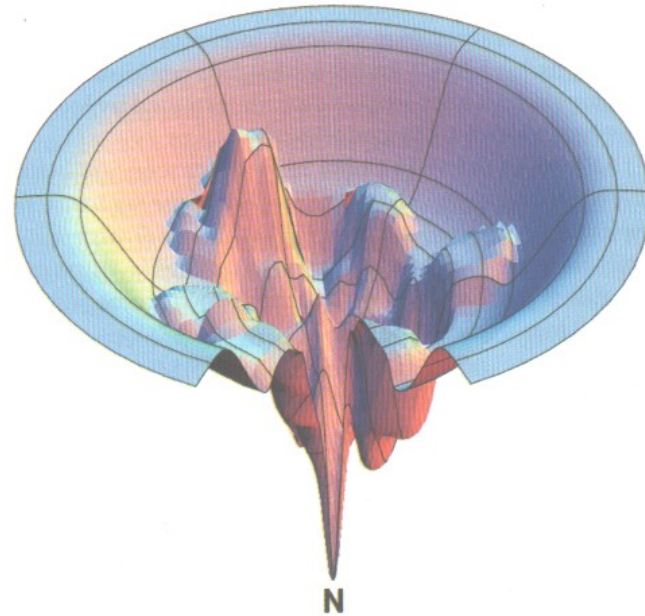


Chapter 9: Protein Folding, Dynamics and Structural Evolution

**Voet & Voet:
Pages 276-299**



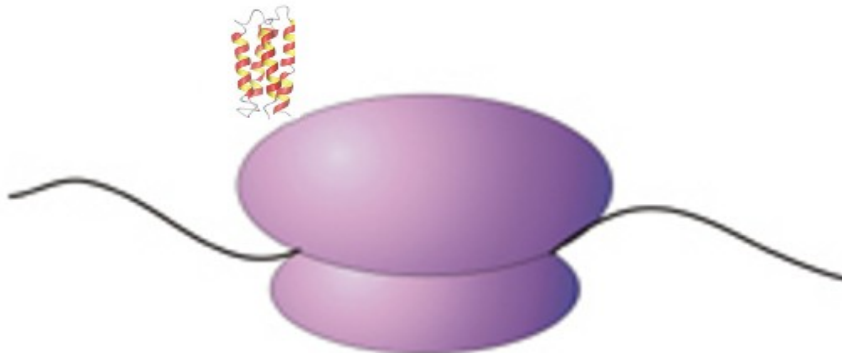
Protein Folding

Proteins are synthesized as linear chains of residues that rapidly and specifically assume their native, folded state

- Proteins fold as they are translated (*in vivo*)

(Most) proteins spontaneously fold into their native conformations under physiological conditions (*in vivo* or *in vitro*)

- self-assembly indicates no requirement for an external template to guide folding
- also implies that protein primary sequence contains information required to adopt 3D structure



**Present understanding of the
“folding” process is incomplete**

Spontaneous Refolding (RNase)

RNase A can be denatured (becomes inactive) and when renatured will maintain its biological function

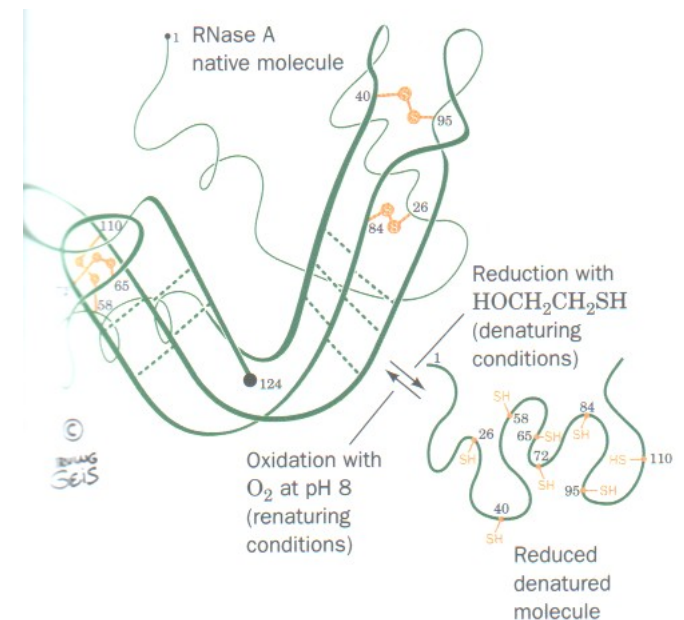
8 M urea: denatures proteins
2-mercaptoethanol reduces (cleaves) disulfide bridges

Expt I:

- a) Add 8 M urea and 2-mercaptoethanol**
- b) Remove urea by dialysis**
- c) Remove 2-mercaptoethanol by dialysis and allow oxidation of disulfides**

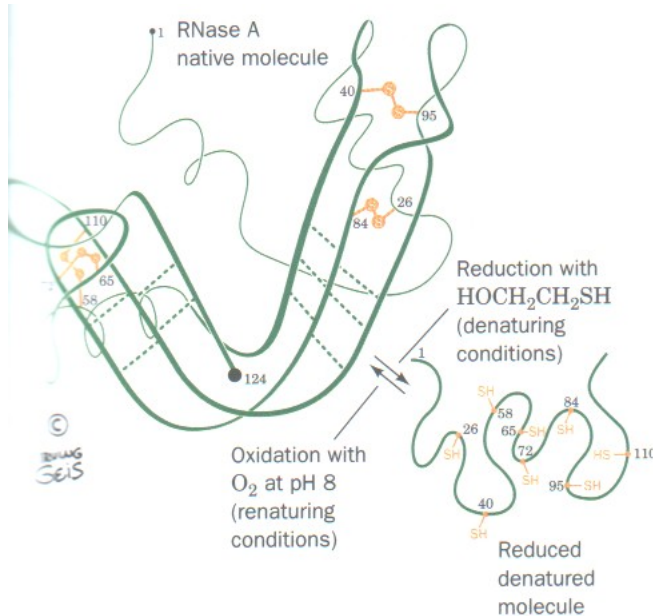
Result:

Protein is indistinguishable from the original. It retains full enzymatic activity and contain identical disulfides bridges.



RNase A
- 124 residues
- 4 disulfide bridges

Spontaneous Refolding (RNase)



RNase A
 - 124 residues
 - 4 disulfide bridges

Expt II:

- Add 8 M urea and 2-mercaptoethanol
- Remove 2-mercaptoethanol by dialysis and allow oxidation of disulfides
- Remove urea by dialysis

Result:

Only 1% of protein retains enzymatic activity!!

Note: Random chance of correctly forming all 4 disulfides is ~1% (same as the observed activity)

$$1/7 \times 1/5 \times 1/3 = 1/105$$

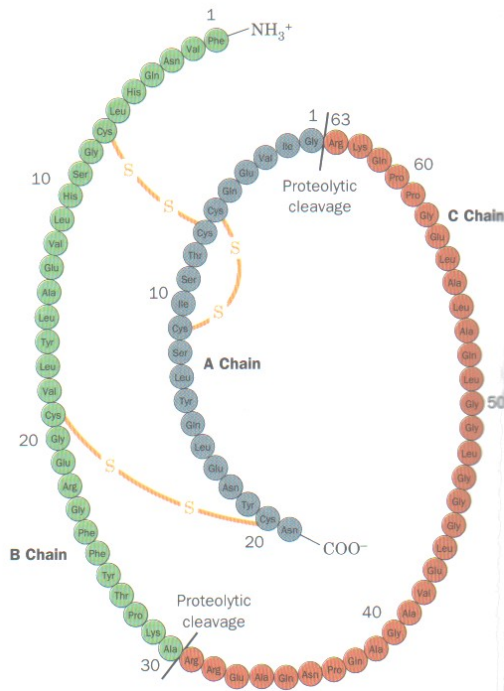
RNase A can be 'refolded' and will retain its structure (disulfides form correctly) and function (full enzymatic activity)

Refolding Insulin

(a post-translationally modified protein)

Insulin is translated as a single polypeptide precursor (proinsulin)

- Proinsulin (3 disulfides) folds and refolds similar to RNase A
- Insulin is post-translational modification of proinsulin arising from the proteolytic removal of a 33 residue segment (chain C, red)



Expt I: (same as slide 3)

Result: Only 25% of reduced, denatured insulin will spontaneously refold

Expt II: As above except insulin is chemical cross-linked prior to reduction & denaturation

Result: Up to 75% of insulin refolds

Primary sequence determines fold

Disulfides stabilize folded state but do not guide refolding

Note: Chain C mutation rates are 8 fold higher than in the A or B. Role in folding & activation less stringent than role in function?

Protein Folding (Determinants)

Primarily driven by (internal) residues contributing to the hydrophobic core

- **Systematic modification (with an octapeptide) of primary amines on the RNase A exterior does not affect folding**
- **Mutation rates for surface residues are greater than for internal residues**
- **Folding is perturbed by denaturants that primarily interact with hydrophobic core residues**

Protein Folding is driven by hydrophobic forces

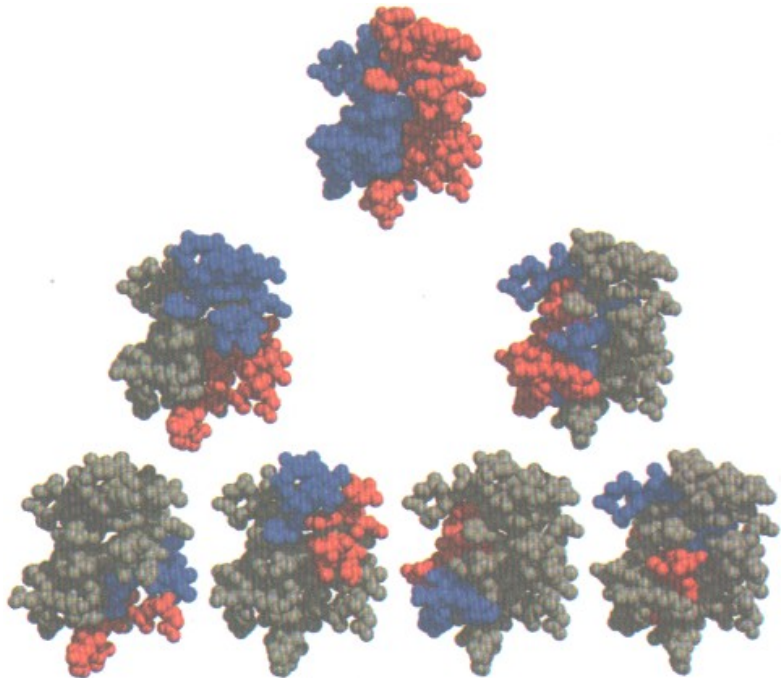
Recall: Helices and Sheets fill space efficiently

- **Secondary structures contribute virtually all residues of the hydrophobic core**
- **Helix and sheet content increases dramatically with the number of intrachain contacts in a folded state**

Protein Structure

Protein structure is hierarchical

- Local segments fold together and associate with adjacent segments until the entire structure is folded



- Domains can further be divided into subdomains and then smaller divisions
- Demonstrates hierarchical organization of the final folded state

**Consistent with knowledge of
hydrogen bond distributions**

Protein Adaptability

Packing density of the hydrophobic core is high

- Despite the extreme complementarity of the core, there is no detectable preference among side chain to side chain contacts

Suggests native fold determines packing but packing does not determine native fold

Large number of ways for the core to efficiently pack non-polar side chains

- Systematically introduced over 300 mutations into T4 lysozyme
 - Mutations to core generally resulted in small local rearrangements to accommodate size differences and structural constraints
- Conservative mutations had limited effect

Large hydrophobic core is significant, not the precise interactions between non-polar side chains.

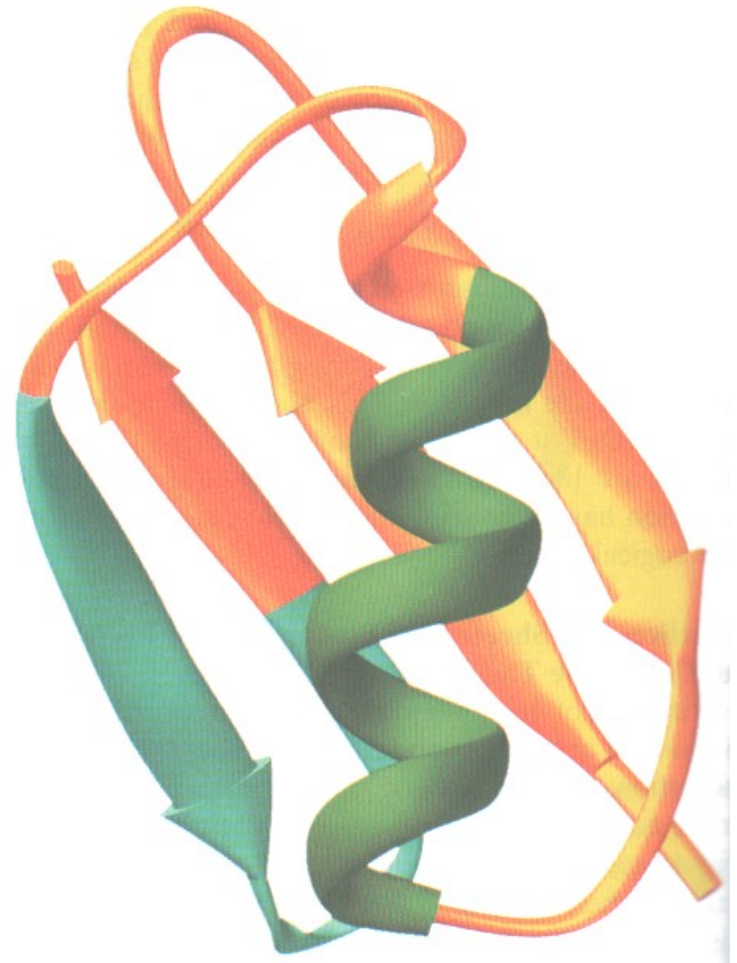
Context Dependence

Especially for small proteins:

Native fold may only be adopted when interacting with another protein or ligand

- GB-1 is partially unfolded when alone in solution as determined by NMR and fluorescence
- In the presence of IgG, GB-1 simultaneously binds IgG and adopts its native fold (right)

*Suggests that in some cases -
native fold can be determined by non-
local forces*



Folding Pathways

Levinthal Paradox

- Protein cannot randomly explore all possible conformations searching for the lowest energy state
 - A 100 residue protein with severe conformational limitations (2 state residues) would take > 40 million years to fold

Proteins must fold by some ordered pathway(s)



Earliest steps of protein folding have a sub-millisecond time scale while complete folding takes seconds

- requires specialized equipment and techniques to study

Techniques (rapid mixing)

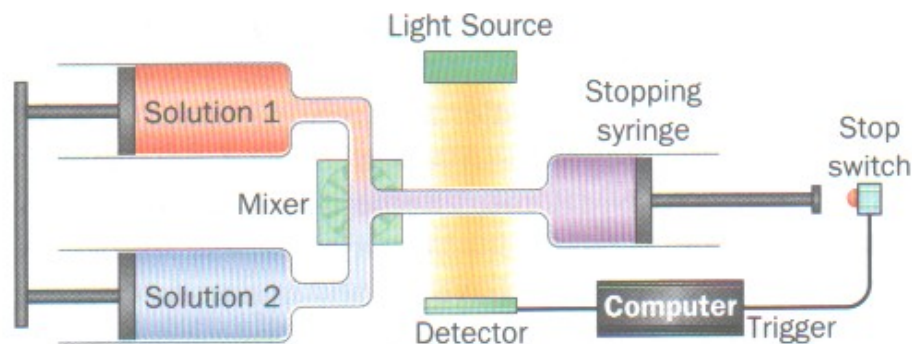
How do you observe such a rapid event (sub-millisecond)?

- Stopped-flow device

- Mix two solutions, stop and monitor earliest events (folding, catalysis, *etc.*)
- Use pH or denaturant concentration that unfold proteins and then rapidly dilute protein to start folding process

- Cold denatured proteins

- cold sensitive proteins may be utilized and folding initiated by a **temperature jump**
- laser-pulse can heat to 10-30 °C in 0.1 ms



Techniques (stable folding intermediates)

Are there 'stable' folding intermediates that can be studied?

- Yes ... at least, many proteins (in the presence of denaturants) can be stabilized in a partially folded state
 - Allows experimental detection methods that are slower to acquire data
- Techniques for studying folding intermediates
 - Circular Dichroism – detects secondary structure content
 - H/D Exchange NMR – detects accessible proton positions within structure
 - Fluorescence Resonance Energy Transfer – can provide some distance information

Work best when combined with existing structural information

Typically provide more structural information than rapid-mixing techniques

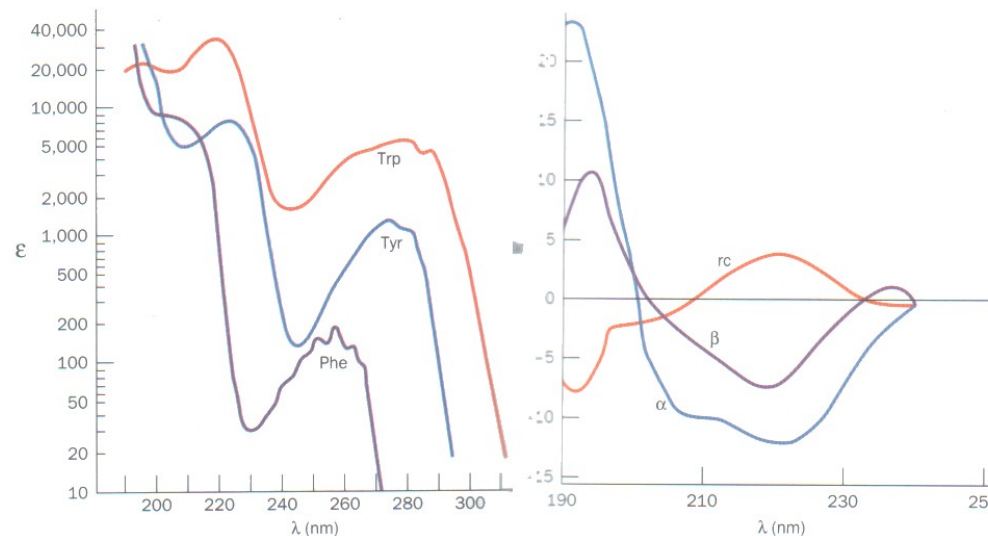
Techniques (detection)

Circular Dichroism (CD) Spectroscopy

Utilizes both right- and left-circularly polarized light

- Chiral molecules and structures absorb one hand more strongly than the other
- Dichroism refers to the difference in absorption
- Proteins absorb in ultraviolet range and have specific wavelengths that are characteristic of helices (209 or 222 nm) and sheets (215 nm)

UV absorption
spectra of aromatic
amino acids



Standard CD spectras
for helix, sheet and coil

- comparison of sample CD
spectra to standard spectra
allows estimation of helix
and sheet content (%)

Techniques

NMR – Pulsed H/D exchange

- Many protons on the surface of proteins will exchange with protons in solvent
 - Follow changes in NMR spectra as hydrogen (protein) is replaced or exchanged by deuterium (solvent)
 - Deuterium isotope is “invisible” to NMR
- Protons involved in hydrogen bonds do not (or very slowly) exchange
 - Allows one to follow formation of secondary structures as mainchain amide resonances will remain following deuterium exchange
 - Powerful when combined with knowledge of native folded state

In general, a method for producing folding intermediates (stopped-flow or cold denaturation) is paired with a spectroscopic technique

Burst Phase

Earliest protein folding event is a hydrophobic collapse (**Burst phase**)

(1) radius of gyration decreases dramatically

- indicates formation of hydrophobic core

(2) most secondary structure in a small protein form in the first few milliseconds

(3) ends with formation of “molten globule”

“**Molten Globule**”

(1) only slightly larger than native radius of gyration and most secondary structures formed

(2) sidechains are extensively disordered and structure is marginally stable

Following stages are much longer (up to seconds) and involve ordering of domains and sidechains

Landscape Folding Theories

Energy landscapes are 3D plots of the number of conformations (horizontal plane) and their relative free energy (vertical)

(A) Simplest theory involves well-defined intermediates (few proteins)

- experimental evidence suggests there are few (no?) well-defined intermediates

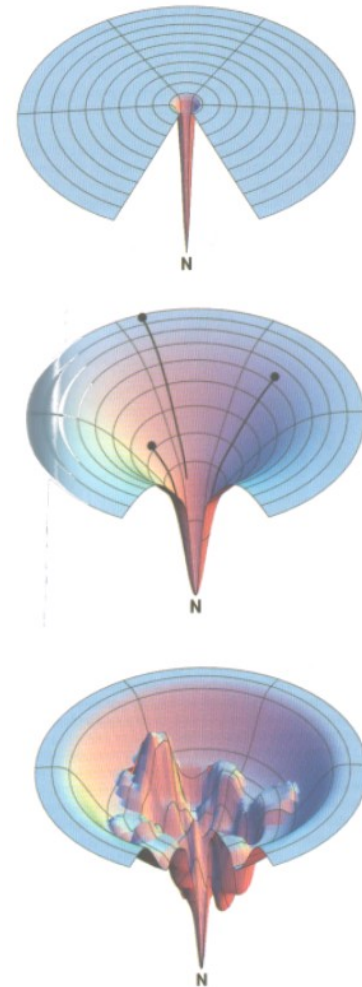
(B) Cooperative folding

- (1) secondary structures form and nucleate into a hydrophobic core, (2) cooperatively fold into native like structure and (3) small adjustments to adopt final structure
- experimental evidence suggests final phase of folding is relatively slow

(C) “Molten Globule” (most proteins)

- Cooperative folding into “molten globule” intermediate followed by series of conformational adjustments (slow) that reduce free energy and entropy until they reach their native state
- Agrees with experimental evidence

Key point is there are many paths that lead to a folded state



Hierarchical Folding

Final stages of folding (seconds timescale) are hierarchical

Experimental evidence final stages are hierarchical includes:

- **peptide fragments excised from a protein have a tendency to adopt native like folds**
- **intermediates in protein folding are consistent with a hierarchical process**
- **boundaries of helices are set by flanking sequences not tertiary structure**
- **secondary structure prediction can be successful (~75% accuracy for three state model (helix, sheet, coil))**

Sequence information specifying a particular fold is distributed throughout the polypeptide chain and is highly overdetermined

Folding Mutant

Mutants of the “tail spike protein” affect folding but not function

- “tail spike” efficiently renatures and forms a homotrimer with $T_m = 88\text{ }^{\circ}\text{C}$
- mutants of “tail spike” prevent folding at $39\text{ }^{\circ}\text{C}$ without affecting the stability of the folded state
 - apparently, these mutants affect the stability of a folding intermediate
 - same mutants refold efficiently at $30\text{ }^{\circ}\text{C}$

Primary sequence of proteins dictates its native structure and how it folds into that native conformation

- Corroborated by the large number of polar residues that occupy helix capping positions but do not make helix cap interactions in the final folded state
 - presumably, these residues do make helix capping interactions during the folding process

BPTI Folding

BPTI is an example of a small monomeric protein that folds via a relatively ordered pathway

- the 3 disulfides of the folded state involve residues 30-51, 5-55 and 14-38
- time course of BPTI folding shows the formation of disulfide bonds occurs in an ordered fashion

Detectable intermediates in BPTI folding

- the single disulfide intermediate is a mixture of 30-51 (correct) and 5-30 (incorrect)
- the two disulfide intermediates arise solely from 30-51 and form 30-51, 14-38 and 30-51, 5-55
- fully folded BPTI arises solely from 30-51, 5-55

BPTI Folding

