Chapter 3. Structure of Enzymes

Enzyme Engineering

3.1 Introduction

With purified protein,

- Determining M_r of the protein
- Determining composition of amino acids and the primary structure
- Determining the secondary and tertiary structure
- Determining the quaternary structure

→ Bioinformatics (Protein Data Bank) tools

3.2. Determining M_r of the protein

- M_r ~ 10,000 to several millions
- Methods to determine M_r
  1. Ultracentrifugation
  2. Gel filteration
  3. SDS PAGE
  4. Mass Spectrometry
- Method 2 and 3 provide relative mass
- Absolute Mass with method 1
- Method 4 is a new methodology and extremely fast and accurate

3.2.1 Ultracentrifugation

- Ultracentrifuge Sedimentation (at very high speed)

$$M_r = \frac{RTs}{D(1-\bar{v}\rho)}$$

R : gas constant, T: temp., s : sedimentation coefficient, D : diffusivity, ρ : density, $\bar{v}$: partial specific volume of a solute

- Ultracentrifuge Equilibrium

$$M_r = \frac{2RT}{(1-\bar{v}\rho)w^2}\frac{d\ln c}{dr^2}$$

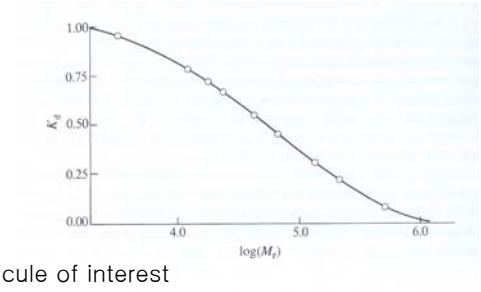
r : distance along the cell, ω : angular velocity, c : concentration

- Useful technique to study cofactors, ligands, etc.

3.2.2 Gel filtration

- Distribution coefficient of proteins in cross-linked polymer is correlated with their M_r's

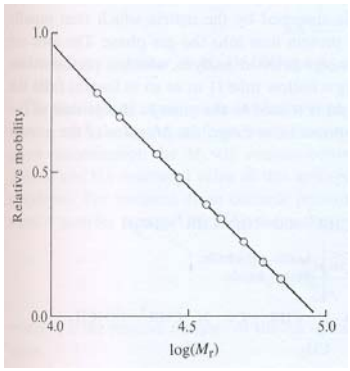
$$K_d = \frac{V_e - V_o}{V_i - V_o}$$



V_e : elution volume of the molecule of interest
V_i : elution volume of small molecule
V_o : elution volume of excluded molecule

3.2.3 Electrophoresis

- SDS (Sodium dodecylsulphate)
  - Making the proteins have constant charge
  - Denaturing the proteins, so they have the same shape
  - SDS dissociate the subunits of the protein

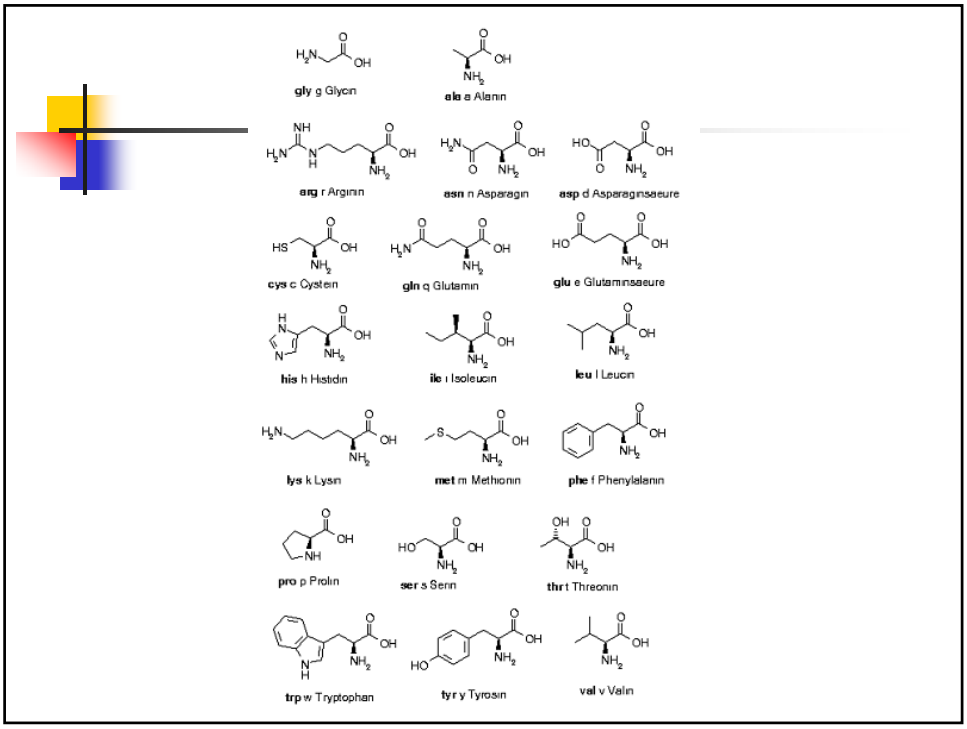


3.2.4 Mass Spectrometer

- Conventional ionization method could be used for the protein with 500 or less amino acids
- Soft ionization methods have been developed
  - MALDI (Matrix-assisted laser desorption ionization)
  - Good for large proteins
  - Protein dissolved in small volume of organic acids as matrix → UV laser → stream of protein ions → Analyzed by time of flight (TOF)
  - ESI (electrospray ionization)
  - Extremely accurate (Good for studying post-translational modification)
  - Protein in organic solvent is passed through a needle with high electron charge → Solvent evaporated and protein with charge is vaporized → Mass is analyzed by quadrupole mass analyzer
  - MS/MS : fast sequencing of protein

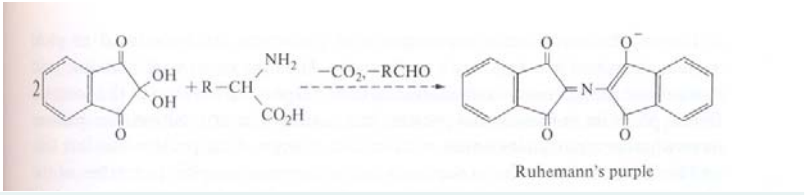
3.3 Amino acids composition and primary structure

- Amino Acids
  - 20 amino acids
  - Zwitterion: $^{+}NH_3-CHR-CO_2^{-}$ is natural form at pH 7
  - Chiral compounds except glycine; only L-form is used
  - Amide (peptide) bond can be formed between NH_3^{+} and CO_2^{-}; polypeptide or protein can be made
  - For convenience N-terminal is numbered first
  - Cysteine residue can form covalent bond (Cystine): important for extracellular proteins



3.3.1 Amino acids composition

- Amino acids composition
  - Hydrolyze the amide bonds: 6 mol dm^{-3} HCl or p-toluene sulfphonic acid at 110°C
  - Separate 20 amino acids with pH gradient ion exchange chromatography
  - Measure the quantity of amino acids with ninhydrin or phthalaldehyde reaction



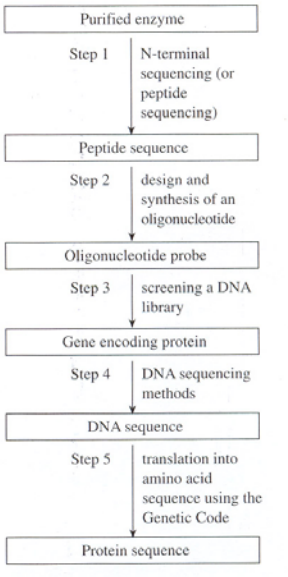
3.3.1 Amino acids composition

- Why amino acid composition?
  1. Compare with protein sequence
  2. Estimate the function or property of proteins
  3. Calculate the molecular weight
  4. Evaluate the concentration of a purified protein

3.3.2 Determining the primary structure of protein

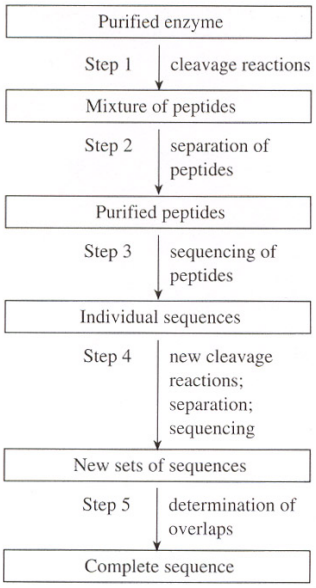
- Indirect method

  – Sequencing of DNA is much easier than protein
  - 15–20 N–terminal peptide sequence
  - Probing cDNA
  - If chromosome sequence is available, 15–20 peptide sequence is enough to get the full information



3.3.2 Determining the primary structure of protein

- Direct method
  - Divide the protein into several fragments
  - Sequence the peptides
  - Determine the overlaps



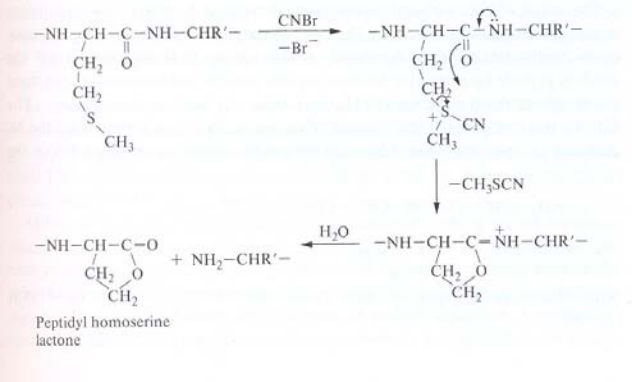
3.3.2 Determining the primary structure of protein

1. Fragmentation with different enzymes

Protease	Nature of amino acid side-chains at cleavage site in polypeptide chain: amino — X Y ↓ Z — carboxyl		
	X	Y	Z
Trypsin[72]	—	Basic side-chain (Lys, Arg)	—
Clostripain[73] (from *Clostridium histolyticum*)	—	Arg	—
Chymotrypsin[74]		Aromatic (Trp, Tyr, Phe) or hydrophobic (e.g. Leu, Met) side-chain	
Elastase[75]	—	Small hydrophobic side-chain (e.g. Ala)	— —
Papain[76,77]	Phe (preference)	Preference for Lys over Ala	Leu or Trp (preference)
Thermolysin[78] (from *Bacillus thermoproteolyticus*)	—	—	Bulky hydrophobic side-chain (e.g. Val, Leu, Ile, Phe)
Endoprotease Glu-C[79] (from *Staphylococcus aureus* V8)	—	Glu	—
Pepsin[80]	—	Phe (preference)	Phe, Trp, Tyr (preference)

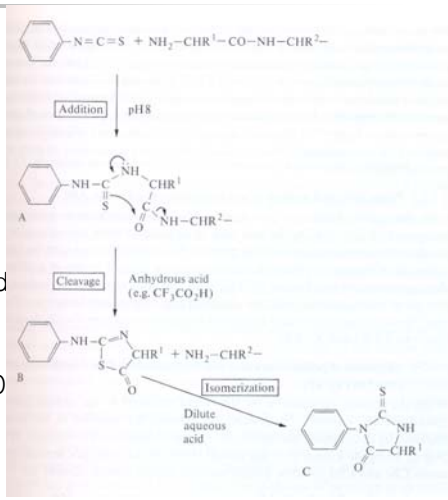
3.3.2 Determining the primary structure of protein

or with chemical



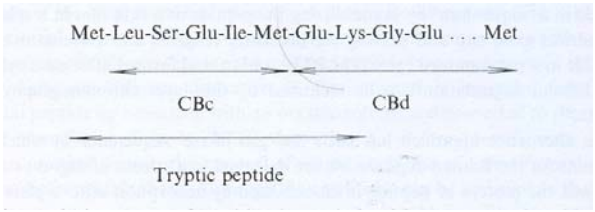
3.3.2 Determining the primary structure of protein

2. Sequencing each fragments
   - Reverse phase HPLC for separating the fragments
   - Edman degradation method
   - Gas sequenator can read through 150 aa
   - Mass spec has been developed to read up to 50 aa



3.3.2 Determining the primary structure of protein

3. Aligning the sequences



The importance of protein sequence
1. It may tell us the function of the protein
2. It provides the evolutionary relationship between organisms

3.4 Determining secondary and tertiary structure

3.4.1 X-ray crystallography

- Crystals of enzyme : Ammonium sulfate or other salts are used.
- Isomorphous heavy metal derivatives : Crystal is soaked to heavy metal solution. (ex. Reaction of cysteine with mercurial compound) Used for calculating phase
- Computing facilities : Electron density is calculated based on positions, intensities, and phase of the protein
- Primary structure

3.4 Determining secondary and tertiary structure

3.4.2 Structure in solution

- Proteins are flexible in solution
- NMR calculates the average distance between hydrogen atoms or between hydrogen and ^{13}C, ^{15}N
- High concentration of protein required

- Importance of flexibility study
  - To understand the induced fit mechanism
  - To find the rate-limiting step in the reaction
  - To understand the thermophilic and mesophilic enzymes

    : The activity of thermophilic enzyme is no greater than mesophilic enzymes

3.4 Determining secondary and tertiary structure

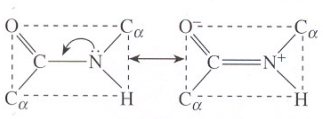
3.4.3 Importance of 3-D structure

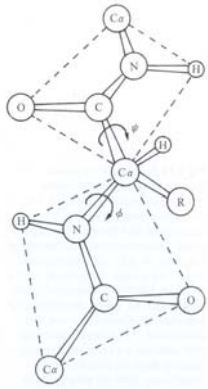
- To test mathematical models aiming the prediction of structure without experiment
- To propose catalytic mechanism
- To explore similarities between enzymes
  - Enzymes with similar function may have similar structure
- To assist rational drug design

3.4 Determining secondary and tertiary structure

3.4.4 Features of secondary structure

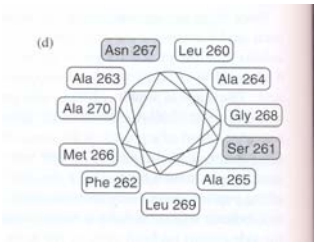
- Peptide bond
  - Partially double bond – not rotating
  - Trans form is more stable than cis except proline
  - The bonds with α-carbon are rotating (ψ, ϕ)





3.4 Determining secondary and tertiary structure

- α-helix
  - Stabilized by hydrogen bond between carbonyl groups of n and n+3 amide bonds
  - Proline is helix breaker
  - 3.6 amino acids, 13 atoms per turn, which is one in $100^{\circ} \rightarrow 3.6_{13}$ helix
  - 3_{10} helix is also found
  - Helical wheel



3.4 Determining secondary and tertiary structure

- β-sheet
  - Hydrogen bond between carbonyl groups between chains stabilizes the structure
  - Slightly twisted to right hand rather than flat
- β-turn
  - 180° turn, which make proteins globular
  - 2 to 5 amino acids turn is most common

3.4 Determining secondary and tertiary structure

- General rules for enzyme structure
  - Very closely packed globular structure
  - Nonpolar amino acids are located in the interior of the enzyme, while polar ones are in the exterior
  - Polar groups in the interior have special functions and are paired by hydrogen bonds
  - Large enzymes consist of multiple domains (ex) phosphoglycerate kinase has two domains, one for substrate binding, the other for interaction with other proteins, while Factor IX has four
  - Large enzymes may consist of multiple subunits

3.4 Determining secondary and tertiary structure

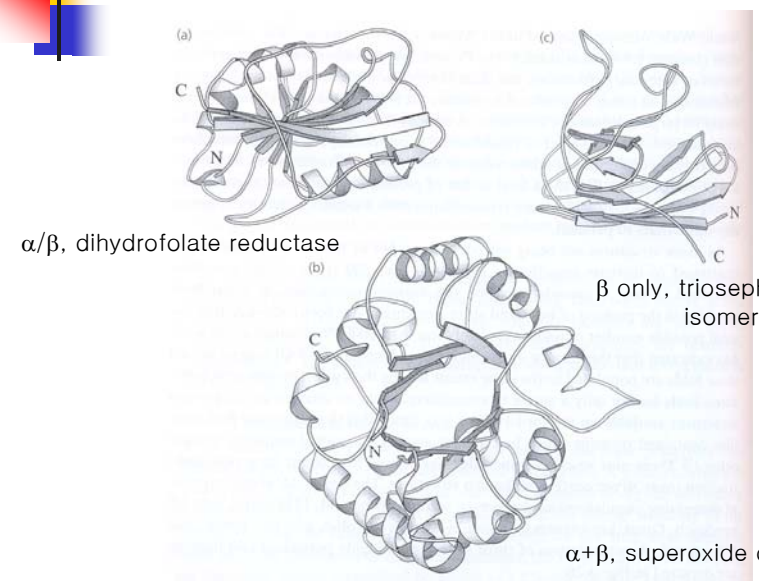
- Classification of enzyme structure

1. All α proteins : myoglobin and citrate synthase
2. All β proteins : chymotrypsin and immunoglobulins
3. α/β proteins : triosephosphate isomerase and etc.
4. α+β proteins : The class is rather small, and sometimes regarded as sub-class of α/β (SCOP)

→ computer-based arrangement of secondary structure elements: fold

- There may be less than 1000 folds existed
- Most fold family contains functionally related proteins

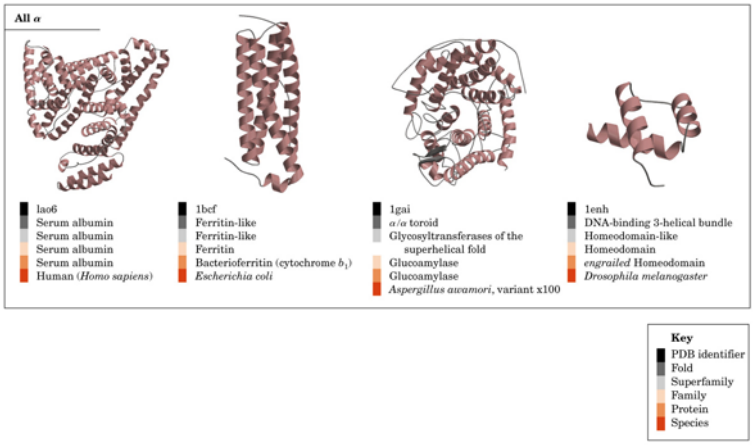
Examples of folds

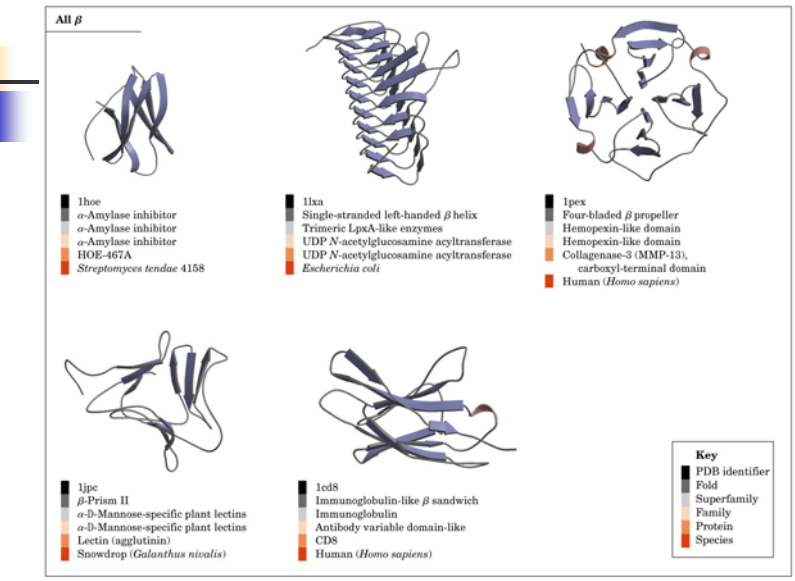


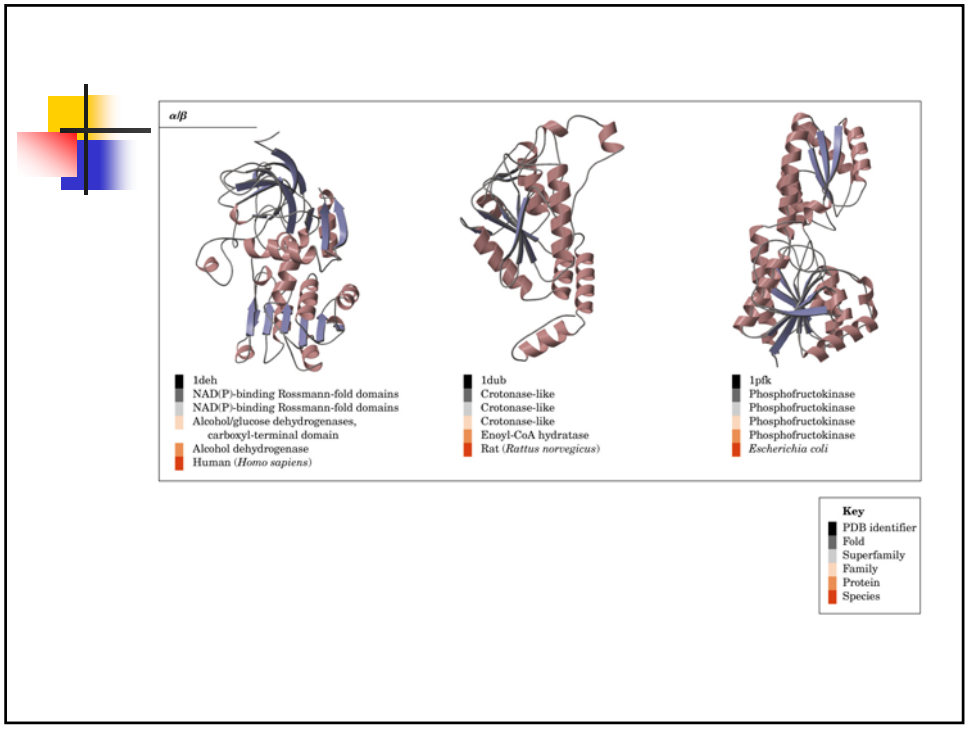
α/β, dihydrofolate reductase

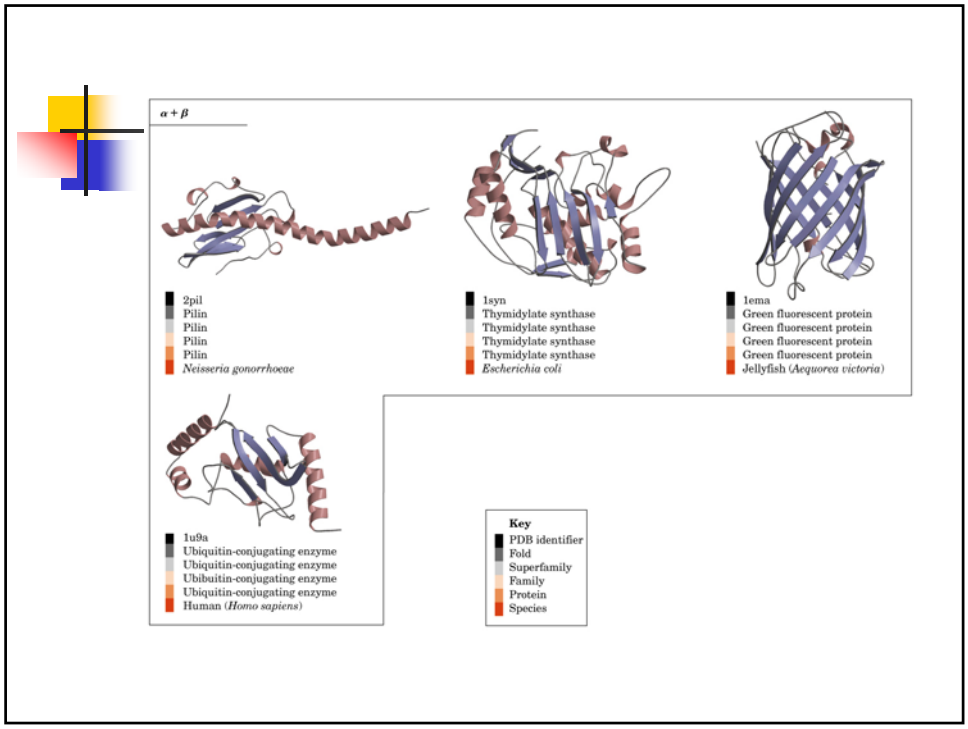
β only, triosephosphate isomerase

α+β, superoxide dismutase


All α
lao6
Serum albumin
Serum albumin
Serum albumin
Serum albumin
Human (Homo sapiens)
1bcf
Ferritin-like
Ferritin-like
Ferritin
Bacterioferritin (cytochrome b1)
Escherichia coli
1gai
α/α toroid
Glycosyltransferases of the superhelical fold
Glucoamylase
Glucoamylase
Aspergillus awamori, variant x100
1enh
DNA-binding 3-helical bundle
Homeodomain-like
Homeodomain
engrailed Homeodomain
Drosophila melanogaster
Key
PDB identifier
Fold
Superfamily
Family
Protein
Species


All β
1hoe
α-Amylase inhibitor
α-Amylase inhibitor
α-Amylase inhibitor
HOE-467A
Streptomyces tendae 4158
1lxa
Single-stranded left-handed β helix
Trimeric LpxA-like enzymes
UDP N-acetylglucosamine acyltransferase
UDP N-acetylglucosamine acyltransferase
Escherichia coli
1pex
Four-bladed β propeller
Hemopexin-like domain
Hemopexin-like domain
Collagenase-3 (MMP-13), carboxyl-terminal domain
Human (Homo sapiens)
1jpc
β-Prism II
α-D-Mannose-specific plant lectins
α-D-Mannose-specific plant lectins
Lectin (agglutinin)
Snowdrop (Galanthus nivalis)
1cd8
Immunoglobulin-like β sandwich
Immunoglobulin
Antibody variable domain-like
CD8
Human (Homo sapiens)
Key
PDB identifier
Fold
Superfamily
Family
Protein
Species


α/β
1deh
NAD(P)-binding Rossmann-fold domains
NAD(P)-binding Rossmann-fold domains
Alcohol/glucose dehydrogenases, carboxyl-terminal domain
Alcohol dehydrogenase
Human (Homo sapiens)
1dub
Crotonase-like
Crotonase-like
Crotonase-like
Enoyl-CoA hydratase
Rat (Rattus norvegicus)
1pfk
Phospbofructokinase
Phosphofructokinase
Phosphofructokinase
Phosphofructokinase
Escherichia coli
Key
PDB identifier
Fold
Superfamily
Family
Protein
Species


α + β
2pil
Pilin
Pilin
Pilin
Pilin
Neisseria gonorrhoeae
1syn
Thymidylate synthase
Thymidylate synthase
Thymidylate synthase
Thymidylate synthase
Escherichia coli
1ema
Green fluorescent protein
Green fluorescent protein
Green fluorescent protein
Green fluorescent protein
Jellyfish (Aequorea victoria)
1u9a
Ubiquitin-conjugating enzyme
Ubiquitin-conjugating enzyme
Ubibuitin-conjugating enzyme
Ubiquitin-conjugating enzyme
Human (Homo sapiens)
Key
PDB identifier
Fold
Superfamily
Family
Protein
Species

3.4 Determining secondary and tertiary structure

3.4.5 Forces stabilize the structure of enzyme

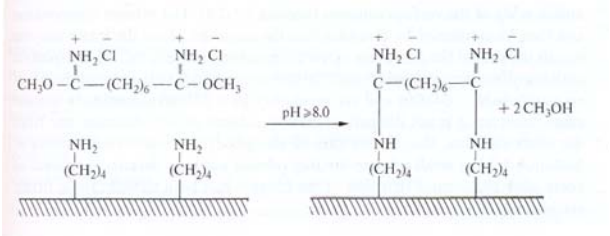
1. Hydrogen bonds ; especially important for α-helix and β-sheet structure (0.28-0.3 nm)
2. Electrostatic forces ; $-NH_3^+$ (Lys) and $-COO^-$ (Asp), also known as salt bridge
3. Van der Waals focrces ; molecules in close contact with one another (less than 0.4 nm)
4. Hydrophobic forces ; greatest contribution in energy for structure stabilization, but no specificity

3.5 Determining quaternary structure

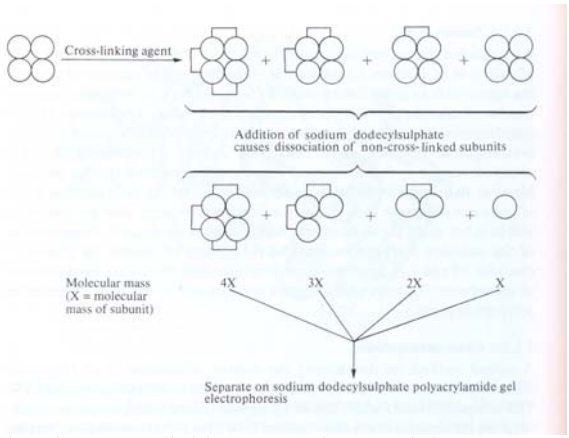
Many enzymes are oligomers → quaternary structure

3.5.1 Number and type of subunits

1. **Studies of M_r** with/without denaturing agent, such as guanidinium chloride
   - Proteases must be absent
2. **Cross-linking studies** with dimethylsuberimidate



3.5 Determining quaternary structure



- Intersubunit cross-linking must be partial
- Intermolecular cross-linking must be avoided

3.5 Determining quaternary structure

3. Studies of ligand binding
4. Studies of symmetry
5. Identity of subunits

   If the subunits are non-identical, more complicated pattern will appear.

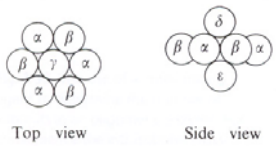
Table 3.5 Some enzymes composed of non-identical subunits

Enzyme	Source	Subunit composition	Comments
Lactose synthase	Bovine mammary tissue	$\alpha\beta$	
Haemoglobin[a]	Human red blood cells	$\alpha_2\beta_2$	
Tryptophan synthase (see Chapter 7, Section 7.9.1)	*E. coli*	$\alpha_2\beta_2$	
cAMP-dependent protein kinase (see Chapter 6, Section 6.4.2.1)	Rabbit skeletal muscle	$\alpha_2\beta_2$	α and β represent catalytic and regulatory subunits, respectively
Aspartate carbamoyltransferase	*E. coli*	$\alpha_6\beta_6$	α and β represent catalytic and regulatory subunits, respectively. The molecule is assembled as $(\alpha_2)_2(\beta_2)_3$
Ribulosebisphosphate carboxylase	Spinach	$\alpha_8\beta_8$	
20S proteasome (see Section 3.4.6.1)	*Thermoplasma acidophilum*	$\alpha_{14}\beta_{14}$	The molecule is assembled as $\alpha_7\beta_7\beta_7\alpha_7$
RNA polymerase (see Chapter 7, Section 7.3)	*E. coli*	$\alpha_2\beta\beta'\sigma$	
Phosphorylase kinase (see Chapter 6, Section 6.4.2.1)	Rabbit skeletal muscle	$\alpha_4\beta_4\gamma_4\delta_4$	

3.5 Determining quaternary structure

3.5.2 Arrangement of subunits

- Determined by structural studies
- Usually arranged to maximize the interactions between subunits



3.5.3 Forces between subunits

- Nonpolar interaction dominant
- If protein has more than 30% nonpolar aa, it likely form an oligomer

3.5 Determining quaternary structure

3.6.4 Why multiple subunits?

1. To regulate catalytic activity
2. To get variation in the catalytic activity
3. To increase the stability of enzyme
4. To get a large geometry, which is necessary for activity

3.6 Folding and unfolding of enzymes

- Unfolding of enzymes
  - The folded form is thermodynamically stable condition by a small margin
  - Easily unfolded = denatured
    - Heat ; Increasing vibration or rotation
    - pH ; Affecting ionization state
    - Organic solvent or detergent ; Breaking non-polar interaction
    - Guanidinium chloride or urea ; Breaking non-polar interaction while maintaining H-bond

3.6 Folding and unfolding of enzymes

- Folding of enzymes
  - Folding achieved *in vitro* from unfolded peptide → Refolding
  - Several intermediates found
  - *In vivo*, chaperones (chaperonines) help correct folding
  - Disulfide isomerase and peptidyl prolyl isomerase are also important for