The Proteome

- Proteome is the protein equivalent of the genome
- The proteome consist of all of the proteins expressed by a cell under specific conditions
- The proteome of a cell depends on the cell type, its developmental stage, environment/stimuli, nutritional and metabolic status etc
- The genome of a cell is fixed, the proteome is dynamic
- The proteome is much larger than the genome. Each gene can translate into mutiple isoforms of proteins

Protein Mass and Concentration

- Protein mass is measured in Daltons (Da) or kDa
- One Dalton is 1/12 the mass of a ^{12}C atom
- On average, the MW of each aa is 110 Da
- Most proteins range from 30 to 80 kDa
- Trp and Tyr have a high ability to absorb light with maximum absorption at 280 nm. Since most proteins contain these aa, protein concentration can be estimated spectrophotometrically.

Characterization of Proteins

- Isolate / Purify (usually multi-step)
- Determine Purity
- Determine amino acid composition
- Determine sequence, 3D structure, presence of additional functional groups (covalent modification)

Protein Purification

- Proteins can be purified from cell or tissue samples
- Samples are homogenized and fractionated by differential centrifugation to isolate the fraction containing protein of interest
- Protein purification is a multi-step procedure
- Need to establish a specific method of identification. Can be enzymatic, binding or activity assay
- With each step, the purification level and specific activity increases, and the yield decreases

Protein Biochemistry Methods

- Methods can be analytical or preparative
- Preparative: purification of protein of interest
- Analytical: identification and characterization
- Methods based on solubility, size, charge, binding affinity, activity

Solubility

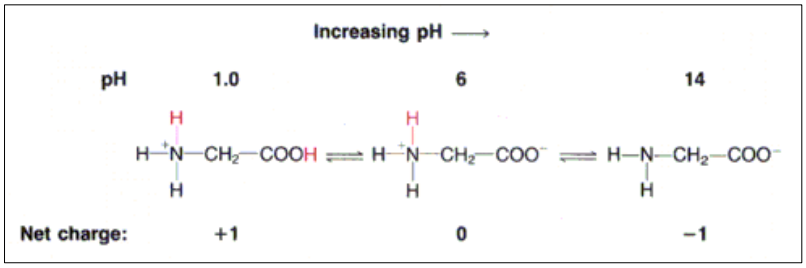
- **Salting Out**: For most proteins, solubility decreases as salt concentration increases.
- Greater the polarity, greater the solubility.
- Proteins can be fractionated by sequential salt precipitation
- **Isoelectric Precipitation**: Proteins are least soluble when pH = pI. (because at pI, net charge is 0)

Ionic forms of Glycine

•Glycine is $H_2N\text{-}CH_2\text{-}COOH$.

•pKa of carboxylate group is 2.3 ; pKa of amino group is 9.6

(Note: glycine can serve as a buffer in 2 different buffer ranges).



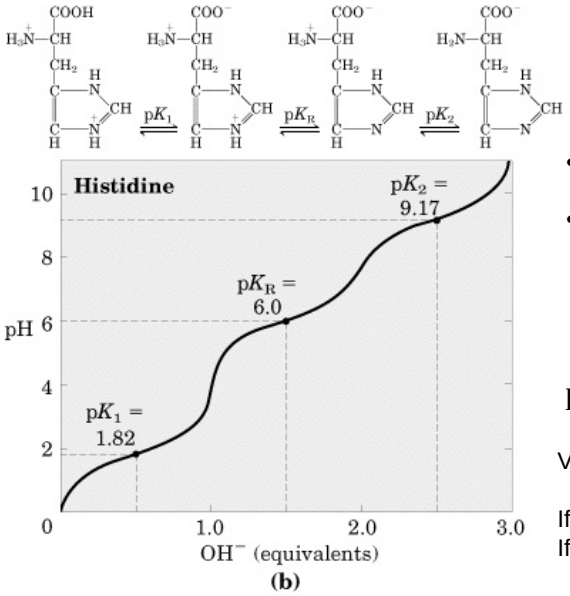
•The ionic form with a net charge of zero is called a zwitterion

•The isoelectric point (pI) is the pH at which the net charge on the ampholyte is zero (or equal number of + and – charged ions).

Calculation of pI for Glycine

- Use the Henderson-Hasselbalch equation to calculate the pI.
- At isoelectric point, pH = pI
- $pI = pK_{COOH} + \log \frac{[H_3N^+CH_2COO^-]}{[H_3N^+CH_2COOH]}$
- $pI = pK_{NH3+} + \log \frac{[H_2NCH_2COO^-]}{[H_3N^+CH_2COO^-]}$
- Adding up: $2pI = pK_{COOH} + pK_{NH3+} + \log \frac{[H_2NCH_2COO^-]}{[H_3N^+CH_2COOH]}$
- When pH=pI, $[H_2NCH_2COO^-]=[H_3N^+CH_2COOH]$
- $2pI = pK_{COOH} + pK_{NH3+}$ or $pI = \{pK_{COOH} + pK_{NH3+}\}/2$

Titration Curve for Histidine



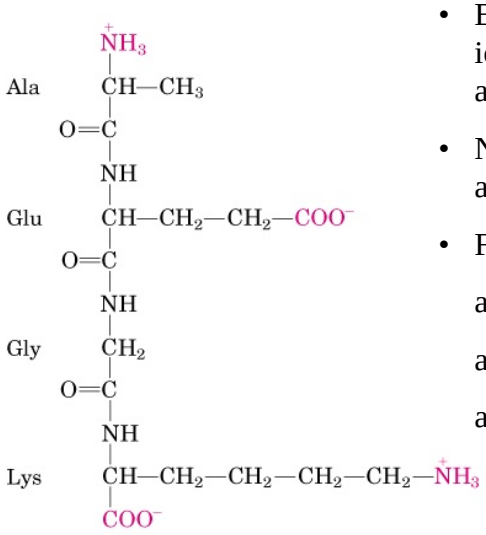
- Triphasic curve
- pI is average of the two pKa values around the zwitterionic form.

$pI = (pK_R + pK_2)/2$

Very Important:

If pH > pKa, [CB] > [WA]
If pH < pKa, [WA] > [CB]

Ionization of Peptides & Net charge



- Each ionizable group ionizes according to its pKa and solution pH.
- Net charge is sum total of all charges on molecule.
- For this tetrapeptide:

  at pH=7.0, net charge = 0

  at pH=3.5, net charge = +1

  at pH=9.0, net charge = -1

Size

- **Dialysis**: Diffusion through a semi-permeable cellulose membrane. Different pore sizes allow removal of molecules smaller than specific MW
- **Gel-Filtration Chromatography**: Column packed with dextran (sephadex) or agarose beads. Smaller molecules get trapped within the porous beads and their flow down the column is retarded. Larger molecules are excluded form the beads and move down between loosely packed bead. Smaller the molecule longer the elution time.

Charge

- **Ion Exchange Chromatography**: Columns are made of charged cellulose particles.
- Carboxymethyl (CM) cellulose: -charge, cation exchange column
- Diethylaminoethyl (DEAE) cellulose: + charge, anion exhange column
- Proteins are eluted using a pH gradient
- Can combine effect of size and charge by using charged sephadex: introduce charge on the surface / interior of spehadex beads. Called functionalized sephadex.

Binding Affinity

- **Affinity Chromatography**: The column matrix is modified by covalent linkage to a compound with high specific binding affinity to protein of interest
- Eg: Lectins, antibodies, ligands, substrates
- A functional group /recognition molecule can be attached to an inert matrix (sepharose). A spacer of $(-CH_2)_6$ is used to prevent spatial hindrance.
- 3 steps: Specific binding of protein, washing unbound proteins, elution of bound protein by specific displacement, high salt or low pH. Can dialyze out excess salt/ions if necessary.

HPLC

- High Pressure Liquid Chromatography
- Enhanced version of all column chromatography techniques
- Column material are very fine and closely packed for better resolution
- High pressure has to be applied to maintain flow
- Clean, rapid separation of very small samples

Electrophoresis

- Primary analytical technique used to determine purity.
- Electrophoresis is the movement of charged proteins in an electric field
- Movement is from the – electrode to + electrode. Migration is related to charge: mass ratio. Generally, smaller proteins migrate further
- Separation on slabs of polyacrylamide cross-linked with methylenebisacrylamide: inert, porous and readily formed
- Visualization by staining (coomassie blue, silver).
- Other methods to determine protein purity include: NMR, mass-spectrometry, finger-printing.

Amino Acid Composition

- Complete hydrolysis for 24 hr at 110 ^{o}C in 6 M HCl
- Separation of amino acids by ion exchange chromatography on sulfonated polystyrene resin
- Elution using a pH gradient
- Detection of amino acids by reaction with ninhydrin or fluorescamine (spectrophotometry)
- Identification of amino acids by position of peak on chromatogram.
- Determination of amino acid ratios by size (height) of each peak.
- Problem: Some Gln and Asn may be decomposed to Glu and Asp, respectively.

Identification of N-terminal amino acid

- The N-terminal aa can be identified by Sanger's method. This method involves modification of the N-terminal residue by flurodinitrobenzene followed by complete hydrolysis of the peptide.
- More recently, fluorescent compounds such as dansyl chloride or dabsyl chloride are used because of their higher sensitivity.
- The N-terminal aa is the only modified aa and it is identified by chromatography.
- The peptide is completely hydrolyzed and cannot be reused
- C-terminal aa hydrolyzed by carboxypeptidase and identified by chromatography.

Amino acid sequencing

- Peptides are sequenced by 'Edman Degradation' method
- Amino acids are removed and identified sequentially one residue at a time from the N-terminus
- The N-terminal amino acid is modified by phenyl isothiocyanate
- Mild hydrolysis releases the tagged amino acid as a cyclic derivative phenylthiohydantion-aa (PTH-aa) which is identified by HPLC ion-exchange chromatography
- The rest of the peptide remains intact, just one aa short
- The next cycle releases residue 2. It is possible to identify ~50 aa from each sample by this method

Specific Cleavage of Polypeptides

- Only few cycles possible by Edman degradation because of depletion of reagents, and byproducts.
- Proteins larger than 50 aa are first hydrolyzed into shorter peptides: Fragmentation
- Chemical or enzymatic methods hydrolyze proteins at specific sites
- Peptides are separated by chromatography
- Peptides generated by 2 or more cleavage methods are each sequenced separately.
- Sequences of individual peptides are overlapped together to deduce the entire protein sequence

Protein Sequencing Example

Method 1 (Trypsin):

ser-glu-phe-his-lys
ala-ile-cys-asp-tyr-thr-ala
gly-leu-pro-arg

Method 2 (staphylococcal protease):

gly-leu-pro-arg-ser-glu
phe-his-lys-ala-ile-cys-asp
tyr-thr-ala

Overall protein sequence:
Gly-leu-pro-arg-ser-glu-phe-his-lys-ala-ile-cys-asp-tyr-thr-ala