Chapter 3-I – Protein Structure and Function

GBME, SKKU
Molecular & Cell Biology

H.F.K.
Chapter 3 – Protein Structure and Function
Chapter 3 - Protein Structure and Function

3.1 Hierarchical Structure of Proteins
3.2 Protein Folding
3.3 Protein Binding and Enzyme Catalysis
3.4 Regulating Protein Function
3.5 Purifying, Detecting, and Characterizing Proteins
3.6 Proteomics
Protein?

- One of the final products in cell
- Various functions
1. Structure
3.1 *Hierarchical Structure* of Proteins

- Protein sequence specifies folding into *secondary and tertiary structures* that either are functional units or can interact with other peptides to form quaternary structure functional units.

- Homologous proteins *evolved from a common ancestor*, have similar sequences, structures, and functions, and can be classified into families and superfamilies.
Four levels of protein hierarchy.

- (a) **Primary** structure – linear sequence of amino acids linked together by peptide bonds
- (b) **Secondary** structure – Folding of the polypeptide chain into local α helices or β sheets
- (c) **Tertiary** structure
  - structure of a peptide composed of secondary structural elements and various loops and turns
  - may form distinct, independently stable domains
- (d) **Quaternary** structure – some functional proteins are composed of more than one polypeptide
Structure of a polypeptide.

- Proteins – unbranched polymers constructed out of 20 amino acids with different R group side chains.
- Amino acid side chains determine the distinct properties of individual proteins.
Secondary structure

The α helix, a common secondary structure in proteins.

- Secondary structures – stable spatial arrangements of polypeptide chain segments held together by hydrogen bonds between backbone amide and carbonyl groups

- α helix:
  - Polypeptide backbone (ribbon) folds into a spiral/helix with 3.6 amino acids per turn (0.54 nm).
  - Helix is stabilized by hydrogen bonds between backbone oxygen and hydrogen atoms (more bonds-more stable).
  - R groups project outward from the surface of the helix – determine chemical nature of helix faces.
  - Prolines – can’t participate in hydrogen bonding and usually are excluded from an α helix.

*Do you remember the proline structure? Please check it!*
Alpha helix

https://www.youtube.com/watch?v=eUS6CEn4GSA
The **β sheet**, another common secondary structure in proteins.

- **β sheet** – laterally packed β strands, each of which is a nearly fully extended polypeptide segment

- (a) Three-stranded β sheet – antiparallel β strands with connecting loops (top view):
  - Stabilized by hydrogen bonds between backbone oxygen and hydrogen atoms in amino acids on different strands

- (b) Antiparallel β sheet (side view):
  - α carbon bond angles produce a pleated polypeptide backbone contour.
  - Alternate R groups project above and below the plane of the sheet.

- (c) Parallel β strand sheet: same N-to-C strand orientations with connecting loops.
Beta sheet

https://www.youtube.com/watch?v=wM2LWCTWlrE
Structure of a β turn.
• Composed of four residues.
• Reverses direction of a polypeptide chain (180° U-turn).
• Cα carbons of the first and fourth residues are usually less than 0.7 nm apart and linked by a hydrogen bond.
• β turns facilitate the folding of long polypeptides into compact structures.
• Glycine (smallest R group) and proline (built in bend) are commonly found in β turns.

*Think about the reason in discussion!*
Protein folding?
Three models of protein folding

- Framework model
- Nucleation model
- Hydrophobic-collapse model

No single mechanism...

https://www.youtube.com/watch?v=B275XtegrJ4
The oil drop model of protein folding.

- Tertiary structure – stabilized primarily by hydrophobic interactions between nonpolar side chains and hydrogen bonds involving polar side chains and backbone amino and carboxyl groups.
- Hydrophobic residues (blue): cluster together like drops of oil in the folded protein core, driven away from the aqueous surroundings by the hydrophobic effect.
- Charged and uncharged polar side chains (yellow): form stabilizing interactions with surrounding water and ions on the protein surface.
Four ways to visualize protein structure.

- Ras, a monomeric (single polypeptide chain) protein that binds to guanosine diphosphate (GDP, in blue).
- (a) Cα backbone trace – depicts how the polypeptide is tightly packed into a small volume
- (b) Ball-and-stick representation – reveals locations of all atoms
- (c) Ribbon diagram – emphasizes how β strands (light blue) and α helices (red) are organized in the protein
- (d) Water-accessible surface model – reveals protein surface topology with positive charge (purple) and negative charge regions (red)
How to find?

- http://www.uniprot.org/
Find structure of CREB!
Similar functions by similar structures!!!

- Structural motifs:
  - *Regular combinations* of secondary structures usually with a specific type of function
  - Can be encoded by a *highly conserved sequence motif*
Motifs of protein secondary structure.

- Two $\alpha$ helices wound around each other

- A type of helix-loop-loop-helix motif in many proteins, including many calcium-binding and DNA-binding regulatory proteins

- Present in many DNA-binding proteins that help regulate transcription
The Alpha-Helical Secondary Structure of Myoglobin (Mb)

- These helical segments are labeled A-H and span the following residues of this 153 amino acid residue polypeptide chain of Mb. Segments (residues): **A** (3-18), **B** (20-35), **C** (36-42), **D** (51-57), **E** (58-76), **F** (83-95), **G** (100-118), & **H** (124-149).

https://biosci.mcdb.ucsb.edu/biochemistry/tw-prt/myoglobin/alphahelixf.htm
Zinc finger domain

https://www.youtube.com/watch?v=WyU2v7HT6bw
Leucin zipper domain

https://www.youtube.com/watch?v=2-qFLfVymnw
Tertiary and quaternary levels of structure.
Tertiary and quaternary levels of structure.

The tertiary structure will have a single polypeptide chain "backbone" with one or more protein secondary structures, the protein domains.

Protein quaternary structure is the number and arrangement of multiple folded protein subunits.
Modular nature of protein domains.

• Epidermal growth factor (EGF) precursor: generated by proteolytic cleavage generates multiple EGFs (green)
• Neu: EGF domain plus other domains
• Tissue plasminogen activator (TPA): EGF domain plus other domains
Several proteins work together in Cell!
A molecular machine: the transcription initiation complex.

Supramolecular complexes
Molecular evolution?
Evolution of the globin protein family.

Myoglobin (symbol Mb or MB) is an iron- and oxygen-binding protein found in the muscle tissue of vertebrates in general and in almost all mammals.

Leghemoglobin (also leghaemoglobin or legoglobin) is a nitrogen or oxygen carrier and hemoprotein found in the nitrogen-fixing root nodules of leguminous plants.
Protein Structure and Function

• 3.2 Protein Folding
• Protein amino acid sequence determines its 3D structure and function.
• ATP-dependent molecular chaperones and chaperonins assist protein folding in vivo.
• Misfolded/denatured proteins can form well-organized amyloid fibril aggregates that can cause diseases, including Alzheimer’s disease and Parkinson’s disease.
Think the protein folding!

What’s the important physical movement for the folding?
Rotation!
Rotation between planar peptide groups in proteins.

Polypeptide backbone steric restraints and amino acid side chain properties severely restrict $C_\alpha$–amino nitrogen bond (the $\Phi$ angle) and $C_\alpha$–carbonyl carbon bond (the $\Psi$ angle) rotations.
What are the trans and cis?

- Planar peptide bonds **limit** the shapes into which proteins can fold.

*Which R group of amino acid has the tras- and cis-forms?*
Proline cis/trans isomerizations influence protein folding and structure.

• Cis/trans isomerization of a single proline:
  • Alters structure of a protein SH$_2$ domain.
  • Can influence a protein’s activity.
  • Proline isomerases may act as switches that regulate protein activity.
Hypothetical protein-folding pathway

primary (a) $\rightarrow$ secondary (b–d) $\rightarrow$ tertiary (e) structure
But,
the folding is not always
‘easy-going’ process
Molecular chaperone–mediated protein folding.
Chaperonin-mediated protein folding.

It’s a protein chamber...
Chaperone and Chaperonin

https://www.youtube.com/watch?v=d1QIEQEyYRo
Misfolded proteins can form ordered amyloid aggregates based on a cross-β sheet structure.
Protein Structure and Function

3.3 Protein Binding and Enzyme Catalysis

• Protein function **depends on binding other molecules** (ligands).

• **Enzymes accelerate rates of cellular reactions** by lowering activation energy and stabilizing transition-state intermediates.

• Enzymes often use acid-base catalysis mediated by one or more amino acid side chains.

• Metabolic pathway enzymes may be associated as domains of a monomeric protein, subunits of a multimeric protein, or components of a protein complex assembled on a common scaffold.
Protein-ligand binding of antibodies.

- Ligand: molecule to which a protein binds
- (a) IgG antibody (ribbon model): two identical heavy chains (light and dark red) and two identical light chains (blue) covalently linked by disulfide bonds
- CDR (complementarity-determining regions): six highly variable loops form the antigen-binding sites
- (b) Hand-in-glove fit: molecular complementarity between the antibody CDR and the site to which it binds (epitope) on its target antigen
Remind the role of enzyme
Active site of the enzyme trypsin.

- Enzymes (proteins or RNAs) catalyze making or breaking substrate covalent bonds.
- (a) Trypsin (serine protease) active site:
  - Substrate-binding pocket – binds specific substrate
  - Catalytic site – contains side chains of the catalytic triad Ser-195, Asp-102, and His-57 that breaks peptide bonds.
- In some enzymes, the catalytic and substrate-binding sites overlap; in others, the two regions are structurally distinct.
Schematic model of an enzyme’s reaction mechanism.
Next class...

- Real enzyme action!
- Structural change by ligand binding & tech.
- Practical protein works learned by text (!?!?)

Kim et al., Neuron 2006
Discussion with friends

• Think about what cause misfolding of proteins.
• Then discuss how the chaperone and chaperonin help the proper folding of proteins (mechanisms).
• Practically protein misfolding can be a big problem in protein production in basic research field and industry. How can you solve this misfolding problem? Please find plausible ways except the chaperone and chaperonin.
Further questions

• How do the chaperon and chaperonin recognize the misfolding proteins?
• Can the human chaperon be used in other animals such as rat and mouse?
• GroEL belongs to the chaperonin family of molecular chaperones, and is found in a large number of bacteria. To function properly, GroEL requires the lid-like cochaperonin protein complex GroES. In eukaryotes the proteins Hsp60 and Hsp10 are structurally and functionally nearly identical to GroEL and GroES, respectively.

• Unfolded substrate proteins bind to a hydrophobic binding patch on the interior rim of the open cavity of GroEL, forming a binary complex with the chaperonin.
Conformational specificity of the chaperonin GroEL for the compact folding intermediates of α-lactalbumin

The chaperonin GroEL binds unfolded polypeptides, preventing aggregation, and then mediates their folding in an ATP-dependent process. To understand the structural features in non-native polypeptides recognized by GroEL, we have used α-lactalbumin (αLA) as a model substrate. αLA (14.2 kDa) is stabilized by four disulfide bonds and a bound Ca^{2+} ion, offering the possibility of trapping partially folded disulfide intermediates between the native and the fully unfolded state. The conformers of αLA with high affinity for GroEL are compact, containing up to three disulfide bonds, and have significant secondary structure, but lack stable tertiary structure and expose hydrophobic surfaces. Complex formation requires almost the complete αLA sequence and is strongly dependent on salts that stabilize hydrophobic interactions. Unfolding of αLA to an extended state as well as the burial of hydrophobic surface upon formation of ordered tertiary structure prevent the binding to GroEL. Interestingly, GroEL interacts only with a specific subset of the many partially folded disulfide intermediates of αLA and thus may influence in vitro the kinetics of the folding pathways that lead to disulfide bonds with native combinations. We conclude that the chaperonin interacts with the hydrophobic surfaces exposed by proteins in a flexible compact intermediate or molten globule state.
The Region of $\alpha$-Lactalbumin Recognized by GroEL

Akio Shimizu, Takahiro Tanba, Isamu Ogata, Masamiichike Ikekuchi, and Shintaro Sugai

Department of Bioengineering, Faculty of Engineering, Soka University, 2-235 Tsurumi-cho, Hachioji, Tokyo 192-8577

Received for publication, February 17, 1998.

The binding constants between disulfide-intact or various disulfide-reduced bovine $\alpha$-lactalbumins and an Escherichia coli chaperonin, GroEL, were determined by using the equilibrium dialysis method. The disulfide-intact and one-disulfide (Cys6-Cys120)-reduced $\alpha$-lactalbumins were shown not to bind with GroEL both in the presence and absence of Ca$^{2+}$. The two-disulfide (Cys6-Cys120 and Cys28-Cys111)-reduced $\alpha$-lactalbumin, which has the native-like tertiary structure in its $\beta$-domain region and an unfolded $\alpha$-domain in the presence of Ca$^{2+}$, showed considerable binding with GroEL. The binding free energy of the two-disulfide-reduced $\alpha$-lactalbumin in the presence of Ca$^{2+}$ is close to that of the molten globule state of disulfide-intact $\alpha$-lactalbumin. This result suggests that GroEL binds to the unfolded $\alpha$-domain of $\alpha$-lactalbumin regardless of the conformation of the $\beta$-domain. The fully disulfide-reduced and two-disulfide-reduced $\alpha$-lactalbumins were found to bind more strongly with GroEL in the absence of Ca$^{2+}$ than the two-disulfide-reduced $\alpha$-lactalbumin in the presence of Ca$^{2+}$, thus indicating that the unfolding of the $\beta$-domain of $\alpha$-lactalbumin leads to stronger interaction with GroEL.

Key words: equilibrium dialysis, GroEL, $\alpha$-lactalbumin, binding constant, binding site.
Roles of Cytosolic Hsp70 and Hsp40 Molecular Chaperones in Post-translational Translocation of Presecretory Proteins into the Endoplasmic Reticulum*

Jantra Ngosuwan, Nancy M. Wang, Katie L. Fung, and William J. Chirico†
From the Department of Anatomy and Cell Biology, State University of New York Downstate Medical Center, Brooklyn, New York 11203

Hsp70 molecular chaperones and their co-chaperones work together in various cellular compartments to guide the folding of proteins and to aid the translocation of proteins across membranes. Hsp70s stimulate protein folding by binding exposed hydrophobic sequences thereby preventing irreversible aggregation. Hsp40s stimulate the ATPase activity of Hsp70s and target unfolded proteins to Hsp70s. Genetic and biochemical evidence supports a role for cytosolic Hsp70s and Hsp40s in the post-translational translocation of precursor proteins into endoplasmic reticulum and mitochondria. To gain mechanistic insight, we measured the effects of *Saccharomyces cerevisiae* Ssa1p (Hsp70) and Ydj1p (Hsp40) on the translocation of histidine-tagged prepro-α-factor (ppαF6H) into microsomes. Radiolabeled ppαF6H was affinity purified from wheat germ translation reactions (or *Escherichia coli*) to remove endogenous chaperones. We demonstrated that either Ssa1p or Ydj1p stimulates post-translational translocation by preventing ppαF6H aggregation. The binding and/or hydrolysis of ATP by Ssa1p were required to maintain the translocation competence of ppαF6H. To clarify the contributions of membrane-bound and cytosolic Ydj1p, we compared the efficiency of chaperone-dependent translocation into wild-type and Ydj1p-deficient microsomes. Neither soluble nor membrane-bound Ydj1p was essential for post-translational protein translocation. The ability of Ssa1p, Ydj1p, or both chaperones to restore the translocation competence of aggregated ppαF6H was negligible.
HSP70 overexpression

Heat Shock Protein 70 Chaperone Overexpression Ameliorates Phenotypes of the Spinal and Bulbar Muscular Atrophy Transgenic Mouse Model by Reducing Nuclear-Localized Mutant Androgen Receptor Protein

Spinal and bulbar muscular atrophy (SBMA) is an inherited motor neuron disease caused by the expansion of the polyglutamine (polyQ) tract within the androgen receptor (AR). The nuclear inclusions consisting of the mutant AR protein are characteristic and combine with many components of ubiquitin–proteasome and molecular chaperone pathways, raising the possibility that misfolding and altered degradation of mutant AR may be involved in the pathogenesis. We have reported that the overexpression of heat shock protein (HSP) chaperones reduces mutant AR aggregation and cell death in a neuronal cell model (Kobayashi et al., 2000). To determine whether increasing the expression level of chaperone improves the phenotype in a mouse model, we cross-bred SBMA transgenic mice with mice overexpressing the inducible form of human HSP70. We demonstrated that high expression of HSP70 markedly ameliorated the motor function of the SBMA model mice. In double-transgenic mice, the nuclear-localized mutant AR protein, particularly that of the large complex form, was significantly reduced. Monomeric mutant AR was also reduced in amount by HSP70 overexpression, suggesting the enhanced degradation of mutant AR. These findings suggest that HSP70 overexpression ameliorates SBMA phenotypes in mice by reducing nuclear-localized mutant AR, probably caused by enhanced mutant AR degradation. Our study may provide the basis for the development of an HSP70-related therapy for SBMA and other polyQ diseases.

Key words: HSP70; chaperone; polyglutamine; SBMA; transgenic mice; protein degradation
Human HSP70 expression in mouse

There are significantly more 1C2-positive cells in AR-97Q/HSP70 double-transgenic mice stained with a monoclonal antibody (1C2) against abnormally expanded polyQ (16 weeks old). AR-97Q/HSP70 six mice. The differences in 1C2-positive cell populations are not statistically significant between AR-97Q/HSP70 mice and control mice. These observations suggest that 1C2 staining is a more sensitive histological marker for the detection of the small subpopulation of neurons and muscles, particularly in the spinal anterior horn and muscle.

In the present study, we demonstrated that the amount of nuclear-localized mutant AR was significantly reduced in the AR-97Q/HSP70 double-transgenic mice. Immunohistochemical study of the spinal anterior horn and muscle showed low levels of 1C2 staining in the nucleus of AR-97Q/HSP70 (tg/tg) mice, whereas testosterone administration enhanced the expression of HSP70 and decreased nuclear-localized mutant AR in double-transgenic mice. This suggested that HSP70 decreases nuclear-localized mutant AR in double-transgenic mice.