The mechanism(s) of protein folding

U ↔ Black box ↔ N

What is meant by mechanism
Questions:

What is the chain of events when a protein folds

Is there a specified sequence of events or are there parallel pathways

In either case what is the nature of intermediates

How much structure is there in the unfolded state and does it define a mechanism.
To answer those questions, we would like to have a specific marker, both for the backbone and the side chain for each amino acid as the protein folds from the unfolded state.

**Backbone:** Hydrogen/deuterium exchange, far UV CD

**Side chains:** Fluorine labeled amino acids

**Unfolded state:** Fluorescence correlation spectroscopy
Side chain measurements

Proteins containing fluorine labeled amino acids have been used for many years to examine ligand-receptor interactions, the role of specific amino acids in protein function, structure mobility.

More recently we have used fluorine labeled amino acids to examine the mechanism(s) of protein folding and unfolding using $^{19}$F-NMR.
Advantages of using $^{19}$F NMR

The fluorine nucleus is only slightly larger than the hydrogen nucleus. Large structural perturbations would not be expected.

NMR spectra are simple and rate constants can be determined by stopped flow methods or line broadening.

Fluorine is exquisitely sensitive to the microenvironment.

There is a large chemical shift range.

High molecular weight proteins can be examined.

A number of $^{19}$F-labeled amino acids are available.

A fluorine cryoprobe is available.
Disadvantages of using $^{19}$F NMR

The fluorine nucleus has a strong dipole moment that could affect the structure or stability

There is no “good” theory for relating the chemical shift to structural changes, especially for aromatic amino acids

Assignments must be made using site-directed mutagenesis (except for phenylalanine)

The fastest time measurement for refolding using stopped flow NMR is about 1 second
What stabilization of side chains means

We define stabilization as the appearance of NMR peaks with the the same chemical shifts as in the native protein

Meaning that the side chain senses the same stable microenvironment as in the native protein

It is not the same as burial of the side chain

It is not necessarily the same as side chain packing
Side chain stabilization

Scenario 1

Stabilization might occur in nucleating regions assuming that there are cluster forming regions that serve as nucleation centers (i.e., hydrophobic clusters)
Side chain stabilization

Scenario 2

Stabilization occurs during an initial collapse assuming there is an initial collapse to, for example, a molten globule.
Side chain stabilization

Scenario 3

Stabilization occurs with formation of regions of secondary structure

As secondary structure (i.e., helices, β-strands) form the side chains become stabilized. If these structures form sequentially, would expect non-cooperative behavior. The same behavior might be expected for multiple pathways of folding.
Side chain stabilization

Scenario 4

Stabilization occurs at the last step of folding or close to the last step. Would expect stabilization to be highly cooperative. Might also expect early destabilization on unfolding.
The *E. coli* dihydrofolate reductase

MWt 18,000

5 tryptophan, 6 phenylalanine residues

4 phases on refolding as measured by fluorescence
ΔF on refolding DHFR in the presence of substrate
What’s needed:
Fluorine labeled protein
Reversible folding/unfolding system
Protein concentrations of about 100 µM (final)
Reasonably slow folding >30 sec

What’s available
Fluorine cryoprobe
NMR Stopped flow apparatus
Refolding into buffer containing NADP

Signal averaging 41 injections with a final protein conc’n of 0.61 mM after dilution of 5.5 M urea to 2.75 M.

Phenylalanines
$^{19}$F-phenylalanine spectra of \textit{E. coli} DHFR
Residues colored red are stabilized first
DHFR Conclusions

Major changes in 2° structure occur early during folding

Native peaks appear (side chains stabilize) slowly and cooperatively although in two phases

No stable misfolded intermediate structures

On unfolding, native peaks disappear early but denatured peaks appear slowly
PapD

MWt 25,000

Two domains, each with 1 tryptophan

6 phenylalanines

A chaperone for the formation of type P pili
p-fluoro phenylalanine
Assignments of p-F-Phe resonances
Expansion of the genetic code

Fully labeled
PapD  $p^{-19}\text{F-Phe}$ chemical shifts as a function of urea concentration
Refolding of PapD $^{19}$F-Phe-labeled 4.5M to 2.25M urea jump
PapD Conclusions

The C-terminal domain forms early as an intermediate.

Folding of the N-terminal region is slow probably due to trans-cis proline isomerization.

The domains fold separately but interact to give the final native structure.

On unfolding, native peaks disappear early
The Intestinal Fatty Acid Binding Protein

15 KDa protein, 10 antiparallel β sheets, 131 amino acids, no proline or cysteine
Two tryptophans (W6 and W82), Helical region moves to allow ligand binding
Trp82 is responsible for most of the fluorescence change on unfolding/refolding
-x- is H/D exchange of Phe68
-◊- is loss of HSQC intensity
-□- is loss of Phe68 intensity
-o- is loss of fluorescence
Unfolding with urea
  loss of backbone hydrogen bonds
  destabilization of the backbone
  loss of hydrophobic clusters

Refolding
  formation of hydrophobic clusters
  stabilization of the backbone
  hydrogen bond stabilization
Murine adenosine deaminase, a 40 Kd protein
Assignment of $6^{-19}$F-Trp Resonances
Urea-induced unfolding of mADA: ~ 2.5hr time course

100 μM protein unfolding in 8 M urea
20 mM Tris-HCl, 2 mM DTT, pH 7.4, 20°C
Slow Refolding Kinetics of holo mADA

Urea jump from 7 to 0.7 M
Final ADA conc. 140 µM
ADA Conclusions

On unfolding, native peaks disappear slowly, probably because tightly bound Zn dissociates slowly.

On refolding, there are intermediates. Refolding is very slow possibly because the protein is trapped in a misfolded form. Proper binding of Zn could be a problem.
Some general comments

There are very few ways to measure the kinetics of Stabilization of specific side chains

$^{19}$F-NMR can be used to measure the kinetics of side chain stabilization as well as packing and the appearance or loss of denatured peaks during folding or unfolding

$^{19}$F-NMR can identify the nature of intermediates
For those proteins studied here:

There is rapid loss of some intensity of denatured peaks presumably forming a collapsed state. In this state the side chains are not in a stable microenvironment. Scenarios 1 and 2 are not applicable.

Side chains are not stabilized concurrent with secondary structure formation. Scenario 3 is not applicable.

Side chain stabilization occurs at the last stages of folding and shows a high degree of cooperativity (scenario 4). This is where the cooperativity of folding/unfolding is found.
Caveats

Very little information about the early steps

So far, have only used aromatic amino acids

Is it just proline isomerization
Is it just proline isomerization

Very possible, but so what. Most proteins contain proline

Even non-proline containing proteins (i.e., IFABP) can fold slowly

How does ~10% cis-proline (in the denatured protein) control overall stability on conversion to trans form

Intermediate forms are also seen at equilibrium so it is not just a kinetic effect

Does not appear to be important in unfolding
Is it just proline isomerization (continued)

Slow folding will allow time-dependent distance measurements

$^{19}$F-proline shows different chemical shifts for the cis and trans forms

Other mechanisms for slow folding: ADA folding appears too slow for proline isomerization (incorporation of Zn?)

Would not be asking this question if not for ability to measure the kinetics of side chain stabilization.

May help define role for proline isomerases.
An approach to fast steps in folding

Probing protein dynamics using fluorescence methods
Time scales of protein motions
-in order of fast to slow time regimes

Side chain rotation
Backbone flexing
Loop or domain movements

Time scales in protein folding

Motions in the unfolded state
Collapse to intermediates
Secondary/tertiary structure formation
Proline isomerization
Side chain stabilization
The Intestinal Fatty Acid Binding Protein

15 KDa protein, 10 antiparallel β sheets,
131 amino acids, no proline or cysteine
Two tryptophans (W6 and W82),
Helical region moves to allow ligand binding
Trp82 is responsible for most of the
fluorescence change on unfolding/refolding
FCS is one of the many different modes of high-resolution spatial and temporal analysis. In contrast to other fluorescence techniques, the parameter of primary interest is the spontaneous intensity fluctuations caused by the minute deviations of the small system from thermal equilibrium. In general, all physical parameters that give rise to fluctuations in the fluorescence signal are accessible by FCS. It is, for example, rather straightforward to determine diffusion coefficients or characteristic rate constants of inter- or intramolecular reactions of fluorescently labeled biomolecules at nanomolar concentrations.

Magde, Elson and Webb (1972) Phys Rev Lett. 29, 705-708 Announcement
Elson and Magde (1974) Biopolymers 13, 1-28 Theory
Magde, Elson and Webb (1974) Biopolymers 13, 29-61 Experimental application
Diffusion kinetics

Chemical kinetics

Fluorescence-diffusion + chemistry

Fluorescence-diffusion only
Advantages of FRET-FCS over single molecule experiments

No need to attach a tether to the protein

No need to attach the molecule to a solid surface

Faster time regimes (µsec vs. sec)

Nanomolar concentrations

Disadvantages
Still have to label the protein with fluoroprobes and there has to be a fluorescence change associated with any isomerization event.

Any isomerization event has to be faster than the diffusion time.
There are no cysteine residues in native intestinal fatty acid binding protein
One cysteine labeled
Unfolded IFABP
\( \tau_D = 225 \ \mu\text{sec} \)
Both cysteine residues labeled
Unfolded IFABP

$\tau_D = 225 \mu\text{sec}$

$\tau_R = 1.6 \mu\text{sec}$
Questions/Issues

How fast can a protein fold.
What is the role of diffusion-loop formation/hairpin formation- is it dependent on the side chains
predicted speed limits-for a generic protein = N/100 μs
Is the dynamic motion observed on or off pathway
If off, could slow folding
If on, can we deduce a folding mechanism
Can we tell?
Is the key to solving the protein folding problem an accurate description of the unfolded state?