Biology of Cells & Tissues

Lecture 3: Protein Structure & Function II

Dan Hardy
Department of Cell Biology & Biochemistry

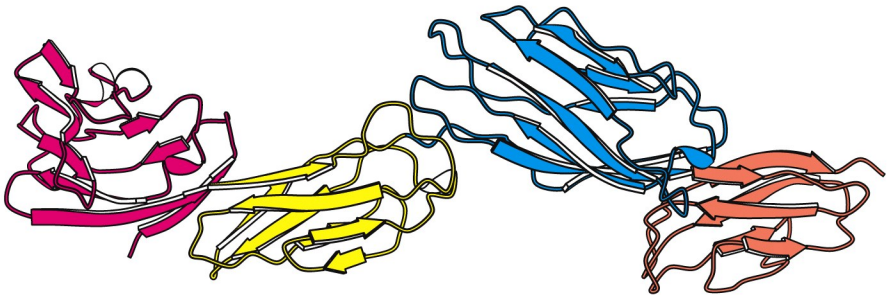
October 29, 2018

Lecture Plan

1. How proteins evolve
2. Protein solubility and stability
3. Protein denaturation
4. How proteins fold
5. Protein characterization and identification
6. Diseases of protein structure

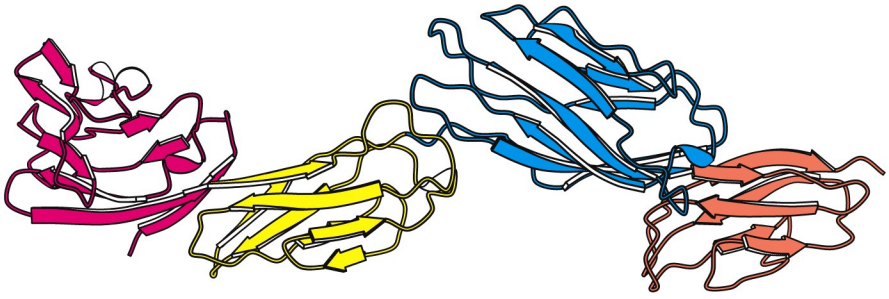
Protein molecular evolution

Some proteins are made up of tandem globular regions called **domains**



Ribbon structure of CD4, a protein on the surface of helper T cells that is the receptor for HIV infection.

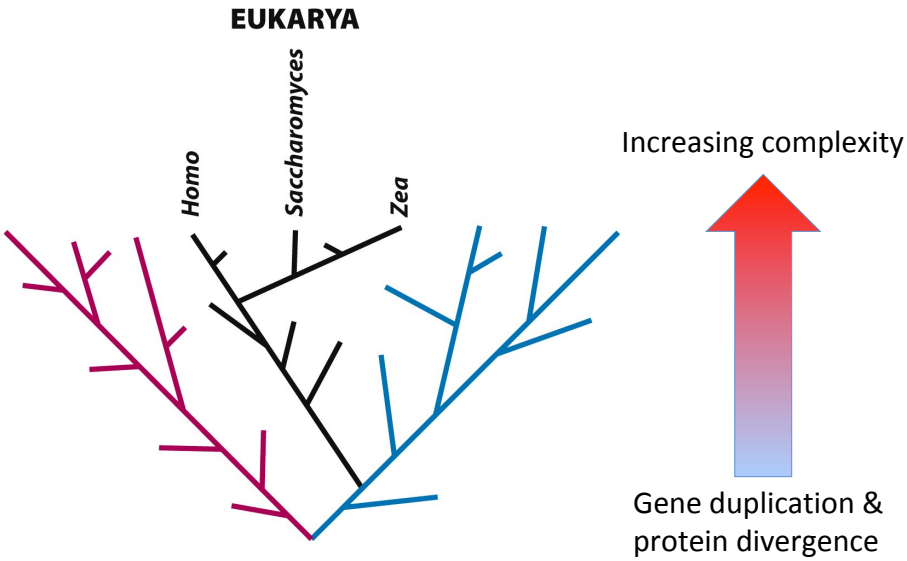
Some proteins are made up of tandem globular regions called **domains**



Ribbon structure of CD4, a protein on the surface of helper T cells that is the receptor for HIV infection.

Duplication and divergence of stable protein domains is a primary mechanism underlying the evolution of biological complexity.

Protein molecular evolution



Gene duplication and divergence produces families of related proteins

Some large gene families:

G-protein coupled receptors
(thousands known, including adrenergic, serotonergic, olfactory, glucagon, cannabinoids)

Gene duplication and divergence produces families of related proteins

Some large gene families:

G-protein coupled receptors
(thousands known, including adrenergic, serotonergic, olfactory, glucagon, cannabinoids)

Serine proteases
(trypsin, chymotrypsin, thrombin, plasmin, etc.)

Gene duplication and divergence produces families of related proteins

Some large gene families:

G-protein coupled receptors
(thousands known, including adrenergic, serotonergic, olfactory, glucagon, cannabinoids)

Serine proteases
(trypsin, chymotrypsin, thrombin, plasmin, etc.)

Protein kinases (all serine and tyrosine kinases)

Gene duplication and divergence produces families of related proteins

Some large gene families:

G-protein coupled receptors
(thousands known, including adrenergic, serotonergic, olfactory, glucagon, cannabinoids)

Serine proteases
(trypsin, chymotrypsin, thrombin, plasmin, etc.)

Protein kinases (all serine and tyrosine kinases)

Immunoglobulin domain proteins (antibodies, MHC)

How do duplicated genes diverge?

Nucleotide substitutions occur by chance

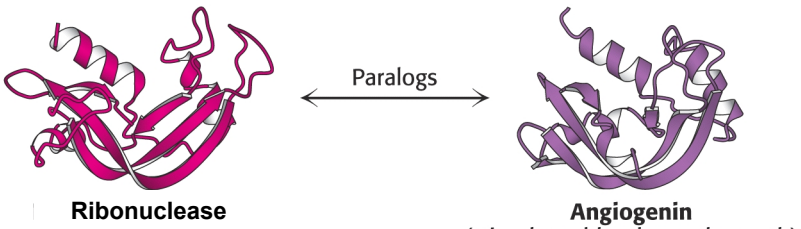
Some Nt substitutions in codons cause an amino acid change ("non-synonomous")

The amino acid change must produce a stable protein

Accumulating changes are subjected to selection

Evolution of Paralogs

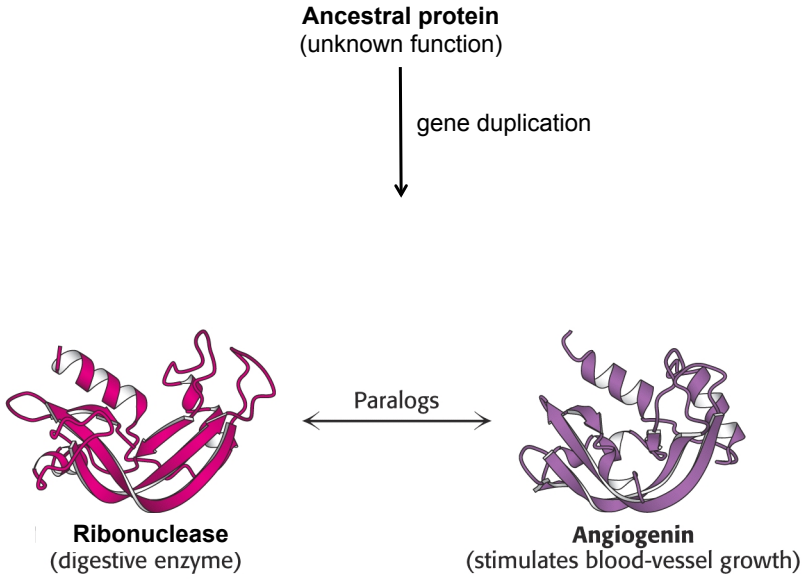
Ancestral protein
(unknown function)



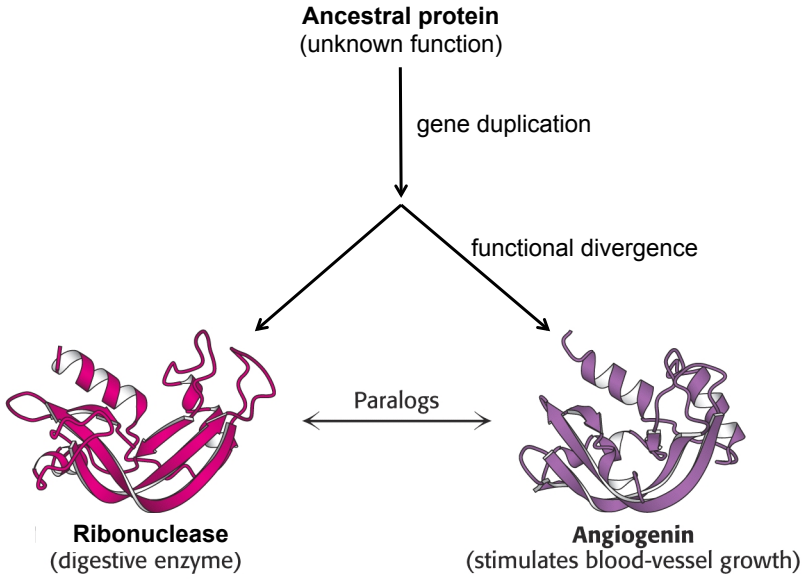
Ribonuclease
(digestive enzyme)

Angiogenin
(stimulates blood-vessel growth)

Evolution of Paralogs



Evolution of Paralogs



How do duplicated genes diverge?

Nucleotide substitutions occur by chance

Some Nt substitutions in codons cause an amino acid change ("non-synonomous")

The amino acid change must produce a stable protein

Accumulating changes are subjected to selection

Question: Based on your knowledge of amino acid side chain structure and chemistry, which two amino acids would be least likely to change in the ongoing molecular evolution of a protein?

Evolution of multi-domain proteins

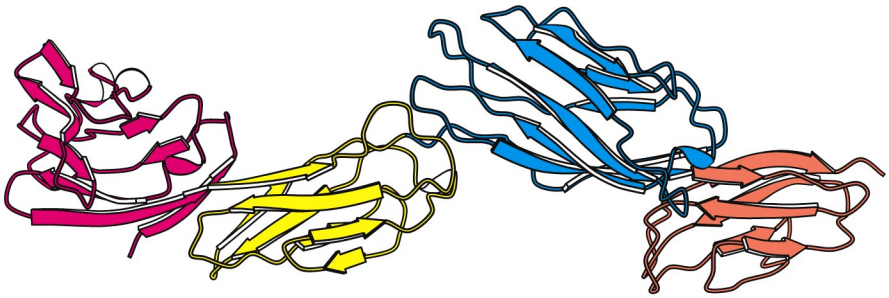
Duplicated genes can fuse to form multi-domain proteins.

Evolution of multi-domain proteins

Duplicated genes can fuse to form multi-domain proteins.

Fused domains can be homologous (=> tandem repeats) or heterologous (=> "mosaic" proteins)

Evolution of multi-domain proteins



Example: CD4 has four tandem Ig domains.

HIV infects and kills $CD4^+$ T-cells, causing AIDS

Evolution of multi-domain proteins

Duplicated genes can fuse to form multi-domain proteins.

Fused domains can be homologous (=> tandem repeats) or heterologous (=> "mosaic" proteins)

Recall from Lecture 2 that flux through pathways can be accelerated by association of enzymes catalyzing sequential reactions into multienzyme complexes.

Evolution of multi-domain proteins

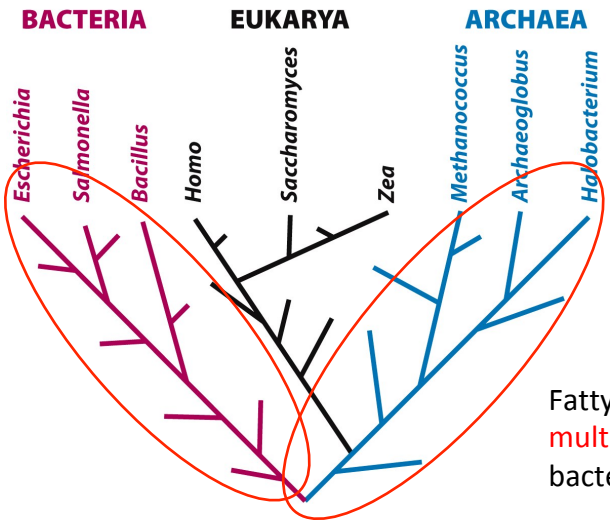
Duplicated genes can fuse to form multi-domain proteins.

Fused domains can be homologous (=> tandem repeats) or heterologous (=> "mosaic" proteins)

Recall from Lecture 2 that flux through pathways can be accelerated by association of enzymes catalyzing sequential reactions into multienzyme complexes.

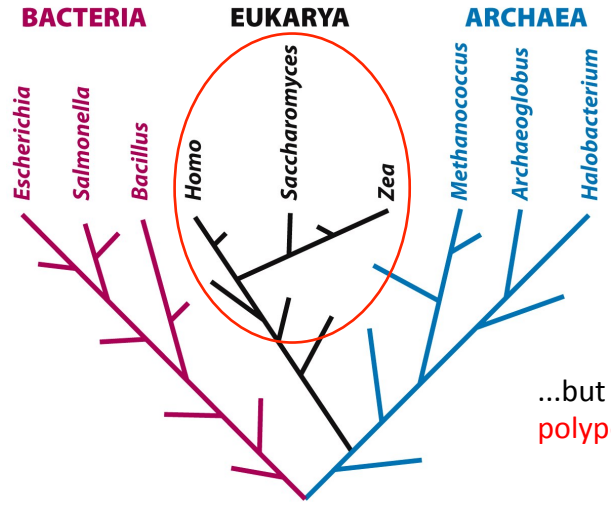
Sometimes genes for different subunits of multienzyme complexes can fuse to produce a single gene encoding a multienzyme polypeptide.

Evolution of multi-domain proteins



Fatty acid synthase is a multienzyme complex in bacteria and archaea...

Evolution of multi-domain proteins



...but a single multienzyme polypeptide in eukaryotes.

Protein solubility and stability

Isoelectric point (pI)

1. pI is the pH at which all positive and negative charges balance out (i.e. net charge is zero)
2. pI is primarily determined by the balance of (Glu + Asp) to (Lys + Arg)
3. Most proteins have a pI below 7
4. At pH = pI, a protein will not migrate in an electric field
5. At pH = pI, the protein is still charged, the net charge is simply balanced

Protein solubility

1. Proteins are generally least soluble at pH = pI
2. Solvent-exposed side chains interact with water and ions
3. Neutralization of charged, exposed side chains by counterions prevents aggregation by ionic interaction. Therefore a minimal amount of "salt" is required for most proteins to be soluble ("salting in").

Protein solubility

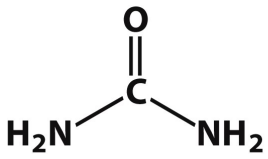
4. Excess salt decreases solubility by interfering with water interaction with the protein ("salting out").
5. Drastically altering pH disrupts normal ionic interactions of a protein, often causing irreversible denaturation (example: stomach acid)
6. Protein solubility is a strong indicator of native structure

Protein denaturation

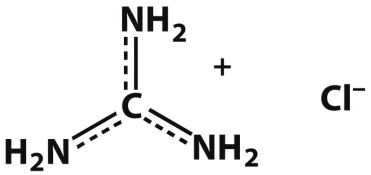
Protein denaturation

1. Denaturation is the loss of native structure elements determined by non-covalent interactions
2. Physical denaturation
   --Heat
   --Freeze-thaw
   --Hydrophobic surface
3. Chemical denaturation
   --Extremes of pH
   --Organic solvents
   --Direct disruption of non-covalent interactions

Chemical denaturants



Urea



Guanidinium chloride

SDS
(an ionic detergent)

Protein folding

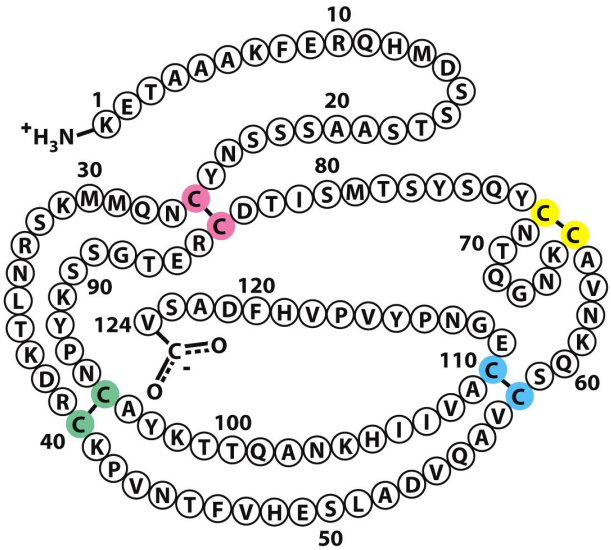
How do proteins acquire their higher order structures?

1. The Anfinsen experiment showed that a protein's sequence specifies its three dimensional structure:

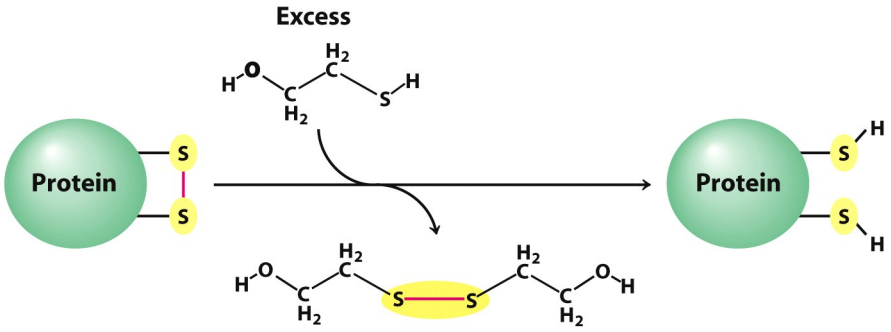
2. Levinthal's Paradox shows that protein folding cannot occur by random trial of allowed conformations:

Protein folding must therefore be a concerted, sequential process.

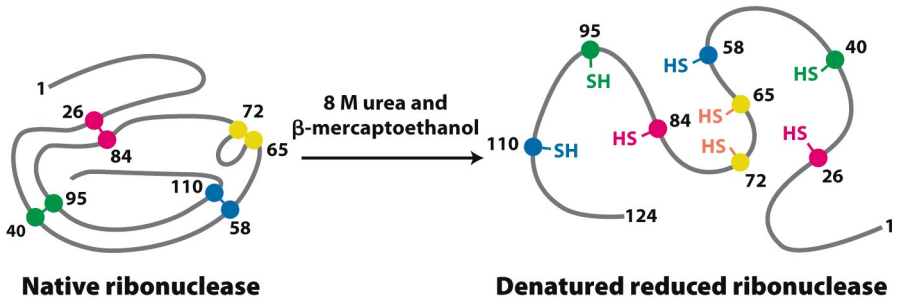
Primary structure of ribonuclease



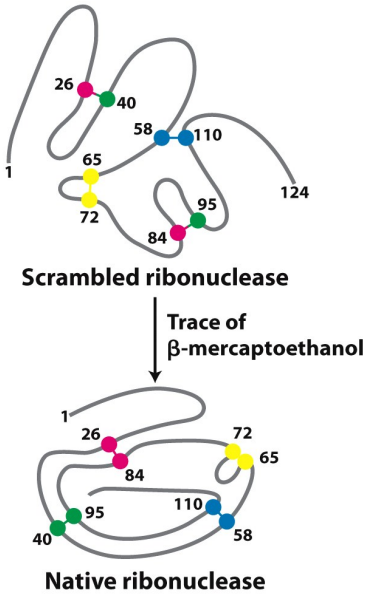
Reduction of disulfide bonds



Disruption of secondary and tertiary structure with a denaturant



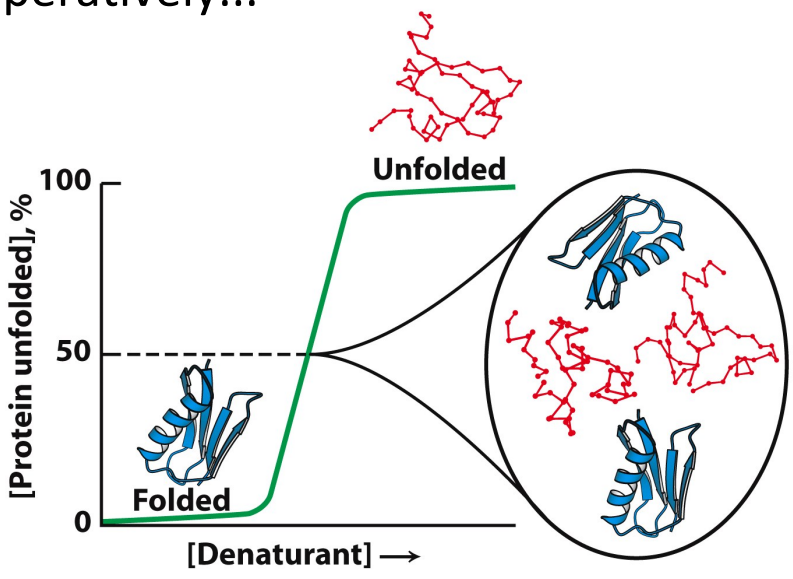
Spontaneous refolding of denatured RNAse



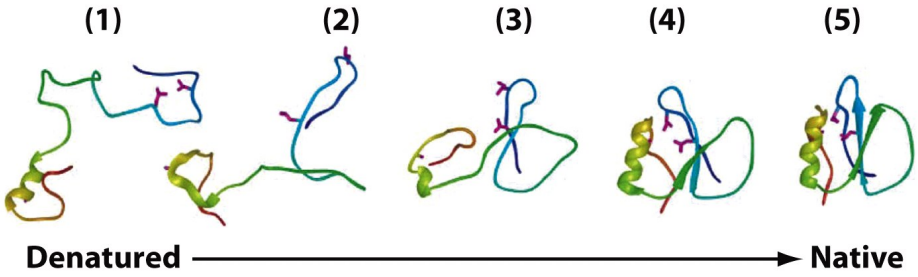
Purified enzyme, after removal of denaturant, spontaneously reformed its native structure as determined by recovery of enzyme activity.

A trace of reductant was needed to reduce incorrect disulfides so the polypeptide could fold properly.

Protein folding and unfolding occur cooperatively...

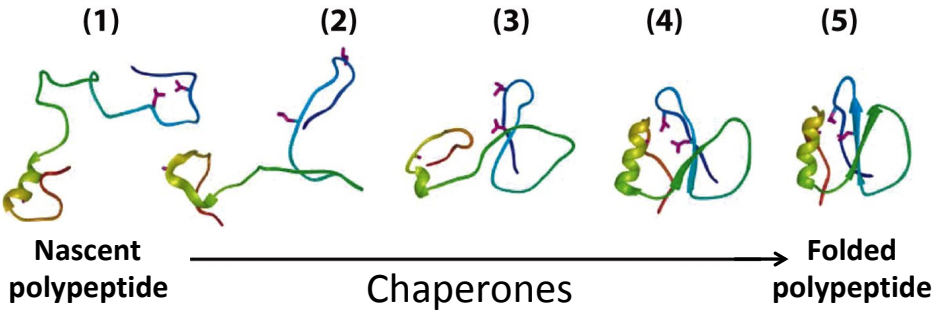


...but partially folded intermediates must exist at least transiently...



...to resolve Levinthal's paradox.

Folding *in vivo* is further aided by chaperone binding to intermediates

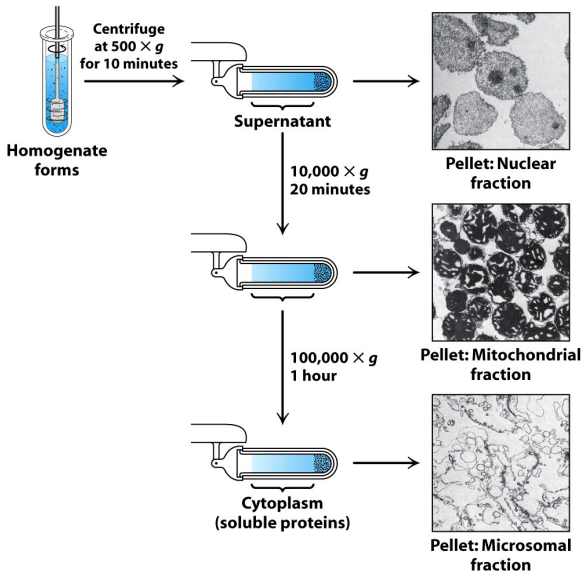


Protein separation and identification

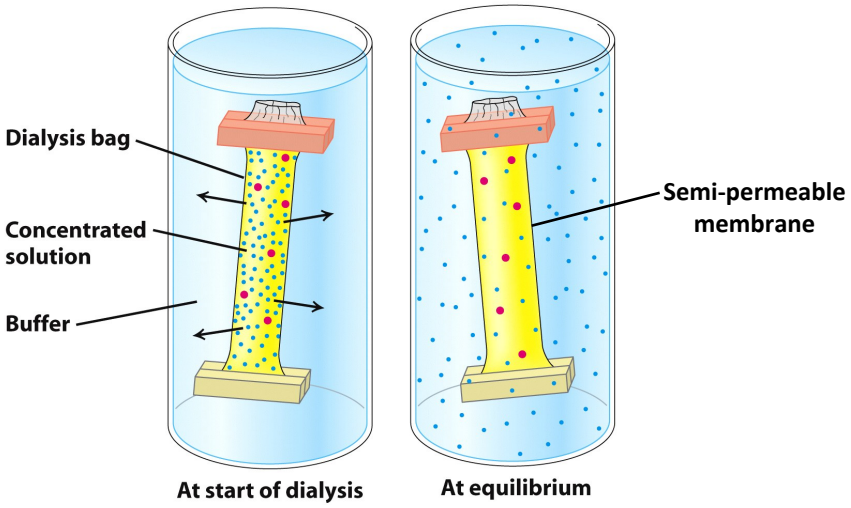
Protein separation

1. Extract the protein from its biological source
2. Exchange the protein into a suitable buffer for fractionation
3. Fractionate based on the target protein's unique characteristics
   --Size
   --Charge
   --pI
   --Binding activity

Cell fractionation by differential centrifugation



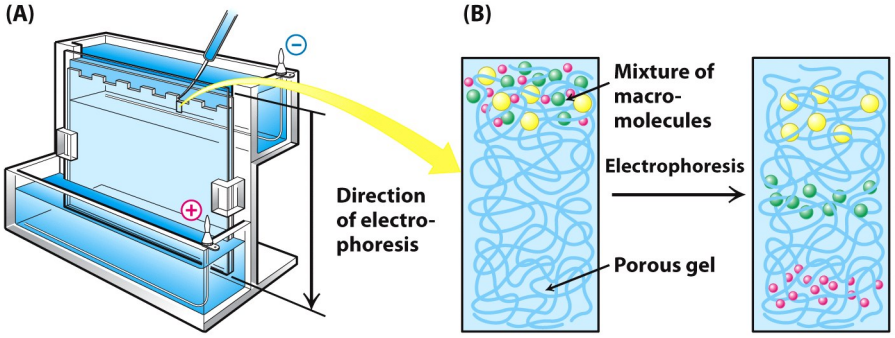
Removal of small molecules by dialysis



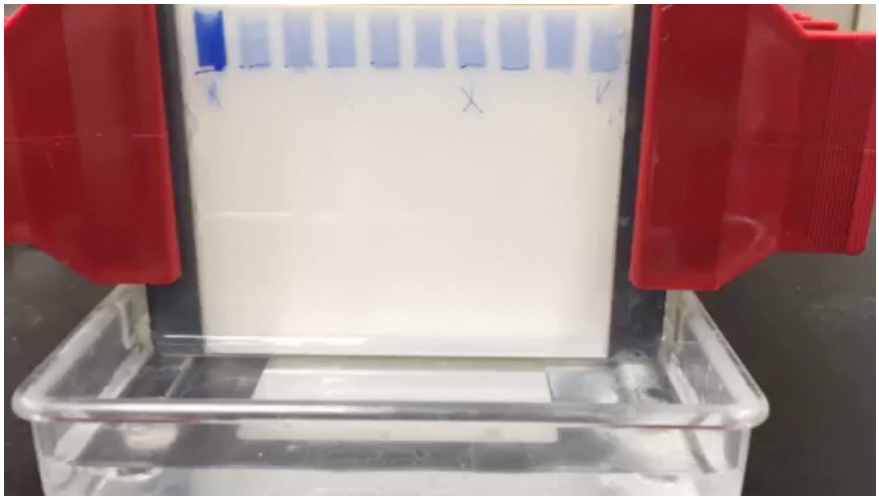
Protein fractionation methods

Basis	Methods
Size	SDS-gel electrophoresis, gel filtration chromatography
Charge	Native gel electrophoresis, ion exchange chromatography
pI	Isoelectric focusing (electrophoresis)
Binding activity	Affinity chromatography

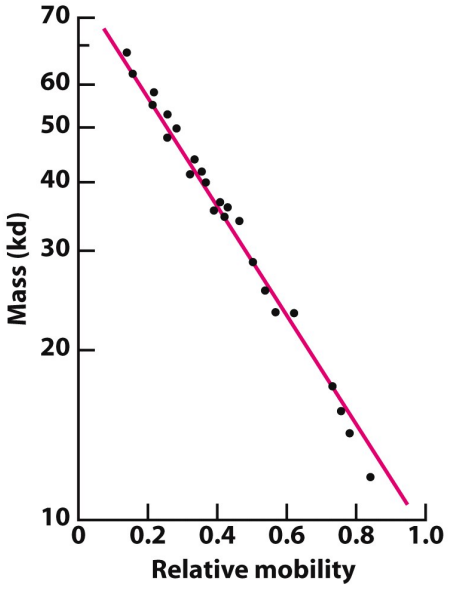
Polyacrylamide gel electrophoresis



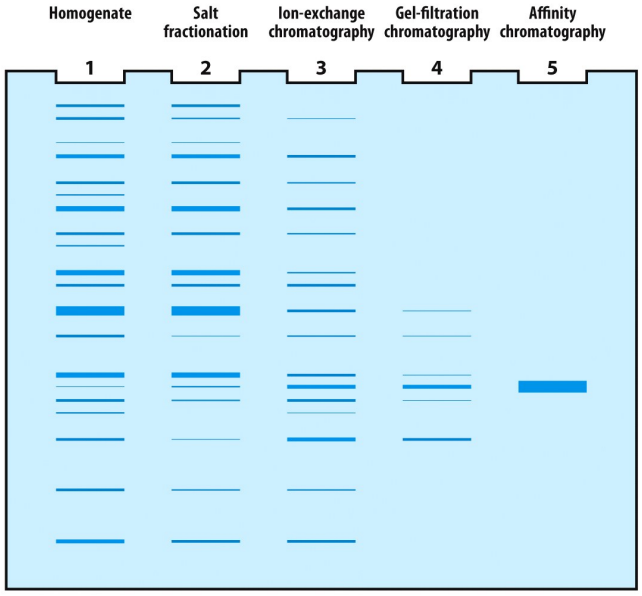
With or without SDS
With or without disulfide reduction



Proteins separate by size in SDS-PAGE



SDS-PAGE to monitor a protein purification



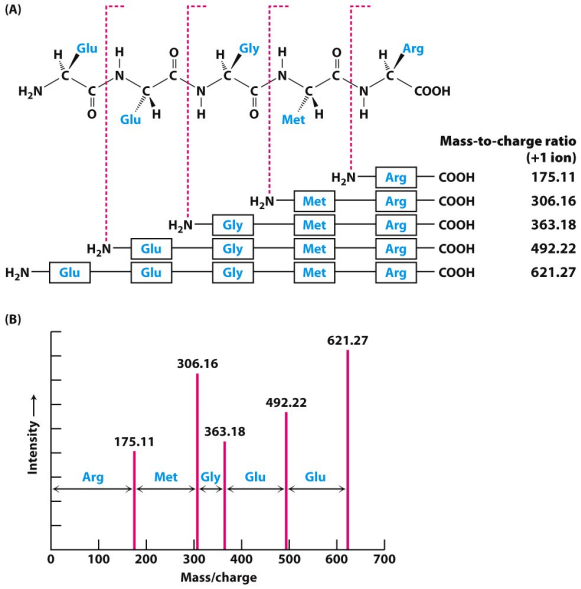
Protein identification by mass spectrometry

1. Optional protein fractionation, often by a high resolution method such as two-dimensional electrophoresis
2. Complete tryptic digestion of proteins
3. Positive identification of peptides by comparison of exact masses to a database of masses calculated from genomic sequence data

-OR-

4. Positive identification of proteins by comparing *de novo* peptide sequences to sequence databases

Peptide sequencing by mass spectrometry



Protein separation and identification

1. Protein constituents in a mixture may be identified thoroughly and rapidly by mass spectrometry
2. Large scale analysis of complex protein mixtures is called proteomics
3. Proteomes typically differ between normal and diseased tissues or body fluids
4. High throughput proteome analysis has the potential to become a part of routine diagnostic testing
5. Proteomics is complementary to genome sequencing

Diseases of protein structure

Protein defects in disease

1. Protein is made but is non-functional (loss of function)
   Example: p53 mutations in cancer
2. Protein is made and is constitutively active (gain of function)
   Example: RAS mutations in cancer
3. Protein is made with altered function
   Example: mutant (E7V) β-globin in sickle cell anemia

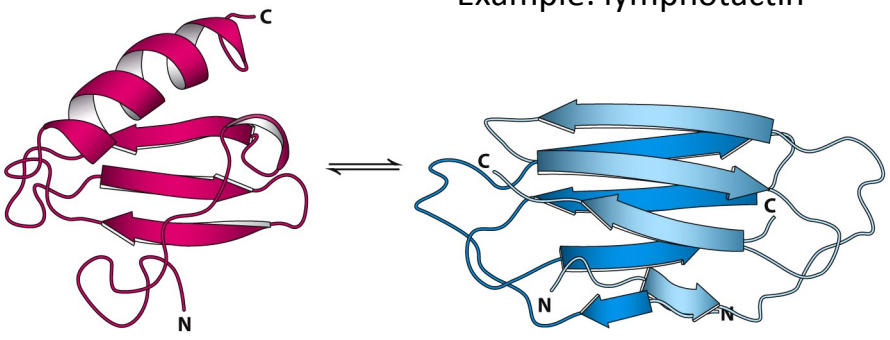
Protein defects in disease

4. Protein does not fold and traffic correctly
   Example: mutant (ΔF508) of CFTR in cystic fibrosis
5. Structure is altered after the protein is made
   Examples: neurodegenerative diseases

Many amino acid substitutions are not deleterious, and reflect normal polymorphisms in a population (hence the neutral theory of evolution).
Mutations that produce a significant change in side chain properties (e.g. E7V) are more likely to affect protein function.

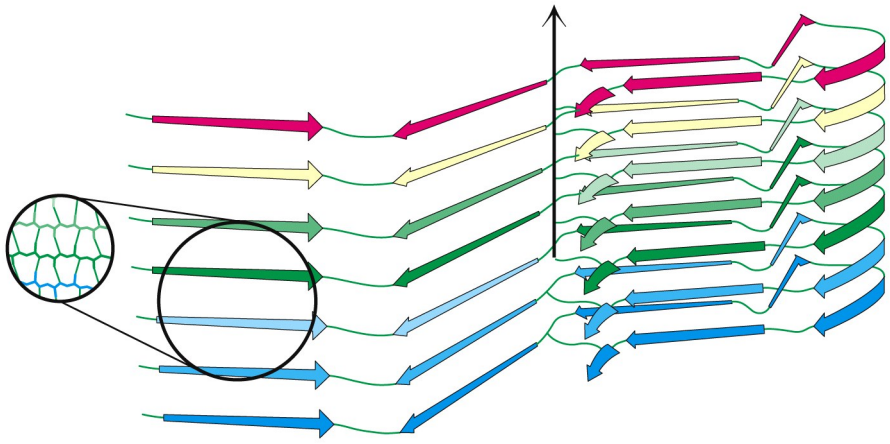
Some proteins can have more than one stable folded structure

Example: lymphotactin



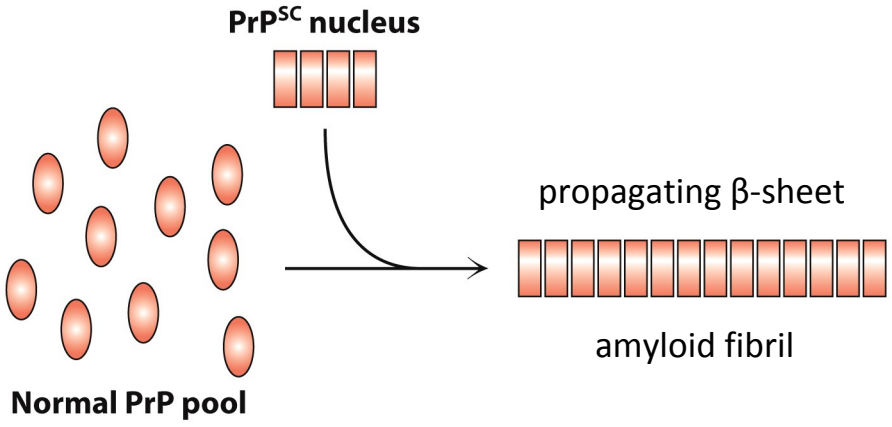
Chemokine structure **Glycosaminoglycan-binding structure**

Alteration of a protein's native shape can cause disease: amyloidoses



Alternately folded prion proteins aggregate into amyloid fibrils with extensive β-sheet structure, causing mad cow disease.

A protein can cause an infectious disease



Lecture 3 Recap

1. Protein duplication and divergence in the evolution of complexity
2. Protein solubility and native structure
3. Physical and chemical denaturation
4. Fractionation and identification
5. Protein structure in disease