HANDOUT

LECTURE-05

AMINO ACIDS AND STRUCTURAL LEVELS OF PROTEINS

Slide 1

Before we move on to proteomics and discuss what are all techniques and concepts are involved in proteomics, let's first start the basic concepts on proteins. So today we will talk about proteins, amino acids and structural levels of proteins.

Slide 2

Lecture outline

We will first talk about:

- Proteins and their functions
- Amino acids, the building blocks
- Different level of protein structure- primary, secondary, tertiary and quaternary.

Slide 3

Proteins and their functions

'Protein' was derived from Greek word "Proteios" which means "of the first rank" or very important. This term was coined by J. Berzelius in 1833.

These are linear polymers built of monomers (amino acids). These are most versatile macromolecules in living systems. They are crucial for various essential functions for all the biological processes and they play very critical role both from structural and functional point of view.

Slide 4

If you look at the central dogma starting from DNA to RNA to protein, the proteins can transform the sequence to functional information.

Proteins can play a wide range of functional roles because of their different functional groups, which can accounts for various protein functions and its activity.

Protein-protein or protein-other biomolecular interaction, they are generated because of synergetic capability of these protein which cannot be obtained from the any individual protein.

Slide 5

Proteins can perform various types of functions –

- Enzymatic catalysis - the enzymes catalyze biochemical reactions by increasing rates of reactions.
- Transport & storage - proteins transport small molecules -like oxygen, iron.
- Proteins are involved in the movement - muscle contraction, if you talk about microorganisms, in bacterial the chemotaxis.
- They are responsible for mechanical strength. For example skin and bones: collagen; Hair: keratins, all of these are different example of mechanical strength.
- Protein are also present as immunoglobulins responsible for the immunity; and antibodies are used for various type of protein-protein and protein-ligand interactions.
- Growth and differentiation – Transcription factors: gene expression during growth and development for exampled growth factors: nerve growth factor; hormones such as insulin.

Slide 6

Amino acids:

- These are the building blocks of proteins. Amino acids constitute the basic monomeric units of proteins, which are joined together by peptide bonds.
- The twenty standard amino acids can be arranged in several ways giving rise to numerous proteins having different structures and properties.
- The diversity & versatility of 20 amino acids enables range of protein functions. Due to the side chains which can vary in – Size, shape, H-bonding capacity, hydrophobic characters, charge and chemical reactivity, proteins perform much diverse function as compared to the DNA.

Slide 7

You have already studied about different amino acids in your undergraduate education. I will again try to refresh you on some of those concepts in a nutshell. So here I have shown various amino acids, which are non-polar and aliphatic R groups.

- Starting with the glycine if you see on the left side top most which is the simplest and a chiral.
- Now, next is alanine which contain methyl group.
- Proline which has aliphatic side chain, proline has very unique feature. It has no free amino group and the side chain is bonded to the α-carbon of the α-carbon atoms. The ring structure provides more conformational restriction and therefore proline plays very crucial role in unique properties of many proteins.
- Valine is branched chain amino acid.
- Leucine on the left hand side bottom panel that is hydrophobic amino acid with isobutyl R-group.
- Isoleucine it also has the hydrophobic amino acid characteristics and it contains chiral side chain.
- The last in the group is methionine which includes a thioether group. Again, there are only two amino acids which contain sulfur play very critical roles.

Slide 8

- Our next category is polar, uncharged R- groups. Let's start with the serine which resembles in the structure -like alanine but contains hydroxyl group. Threonine it resembles in the structure -like valine and it contains hydroxyl group. It has an addition asymmetric center.
- Cysteine is similar to serine but it contains a sulfhydryl or thiol group. Two cysteine molecule forms a disulphide bond.
- Let's talk about asparagine which shown in the left side lower panel. It contains corboxy-amide side chain as a functional group.
- Glutamine, the side chain called as amide of glutamic of acid which is formed by replacing the side chain hydroxy group glutamic acid with an amino functional group.

Slide 9

- The next category is positively charged R groups. Three amino acids here are lysine, arginine and histidine. Lysine is a base, it contains capped primary amino group. whereas arginine contains guanidium group. Histidine has a functional imidazole group which is an aromatic ring that can be positively charged. Histidine play a very critical role in many enzymatic activities.

Slide 10

- Next group is negatively charged R groups- aspartate and glutamate or aspartic acid and glutamic acid. The name aspartic acid or glutamic acid is so because at

physiological pH their side chains lacks a proton present in acid form. Therefore, these amino acids are negatively charged.

- Aspartate is a carboxylate anion of aspartic acid known as aspartate whereas carboxylate anions and salts of glutamic acid are known as glutamates.

Slide 11

- Next category is aromatic R groups, in this category- three amino acids phenyalanine, tyrosine and tryptophan. Phenyalanine contains pheny ring, tyrosine has one reactive hydroxyl group and tryptophan contains indole ring, two rings which are fused. Now if you look at the hydrophobic or hydrophilic characteristic phenylalanine is hydrophobic whereas tyrosine and tryptophan are hydrophilic due to the side chain containing hydroxyl and –NH reactive group.

Slide 12

- Aromatic amino acids have a unique property, they can absorb UV light.
- So all the amino acids we just discussed, Tryptophan, Tyrosine and Phenyalanine can absorb UV light. Tryptophan absorption max at 280 nm, tyrosine absorption max is at 276 nm, while Phenylalanine absorbs light less strongly & at shorter wavelengths. Light absorption at 280 nm can be used for protein concentration determination.

Slide 13

- Amino acids are the building blocks or monomers that make up proteins. They consist of a central alpha carbon atom bonded covalently to an amino group, a carboxyl group, a hydrogen atom and a variable side chain, also called the R group.
- Amino acids are the basic monomeric constituents of proteins found in varying amounts depending upon the type of protein. They are classified based on the properties of their side chains or R groups which vary in size, structure and charge. The polarity of the side chains is one of the main basis for classification.
- Amino acids having non-polar, aliphatic side chains include glycine, alanine, proline, valine, leucine, isoleucine & methionine. Essential amino acids are those that cannot be synthesized *de novo* in the organism and therefore must be included in the diet. Non-essential amino acids on the other hand, can be synthesized from various precursors.
- Serine, threonine, asparagine, glutamine & cysteine consist of polar but uncharged side chains.

- Lysine, Arginine, Histidine - These amino acids have positively charged side chains.
- Aspartic acid and glutamic acid are polar and negatively charged amino acids. Tryptophan, tyrosine and phenylalanine are all essential amino acids having an aromatic side chain.

Slide 14

- After having discussed the different types of amino acids, let's look at the basic constituents of amino acids and different isoforms it can form.
- So amino acids having four different groups connected to α-carbon atom, it can form two mirror images which can exist in L/D isolmers which are shown in the slide here.
- The α-amino acids are chiral.

Slide 15

- There could be R or S configuration in the amino acids depending upon the priority groups but only L amino acids, having the S configuration are found in proteins.
- A counterclockwise direction from highest to lowest priority groups is indicative of chiral center with S- configuration.

Slide 16

- So the isomerism property of amino acids will be discussed in following animation. Before learning about the isomerism let us first know what is chirality. The term 'chirality' arises from the Greek term cheir meaning 'handedness'. Just - like the two hands are non-superimposable mirror images of each other, amino acid molecules are also non-superimposable due to their chiral alpha-carbon centre.
- All amino acids except glycine contain an asymmetric centre that makes them chiral in nature due to which they can rotate the plane of polarized light. The two enantiomers, designated as D and L, rotate the plane of polarization in opposite directions.
- The two enantiomers of amino acids are non-superimposable mirror image due to the spatial arrangement of four different groups about the chiral carbon atom. Rotation of either isomer about its central axis will never give rise to the other isomeric structure.

Slide 17

- Ionization state of an amino acid- the ionization state of an amino acid varies with its pH. In the acidic solution (if you follow the slide from left to right) amino group is protonated (-NH3+), carboxylic group is non-dissociated (-COOH).
- At neutral pH amino acids exist as dipolar ions (or zwitterions) amino group is protonated (-NH3+),Carboxyl group is deprotonated (-COO-). Now this dipolar form exist till pH=9. Now when you move to the basic pH protonated amino group loses its proton and forms (-NH2).

Slide 18

Lets now talk about peptide bond which links amino acids to form polypeptide protein.

The α-carboxylic group of one amino acid linked to α-amino group of another amino acid. As you can see here these two amino acids are forming a peptide bond and peptide bond formation accompanies loss of water molecule.

Slide 19

- When many amino acids are linked together they form a polypeptide as you can see in this slide.
- Multiple peptide bonds are present.

Slide 20

- The polypeptide chains conformationally restricted therefore the peptide bond is planer.
- Amino acid pairs their link by the peptide bond and all the six atoms lie in the same plane as you can see here- α-carbon, carbon, oxygen, nitrogen, hydrogen and another α-carbon.

Slide 21

- Peptide bond can be stabilized by the resonance structure.
- Peptide bond is rigid because of its partial double bond character which arises due to resonance structure present in peptide bond.

Slide 22

- There could be two forms- cis form or trans form. But peptide bonds in proteins exist in the trans form. If you see in the top panel, the trans configuration shows two C-α on the opposite side of the peptide bond.
- This configuration is allowing less steric clashes whereas if you look at the bottom panel the cis configuration, there are two C-α on the same side of the peptide bond so there is more probability of having steric clashes.
- Therefore peptide bonds in the protein exist in the trans form.

Slide 23

- Now, proline is a unique amino acid as we discussed earlier.
- Peptide bonds of Proline can exist in both cis and trans form. So as you can see here it can avoid steric clashes and both cis and trans configurations are possible.

Slide 24

So some of the concepts of peptide bond will described in the following animation-

Animation 1

- Amino acids are the building blocks or monomers that make up proteins.
- Amino acids are oriented in a head-to-tail fashion and linked together such that the carboxyl group of one amino acid combines with the amino group of another.
- Two amino acids joined together by means of such a condensation reaction with the loss of a water molecule forms a dipeptide. Many such amino acids linked together form a polypeptide.
- The peptide bond is rigid due to its partial double bond character arising from resonance structures. However, the bonds between the a-carbon and amino and carboxyl groups are pure single bonds that are free to rotate.

Slide 25

Primary Structure of Proteins

- Amino acids constitute the basic monomeric units of proteins, joined together by peptide bonds. The twenty standard amino acids can be arranged in several ways giving and therefore can give rise to numerous proteins having different structures and properties.
- So primary structure refers to the sequence of the amino acid.

Slide 26

- As you can see on the left side, different amino acids can come together and linear sequence of amino acid constitutes the primary structure with lot of water molecule.

Slide 27

- What is the directionality of primary structure? The polypeptide chains has polarity so one end is α-amino group and other end is α-carboxylic group. The amino end marks the start of a polypeptide chain.

Slide 28

- So what is the significance of amino acid sequence, which determines a protein's primary structure? So amino acid sequence is essential for elucidation of its mechanism of action
- An enzyme's catalytic action can be determined
- It determines 3-D structure of the proteins
- It links the functional 3-D protein structure & genetic obtained from message from DNA
  Amino acid alteration can produce different type of protein abnormalities, for exampleg. cystic fibrosis – change in single amino acid can give rise to abnormality
- Sequence can also tell us an evolutionary history of protein

Slide 29

Now let's move on to concept of Ramachandran plot but before that it is important two know Phi (φ) and Psi (ψ) angles. As you can see here in the slide, rotation of two single bonds adjusts structure of each amino acid in polypeptides.

–Phi (φ) angle – between nitrogen and C_α

–Psi (ψ) angle – between C_α and carbonyl carbon φ and ψ angles determines path of polypeptide chain

- All combinations of φ and ψ angles are not possible
- So allowed combinations can be viewed on 2-D plot known as Ramachandran diagram or Ramachandran plot.
- Many combinations are not allowed due to steric collisions between atoms. Therefore, steric exclusion is a powerful organizing principle for such plot.

Slide 30

Now, you can see it more clearly in this slide. The more favorable regions are shown in dark green color and less favored regions are shown in light green.

Slide 31

The structural levels of proteins-the primary structure, few concepts will be discussed in following animation.

Animation 2

- Amino acids are joined together in a head-to-tail arrangement by means of peptide bonds with the release of water molecules.
- This linear sequence of amino acids constitutes the primary structure.

Slide 32

Secondary Structure

- This refers to locally folded regions.
- The folding of polypeptide/protein chain into regular structures -like a-helices, b-sheets, turns and loops, all this represents the secondary structure.

Slide 33

Let's first start with α-helices-

- Proteins have variable helix content. An alpha helix is a rod--like structure.
- It has a main chain and a side chain. The main chain is tightly coiled around helical axis and side chains is extended outward away from helical axis.
- As you can see in the ribbon diagram on the right side and ball and stick diagram on the left side. So specific hydrogen bonds stabilize helical core.

Slide 34

- The a-helix can be stabilized by H-bond so the –CO group of each amino acid with NH group of amino acid which are four residues ahead in sequence, they form a these hydrogen bond,as you can see in the figure here.

Slide 35

There are special types of α-helices where two α-helices can wrap up or three α-helices can come together. So the first example is

α-Keratin

- where two α helices wrap to form a stable structure
- α keratin is primary component of hair
- It consists of two helices coiled around each other and forms a left handed super-helix known as α *coiled-coil*

Collagen

- It is a fibrous component of skin & bone.
- it is most abundant protein in mammal
- It contains 3 helical polypeptide chains

Slide 36

- β sheets is another common, periodic structural motif discovered by Pauling and Corey. It is fully extended structure (un-like tightly coiled α-helices). It can have two directions parallel (running in same direction) or anti-parallel (running in opposite directions).

Slide 37

Another category of secondary structures turns and loops. These are more elaborate structures loops or omega turns, which also perform chain reversal.

• Loops are rigid, well-defined

• Loops are not periodic structures

• Turns and loops are present on surfaces

• Participate in various properties of proteins and other bimolecular interactions

Slide 38

The differences in α-helix and β-sheet are summarized in this slide.

- The α-helix, polypeptide chain is tightly coiled whereas in β-sheet, it is fully extended.
- α-helix is rod -like structure and β-sheet is sheet -like structure. In α-helix, axial distance between adjacent atoms is 1.5 A whereas it is 3.5 A in β-sheet. In α-helix, H-bond between NH and CO groups in same polypeptide chains whereas in β-sheet, it is in different polypeptide chains. α-helix examples include ferritin, keratin and collagen in β-sheet it is fatty acid binding protein.

Slide 39

- The folding of the primary structure into the secondary is governed by the permissible rotations about the phi and psi angles. Not all values of these angles lead to sterically favorable conformations. The Ramachandran's plot defines these regions of favorability.
- Amino acids along the polypeptide backbone interact through hydrogen bonds leading to secondary structures.
- The a-helix has intra-chain hydrogen bonds between the 'H' of NH and 'O' of CO in every 4th residue. Most alpha helices are right handed since this conformation is energetically more favorable.
- The amino acid proline which has a cyclic side chain does not fit into the regular alpha helix structure and thereby limits flexibility of the backbone. It is commonly referred to as the 'helix breaker'.
- Alpha helices can also wind around each other to form stable structures such that their hydrophobic residues are buried inside while their polar side chains are exposed to the aqueous environment.
- α-keratin, the major protein component of hair consists of two such coiled coils forming a left-handed superhelix.
- Collagen which is the fibrous component of skin, muscle etc. consists of three such coiled alpha helices.
- It has a characteristic recurring amino acid sequence of glycine-proline-hydroxyproline with glycine appearing at every 3rd residue.
- β-pleated sheets, discovered by Pauling & Corey, are another common secondary structure with periodic repeating units.
- They are composed of two or more polypeptide chains with their side chains oriented above & below the plane.
- A beta pleated sheet is an extended structure with hydrogen bonds between the chains stabilizing it.
- Amino acids in parallel beta-sheets, which run in the same direction, interact with two different amino acids on the adjacent strand through hydrogen bonds.
- Amino acids in antiparallel strands on the other hand interact with only one amino acid on an adjacent strand.
- Almost all proteins exhibit a compact, globular structure which is possible only if there are turns or loops between the various regions.
- β-turns, which are the most commonly observed turn structures, consist of rigid, well-defined structures that usually lie on the surface of the protein molecule and interact with other molecules.

Slide 40

Protein tertiary Structure

- Let's now move on to tertiary structure which refers to overall folded structure.
- NMR and X-ray crystallography provides detailed three dimensional structure.

Slide 41

- The three dimensional compactly folded structure of proteins and it represents overall organization of secondary structural elements in 3-D space.

Slide 42

- The numerous interactions which stabilize the tertiary structure of protein will be discussed using myoglobin as an example and describe some of the properties of tertiary structure in following animation.
- Amino acids located far apart on the polypeptide chain interact with each other by means of hydrogen bonds, electrostatic interactions, disulphide bridges etc., allowing the protein to fold three dimensionally in space, giving rise to the tertiary structure.
- Folding takes place such that the hydrophobic residues are buried inside the structure while the polar residues remain in contact with the surroundings.
- The tertiary structure of myoglobin, determined by John Kendrew, clearly revealed that the nature of amino acid side chains dictate their location in the tertiary structure.
- Hydrophobic residues are found buried inside the structure while the polar amino acids are found on the surface. 70% of the main chain of myoglobin is folded into alpha-helices with the rest being present in the form of turns & loops which are essential to give it a compact structure.

Slide 43

Quaternary structure

- It refers to interaction between individual protein subunit in a multi subunit complex.

Slide 44

- The quaternary structure represents final level of protein structure which is spatial arrangement of subunits and their interactions.

- The polypeptide chains assemble to form multisubunit structure wherein each polypeptide chain is a as subunit. The different examples-
- DNA binding Cro protein of bacteriophage λ
    - Simplest quaternary structure
    - homodimer: two identical subunits

Slide 45

Hemoglobin: Tetramer Quaternary structure

- It has two α subunits and two β subunits. The individual polypeptide chains come together and form quaternary structure.

Slide 46

- Properties of quaternary structure and some details about hemoglobin and comparisons with myglobin and hemoglobin will be discussed with following animation.
- Different subunits or polypeptide chains interact with one another and are held together by means of ionic, electrostatic, Van der Walls etc. interactions. Such multisubunit proteins having a quaternary structure which is the final level of protein structure.

Slide 47

After discussing various properties of amino acids and different level of structures of proteins, now let's just touch upon why understanding protein function is key to the biology.

- Diseases are the result of protein malfunction therefore all current drugs either alter protein function or are proteins themselves which demonstrates the significance of studying about proteins.
- Let me describe you this slide where I have shown the various biomolecules of central dogma DNA, RNA and protein. So for example if you look at the map here where IIT is located in Powai area so it is -like DNA which is genetic blueprint. It contains only information, now if you want to make a building in this area you have to define that area which is -like RNA molecule which is a molecular photocopy and it is used at the site of construction by the cell's contractors. Now proteins are -like building which you want to create on that site, these are the building blocks or building material. They are bricks and mortar and engine of biology. So it shows you how various types of biomolecules have their significance but it's the protein which ultimately defines the function.

Slide 48

The malfunction of proteins can be responsible for in various diseases and some of the diseases will be described in following animation-

Animation 3

A large number of mutations have been described in global genome:

- The mutation causing sickle cell anemia is single nucleotide substitution of A to T in the codon for amino six in the β chain. This change converts amino acid glutamic acid to valine in the corresponding amino acid sequence. Replacement of glutamic acid by valine creates a sticky hydrophobic contact point at position six of β chain. These sticky spots cause deoxyhemboglobin s molecules to associate abnormally with each other leading to clumping of the cell. Their oxygen carrying capacity is greatly reduced and these patients requires frequent transfusions.
- Thalassemia is the result of abnormalities in hemoglobin synthesis. Deficiency in synthesis of β globin result into β thalassemia. Mutation of a single base from G to A in an introns of β globin gene generates a new splice site. The resulting mRNA contains a stop codon further upstream and leads to premature translation termination thereby producing abnormal protein. Deficiency in α globin gene due to inactivation of one or all the α globin gene result into α thalassemia.
- Parkinson's and Alzheimer's disease: The structure of certain normal cellular proteins which are normally rich in alpha helical regions are believed to be converted into beta strand conformations which can further link with each other to form beta sheets aggregates known as amyloids. These amyloid plaques, found in the brain of patients with these diseases, are essentially made up of a single polypeptide chain known as Ab. Clinical manifestation include Neurodegenerative, tremors, stiffness, memory loss, confusion and dementia.
- Lathyrism: This is caused by ingestion of seeds from sweet pea *Lathyrus odoratus* leads to disruption of cross-linking in the muscle protein, collagen. Collagen is an important structural protein having a triple helical structure. The cross-links formed are due to the oxidation of some lysine residues by the enzyme lysyl oxidase. β-aminopropionitrile, present in abundance in sweet pea, deactivates this enzyme by binding to its active site. Clinical manifestation- Reduced cross-linking causing increased fragility of the collagen fibers.

Slide 49

So in summary today we talked about proteins and its function. We refreshed our concepts of amino acids, which are the building blocks. We have talked about different levels of protein structure-Primary, Secondary, Tertiary, and Quaternary. We discussed in little more detail about myoglobin and hemoglobin: the model proteins. And then we briefly touched upon significance of studying proteins and how their malfunctions may result in various diseases.

We will continue our discussion on some basic concepts of proteins in the next class.