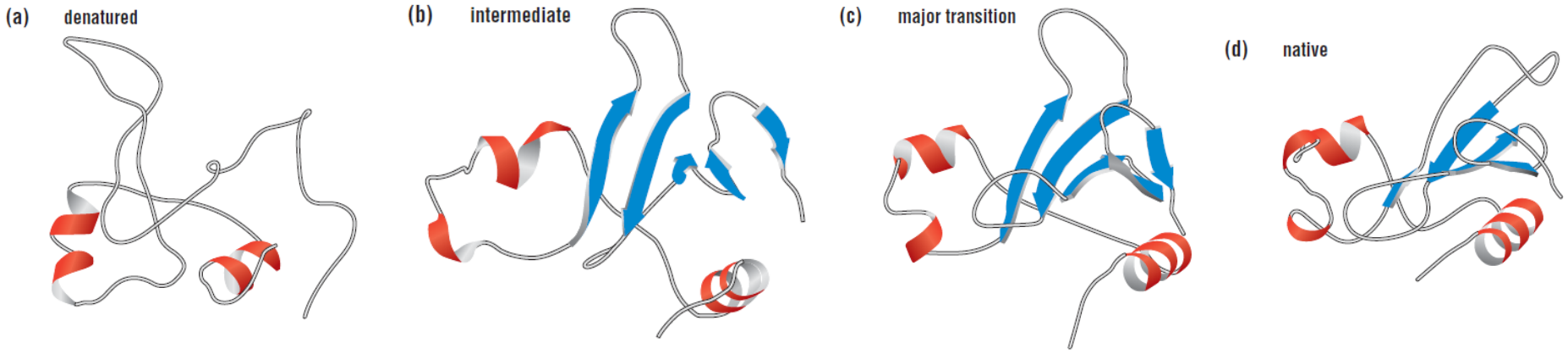
PROTEIN FOLDING AND FLEXIBILITY

Dr. Zhiyi Wei
SUSTC

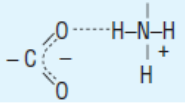
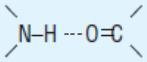
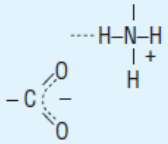
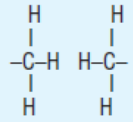
Protein Folding



- The process by which a polypeptide chain acquires its correct 3D structure to achieve the biologically active native state
- Determined by its primary structure
- Driven by competition between self-interactions and interactions with water

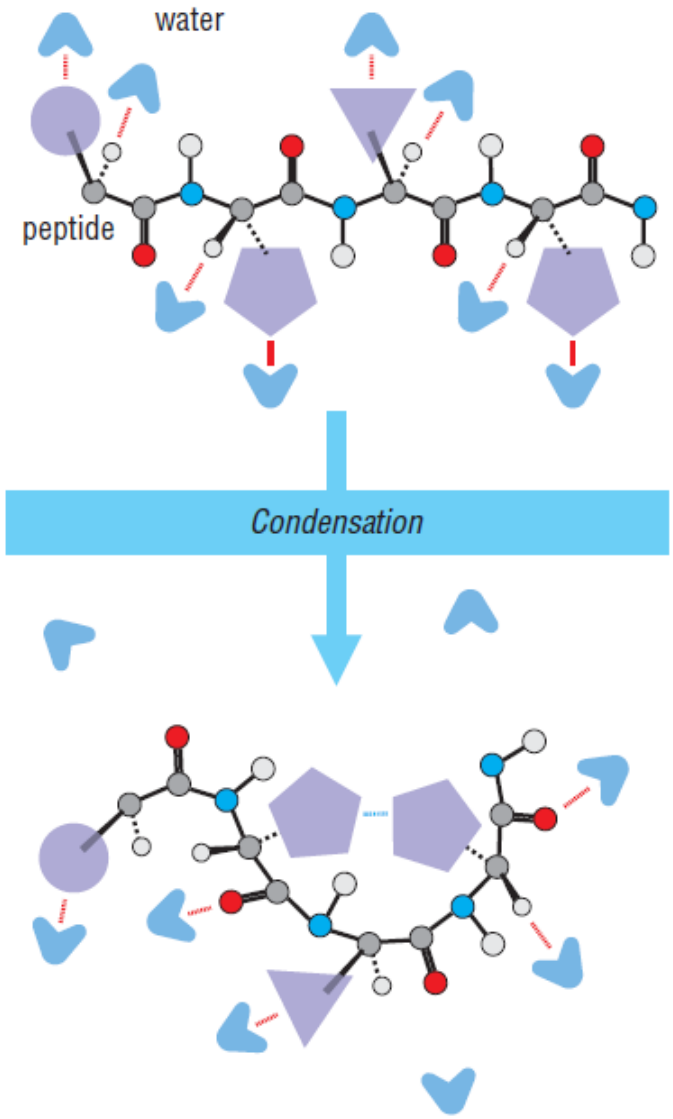
Bonds that stabilize folded proteins

Chemical Interactions that Stabilize Polypeptides

Interaction	Example	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)
Covalent bond	–Cα–C–	-	1.5 Å	356 kJ/mole (610 kJ/mole for a C=C bond)
Disulfide bond	–Cys–S–S–Cys–	-	2.2 Å	167 kJ/mole
Salt bridge		Donor (here N), and acceptor (here O) atoms <3.5 Å	2.8 Å	12.5–17 kJ/mole; may be as high as 30 kJ/mole for fully or partially buried salt bridges (see text), less if the salt bridge is external
Hydrogen bond		Donor (here N), and acceptor (here O) atoms <3.5 Å	3.0 Å	2–6 kJ/mole in water; 12.5–21 kJ/mole if either donor or acceptor is charged
Long-range electrostatic interaction		Depends on dielectric constant of medium. Screened by water. 1/r dependence	Variable	Depends on distance and environment. Can be very strong in nonpolar region but very weak in water
Van der Waals interaction		Short range. Falls off rapidly beyond 4 Å separation. $1/r^6$ dependence	3.5 Å	4 kJ/mole (4–17 in protein interior) depending on the size of the group (for comparison, the average thermal energy of molecules at room temperature is 2.5 kJ/mole)

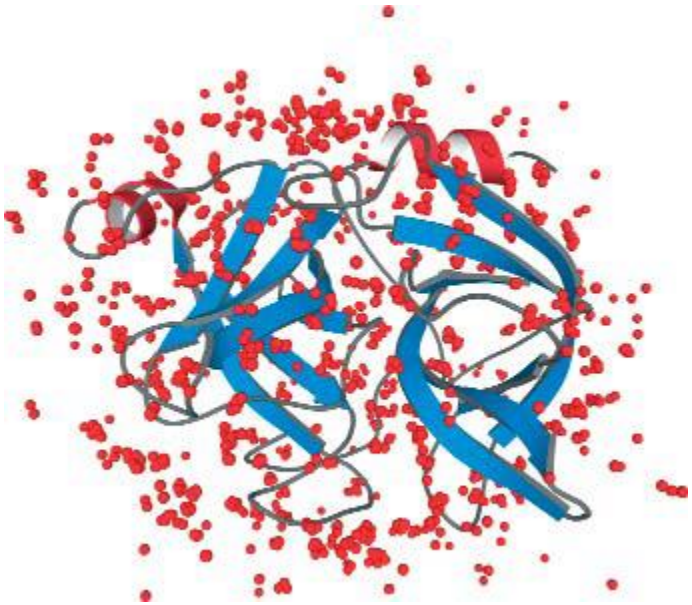
Hydrophobic effect is the driven force

- Generally, hydrophobic core and hydrophilic surface
  - How to bury hydrophilic main chain atoms
- Exposed hydrophobic residues on protein surface
  - Specific binding sites
- Membrane proteins
  - Folding in lipid bilayer (hydrophobic environment)
  - Hydrogen bonds are formed between the polar N–H and C=O groups of a peptide backbone first
    - Secondary structures form in early stages
  - Preformed secondary structure elements are condensed.



Bound water molecules on tertiary structure

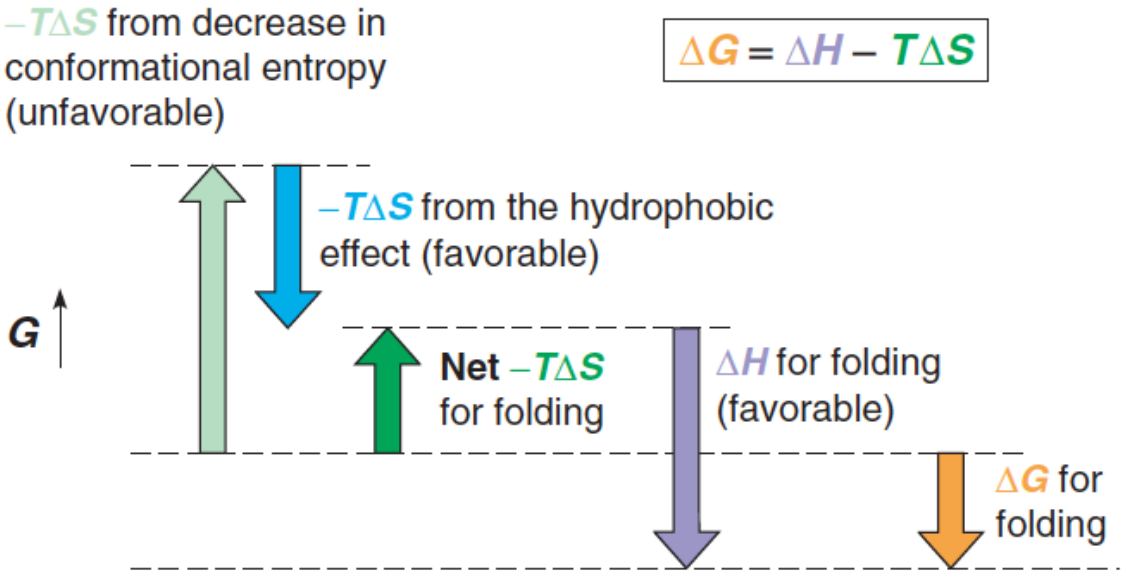
- Binding with both backbone and side-chain polar groups on the surface of a folded protein
- Fix position / unique position
  - often being involved in forming H-bond network for protein stability and function
  - A few are trapped inside the protein in internal cavities.
- Non-unique position
  - for the hydration of the protein surface



Porcine pancreatic elastase showing the first hydration shell surrounding the protein.

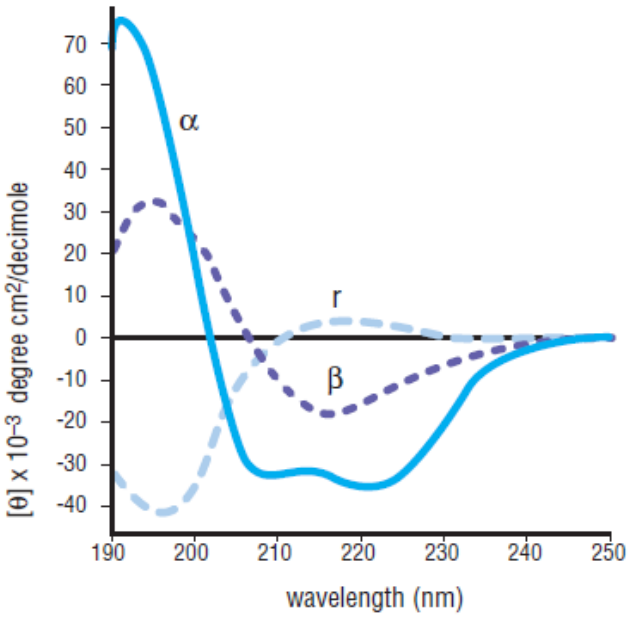
Thermodynamics of folding

- Protein stability refers to the thermodynamic stability.
- Enthalpy and entropy

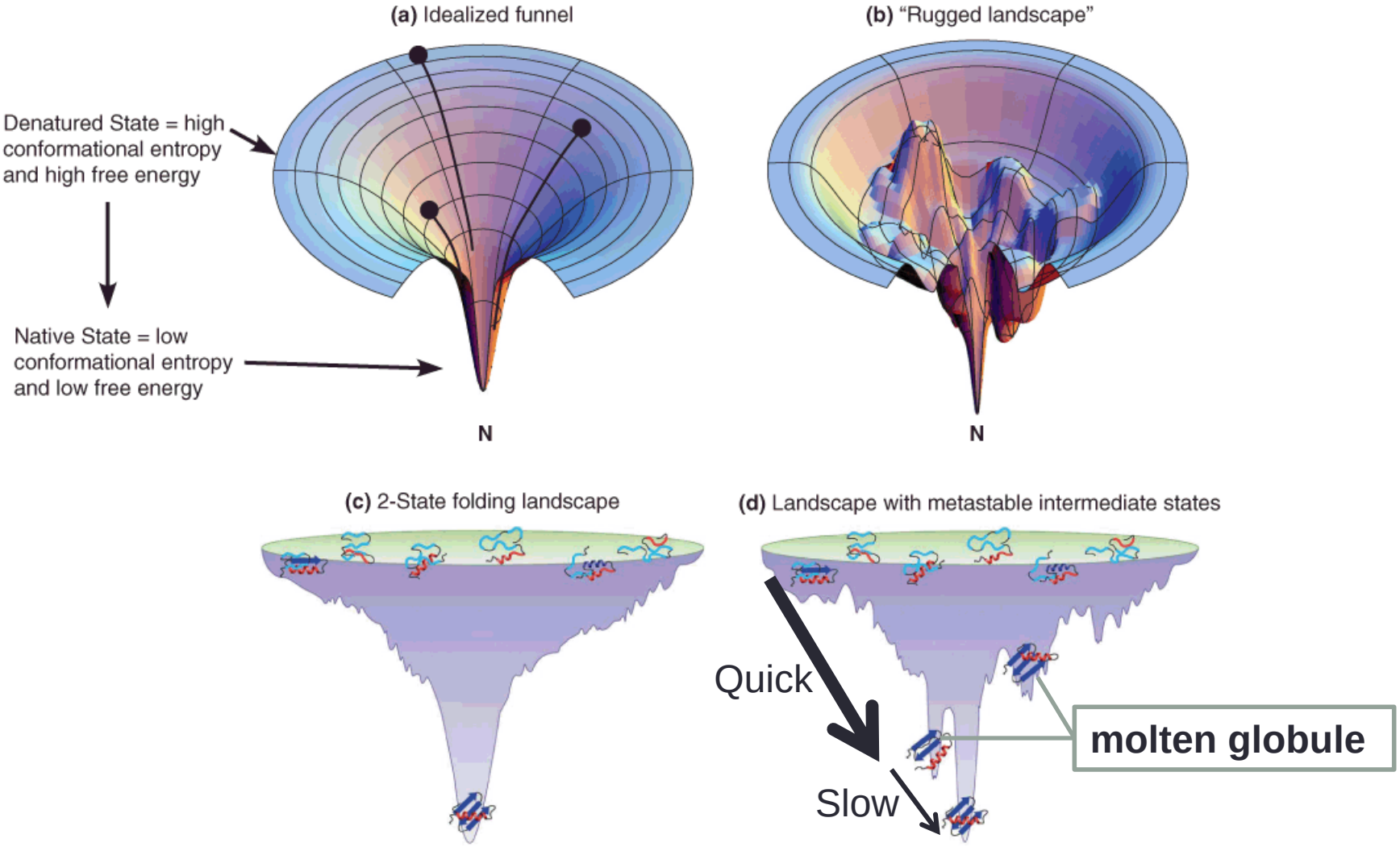


Thermodynamics and protein stability

- Globular proteins are marginally stable.
  - The average stability ($\Delta G = G_{folded} - G_{unfolded}$) of a folded small protein: ~5-10 kcal/mol (equals to the bonding energy of just one H-bond!)
    - Even thought, in aqueous solution and room temperature, the ratio of folded/unfolded protein molecules is ~10^7!
  - A single site mutation may destabilize the whole protein
  - Good for cellular control and activity
- Factors affecting protein stability
  - Temperature
    - Thermophilic proteins
  - pH
  - Chemical denaturants
    - stabilizing unfolded state
  - Other factors (ligand, etc.)
- Measurement of protein stability



Kinetics of folding

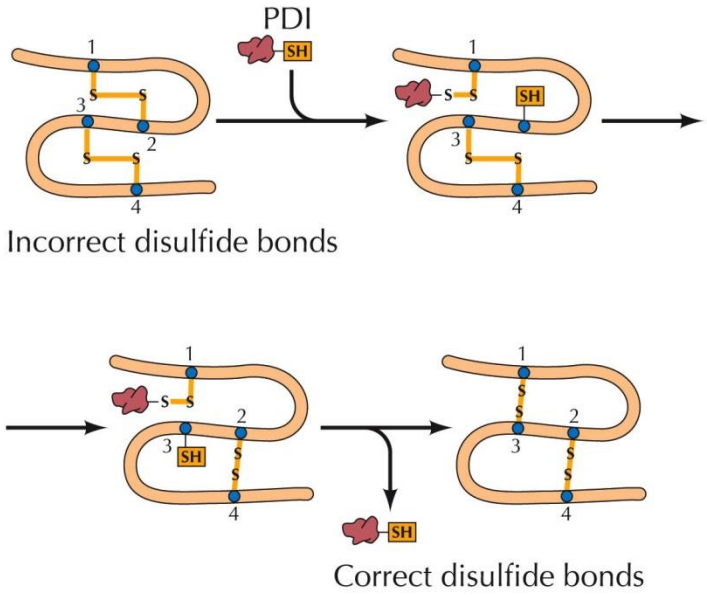


Kinetics and protein stability

- Kinetically stable state
  - The folded structure may not be the one with the lowest free energy but rather the most stable of those conformations that are kinetically accessible.
- Molten globule
  - High energy barrier prevent it from reaching the global energy minimum
  - Not a single structural entity but an ensemble of related structures that are rapidly interconverting
- The common obstacles to correct folding
  - Aggregation of the intermediates through exposed hydrophobic groups
  - Formation of incorrect disulfide bonds
  - Isomerization of proline residues (cis/trans)

Proper formation of disulfide bonds during folding

- Assisted by enzymes *in vivo*
- Disulfide bridge-forming enzymes (Dsb)
  - In periplasmic space of bacteria
- Protein disulfide isomerase (PDI)
  - In endoplasmic reticulum of eukaryotic cells

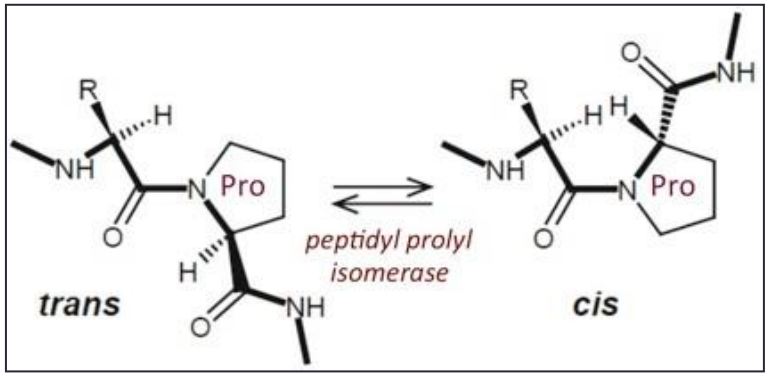


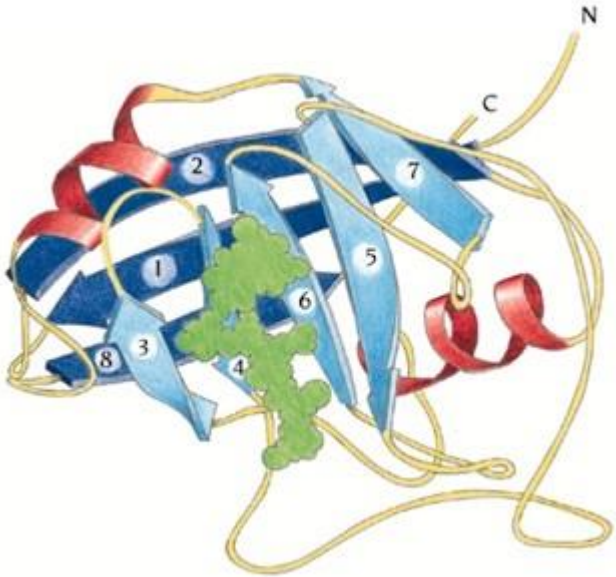
Thioredoxin domain

DsbA (PDB **1A2M**)

Catalyzing isomerization of proline residues

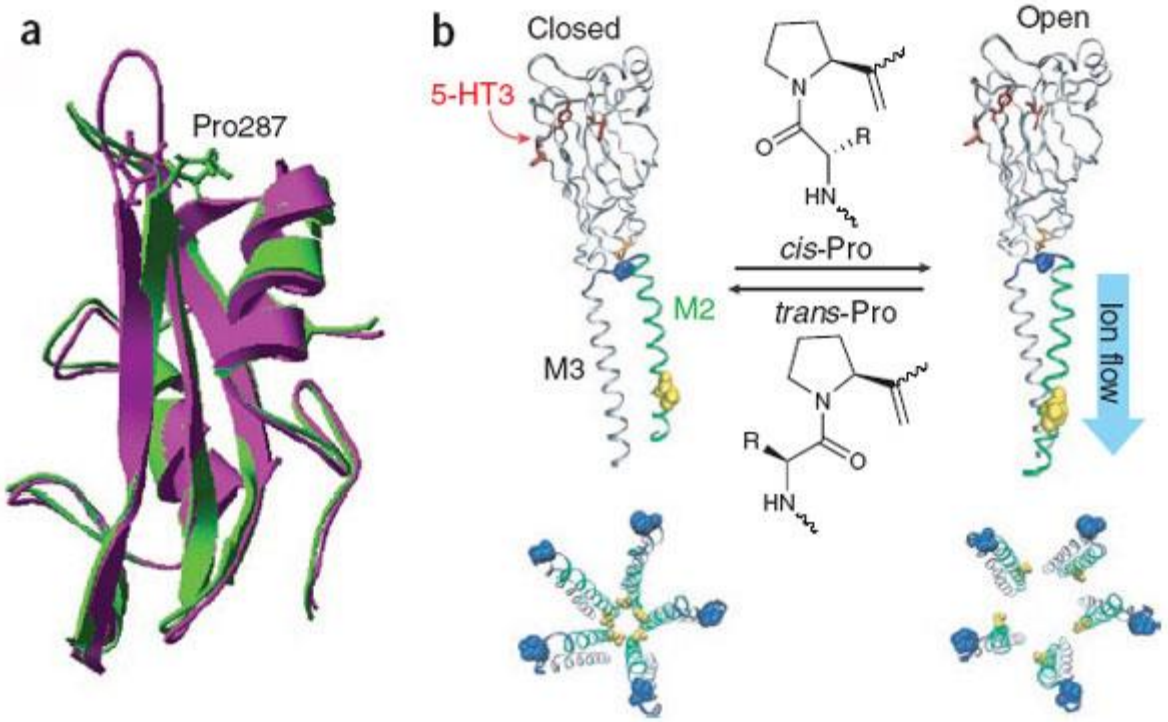
- Isomerization of proline residues can be a rate-limiting step in protein folding.
- In cell, the rate is increased by enzymes called peptidyl prolyl isomerases.





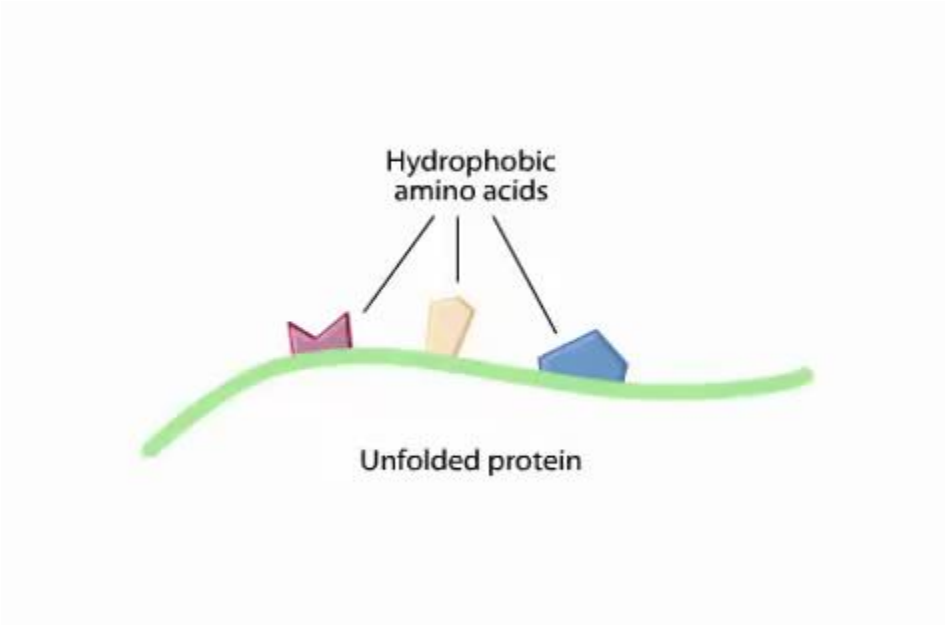
Cyclophilin in complex with a tetrapeptide

Isomerization of proline is also a regulatory mechanism for protein activity



Molecular chaperon

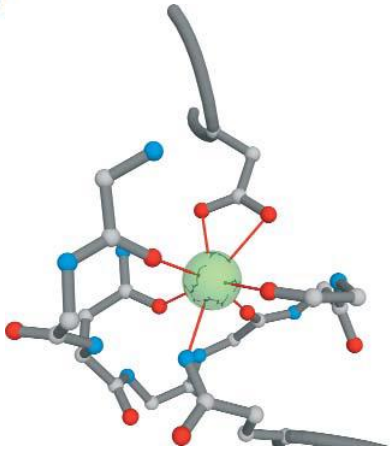
- Assisted folding *in vivo*
- Usually Heat Shock Proteins (HSP)
  - Prevent improper folding
  - Prevent and disrupt aggregation



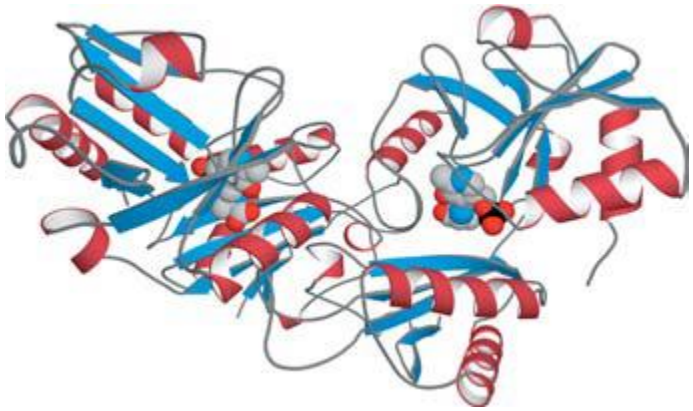
Covalent bonds can add stability to tertiary structure

- Disulfide bridge
  - The most common one
  - Required cysteines and oxidized conditions
  - Abundant in extracellular part of membrane protein and secretion protein
- Coordinate covalent bond
  - The coordination of a metal ion to several side chains
  - Binding strength depending on the nature of the metal ion and the protein ligands
  - Water molecules can also occur in the coordination
  - Common chelates including Ca^{2+} and Zn^{2+}
  - Application of His_6-tag for Ni-affinity purification
- Cofactor
  - the covalent binding of a dissociable organic or organometallic molecule at the active site
  - the formation of a covalent cross-link between amino-acid side chains
  - Apoprotein is less stable

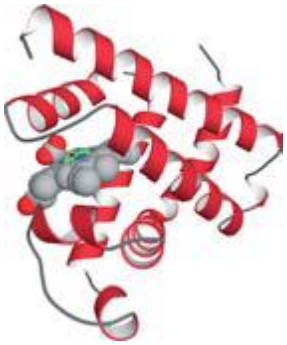
Bovine pancreatic trypsin inhibitor, BPTI (PDB **IBPI**)



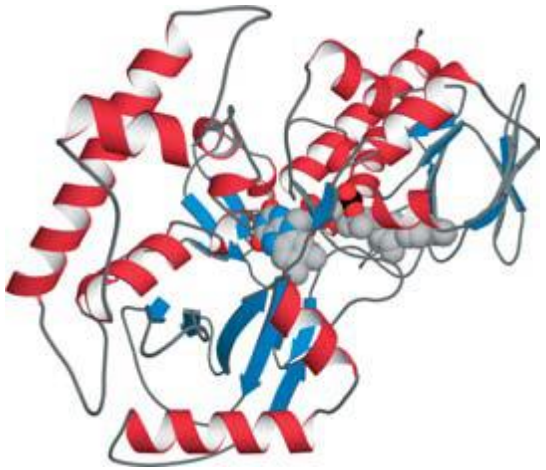
A calcium ion binding sites in the bacterial protein subtilisin (PDB **1SCA**)



DaAT is a dimeric protein with covalently bound pyridoxal phosphate bound in each subunit. (PDB **3DAA**)



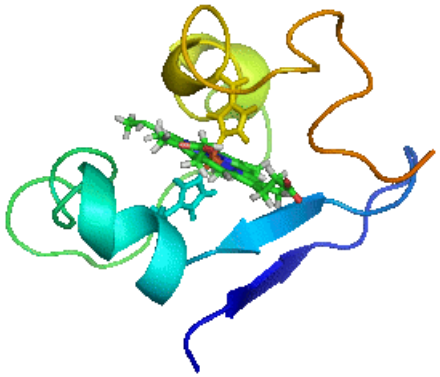
The heme group of myoglobin is attached to the protein by a coordinate covalent bond between a histidine side chain and the heme iron. (PDB **1A6K**)



The cofactor PQQ in polyamine oxidases is formed by the reaction of two side chains with each other, leaving it attached to the protein. (PDB **1B37**)

Stability is a “double-edge sword”

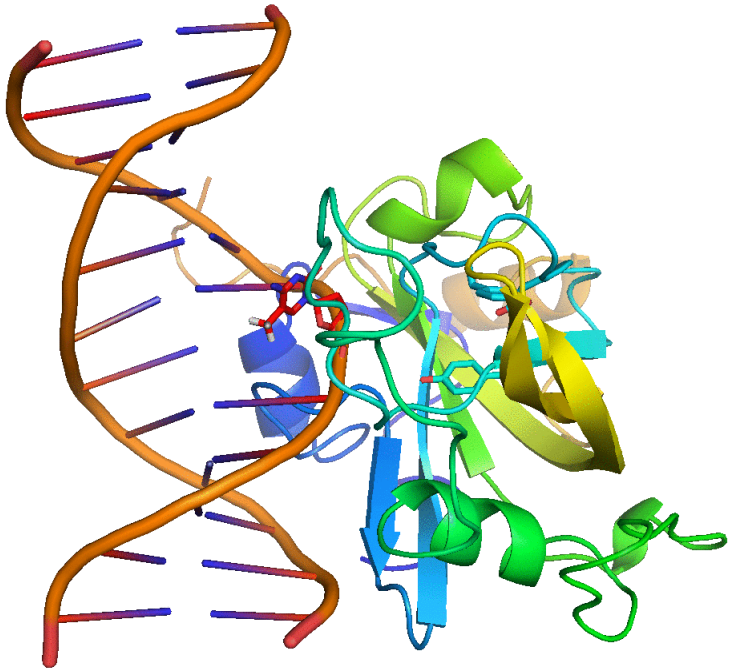
- Some thermophilic enzymes with high stability show low activities at lower temperature.
- Stabilizing the active site decrease may lead to loss of enzymatic activity.
- Flexibility (dynamics) is the important factor for protein function.



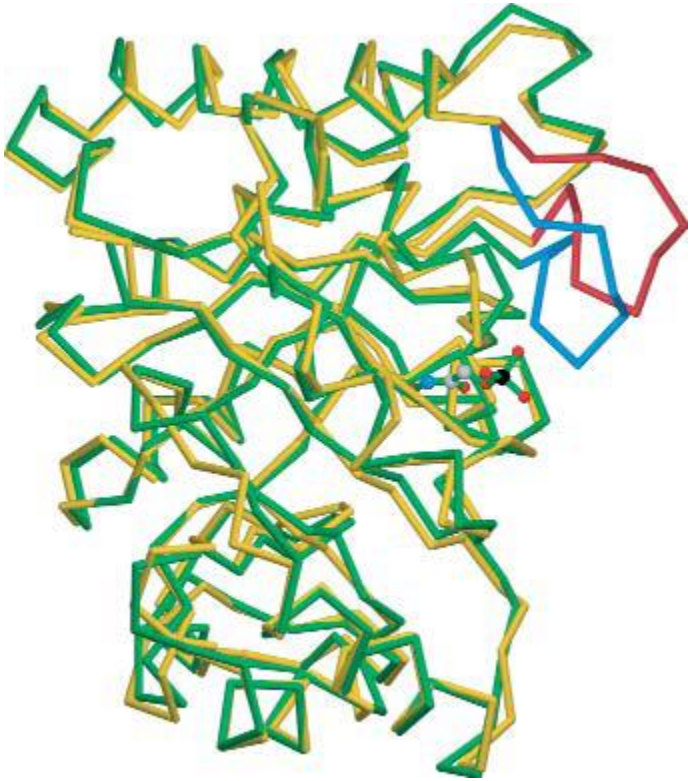
Protein Flexibility

- Proteins are flexible molecules
  - The folding forces are weak
  - Protein motion

Motion	Spatial displacement (Å)	Characteristic time (s)	Energy source
Fluctuations (e.g., atomic vibrations)	0.01 to1	10^{-15} to 10^{-11}	k_bT
Collective motions (A) fast, infrequent (e.g., Tyr, Phe ring flips) (B) slow (e.g., domain movement; hinge-bending)	0.01 to > 5	10^{-12} to 10^{-3}	k_bT
Triggered conformational changes	0.5 to > 10	10^{-9} to 10^{3}	Binding interactions



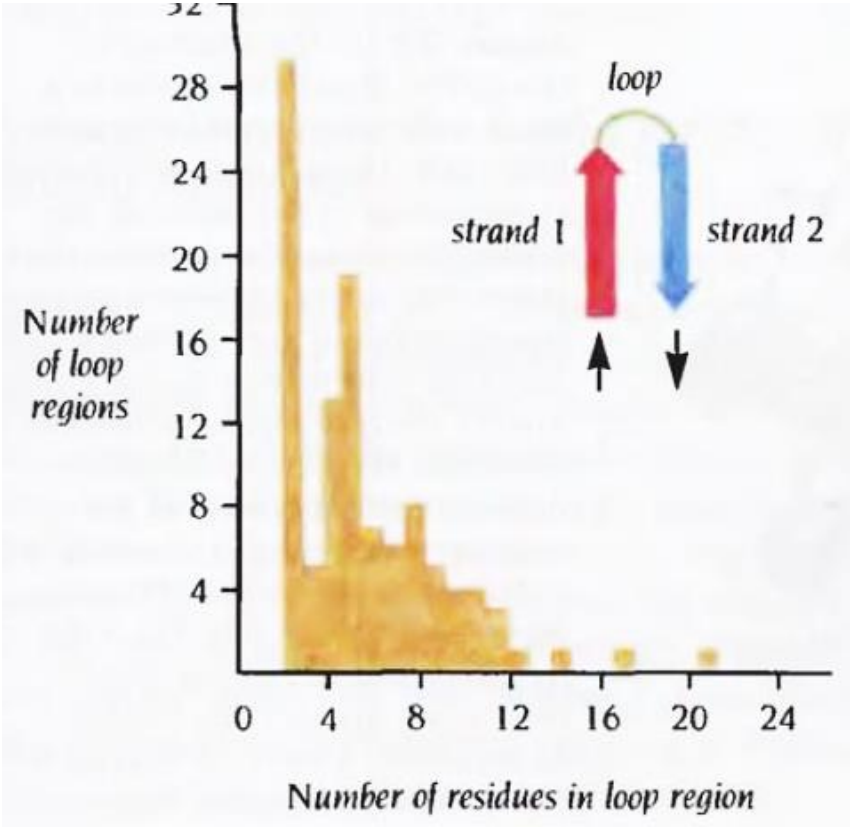
Conformational fluctuations in domain structure



Loop!

Triosephosphate isomerase. Binding of substrate or inhibitor to the active site of the enzyme triosephosphate isomerase induces a 10 Å rigid-body movement in an eight-residue loop.

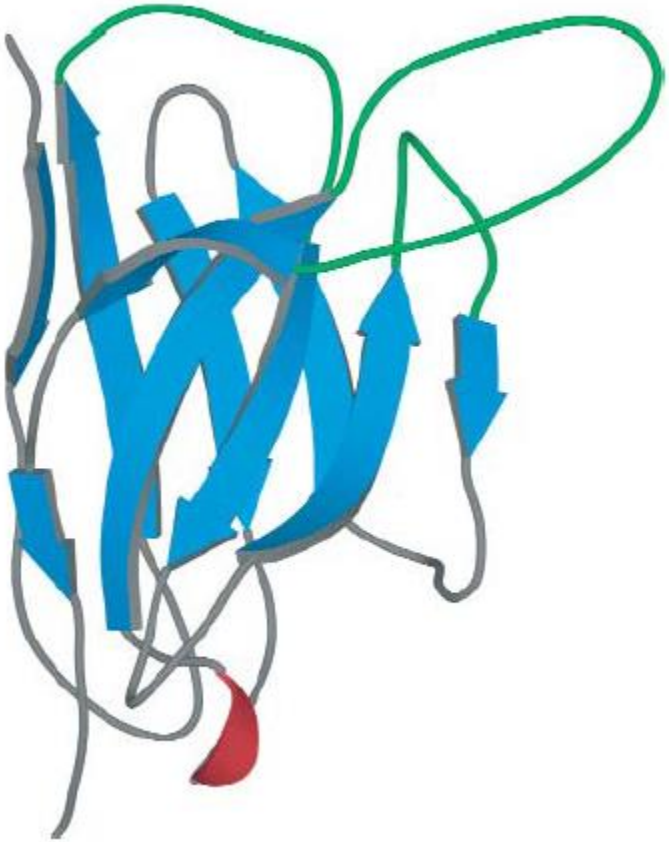
Variable loops in tertiary structures



- Secondary structure elements are more often connected by loops.
  - Variable length (a few to more than 20 residues)

- Loops are usually flexible and localized on protein surface.
  - Mainly charged or polar residues
  - Contribute little to overall folding

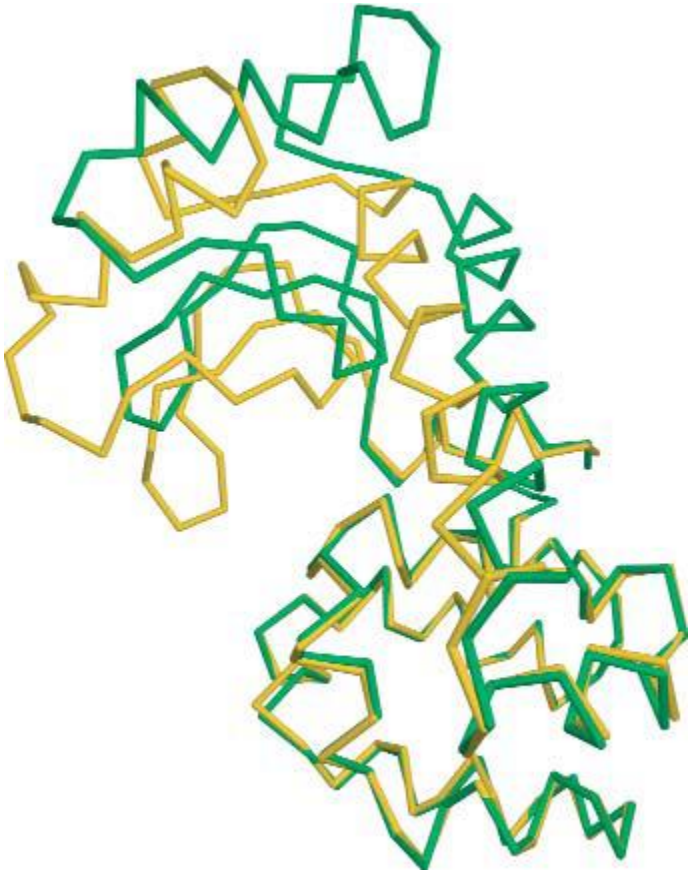
molecular dynamics simulation



Three-dimensional structure of the V domain of an immunoglobulin light chain (PDB **1OGP**). The hypervariable loops protruding from the ends of a sandwich formed by two antiparallel beta sheets form the antigen-binding site.

- Loops can tolerate mutations more readily than can the core of the protein.
  - Less conserved (used for secondary structure prediction)
  - Insertion and deletion
  - Evolution of new functions

Collective motion of covalently bonded atoms



The enzyme T4 lysozyme contains two domains connected by a hinge.
(PDB **1L96** and **1L97**)

Triggered conformational changes (Induced fit)

Open and closed forms of aspartate aminotransferase.
(PDB **1ARS** and **1ART**)