Protein Purification and Analysis

Numbers of genes:
Humans ~40,000 genes
Yeast ~6000 genes
Bacteria ~3000 genes

Solubility of proteins important for purification:
______% soluble, ______% membrane

Some proteins expressed at high levels (collagen, hemoglobin)
Some proteins expressed at low levels (repressors, signaling)

Fibrous proteins - structural (collagen, elastin, keratin)
Globular proteins - structure and/or function (actin, enzymes)

Steps of purification and analysis
(1) Choose protein to purify
(2) Choose source (natural or expressed)
(3) Soluble in aqueous solution?? (problem with membrane proteins)
(4) Stability
(5) Purify
(6) Study (activity, structure, mechanism of action, etc.)
Protein Purification and Analysis

Numbers of genes:
Humans  ~40,000 genes
Yeast    ~6000 genes
Bacteria ~3000 genes

Solubility of proteins important for purification:
60-80% soluble, 20-40% membrane

Some proteins expressed at high levels (collagen, hemoglobin)
Some proteins expressed at low levels (repressors, signaling)

Fibrous proteins - structural (collagen, elastin, keratin)
Globular proteins - structure and/or function (actin, enzymes)

Steps of purification and analysis
(1) Choose protein to purify
(2) Choose source (natural or expressed)
(3) Soluble in aqueous solution?? (problem with membrane proteins)
(4) Stability
(5) Purify
(6) Study (activity, structure, mechanism of action, etc.)
Protein Purification and Analysis

(1) Choose protein to purify - YOUR BOSS TELLS YOU!

(2) Choose source (natural or expressed)

Source of protein for study

Early biochemistry (1970’s)
utilized proteins that were abundant from natural sources
(myoglobin, lysozyme, hexokinase)

Middle biochemistry (1980’s to mid 1990’s)
isolated small amounts of proteins, get gene, express and
purify from bacteria, yeast, insect cells, mammalian cells

Now (2000s)
get gene from library based on homology
choose gene and express and study it

Still problems with:
membrane proteins and solubility
Protein Purification and Analysis

(2) Choose source (natural or expressed)
Break open cells by destroying membranes and releasing cytosolic protein mix - crude extract
If nuclear or membrane protein - more work!

(3) Soluble in aqueous solution?? (problem with membrane proteins)

(4) Stability (perform purification/analyses in cold)

(5) Purify
Separate proteins using fractionation based on physical characteristic:
1. solubility
2. electrical charge
3. size + shape
4. affinity for other molecules
5. polarity
Protein Purification and Analysis

(5) Purify

Solubility used for many years - lower solubility at high salt conc. called “salting out”

selectively precipitate proteins using $(\text{NH}_4)_2\text{SO}_4$

To remove excess salt then perform:
Dialysis - separate proteins from solvents, remove $(\text{NH}_4)_2\text{SO}_4$
Protein Purification and Analysis

(5) Purify
Characteristic: Procedure:

_______ Ion exchange
1. Ion exchange
2. Electrophoresis
3. Isoelectric focusing

_______
ultracentrifugation
1. Dialysis and
2. Gel electrophoresis
3. Gel filtration (size exclusion) chromatography

_______
1. Affinity chromatography

_______
1. Adsorption
2. Paper chromatography
3. Reverse-phase chromatography
4. Hydrophobic chromatography

chromatography
chromatography
# Protein Purification and Analysis

(5) Purify

<table>
<thead>
<tr>
<th>Characteristic:</th>
<th>Procedure:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Charge</strong></td>
<td>1. Ion exchange chromatography</td>
</tr>
<tr>
<td></td>
<td>2. Electrophoresis</td>
</tr>
<tr>
<td></td>
<td>3. Isoelectric focusing</td>
</tr>
<tr>
<td><strong>Size:</strong></td>
<td>1. Dialysis and ultracentrifugation</td>
</tr>
<tr>
<td></td>
<td>2. Gel electrophoresis</td>
</tr>
<tr>
<td></td>
<td>3. Gel filtration (size exclusion) chromatography</td>
</tr>
<tr>
<td><strong>Specificity:</strong></td>
<td>1. Affinity chromatography</td>
</tr>
<tr>
<td><strong>Polarity:</strong></td>
<td>1. Adsorption</td>
</tr>
<tr>
<td>chromatography</td>
<td>2. Paper chromatography</td>
</tr>
<tr>
<td></td>
<td>3. Reverse-phase chromatography</td>
</tr>
<tr>
<td></td>
<td>4. Hydrophobic chromatography</td>
</tr>
</tbody>
</table>
Protein Purification and Analysis

Chromatography

Important steps in chromatography
1. Pack column - Column is packed with material (resin) that can absorb molecules based on some property (charge, size, binding affinity, etc.)
2. Wash column - Molecules washed through the column with buffer
3. Collect fractions - Fractions are taken, at some point your molecule will elute
Protein Purification and Analysis

Ion exchange chromatography
Separate by charge

Elute protein
• Increase salt or pH to elute protein of interest

Proteins move through the column at rates determined by their net charge at the pH being used. With cation exchangers, proteins with a more negative net charge move faster and elute earlier.
Protein Purification and Analysis

Ion exchange chromatography

Carboxymethyl (CM)
Negatively charged resin

Column- \( \text{CH}_2\text{C} - \text{O} \)

Diethylaminoethyl (DEAE)
Positively charged resin

Column- \( \text{CH}_2\text{CH}_2\text{-NH} + \text{C}_2\text{H}_5 \)

Proteins move through the column at rates determined by their net charge at the pH being used. With cation exchangers, proteins with a more negative net charge move faster and elute earlier.
Protein Purification and Analysis

Size exclusion (gel filtration) chromatography
Separate by size

As wash with buffer:
- Small molecules enter the beads
- Large molecules move between the beads
Protein Purification and Analysis

Affinity chromatography
Separate by specificity

Elution: Bound proteins eluted by adding high concentration of ligand
Protein Purification and Analysis

Additional Chromatography info

HPLC (high-performance liquid chromatography)

Column can be:
hydrophobic, (+) or (-) charged, stereospecific, etc.
Resin needs to have incompressible beads

high pressure pumps speed the movement of proteins down the column

HPLC limits protein band spreading - increase resolution
Protein Purification and Analysis

Gel Electrophoresis
Protein Purification and Analysis

Electrophoresis

Separation of proteins, nucleic acids, etc. by size, shape, charge
Proteins migrate based on their charge-to-mass ratio

Proteins visualized (radioactivity or staining)
Use gels made of crosslinked polymer (polyacrylamide) or solidified agarose
Protein Purification and Analysis

SDS Gel Electrophoresis

Used to estimate purity and molecular weight, separate proteins by size
Denature protein by adding SDS (then separate by size only)

Na\(^+\)\(\text{O} - \text{S} - \text{O} - (\text{CH}_2)_{11}\text{CH}_3\)  
Sodium dodecyl sulfate (SDS)

SDS forms micelles and binds to proteins

Determination of unknown protein molecular weight

<table>
<thead>
<tr>
<th>Protein</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myosin</td>
<td>200,000</td>
</tr>
<tr>
<td>β-Galactosidase</td>
<td>116,250</td>
</tr>
<tr>
<td>Glycogen phosphorylase</td>
<td>97,400</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>66,200</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>45,000</td>
</tr>
<tr>
<td>Carbonic anhydrase</td>
<td>31,000</td>
</tr>
<tr>
<td>Soybean trypsin inhibitor</td>
<td>21,500</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>14,400</td>
</tr>
</tbody>
</table>

Relative migration

Unknown protein

\(\log M_r\)
Protein Purification and Analysis

Isoelectric focusing gel electrophoresis

determine the isoelectric point (pI) of a protein

separates proteins until they reach the pH that matches their pI (net charge is zero)
Protein Purification and Analysis

Centrifugation

Separate proteins by size or density
Differential centrifugation - separates large from small particles
Isopycnic (sucrose-density) centrifugation - separates particles of different densities
Protein Purification and Analysis

TABLE 3–5  A Purification Table for a Hypothetical Enzyme

<table>
<thead>
<tr>
<th>Procedure or step</th>
<th>Fraction volume (ml)</th>
<th>Total protein (mg)</th>
<th>Activity (units)</th>
<th>Specific activity (units/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crude cellular extract</td>
<td>1,400</td>
<td>10,000</td>
<td>100,000</td>
<td>10</td>
</tr>
<tr>
<td>2. Precipitation with ammonium sulfate</td>
<td>280</td>
<td>3,000</td>
<td>96,000</td>
<td>32</td>
</tr>
<tr>
<td>3. Ion-exchange chromatography</td>
<td>90</td>
<td>400</td>
<td>80,000</td>
<td>200</td>
</tr>
<tr>
<td>4. Size-exclusion chromatography</td>
<td>80</td>
<td>100</td>
<td>60,000</td>
<td>600</td>
</tr>
<tr>
<td>5. Affinity chromatography</td>
<td>6</td>
<td>3</td>
<td>45,000</td>
<td>15,000</td>
</tr>
</tbody>
</table>

Note: All data represent the status of the sample after the designated procedure has been carried out. Activity and specific activity are defined on page 94.

Activity - total units of enzyme in solution
Specific activity - number of enzyme units per mg of total protein
Protein Sequencing

Function of protein depends on its ______________
Proteins with different functions always have ____________ sequences
Changing just 1 amino acid can make a protein defective
Functionally similar proteins from different species have __________
sequences

Steps for sequencing large protein:

1. Cleave S-S bonds
2. Separate subunits
3. Determine N-terminus of protein
4. Determine amino acid composition
5. Use cleavage agents to digest protein into smaller fragments
6. Amino acid composition and sequence of fragments
7. Use overlapping fragments to get full sequence
Protein Sequencing

Function of protein depends on its amino acid sequence
Proteins with different functions always have different sequences
Changing just 1 amino acid can make a protein defective
Functionally similar proteins from different species have similar sequences

Steps for sequencing large protein:

1. Cleave S-S bonds
2. Separate subunits
3. Determine N-terminus of protein
4. Determine amino acid composition
5. Use cleavage agents to digest protein into smaller fragments
6. Amino acid composition and sequence of fragments
7. Use overlapping fragments to get full sequence
Protein Sequencing

1. Cleave S-S bonds
   To sequence large protein, first break disulfide bonds
Protein Folding and Denaturation

2. Separate subunits

Denaturation = loss of 3D structure resulting in loss of function
Denaturation affects weak interactions, such as H-bonds
Denature proteins by:
    Heat, extreme pH, add organics (alcohol, acetone)
    Add urea, guanidine hydrochloride, detergent

separate subunits by gel electrophoresis, chromatography, etc.
Protein Sequencing

3. Determine N-terminus of protein
Protein Sequencing

4. Determine amino acid composition

Polypeptide

| 6 M HCl heat | Free amino acids | HPLC or Ion-exch. chromatography | AA composition | Determine types and amounts of amino acids |

Phenylisothiocyanate

Trifluoroacetic acid

Phenylisothiocyanate derivative of amino-terminal AA

+ Free amino acids

Identify amino-terminal residue of protein

Purify and recycle remaining peptide fragment through Edman process
5. Use cleavage agents to digest protein into smaller fragments

TABLE 3–7 The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

<table>
<thead>
<tr>
<th>Reagent (biological source)*</th>
<th>Cleavage points†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin (bovine pancreas)</td>
<td>Lys, Arg (C)</td>
</tr>
<tr>
<td><em>Submaxillaris</em> protease (mouse submaxillary gland)</td>
<td>Arg (C)</td>
</tr>
<tr>
<td>Chymotrypsin (bovine pancreas)</td>
<td>Phe, Trp, Tyr (C)</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em> V8 protease (bacterium <em>S. aureus</em>)</td>
<td>Asp, Glu (C)</td>
</tr>
<tr>
<td>Asp-N-protease (bacterium <em>Pseudomonas fragi</em>)</td>
<td>Asp, Glu (N)</td>
</tr>
<tr>
<td>Pepsin (porcine stomach)</td>
<td>Phe, Trp, Tyr (N)</td>
</tr>
<tr>
<td>Endoproteinase Lys C (bacterium <em>Lysobacter enzymogenes</em>)</td>
<td>Lys (C)</td>
</tr>
<tr>
<td>Cyanogen bromide</td>
<td>Met (C)</td>
</tr>
</tbody>
</table>

*All reagents except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.
Protein Sequencing

6. Amino acid composition and sequence of fragments
7. Use overlapping fragments to get full sequence

**Procedure**
- Hydrolyze; separate amino acids
- React with FDNB; hydrolyze; separate amino acids
- Reduce disulfide bonds (if present)

**Result**
- A 5 H 2 R 1
- C 2 I 3 S 2
- D 4 K 2 T 1
- E 2 L 2 V 1
- F 1 M 2 Y 2
- G 3 P 3

**Conclusion**
- Polypeptide has 38 amino acid residues. Trypsin will cleave three times (at one R (Arg) and two K (Lys)) to give four fragments. Cyanogen bromide will cleave at two M (Met) to give three fragments.
- E (Glu) is amino-terminal residue.

**Polypeptide**
- Cleave with trypsin; separate fragments; sequence by Edman degradation
- Cleave with cyanogen bromide; separate fragments; sequence by Edman degradation

**Amino terminus**
- EGAAYHDFEPIDPRGASM
  - T-2 overlaps with T-1 and T-4, allowing them to be ordered.
  - T-2 placed at amino terminus because it begins with E (Glu).
  - T-3 placed at carboxyl terminus because it does not end with R (Arg) or K (Lys).
Protein Sequencing

Clues about functions of proteins/role of specific sequences
Elucidate history of life on earth

**E. coli**

**B. subtilis**

**Archaebacteria**
- *Halobacterium halobium*
- *Sulfobolus solfataricus*

**Eukaryotes**
- *Saccharomyces cerevisiae*
- *Homo sapiens*

**Gram-positive bacterium**
- *Bacillus subtilis*

**Gram-negative bacterium**
- *Escherichia coli*

**Signature sequence**

- *E. coli* signature sequence
- *B. subtilis* signature sequence

**Proteobacteria**

- *Chlamydia*
  - *Chlamydia trophomatis*
  - *Chlamydia psittaci*

- *Spirochaetes*
  - *Borrelia burgdorferi*
  - *Leptospira interrogans*

- *Thermophilic bacterium PS-3*
  - *Staphylococcus aureus*
  - *Clostridium acetobutylicum*
  - *Clostridium perfringens*

- *Mycobacterium leprae*
  - *Mycobacterium tuberculosis (operon)*
  - *Mycobacterium tuberculosis*
  - *Streptomyces albus (gene)*

- *Cyanobacteria and chloroplasts*
  - *Cyanidium caldarium chl.*
  - *Synechocystis*
  - *Ricinus communis chl.*
  - *Triticum aestivum chl.*
  - *Brassica napus chl.*
  - *Arabidopsis thaliana chl.*

**0.1 substitutions/site**
Protein Synthesis - additional info

Increasing uses for making proteins (antibodies, hormones, study of proteins)
Three ways to obtain protein: purify from tissue, genetic engineering, direct chemical synthesis
Macromolecular Structure

ATOMS  MACRO- MOLECULES  ASSEMBLIES  CELLS
MOLECULES

C-C bond  Hemoglobin

1 Å  10^2 Å  10^3 Å  10^4 Å  10^5 Å
10^{-10} m  10^{-9} m  10^{-8} m  10^{-7} m  10^{-6} m  10^{-5} m
1 nm  10^{-7} m  10^{-6} m  10^{-5} m

Glucose  Ribosome  Red blood cell

Electron microscopy

X-ray crystallography, Solution NMR

Resolution limit of light microscope
Macromolecular Structure

X-ray crystallography
Need lots of pure protein in crystallized form
Beam of X-rays of given wavelength
Beam diffracted by electrons of atoms in protein
Collect diffracted x-rays on photographic film
Create electron density map using Fourier transform
Macromolecular Structure

Nuclear Magnetic Resonance (NMR)
Need lots of pure protein in solution
Sample must contain atoms that possess specific nuclear spin (\(^1\text{H}, \ ^{13}\text{C}, \ ^{15}\text{N}, \ ^{19}\text{F}, \ ^{31}\text{P}\))
Nuclear spin generates magnetic dipole, apply magnetic field, magnetic dipoles align
Pulse of electromagnetic energy, energy absorbed by nuclei, absorption spectrum collected
Absorption spectrum gives information about identity of nuclei and its environment
2D NMR gives information about interactions of nearby atoms through space and covalent bonds