proteins & their properties

The work horses of cell metabolism
PROTEINS... work horses of cell metabolism

PROTEOME:
entire complement of an organisms proteins:
yeast  ≈  6,000 proteins
human  ≈  32,000 proteins

We'll look at how Structure gives rise to Function

a) structure: primary, secondary, teritary, & quarternary
b) protein folding – chaperones
c) degradation/turnover – proteosomes
d) molecular motors
e) enzyme kinetics

pages 63-78
Proteins - classified by functions

Transport Proteins - bind & carry ligands
Enzymes - catalytic activity and function
Storage Proteins - ovalbumin, gluten, casein, ferretin
Contractile (Motor) - can contract, change shape, elements of cytoskeleton (actin, myosin, tubulin)
Structural (Support): collagen of tendons & cartilage, elastin of ligaments (tropoelastin), keratin of hair, feathers, & nails, fibroin of silk & webs
Defensive (Protect): antibodies (IgG), fibrinogen & thrombin, snake venoms, bacterial toxins
Regulatory (Signal): regulate metabolic processes, hormones, transcription factors & enhancers, growth factor proteins
Receptors (Detect stimuli): light & rhodopsin, membrane receptor proteins and acetylcholine or insulin.
Nomenclature - classes of proteins

Based on **SOLUBILITY of PROTEINS** Two classes - **Simple & Complex**

**SIMPLE PROTEINS** include:

1. **Albumins** - soluble in water, globular, mostly enzymes
2. **Globulins** - soluble in dilute aqueous solutions; insoluble in pure distilled H2O
3. **Prolamins** - insoluble in water; soluble in 50% to 90% simple alcohols
4. **Glutelins** - insoluble in most solvents; soluble in dilute acids/bases
5. **Protamines** - not based upon solubility; small MW proteins with 80% Arginine & no Cysteine
6. **Histones** - unique/structural - complexed w DNA high content basic aa's - 90% Arg, Lys, or His
7. **Scleroproteins** - insoluble in most solvents fibrous structure - cartilage & connective tissue
   - **Collagen** = high Glycine, Proline, & no Cysteine when boiled makes gelatin
   - **Keratins** - proteins of skin & hair high basic aa's (Arg, His, Lys), but w Cys
Complex Proteins:

- lipoproteins - blood, membrane, & transport proteins
- glycoproteins - antibodies, cell surface proteins
- nucleoproteins - ribosomes & organelles

Common terminology:
- peptide = short chain of amino acids (20-30)
- dipeptide = 2 amino acids
- tripeptide = 3 amino acids
- polypeptide = many amino acids (up to 4,000)
- protein = polypeptide with well defined 3D structure
Structure of Proteins

the Variety of Protein Structures may be INFINITE...

average protein has 300-400 amino acid's & has a MW of 30 to 45kD
a PROTEIN of 300 amino acids
made with 20 different kinds of amino acids
can have $20^{300}$ different linear arrays of aa's
that's $10^{390}$ different proteins

1st protein sequenced was Beef Insulin *
by Fred Sanger - 1958 Nobel Prize winner
2 polypeptides [21/30 aa's] Humulin & ADA

to date about 100,000 protein have been sequenced
only about 10,000 structures known [2K/yr]
E. coli make about 3,000 proteins; humans about 100,000
4 levels of protein structure are recognized

primary - linear sequence of aa's

secondary - regular, recurring orientation of aa in a peptide chain due to H-bond

tertiary - complete 3-D shape of a peptide

quaternary - spatial relationships between different polypeptides or subunits
PRIMARY SEQUENCE is…

Linear sequence of amino acids in a polypeptide
repeated peptide bonds form the back bone of the polypeptide chain
R side groups project outward on alternate side

Chain... one end polypeptide chain has free (unlinked) amine group: N-terminus
other end has a free (unlinked) carboxyl group: C-terminus
H₂N-C-C-N-C-N-C-N-C-N-C-N-C-N-C-N-C-COOH

Size… protein size is specified by mass (MW in daltons = 1 amu)
average MW of a single amino acid ≈ 113 Da
thus if a protein is determined to have a mass of 5,763 Da ≈ 51 amino acids
average yeast protein = 52,728 Da [52.7 kDa] with about 466 amino acids

Protein Primary Sequence today is determined by reading GENOME Sequence

Function is derived from the 3D structure (conformation) specified by
the primary amino acid sequence and the local environs interactions
Primary sequence... & some consequences

**Polymorphism**... proteins may vary in primary sequence but have the same function. ex: enzymes $H_2O_2 \rightarrow 2 H_2O + O_2$

- **inter-specific**: between species [diff. aa sequences]
- **intra-specific**: within a species  [ liver vs. kidney ]

**Invariants**... don't vary significantly in aa sequence

examples: ubiquitin (proteosomes) & histones (chromosomes)

**Site Specificity**... sequences determine intra-cellular location

signal sequences, prosthetic binding sites, etc...

**Families of proteins**: different but related functions
evolved from a single ancestral protein, 30% + commonality of sequence... serine proteases (trypsin, chymotrypsin, elastase)

**Homologous Proteins**: evolved in related fashion & perform the same cellular function in different species  
ex: **cytochrome-C** : in duck & chickens = 2 variants & in yeast & horses = 48 variants

**Mutation** - change in primary aa sequence = defective protein - **SICKLE CELL**
Secondary structure - 3D conformation of portions of polypeptide chains

Alpha helix* described by Linus Pauling 1954 Nobel using X-ray* diffraction technique

peptide backbone around long axis core
rigid cylinder
R-groups radiate outward
3.6 aa per 360° turn
single repeat turn of helix (360°) = 0.54 nm
right handed helix - (counterclockwise)
helix formed from H-bond interactions
$H^+$ of N (of any aa) & $-C=O^-$ (of 4th aa)
$\frac{1}{4}$ of aa's in globular proteins occur in alpha helix
flexible - wool is stretchable (breaks H-bonds)

mcb fig 3.4
Secondary structure
- BETA SHEET (fig 4.10) (model = silk protein fibroin)
  a linear extended ZIG-ZAG pleated sheet formed by H-bonds intra- & inter-chain

resist pulling (tensile) forces = strength of silk fibers
non- $\alpha/\beta$ regions = hinges, turns, loops, etc = flexibility

ribons & sheets*

turns - mcb 3.6

MOTIFS combos of recurring arrangements of $\alpha$-helix and/or $\beta$-sheets in unrelated proteins.... such as:

hairpin beta motif... antiparallel beta-sheets joined by
**Structural MOTIFS**: regular 3D conformations or folds within secondary or tertiary structure common to many different proteins... indicative of a particular 3-D architecture & associated with specific function... same structure is present in different proteins that have similar functions; recurring arrangements of \( \alpha \)-helix and/or \( \beta \)-sheets in unrelated proteins.... such as:

**EF hand**  two short helices connected by a loop; a Ca\(^{2+} \) ion binder region of hydrophilic residues present in over 100 Ca\(^{2+} \) binding proteins.

**helix loop helix**... commonly bind gene transcription factors to DNA

**zinc finger**... 1 \( \alpha \) and 2 \( \beta \) strands with antiparallel orientations.
forms fingers bound by Zn ion that often link to DNA & RNA

**coiled coil**... \( \alpha \) helices, where the hydrophobic amino acids wind together forming a coil; also called leucine zippers: common to transcription factors.
Tertiary level

most responsible for 3-D orientation of proteins in space
- thermodynamically most stable conformation
- weak non-covalent interactions \([\text{fig 4.4}]*\) & S-S bridges \([\text{fig 4.29}]*\)
- hydrophobic interior & hydrophilic exterior

Protein Folding... forms 3D shapes & binding sites
occurs via H-bonds* \([\text{fig 4.31}]*\) & \([\text{fig 4.9}]*\)

some examples:
- **Myoglobin** MW 16,700 - animal muscle protein - stores O2
- **Cytochrome-C** MW 12,400 - heme binding single polypeptide of 100 aa's in ETS of mitochondria
- **Lysozyme** MW 14,600 enzyme; egg white & human tears \([\text{pdb-lysozyme}]*\)
  124 aa's with 4 S-S; that hydrolyses polysaccharides in bacterial cell walls = bactericidal agent \([\text{pic catalog}]*\)
- **Ribonuclease** MW 13,700 enzyme of 124 aa w 4 S-S
DOMAINS - distinct modules or structural element of the tertiary level of protein structure...
compact folded regions in a polypeptide of 100-150 amino acids, often self-forming, self-stabilizing, that often fold independently.  

3 classes of domains:

- **functional domain** - region with particular activity characteristic of a protein CATALYSIS: ex: kinase domains add PP to other molecules.

- **structural domain** - region of 40+ aa's in a stable 2nd or 3rd-ary conformation (repeatable). ex: 1. hemagglutinin: - a surface protein on influenza viruses, that is made of 3 quaternary identical subunits composed of 2 polypeptides (HA₁ & HA₂); each HA peptide has two domains... globular domain & a fibrous domain

  - **mcb 3.10a**  

  2. EGF (Epidermal Growth Factor) domain - small soluble peptide hormone that binds to embryonic cells in skin/connective tissue & promotes cell division. EGF is generated by proteolytic hydrolysis of the domain from several other proteins, all of which have an EGF domain as a structural part.

  - **mcb 3.11**

- **topological domain** - distinctive spatial relationships to rest of a protein; ex: membrane proteins with extrinsic (cytoplasmic domain) and intrinsic (membrane spanning domain).
PROTEIN FAMILIES –
proteins with a common evolutionary ancestry
function derives from 3D structure that is due to primary sequence, thus some proteins have many identical or chemically similar amino acids in identical sequence positions each may contain domains that closely resembles that of other proteins

Proteins with common ancestors are known as homologs and homologous proteins belongs to a "family"
taxonomic cladistics (tree diagrams) of sequence analysis are used to show homologies

ex: 1. **serine proteases** ecb fig 4.21- proteolytic enzymes with nearly identical amino acid sequences all with a SER at the active site

2. **globins** - gene slowly diverged into animal and plant lineages
myoglobin - monomeric oxygen binder of muscle fig 3.13a*
hemoglobin - tetrameric oxygen binder of blood fig 3.13b*

Today, computer modeling is used to predict function of yet unisolated proteins by comparing known sequence homologies

sequence analysis = 2ndary structure
QUARTERNARY structure:
  multiple polypeptides each with 3-D conformations
  ex: hemoglobin, RNA polymerase, ASP-trans-carbamylase

Some Common Quarternary Level Protein Shapes...
1. dimers - self recognizing symmetrical regions
   - bind together @ identical binding sites
   [Catabolic Activator Protein] homodimers - 2 identical subunits
   heterodimers - non-identical subunits (PDH)

2. tetramers - 4 identical subunits... [neuraminidase]

3. filaments - polymers of subunits each bound together in an identical way
   forming a ring or helix see fig 4.24*

4. colied-coil - 2 parallel helicities forming a stiff filament, linked via
   a stripe of hydrophobic aa's. figure* [keratin-fig 4.16*]

Multi-Enzymes Complexes: pyruvate dehydrogenase picture* & pic
ATP-synthase figure*
Multimeric proteins have Quartnerary Structure...

**3D shape of a protein**.... involving more than one polypeptide or subunits of a protein

**HA (hemagglutinin A)** is a trimer of 3 identical polypeptide subunits held together by the weak electrostatic 3o level forces  fig 3.10a*
creates a globular domain and a fibrous domain

Some proteins form **Macromolecular Assemblies**...
very large > 1m Da in ma ), 30-300 nm in size, & 10-100 individual peptides

**examples** include: viral capsids, some cytoskeletal complexes, molecular machines, & mRNA transcription complex (some 60 proteins - fig 3.9*)
Selected examples of some Molecular Machines can be seen in **Table 3.1***
we will look at some of these in greater detail later
Protein Conformation is critical to Biological Function

**DENATURATION**  loss of 3-D conformation by heat, pH, organic solvents, detergents

![Diagram of protein denaturation and renaturation](image)

*Fig 4.7 p124*

**RENATURATION**  - regaining of biological activity via self-assembly

Proteins & their structure  Mallery  18
protein shape & conformation...

the **NATIVE Protein CONFORMATION** is the...
3-D SPATIAL ORIENTATION
that's **MOST thermodynamically STABLE**
& has the lowest free energy expenditure, and forms spontaneously

3 most common conformations

**HELIX** - a spiral staircase-like shape
**FIBER** - elongated bound monomers
**GLOBULAR** - roughly a sphere

the Native Conformation of **most enzyme proteins is GLOBULAR:**
an interior pocket of **hydrophobics**
 exterior surface of **hydrophilics**
- maximizes the number H-bonds that form  fig 5.5*

the **PHYSICAL forces** include mostly **weak electrostatic bonds***:
non-covalent bonds, H-bonds, hydrophobic & hydrophilic interactions,
& covalent bonds (as in peptide bonds & disulfide bonds)...
results in a variety of protein shapes & sizes - fig 4.9 pg 127*
How does 3D protein folding come about? "FUNCTION follows FORM"

peptide bond is PLANAR (partial double bond character) as are all the atoms bonded to it all occur is same plane* & thus there is no free rotation = restricts protein conformations

the native folded conformation is most stable, i.e., in lowest free energy state, often dictated by R-group properties (size, hydrophobicity) hydrophilicity, ionic strength, etc...

folding involves: changes in 3D conformations:
- by orderly steps in a sequential way, each step facilitating the next -
  - first 20 structure (α & β), then structural motifs & assembly of complex domains, followed by 30 level forces and/or 40 shapes. fig 3.15*

Unless protected during folding, proteins would interact with all the other molecules in a cell.

Cells makes 2 sets of proteins that facilitate folding: CHAPERONES...

Molecular Chaperones - which bind and stabilize newly made unfolded proteins preventing these proteins from self aggregating and/or being denatured before folding.

Chaperonins - which makeup a small folding chamber into which unfolded proteins are moved to provide a proper environment favoring native folding of a protein.
**MOLECULAR CHAPERONES** - are families of proteins to help "properly fold" a new protein...

multiple ones bind to newly made proteins and include:

- **Hsp70** (of cytosol & mitoplasm);
- **BiP** (of the E.R.);
- **DnaK** (of bacteria).

1st discovered by **heat shock treatment** [under temperature elevation (25° --> 32°C)] cells make heat shock proteins (HSPs); mutant bacteria didn't make Hsp's nor assemble normal proteins. when bound with **ATP = OPEN** conform w hydrophobic pocket for new unfolded protein  ADP conform closes around protein and aids native folding...

**Classes of Heat Shock Proteins:**  **Hsp -40, -60, -70, -90 & -100.**

Hsp are named according to the molecular weights (Hsp-70 = 70 kilodaltons)

- **Hsp-40** binds new protein amino acid chains & carries it to Hsp-70

- **Hsp-70** grabs proteins by an open cleft when ATP is bound to Hsp-70;
  
  OPEN conformation has hydrophobic pocket for new unfolded protein...
  
  in its ADP conform closes around protein and aids native folding...

- **Hsp-90** receives partially folded proteins from Hsp-70's and other chaperones...
  
  helps join polypeptides into larger quaternary proteins forming multi-subunit proteins, such as cellular receptors.
CHAPERONINS or Foldase
- small folding CHAMBERS into which unfolded proteins are moved to provide a proper environment favoring native folding.
- a molecular Machines made of chaperone proteins hsp70's & hsp60's form a barrel shaped structure made of 14 polypeptides (from GroEL gene) in 2 donut rings with a cap (from GroES gene) that opens an inner chamber, where a cell's new protein enters & is folded.

barrel chamber has 2 conformations: tight & relaxed;
new polypeptides is inserted into cavity of GroEL chamber & conformational changes favor native protein folding; ATP hydrolysis = relaxed state & release of a native 3D-protein mcb6e-fig 3.17*

Proteins & their structure  MALLERY 22
Misfolded Proteins & Disease

CJD: Creutzfeld-Jacob disease, genetic based or acquired
- (eating "mad cow" tissue)
fatal neurological disease due to misfolded PRPc protein.

Spongiform Encephalopathy (SE) –
vacuolation (holes) in brain nerve tissue

PRION: a defective protein agent (PrPsc) due to mis-coded gene (PRNPC)
native prion protein is PrPc & resides on nerve cell surfaces...
defective protein PrPsc accumulates forming aggregates
that lead to CJD & SE’s

Both PRION proteins can have identical aa sequence,
but may fold differently
[are known as conformers =
proteins differ only in conformation]

A. normal (PrPc) protein...
mostly α-helix foldings - remains soluble
B. abnormal PrPsc protein...
45% β-sheet - insoluble & protease insensitive
produces cell surface aggregates that kill cells
PROTEIN DEGRADATION (Digestion/Turnover)

cells often contain specialized mechanisms or pathways to digest cell proteins...

1. that rapidly turnover proteins with short half-lives

2. that recognize & eliminate damaged or misfolded proteins that can lead to diseases as Huntington's, Alzeheimer's, and Creutzfeldt-Jacob disease.

many proteins are degraded in cytosol using proteases hydrolyze peptide bonds
some proteins are degraded in the lysosomes via phagocytosis,

but most proteins are degraded by large complexes of proteolytic enzymes in structures known as PROTEASOMES by ubiquitin-mediated proteolysis (UMP)

short half-life proteins hold a signal sequence targeting proteins for UMP and misfolded proteins seem to be recognized for degradation by the UMP.
Discovered by Alfred Goldberg & Martin Rechsteiner in 1980’s

PROTEOSOMES are large multi-enzyme complexes (fig 7.36*)

Average human cell holds between 20,000 & 30,000 proteasomes. Each proteasome is a barrel shaped complex (2,400kD) made of 4 parts

1) a Lid of 9 proteins,
2) a Regulatory Cap that lets in only ubiquitinized proteins,
3) a Base of 4 stacked protein rings with protease activities, and
4) a Base Cap.

Protein Digestion... begins when cells add small polypeptide (ubiquitin)-to-protein to be degraded.

Ubiquitin: globular protein of 76 aa (virtually identical aa sequence in bacteria, yeast, or mammals); 3 ubiquitin ligase enzymes [ E1, E2, E3 ] add Ubiquitin to proteins to be degraded, a ubiquitinized protein is targeted for entry into a Proteasome 's central chamber, where proteases with chymotrypic, tryptic, & caspase-like proteolytic activity cleave the protein into peptides. The ubiquitin is recycled.

figure of the UMP
Protein Engineering...
producing novel proteins, with unique shapes, via artificial means

1. use proteomics...
   make artificial proteins of desired sequence to functions as "drugs"
   vaccine protein - binds to viral surface and inactivates it
   simplistic idea - but it's hard to make connection from 1o to 3o

2. modify existing proteins via site directed mutagenesis
   isolate a gene, alter its sequence in precise way,
   clone the protein product
   - can be used to study effect of one amino acid change on 3D-folding
   - often done with clinically useful proteins to enhance efficiency (Km)

3. structure based drug design
   make drug molecules with high binding affinity to known proteins
   [to remove it] use computers to design 'virtual' drug to fit into
   a protein rendering it inactive

4. Bionanotechnology - viruses made to order (A. belcher of MIT)