Slide
1

Chapter 5

Proteins: Biological Function
And Primary Structure

Slide
2

Polypeptides

- Proteins are linear polymers of Amino Acids (See Figure 5.1)
- Each amino acid is a "residue"
    - 2 residues = dipeptide
    - 3 residues = tripeptide
    - 12-20 residues = oligopeptide
    - Many residues = polypeptide
- Each polypeptide has an **amino terminal** and a **carboxy terminal** amino acid (though it may be chemically "blocked")

Insulin, with 51 amino acids, and a MW of 5,733 represents about the lower limit of what we would call a "protein" as compared with a "polypeptide". Glutamine Synthetase, at 600,000 MW is a relatively large protein, but still not one of the largest.

Slide
3

Geometry of Peptide Bond

- Trans configuration (Figure 5.2)
- Length = 0.133 nm, between double and single bond
- Double bond "character" (Figure 5.3)
- Six atoms of the peptide bond are coplanar (Figure 5.4)

Size of Proteins

- Proteins may consist of one or more polypeptide chains
  - Multimeric
  - Homomultimer
  - Heteromultimer (Cf Hemoglobin)
- Size range is enormous (Table 5.1)
  - 51 residues (5733 MW) for insulin
  - 468 residues (600,00 MW) for glutamine synthetase

Multimeric proteins have more than one polypeptide chain. If the chains are identical, it would be **homomultimer**. If the chains were different, it would be a **heteromultimer**. Hemoglobin is an example of a heteromultimer having two **alpha** and two **beta** chains making up a **tetrameric** structure.

Levels of protein structure

- Primary structure (sequence)
- Secondary structure (ordered regular structures along peptide bond—Figure 5.8)
  - Difference between **conformation** and **configuration** (Figure5.11)
- Tertiary structure (overall 3 dimensional structure)
- Quaternary structure (subunit organization)
  - See Figure 5.10—Hemoglobin—for example

Amino Acid Analysis of Proteins

- Hydrolysis in 6 N HCl at 110 ºC
- Correct for destruction of serine, threonine
- Correct for slow hydrolysis of hydrophobic amino acids
- Asn and Gln are hydrolyzed, therefore compositon reported as Asx and Glx
- Chromatographic analysis of amino acids (ion exchange or HPLC—See Fig. 4.21 and 4.22)

Primary Structure—Sequence

- Unique to each protein (Figure 5.6)
- Determines overall structure and function
- Encoded by nucleotide sequence in DNA
- Possible variations are essentially infinite

Illustration of infinite number of sequence variations

- Small polypeptide, 100 residues
- Possible sequences = $20^{100} = 10^{130}$
- **One molecule** of each possible sequence would occupy a volume of

10^{130} molecules x $\dfrac{1 \text{ mole}}{6 \times 10^{23} \text{ molecules}}$ x $10^4 \dfrac{\text{grams}}{\text{mole}}$ x $\dfrac{1 \text{ mL}}{1.2 \text{ grams}}$

= 1.4×10^{10} mL or 1.4×10^{107} L, which is much
   larger than volume of the known universe

Shape of Proteins

- Three general types of shapes (Figure 5.7)
  - Fibrous proteins
  - Globular proteins
  - Membrane proteins

Forces determining structure of proteins

- Primary structure—covalent bonds
- Secondary, tertiary, quaternary structure
  - Weak forces (H-bonding, ionic interactions, hydrophobic interactions, van der Waals interactions)

Different views of proteins

- See Figure 5.9 for example
  - 5.9a give primary structure (sequence)
  - 5.9b gives linear representation of peptide chain
  - Ribbon structure
  - Space filling structure

Biological Functions of Proteins

- Enzymes
- Regulatory proteins
- Transport proteins
- Storage proteins
- Conractile and motile proteins
- Structural proteins
- Scaffold proteins
- Protective proteins
- Exotic proteins

Enzymes are **catalysts** of reactions in metabolism. Virtually every reaction that occurs is catalyzed by a specific enzyme.

Regulatory proteins include **hormones** which activate signaling reactions in the cell, as well as **repressors** and **activators** which affect the expression of a gene.

Transport proteins include both those transporting substances from one tissue to another (as hemoglobin in oxygen transport) as well as those that transport a substance across a membrane (as the glucose transporter)

Storage proteins provide reservoirs of nutrients for seeds and eggs.

Contractile proteins include proteins involved in muscle contraction, cell division, and cell motility.

Structural proteins provide structural integrity to a cell or tissue, such a collagen in bones, teeth, cartilage, and connective tissue.

Scaffold proteins are involved in multimeric interactions with other proteins in certain signaling processes.

Protective proteins include antibodies and blood clotting proteins for example.

Exotic proteins are those with functions that don't fit neatly in the above categories.

Conjugated Proteins

- Proteins containing only amino acids are called **simple proteins**
- Non amino acid constituents of proteins are called **prosthetic groups**
- Protein minus its prosthetic group is called an **apoprotein**
- Protein plus its prosthetic group is called a **holoprotein**

Classes of Conjugated Proteins

- Glycoproteins—carbohydrate residues
- Lipoproteins—associated with lipids
- Nucleoproteins—associated with nucleic acids
- Phosphoproteins—covalently modified by phosphate
- Metalloproteins—metal ion(s)
- Hemoproteins—heme (iron protoporphyrin)
- Flavoproteins—flavin group, involved in oxidation-reduction

Purification of Proteins

- An art as much as a science
- Separations based on different properties
  - Solubility
  - Electrical charge
  - Surface adsorption properties
  - Specific binding characteristics
  - Size and shape

Solubility

- pH dependence
- Ionic strength
  - Salting in and salting out
  - $(NH_4)_2 SO_4$ common for salting out
- Organic solvents (acetone, butanol, etc)

See the Apppendix to Chapter 5 for more discussion of these separation techniques

Proteins tend to be least soluble at their pI. When there is no net negative or positive charge on the protein molecules, the repulsion between them is at a minimum.

Ammonium sulfate, $(NH_4)_2SO_4$, is commonly used because it (a) is soluble up to high concentrations, even in the cold, (b) has larger effect on ionic strength than monovalent ions, and (c) tends to stabilize rather than denature or inhibit enzymes.

Solubility for some proteins is decreased in organic solvents, and for a few it may be increased. The danger is that the organic solvent can disrupt the hydrophobic bonding contributing to the stability of the protein, so these precipitations usually have to be carried out at very low temperatures, sometimes below $0^{o}C$, to minimize the denaturation.

Electrical Charge

- Electrophoresis
  - Non-denaturing
  - Denaturing (with SDS)
  - Isoelectric focusing
    - pH gradient set up by ampholytes
    - Proteins migrate to their pI
  - Two dimensional (See Figure 5A.5)

Non-denaturing electrophoresis is principally a function of the charge/mass ratio of the protein. Below its pI, the protein has a positive charge. Above its pI it has a negative charge. The speed of migration in an electric field will depend both on the size of the charge (faster with higher charges) and the size of the protein (slower with larger proteins, because of frictional drag).

Electrical Charge, con't.

- Ion Exchange Chromatography
  - Binds to cation exchange columns **below pI**
    - Eluted by increasing salt concentration or increasing pH
  - Binds to anion exchange columns **above pI**
    - Eluted by increasing salt concentration or decreasing pH

Denaturing electrophoresis involves reaction with a detergent called SDS (for sodium dodecylsulfate). This strong detergent is able to completely disrupt the hydrophobic bonding of the protein and allow the protein to unfold to an extended structure. The SDS molecules bind to the extended structure at a ratio of about 1 SDS molecule for each amino acid residue, or about 1.4 g SDS per g protein. Rod-like structures are formed, where the charge along the rod is uniformly negative from the sulfate anions, swamping out any charges from protein side chains. Thus the charge to mass ratio of the SDS complex is similar for all proteins, and then the separation depends primarily on size (or length of the rod). Smaller molecules move through the supporting matrix faster. The migration rate in an experiment can be graphically related to a size parameter (molecular weight, effective radius).

Surface Adsorption Properties

- Adsorption to minerals, such as calcium phosphate (hydroxyapatite)
  - Binding strongest at **low salt**; elution by increasing salt
- Hydrophobic Interaction Chromatography (HIC)
  - Proteins bind more at **high salt** and are released by lowering salt concentration
  - Similar to **reversed phase** chromatography, elution by solvents of decreasing polarity

Specific Binding Properties

- Affinity chromatography (See Figure 5A.6)
  - Many biological proteins bind specific ligands
  - Structurally similar ligands are attached to stationary resin in chromatography experiment
  - Specific proteins bind tightly to ligands on resin
  - Elution by changing salt, pH or using free ligand

Size and Shape

- Dialysis (separation of "small" from "large")
- Size exclusion chromatography
  - (See Figure 5A.1)
- Ultracentrifugation
  - Sedimentation rate depends on
    - Size (the larger the faster)
    - Shape (the more assymetric, the greater the frictional coefficient, the slower the movement)

Compare a feather and a ping-pong ball of about the same weight falling from the leaning tower of Pisa. Which would fall faster? The feather would meet more air resistance and would fall more slowly, even though it had the same mass and the same acceleration due to gravity as the ping pong ball.

Size and Shape, con't.

- Frictional coefficient is related to "effective size"
- Can be measured from Diffusion coefficient (D)
  - $dn/dt = -D(dc/dx)$
  - $D = RT/Nf$
  - $f = 6\pi\eta r$
- Stokes radius (r) is radius of an "equivalent sphere with the same frictional coefficient"
- Elution from a size exclusion column is a function of Stokes radius

Rate of diffusion is proportional to the concentration gradient. The proportionality constant is D, the Diffusion coefficient
f is the frictional coefficient. For a sphere, $f = 6\pi\eta r$, where r is the **Stokes radius**

The elution position from a size exclusion column is actually correlated best with Stokes radius.

Size and Shape, con't.

- Ultra centrifugation
  - Centrifugal force

  $$\text{centrifig. force} = m(1 - \bar{V}\rho)\omega^2 r$$

- Frictional force
  - Frictional force = fv, where v is velocity of sedimentation, f is frictional coefficient
- Setting equal, and solving, we get

$$\frac{v}{\omega^2 r} = S = \frac{DM(1 - \bar{V}\rho)}{RT}$$

m = particle weight, V bar = partial specific volume of the protein (or the reciprocal of its density), ρ = the density of the solution. When the density of the protein is less than the density of the solution, this term is negative, and the protein floats. Most proteins have densities of about 1.2 g/mL.

Setting the two forces equal gives $fv = m(1-v\rho)\omega^2 r$. Substitute RT/DN for f and rearrange. Nm = M, the molecular weight of the particle.

The sedimentation coefficient (S) is usually quoted in Svedberg units, which are 10^{-13} seconds.

Centrifugation is commonly used to separate cellular organelles from other cellular constituents because of such large variation in S values.

Ultracentrifugation techniques were used some years ago to measure molecular weights. Simpler and more efficient techniques are now available.

Protein Purification Summary

- Purification utilize some combination of these techniques.
  - (See Table 5.5)
- Large volume batch steps usually done first.
- Usually "trial and error" to work out the best combination of steps.
- Chromatographic steps improved with HPLC (High Performance Liquid Chromatography)

Protein Sequencing

- Pioneering work of Sanger on insulin
  - Laborious process, acid hydrolysis into small fragments, separation by paper chromatography, analysis of peptides, pieces fit back together
  - Earned first of two Nobel Prizes for Sanger
- Current methods involve some automation, making the process easier
- Obtaining sequence from DNA sequence is easier still

Steps in Protein Sequencing

- Cleave disulfide bonds, if any
- Separate and purify peptides
- Determine amino acid composition of each peptide
- Determine N-terminal and C-terminal residues
- Cleave each polypeptide into smaller fragments, enzymatically or chemically.

Steps in Protein Sequencing, con't.

6. Determine amino acid composition and sequence of each fragment.
7. Repeat step 5 and 6with a different cleavage procedure.
8. Reconstruct overall sequence by overlap of fragments from 6 and 7.
9. Determine positions of disulfide bridges

Cleavage of Disulfide Bonds

- Oxidation with performic acid, converting S-S to SO_3^- groups
- Reduction by disulfide exchange with mercaptoethanol or dithiothreitol, followed by alkylation with iodoacetate or 3-bromopropylamine
- See Figure 5.18

Reduction with an SH reagent occurs through disulfide exchange:
R-S-S-R + R'-SH ¾ R-SH + R-S-S-R'
R-S-S-R' + R'-SH ¾ R-SH + R'-S-S-R'
With mercaptoethanol, excess reagent drives equilibrium toward reduction. With dithiothreitol, a cyclic disulfide is formed as the product, and equilibrium lies toward this product even at lower dithiothreitol concentrations. The reduced cysteine residues would quickly re-form disulfide bonds by air oxidation unless they are modified by the alkylation reactions. Note—when amino acid analysis is carried out later, one gets the cysteine derivative rather than cysteine itself in the analysis mixture.

Separation of Peptides

- Some combination of protein purification schemes can be used, though often a single reversed-phase HPLC column is sufficient.
- Earlier methods involved two-dimensional paper chromatography.

Amino Acid Composition

- Peptide hydrolyzed in 6 N HCl, amino acid mixture analyzed as discussed in Chapter 4.
  - Separation on cation exchange columns (Figure 4.21)
  - Chemical derivative made and separated by HPLC (Figure 4.22)
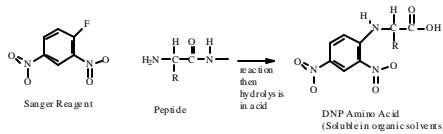
Carboxyl Terminal Analysis

- Carboxypeptidase is an enzyme that specifically cleaves the terminal amino acid
  - Carboxypeptidase A—all residues except pro, arg, and lys
  - Carboxypeptidase B—works on arg and lys
  - Carboxypeptidases C and Y—any residue
- Incubate for several time periods, then analyze for free amino acid

The rates of cleavage of different carboxy terminal amino acids vary, so it is necessary to look at cleavage with several different times of incubation. Once one is released, the second one is release, and so on. Note: If there are really more than one carboxyl terminal amino acids, then the peptide is not pure!

Amino Terminal Analysis
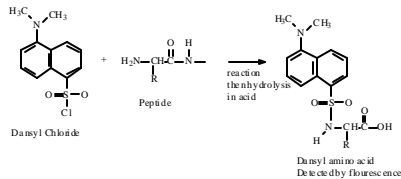
- Older methods involved alkylation followed by acid hydrolysis, and identification of the alkylated residue.
  - Sanger reagent

Amino Terminal Analysis, con't.

- Dansylation

Amino Terminal Analysis, con't.

- Edman Reagent
  - Phenyl isothiocyanate reacts in alkali
  - Cyclization in acid releases the phenylthiohydantoin (PTH) derivative of the amino acid, exposing the next amino acid for a similar reaction.
  - (See Figure 5.19)
  - Called the Edman Degradation, automated machines can cycle up to 50 times to sequence moderate sized peptides

Specific Fragmentation of Peptide

- Enzymatic Cleavage
  - Trypsin (cleaves at Arg and Lys)
    - See Figure 5.20
    - Specificity can be restricted to Arg by making succinyl derivative of Lys
  - Chymotrypsin (cleaves at aromatic residues, but not as cleanly specific)
  - Staphylococcal protease (cleaves at Asp and Glu)

From amino acid analysis, one can predict the number of cleavage fragments that one should get. For example, if the protein contains 5 arginine and 7 lysine residues, there should be 12 cleavage points, and 13 peptides produced.

Methionine is a relatively rare amino acid, so the cyanogen bromide cleavage produces a few large peptide fragments.

Specific Fragmentation of Peptide, con't.

- Chemical Cleavage
  - Cyanogen Bromide (cleaves at methionine)
    - See Figure 5.21
  - Other methods discussed in Table 5.6, not as commonly used.

Separate Fragments and Sequence

- Automated Edman degradation generally used to sequence the individual fragments.
- (Sometimes it may not be necessary to separate a few peptides before carrying out the automated sequencing. See your extra credit problem.)

Fragmentation by Mass Spectrometry

- Several mass spectrometric methods have been developed as very sensitive ways to carry out sequencing, particularly of peptide mixtures. (See Figures 5.23 and 5.24)

Reconstruct the Sequence from Overlaps

- Trypsin fragments:
  - AEGPK + GWASGNA + LCFATR
- Chymotrypsin fragments
  - ASGNA + AEGPKLCF + ATRGW
- Overlap
  - AEGPK LCFATR GWASGNA
  - AEGPKLCF ATRGW ASGNA
- See Figure 5.25 for more extensive example.

GWASGNA must be from the carboxyl-terminal peptide, because trypsin doesn't cleave at alanine.

Determining Disulfide Linkages

- Fragment the polypeptide with trypsin **before** cleaving the disulfide bonds.
- Separate fragments by paper electrophoresis in one dimension
- Treat paper with performic acid
- Separate in the second dimension by the same electrophoresis
- Isolate "off-diagonal" peptides and sequence
  - (See Figure 5.26)

Sequence Databases

- Several electronic databases contain sequence information for proteins. Most sequences have come from DNA gene sequencing.
  - GenBank
  - Protein Information Resource
  - Swiss Protein Database

These are just a few of the databases on the web. Find more by searching on "protein sequences" with your web browser.

Protein Homology

- Similar proteins from different species have similar sequences.
- Sequence similarity gives clues to evolution
  - A phylogenetic tree has been developed just from comparing sequences of cytochrome c from many organisms. (See Figure 5.29)
- Myoglobin and hemoglobin subunits have high degree of homology, and are evolutionarily related. (See Figure 5.30)

Protein Homology, con't.

- Sequence similarity is sometimes found between proteins that are no longer functionally similar.
  - Egg white lysozyme and human lactalbumin
    - See Figure 5.32

Peptide Synthesis

- Chemical synthesis requires first blocking reactive functional groups, then reagents to form the peptide bond.
- The process has been automated for machines.
- For your own interest, read this section of the chapter, but we will not cover it.