

7.88 Lecture Notes - 11
7.24/7.88J/5.48J
The Protein Folding Problem

- **Protein Folding Intermediates and the Failure of Protein Folding**
- **Mutants affecting intermediates**
- **Protein Misfolding and Inclusion Body Formation**

A. Protein Folding Intermediates and the Failure of Protein Folding

We discussed last week the refolding of cytochrome c, whose in vitro refolding pathway is robust.

Considering proteins that successfully refold:

Discovered that a key intermediate had a somewhat unexpected structure; N- and C-terminal helices docked, but interior fold not folded, or not-native-like.

But if you are investigating the properties of cardiac muscle proteins, cytochrome c doesn't help you; need to study the refolding of your protein!!

What you might discover- is that you would clone gene for cardiac actin, purify cardiac actin

Denature in urea, dialyze to refold, fail to obtain soluble actin:

Well maybe purification regime does some subtle damage; Clone coding sequence into expression vector in E.coli; chain synthesized at high levels - but lo? fail to obtain soluble protein with properties of actin

For many purified proteins, attempts to refold from the fully denatured state fails:

- No detectable renaturation - Of course generally these not reported; only successful cases reported. Very difficult, essentially impossible from scientific literature to determine if sequence sufficient to drive in vitro refolding
 - actin
 - tubulin
 - T4 DNA polymerase
 - Rubisco –ribulose biphosphate decarboxylase
 - Collagen
- Yield low and highly variable (diluting from denaturant to buffer)
 - Beta galactosidase
 - Tryptophanase
 - Multi-subunit proteins: beta-galactosidase, lactic dehydrogenase

B. How and Why do In Vitro Folding Reactions Fail??

This was examined systematically by a French group Michel Goldberg and Charles Ghelis and Jeanine Yon (authors of the best textbook on the subject)

One of their early observations was that

- If remove denaturant by dialysis (slow) refolding fails:
- If remove denaturant by dilution (fast) higher refolding yield.

Initial insight from this came from a two stage experiments - Suppose we do dilution out of urea as two stage reaction:

Here is data in the refolding of beta galactosidase: urea:

Dilute from 8M to intermediate concentration;
let incubate 1 hour, then
dilute further

Get sharp trough in % reconstitution minimum at 4M

Longer it sits, deeper trough, wider trough
F (time at intermediate levels:)
Also f(concentration):

Higher concentration in intermediate urea, broader the trough: Thus here is similar data for enzyme tryptophanase: at 100/200/500 ug/ml.

As concentration during refolding is increased, yield is decreased.

Anfinsen originally observed this with ribonuclease, but was unsure as to interpretation; If you remember ribonuclease, plot of yield of refolded protein vs. concentration roughly like so:

Explanation is

- folding intermediate populated at intermediate denaturant population
- associate irreversibly into inactive aggregated state

of folding :

{I} >>> {I}{I} >>>>

Model: Off pathways interactions specific, not non specific

Goldberg experiment:

So: practically to optimize refolding yields for a monomeric protein;

- rapid dilution from denaturant'
- refold at lowest concentration possible

C. Multimeric Proteins and The Concentration Dilemma

Now for a monomeric protein, aggregation requires collisions, folding doesn't:

In attempting to avoid the widespread problem of aggregation of nonnative states, investigators of monomeric proteins simply lower the concentration to suppress second order reactions.

When working with single chain proteins in a laboratory environment this is a feasible strategy. However the:

- Native states of a vast variety of proteins are as multimers or oligomers
- In this case it is not possible to refold the native state by keeping the solution sufficiently dilute to avoid chain-chain interactions.
- Furthermore, in the industrial situation, dilution is not always practical or commercially feasible, given the need for subsequent concentration.
- Refolding versus concentration for a multimeric protein:

One practical methodology: **Refold by drop-wise dilution**; thus at any instant partially folded intermediate concentrations kept low;

Properties of Intermediates Cannot Be Predicted from Structure of Native State!!

How get at these properties:

D. Mutants affecting intermediates

Genetic approach: Isolation of mutations which specifically affect protein folding intermediates;

A very important general class of mutants for identifying functions of genes

Temperature sensitive mutations:

two classes:

TL – mutant protein is thermolabile in native state; replacements lower T_m .

TSF – Native state of mutant protein is not thermolabile! Protein folding intermediates are thermolabile!

Goldenberg, Smith, and King. "Genetic Analysis of the Folding Pathway for the tailspike of Phage P22." *PNAS* 90 (1983): 7060-7064.

P22 Tailspike protein:

- Example of Parallel beta coil fold; simplest beta sheet structure
- But tailspike is a trimer; not in equilibrium with monomer:
- Distinctive feature, very high stability to denaturants
- $T_m = 88^\circ\text{C}$
- Stable in SDS under conditions in which all other proteins denatured (SDS GEL WITH AND WITHOUT HEATING)
- Native state completely resistant to Proteases

Infect cells: divide culture and incubate at two temperatures

- 40°C : Chains synthesized but no native tailspike
- 25°C : Chains synthesized, fold into native trimer
- Label: take samples; lyse: SDS gel

Protocol: Purify mutant proteins at permissive temperature: measure properties at high temp: wild type:

Can actually do this in vivo also: Proteins stable at 37 once folded at 30°C ;

(Sturtevant et al)

So these residues identified as sites of tsf mutations do not stabilize native state;

They stabilize critical folding intermediates, or kinetically control folding pathway, to lead to native state:

But, since proteins fold up at low temperature, not essential for folding; auxiliary folding information

In Coiled coil; heptad repeat essential, e and g contribute, not essential

First note that in vivo folding is not that fast; $t_{1/2}$ of 5 minutes at 30°C ; surprisingly slow;

Now Measure yield as a function of temperature of folding:

Yield very sensitive to temperature!

E. What happens to chains that fail to fold

Form high molecular weight aggregate; Scatter light, Sediment in low speed pellet, insoluble, in active >>> inclusion body:

Is there a pathway for formation of this state?

To answer, we took advantage of the long-lived folding intermediates identifiable in the P22 tailspike pathway;

Inclusion bodies form from partially folded intermediates.

Refold in vitro, under conditions where refolding fails;

Trap intermediates on ice, fractionate in cold by native gel electrophoresis: Here is native gel electrophoresis:

Identify In vivo intermediates: Monomer: Protrimer > Native trimer:

Chain initially rapidly electrophoresing, then detect slower species then much slower, then native

Is off pathway reaction non specific aggregation:

How prove:

Speed, Wang, and King. "Specific Aggregation of Partially Folded Polypeptide Chains." *Nature Biotechnology* 14 (1996): 1283-1287.

Mixing Experiment: Unrelated protein is phage coat protein: monomer of 55,000 mol wt Folds then polyermizes into icosahedral shell of virus; but that requires scaffolding:

Folding intermediates go off pathway and aggregate just like soluble proteins:

Non-specific predicts all mixed multimers: specific, even though refolding aggregating in same tube, each should self-associate:

Denature two proteins: P22 coat and P22 tailspike: dilute under conditions in which both chains aggregate:

Fractionate on native gel: discriminate between species by Western blotting with monoclonal antibodies:

predict mixed species: find each chain only associates with itself: specific polymerization pathway, all chains in inclusion body.

Since native state not formed, inclusion body can't be formed from it; in fact precursor is partially folded monomer.

It is these species that are recognized by molecular chaperones, which we will discuss

Use monoclonal antibodies to identify chain composition of monomers:

What is pathway of this reaction; monomer to 1000-mer. Unlikely to be concerted reaction:

P22 Tailspike protein: Is this specific pathway, that is it encoded in amino acid sequence??