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, $R$, determine the properties of 20 amino acids.
- 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
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**

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>CTG</th>
<th>ACT</th>
<th>CCT</th>
<th>GAG</th>
<th>GAG</th>
<th>AAG</th>
<th>TCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acid</td>
<td>Leu</td>
<td>Thr</td>
<td>Pro</td>
<td>Glu</td>
<td>Glu</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>CTG</th>
<th>ACT</th>
<th>CCT</th>
<th>GTG</th>
<th>GAG</th>
<th>AAG</th>
<th>TCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acid</td>
<td>Leu</td>
<td>Thr</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
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\textsuperscript{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”
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 Sequence 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)
Recombinant Insulin

Insulin Aspart, Lispro, & Glulisine Structure

A-CHAIN

1  5  10  15  20Asn

B-CHAIN

1 Asn  5  10  15  20  25  Pro-Lys  30

LYS

Insulin Glulisine

ASp

Insulin Aspart

LYS-Pro

Insulin Lispro

GLU

Insulin Glulisine
Recombinant Insulin

Insulin Glargine & Detemir Structure

A-CHAIN

1  5  10  15  20 Asn

B-CHAIN

1  5  10  20  25  30

Glargine: Asparagine substituted to Glycine

Detemir: B30 deleted & fatty acid added

14-C fatty acid

Glargine: 2 Arginines added to B30
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 pl of insulin ?
  – Decrease
  – Increase
  – Unchange
Extra 2 Arginine....Added in Human Insulin

• Human insulin has isoelectric pH of 5.4. what can be pl 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 pl (6.7) is nearer to physiological pH (7.4) than human insulin pl (5.4). So what will be effect on it’s solubility?
  – More soluble
  – Less soluble (precipitation)
Glargine Insulin (Lantus)

- Injection site
- Lantus® (clear solution) pH 4
- pH 7.4
- Microprecipitation
- Depot
- Slow release of Lantus®
- Hexamers
- Dimers
- Monomers
- Capillary membrane
- Insulin in blood
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”
Characteristics 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 Å
  - single bond(1.42Å°)
  - double bond(1.27Å°).
- Angles of rotation known as Ramanchandran angles
Characteristics of the peptide bond

Trans peptide bond

Cis peptide bond

Peptide bonds in proteins
- Partial double-bond character
- Rigid and planar
- Trans configuration
- Uncharged but polar
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
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

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):**

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>CTG</th>
<th>ACT</th>
<th>CCT</th>
<th>GAG</th>
<th>GAG</th>
<th>AAG</th>
<th>TCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acid</td>
<td>Leu</td>
<td>Thr</td>
<td>Pro</td>
<td>Glu</td>
<td>Glu</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

**HBB Sequence in Mutant Adult Hemoglobin (Hb S):**

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>CTG</th>
<th>ACT</th>
<th>CCT</th>
<th>GTG</th>
<th>GAG</th>
<th>AAG</th>
<th>TCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acid</td>
<td>Leu</td>
<td>Thr</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Lys</td>
<td>Ser</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>
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.
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.
<table>
<thead>
<tr>
<th>Noncovalent forces</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrostatic forces</td>
<td>Attraction between opposite charges</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
<td>Hydrogen shared between electronegative atoms (N,O)</td>
</tr>
<tr>
<td>Van der Waals forces</td>
<td>Fluctuations in electron clouds around molecules oppositely polarize neighboring atoms</td>
</tr>
</tbody>
</table>
| 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 | ![H°O] (\(\delta^+\text{H}°\delta^-\text{O}°\))  
|                          |                                                                        | ![H°H] (\(\delta^-\text{H}°\delta^+\text{H}°\))  
|                          |                                                                        | ![O°H] (\(\delta^-\text{O}°\delta^+\text{H}°\))  

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
Hydrophobic interactions (clustering of hydrophobic groups away from water) and van der Waals interactions

Hydrogen bond

Disulfide bridge

Ionic bond

Polypeptide backbone
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 (−NH₃⁺) in the side chain of lysine.
van der Waals attractions between atoms (black) in contact

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

(A) Diagram of the alpha helix with labels for amino acid side chain, oxygen, hydrogen, carbon, and nitrogen.

(B) 0.54 nm scale for the alpha helix with carbon and nitrogen labels.

(C) Close-up of the alpha helix structure.

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 1.5Å°.
- Right handed.
- As A.A. in the proteins are of L-variety.
- Proline and hydroxyproline will not allow the formation of α-helix.
Example of Alpha Helix

1. Keratins = fibrous proteins in Hair & Nail
   • It’s rigidity is determined by the number of disulfide bonds

   (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

Hydrogen bonds between chains

Polypeptide chains almost fully extended

B

Antiparallel β-pleated sheet

C

Parallel β-pleated sheet
2. Beta-pleated sheet

- The distance between adjacent a.a. is 3.5Å.
- It is stabilised by hydrogen bonds.
- Adjacent strands in a sheet
  - Parallel
  - Anti parallel
- E.g. Flavodoxin, Carbonic anhydrase
  Triple helical structure in collagen
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
Zink finger Motif
The Leucine Zipper

Original concept

More correct
Figure 7-25. Molecular Biology of the Cell, 4th Edition.
3. TERTIARY STRUCTURE

- three dimensional 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 dimension.
- It is maintained by non-covalent interactions
  - hydrophobic bonds
  - electrostatic bonds
  - Van der Waals forces.
Body Parts

- Head
- Forehead
- Eyebrow
- Eye
- Ear
- Cheek
- Chin
- Neck
- Shoulder
- Waist
- Hand
- Finger
- Thumb
- Elbow
- Arm
- Hip
- Thigh
- Knee
- Calf
- Shin
- Ankle
- Foot
- Toe
- Arch
- Ball
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 do “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

- 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

Normal cleavage of amyloid precursor protein

Abnormal cleavage of amyloid precursor protein leading to excess amyloid accumulation

α-secretase

APP

β-secretase cleavage

γ-secretase

APP mutations increase

PSEN1/PSEN2 mutations increase γ-secretase activity

Aβ peptide

Oligomer aggregate

Extracellular space

Cell membrane

Cytoplasm
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 (PrP⁰) are replaced by β-sheets in the infectious form (PrP²⁰).
- Highly resistant to proteolytic degradation
- Accumulation of insoluble aggregates of fibrils
Infection → PrP^Sc → Elongation, conformational change → Prion replication cycle → Breakage into multiple new seeds → Breakage → Nucleation → PrP^C → Infection
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)

2. These two molecules dissociate, and convert two additional non-infectious PrP molecules to the infectious form.

Non-infectious PrP (contains α-helix) → Infectious PrP (contains β-sheets)

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?
dsRNA → Dicer

(1) dsRNA cleavage → 19-nt duplex

siRNA

(2) RISC formation

RISC

(3) RISC activation

RISC* → Nuclease ?

(4) mRNA cleavage
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 = Immunoglobulin, 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
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