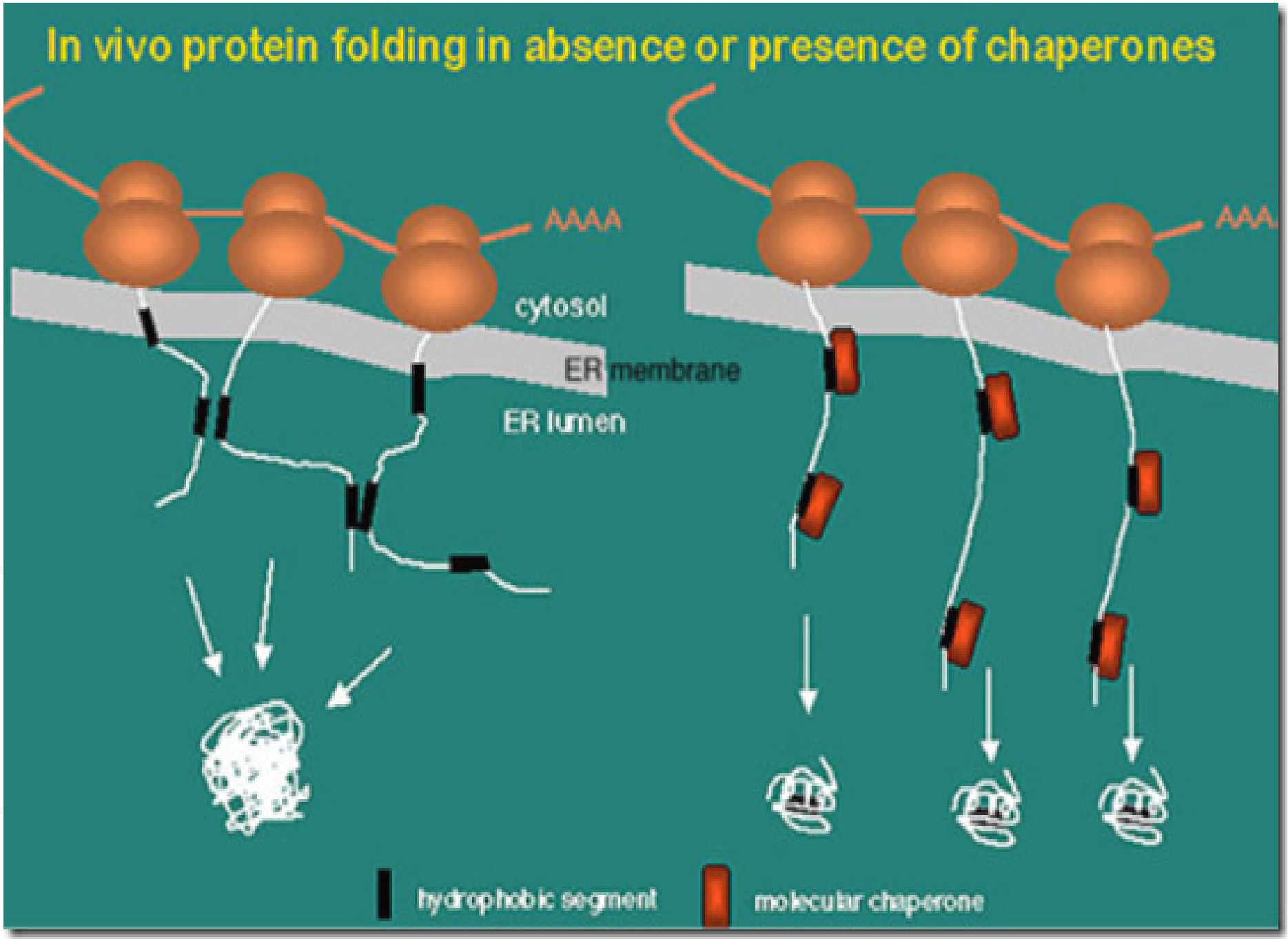
Factor which can does “Denaturation”

- Heating = Cooking, Heat coagulation test
- Urea = Renal Failure
- X-ray
- Ultra - violet rays
- High pressure = Cooking
- Organic solvent = Alcohol as antiseptic
- Metals = Mercury poisoning
- Acid & Alkali = Digestion

Chaperones = In protein folding

- **Specialized group of proteins for the proper folding.**
- Chaperones = "Heat Shock" proteins
- Information for Correct folding is in primary structure of protein.
- Interact with the polypeptide at various stages
- Some chaperones keeps protein unfolded until its synthesis is finished.
- Some Chaperones tends to fold so that their vulnerable, exposed regions do not become tangled(mixed).

- Some denatured protein can not be re-folded properly,
  - Because folding starts with stages its synthesis
  - Folding does not wait for synthesis be totally completed.


In vivo protein folding in absence or presence of chaperones
AAAA
cytosol
ER membrane
ER lumen
hydrophobic segment
molecular chaperone

4.QUATERNARY STRUCTURE

- Certain polypeptides aggregate
- form one functional protein
- known as quaternary structure.
- The protein will lose its function when the subunits are dissociated.
- Stabilised by
    - hydrogen bonds
    - electrostatic bonds
    - hydrophobic bonds
    - van der Waals forces.

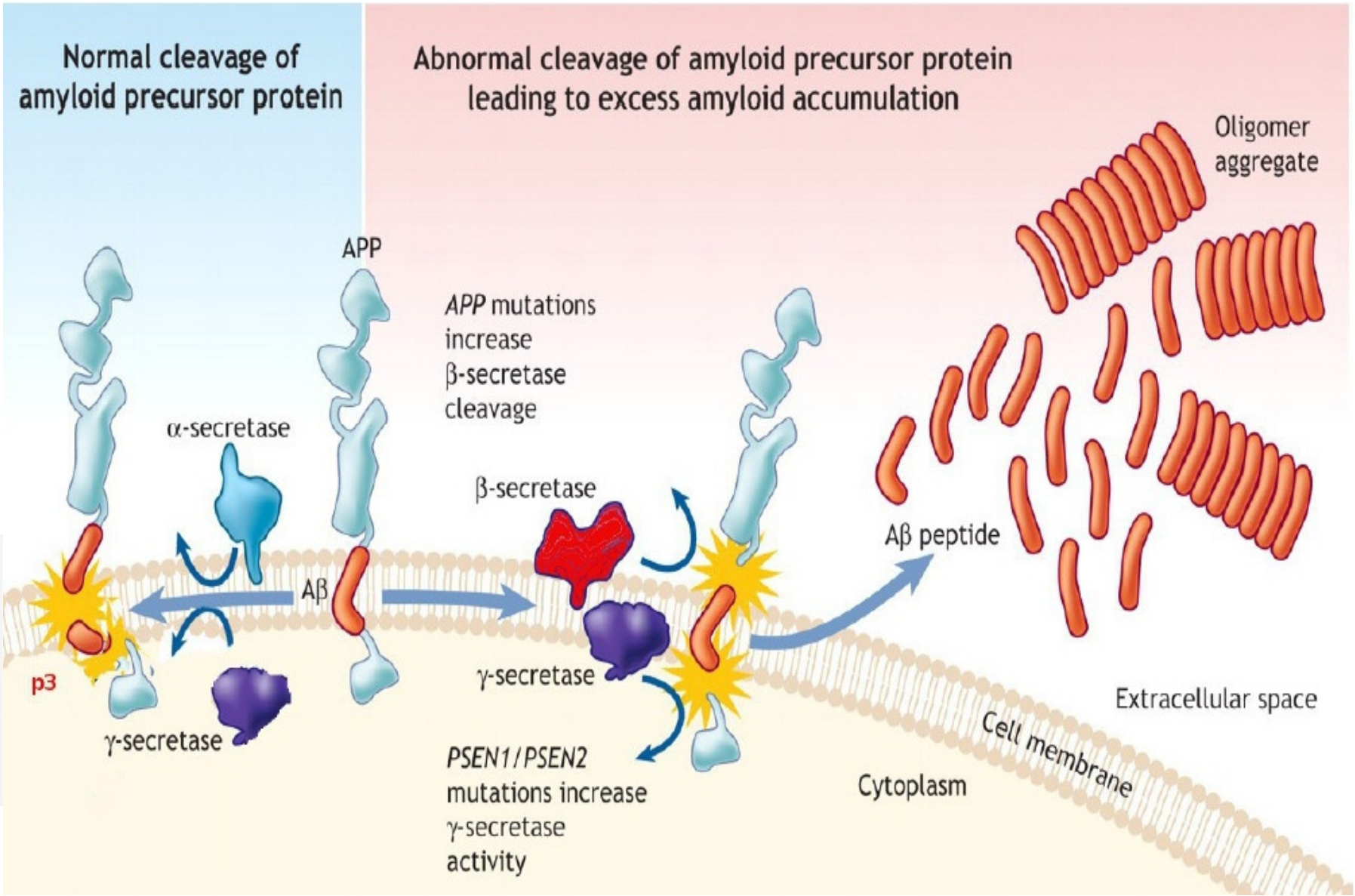
- Depending on the number of the monomers
  - Dimer(2)
  - Tetramer(4).
- Each polypeptide chain is termed as Subunit or Monomer.
- For example,
  - Hemoglobin = 2 alpha and 2 beta chains
  - Immunoglobulin G = 2 heavy & 2 light chains
  - Creatine kinase is dimer
  - Lactate dehydrogenase is a tetramer.

Protein Misfolding

- Complex process
- Trial & Error process
- Misfolded proteins are usually tagged and degraded.
- Misfolding of proteins may occur
  - Spontaneously
  - By mutation in a particular gene

- Accumulation of misfolded proteins can cause disease
- E.g. Amyloidosis

Amyloidosis



Amyloidosis = Alzheimer's disease

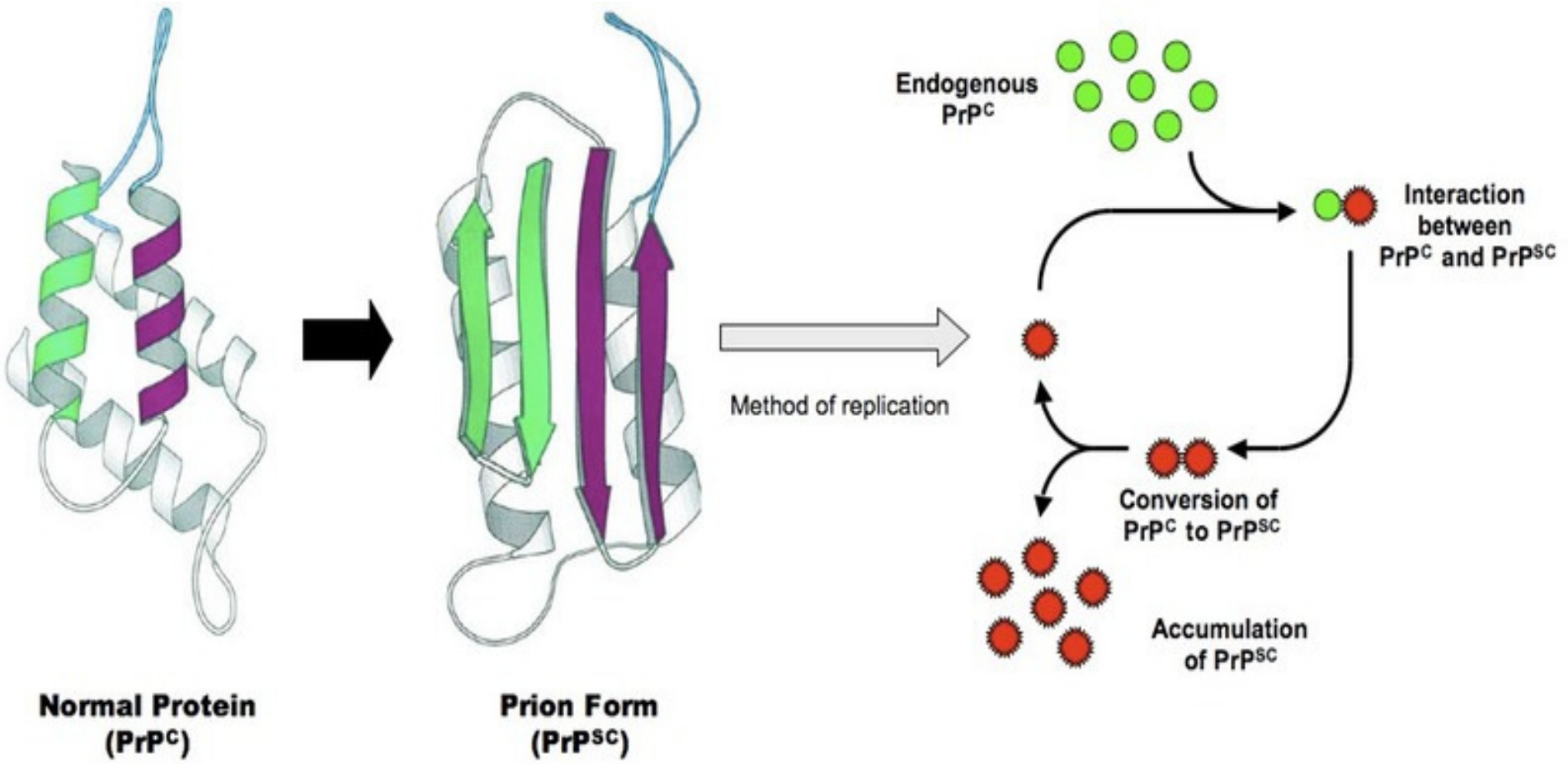
- Accumulation of amyloid β (Aβ), 40 – 42 A.A. = " Amyloids"
- Neurodegenerative disorder = "**Alzheimer disease**".
- Found in the brain parenchyma & around blood vessels.
- Neurotoxic & leading to the cognitive impairment
- Abnormal proteolytic cleavage
- Formation of long fibrillar protein = β-pleated sheets.

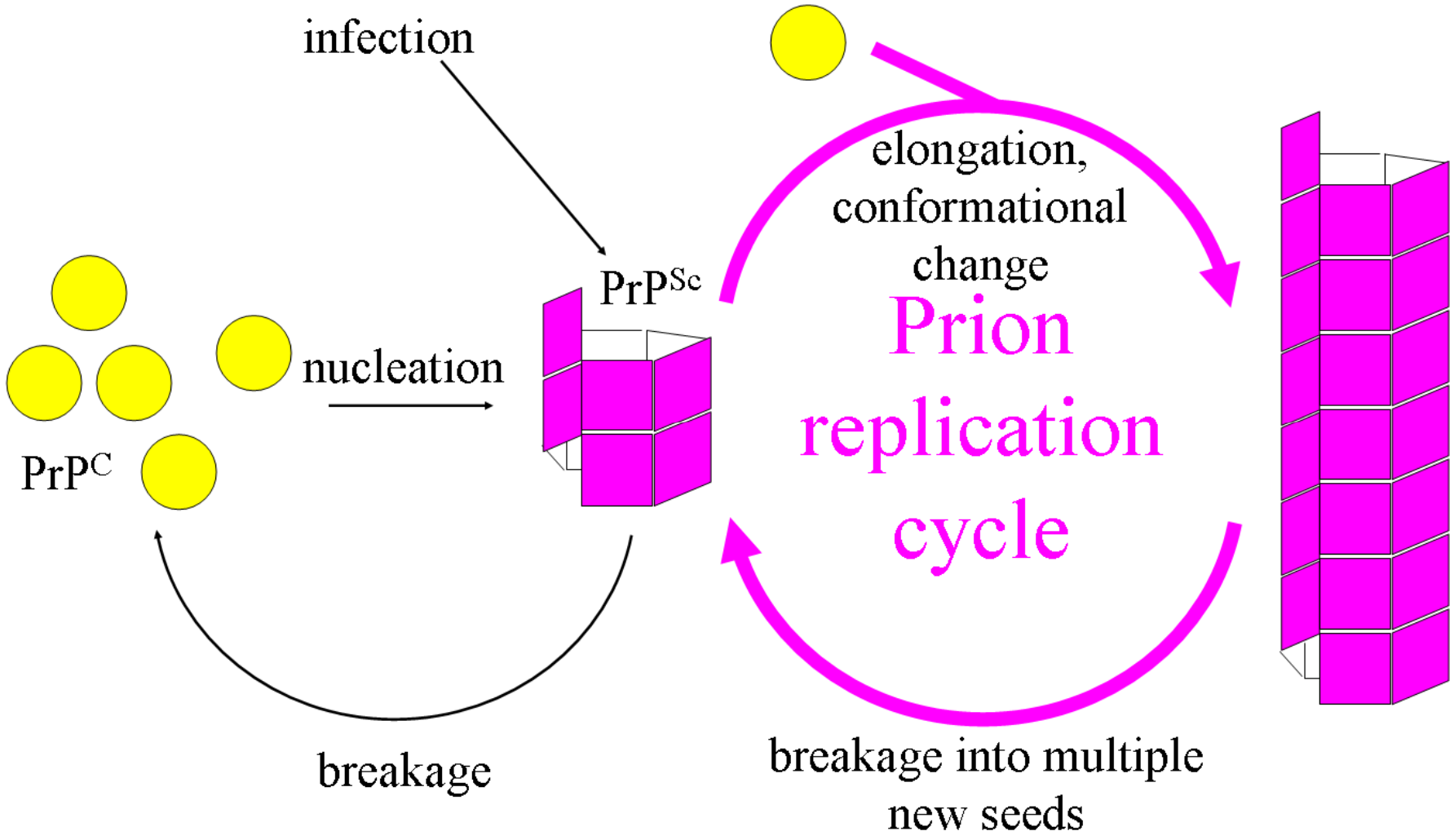
- Second factor = Accumulation of neurofibrillary tangles
- Tangle protein = Role in assembly of the microtubular structure.
- Abnormal form of tangle protein = **tau (τ) protein**
- Abnormal neurofibrine actions

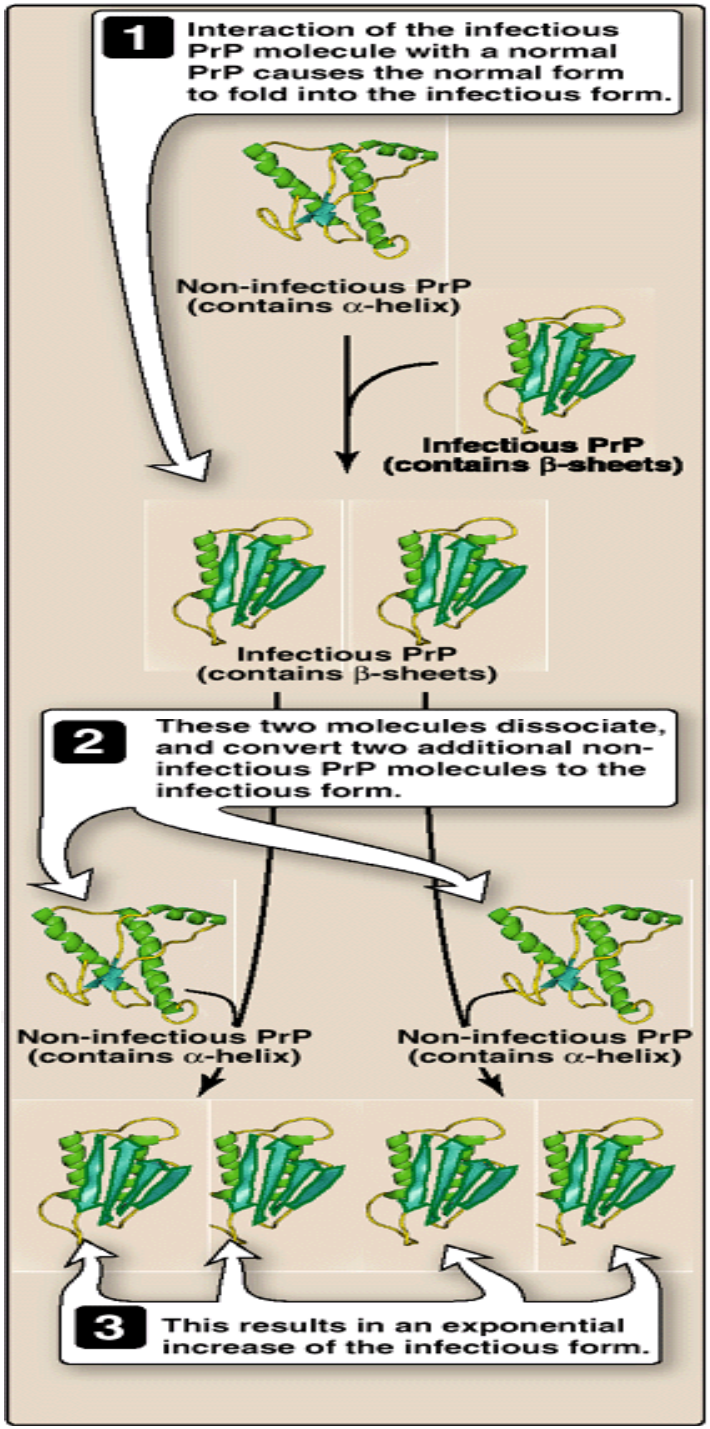
Prion Protein(PrP) & Prion disease

- PrP is a host protein.
- Present on the surface of neurons and Glial cells.
- **Infective Prion Protein**
  - No any change in amino acid and gene sequences
  - No change in primary structure differences
  - No any alternate post-translational modifications
  - changes in the three-dimensional conformation
  - number of α-helices noninfectious PrP (PrP^{c}) are replaced by β-sheets in the infectious form(PrP^{sc}).
  - highly resistant to proteolytic degradation
  - Accumulation of insoluble aggregates of fibrils

Prion Protein




infection
PrP^Sc
nucleation
PrP^C
elongation,
conformational
change
Prion
replication
cycle
breakage
breakage into multiple
new seeds


1
Interaction of the infectious PrP molecule with a normal PrP causes the normal form to fold into the infectious form.
Non-infectious PrP (contains α-helix)
Infectious PrP (contains β-sheets)
Infectious PrP (contains β-sheets)
2
These two molecules dissociate, and convert two additional non-infectious PrP molecules to the infectious form.
Non-infectious PrP (contains α-helix)
Non-infectious PrP (contains α-helix)
3
This results in an exponential increase of the infectious form.

Prion Protein(PrP) & Prion disease

- The Prion Protein (PrP) as the causative agent of
  - Transmissible Spongiform Encephalopathies (TSEs)
  - Creutzfeldt Jakob disease in humans
  - Scrapie in sheep
  - Bovine spongiform encephalopathy in cattle
- Infective agent is thus an altered version of a normal protein
- Which acts as a “template” for converting the normal protein to the pathogenic conformation.
- TSEs are invariably fatal
- No treatment is currently available

Can

dsRNA (double stranded RNA)

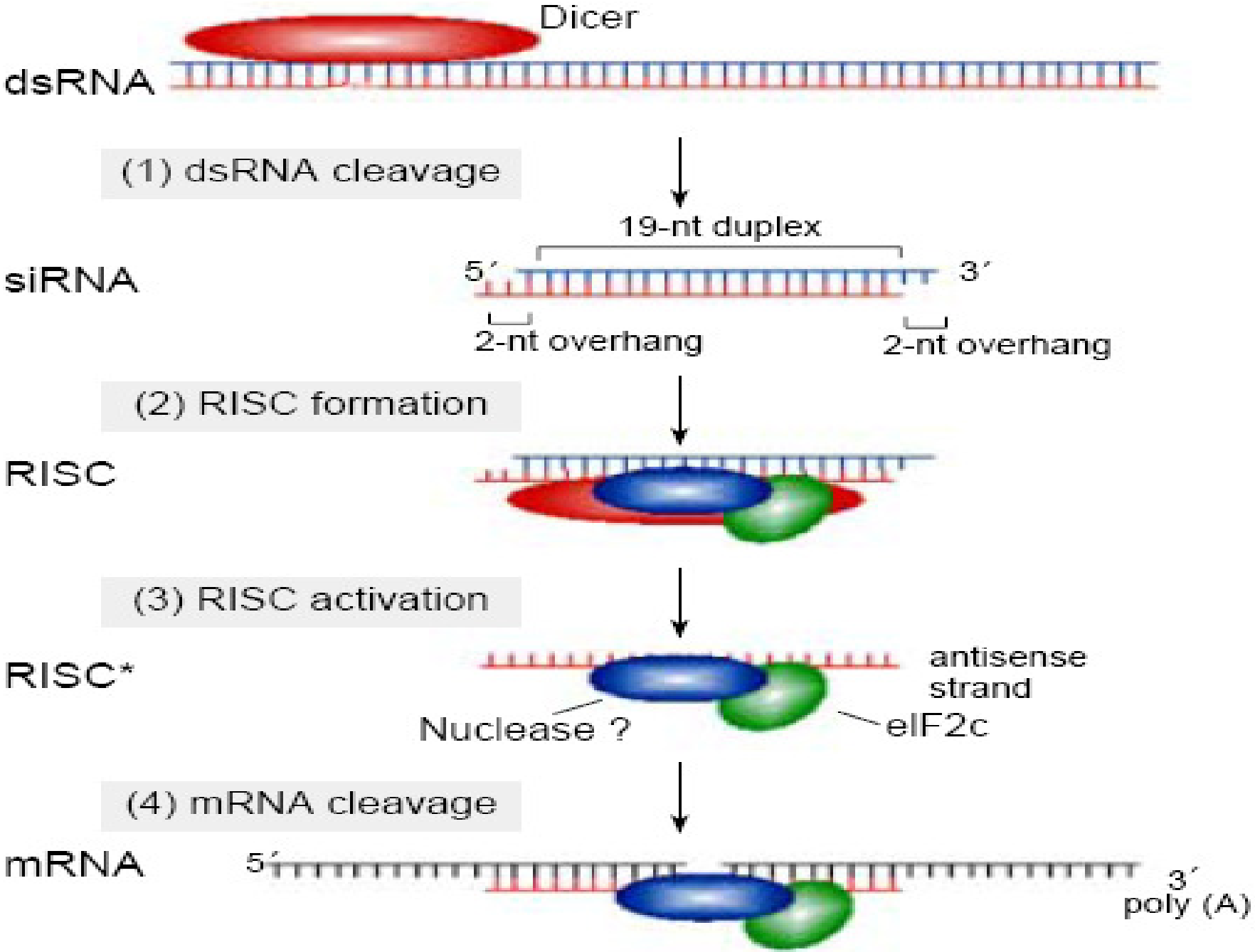
&

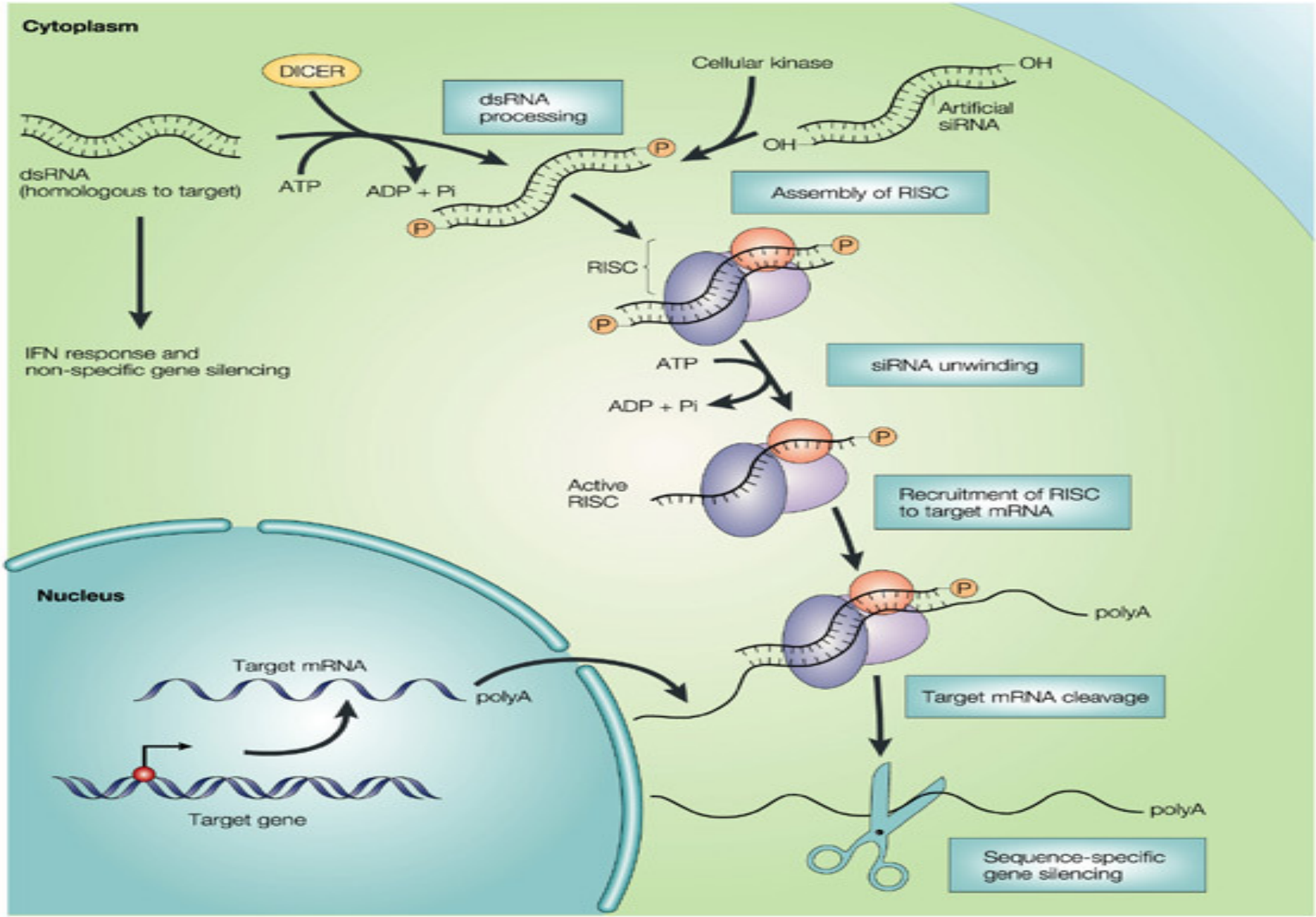
RISC (RNA Induce Silencer Complex)

be useful to treat

Prion Disease

?


Dicer
dsRNA
(1) dsRNA cleavage
19-nt duplex
siRNA
5′
3′
2-nt overhang
2-nt overhang
(2) RISC formation
RISC
(3) RISC activation
RISC*
antisense
strand
Nuclease ?
eIF2c
(4) mRNA cleavage
mRNA
5′
3′
poly (A)


Cytoplasm
DICER
dsRNA processing
Cellular kinase
OH
Artificial siRNA
OH
dsRNA (homologous to target)
ATP
ADP + Pi
P
P
Assembly of RISC
RISC
P
P
IFN response and non-specific gene silencing
ATP
siRNA unwinding
ADP + Pi
P
Active RISC
Recruitment of RISC to target mRNA
Nucleus
P
polyA
Target mRNA
polyA
Target mRNA cleavage
Target gene
polyA
Sequence-specific gene silencing

CLASSIFICATION OF PROTEINS

BASED ON FUNCTIONS

a) Catalytic proteins = Enzymes

b) Structural proteins = Collagen, Elastin

c) Contractile proteins = Actin, Myosin

d) Transport proteins = Hemoglobin, Myoglobin, Albumin, Transferrin.

e) Regulatory proteins = ACTH, Insulin, Growth hormone.

f) Genetic proteins = Histones

g) Protective = Immunoglobing, Interferon, Clotting factors

Type of Protein based on Composition & Solubility

Simple proteins (Only Polypeptide Chain)

1. Albumins

2. Globulins

3. Protamines More of arginine and lysine = Strongly basic

E.g. protamine zinc insulin

4. Prolamines:

Proline but lack in lysine.

E.g. zein from corn, gliadin of wheat

5. Lectins:

High affinity to sugar groups.

human blood group A1 RBCs.

6. Scleroproteins:

Form supporting tissues.

E.g. Collagen = Cartilage & Tendon, keratin of hair,nail

Conjugated proteins (Protein + Non Protein group)

1. Glycoproteins: = Blood group antigens
2. Lipoproteins: = HDL, LDL,IDL,VLDL
3. Nucleoproteins: = Histones.
4. Chromoproteins = Hemoglobin , Flavoprotein
5. Phosphoproteins: = Casein of Milk , Vitellin of Egg yolk.
6. Metalloproteins: = Hemoglobin(iron), Cytochrome(iron), Tyrosinase(copper) , Carbonic anhydrase(zinc)

Derived proteins

- Degradation products of the native proteins.
- Progressive hydrolysis of protein results in smaller & smaller chains
- protein→ peptones→ peptides→ amino acids

Type of Protein Based on Nutritional Value

Nutritionally rich proteins:

- Complete proteins or First class proteins.
- Contains all the essential amino acids in required proportion.
- E.g. Casein of milk.

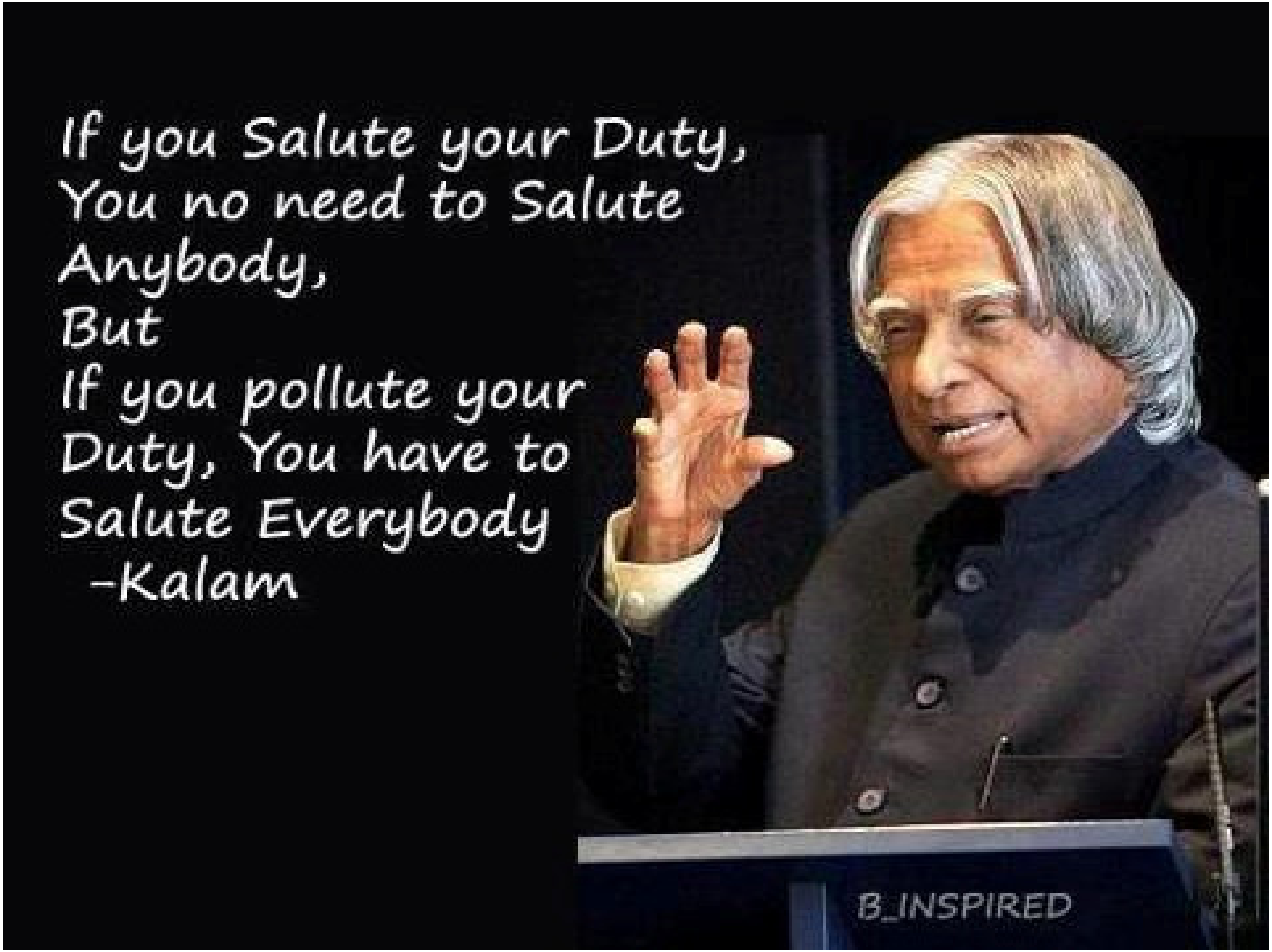
Incomplete proteins:

- Lack one essential amino acid.
- Cannot promote body growth but able to sustain growth in adults.
- Proteins of Pulses = Deficient in Methionine

  Proteins of Cereals = Deficient in Lysine.
- If both are combined in the diet , adequate growth may be obtained.

Poor proteins:

- Lack in many essential amino acids
- Zein of Corn lacks Tryptophan and Lysine.


If you Salute your Duty,
You no need to Salute
Anybody,
But
If you pollute your
Duty, You have to
Salute Everybody
-Kalam
B_INSPIRED