Protein Folding II

Biochemistry 530 David Baker

http://courses.washington.edu/bioc530

Lecture Outline

- I. Are folding pathways designed into sequences by natural selection?
 - Methods:
 - Phage display selections
 - Kinetic analysis
- II. What determines rate of protein folding?
 - Methods:
 - Collect kinetic data from literature
- III. Is there a hierarchy to the folding of small proteins?
 - Develop simple models Methods:
 - Side chain truncation mutations
 - Kinetic analysis to map folding transition state
- IV. Computer based design of protein folding pathways

I. Are folding pathways designed into sequences by natural selection?

- Case Study: SH3 domain/ Protein L
- Methods:
 - Phage display selections
 - Kinetic analysis
- It is often assumed that the folding process is optimized by natural selection.
- To determine whether this is the case, you need to compare folding rates of naturally occurring sequences to sequences created in the laboratory (which have not been under natural selection).
- How do you identify new sequences which fold?
 - Generate large combinatorial libraries of millions of different sequences; select out those which fold.

Simplification of the sequence of the SH3 domain

Can Nature's 20 letter alphabet be reduced while retaining the protein structure and function?

- SH3 domain
 - All beta sheet
 - ~60 residues
 - No disulfides
 - Binds short proline-rich peptides
- Strategy: select functional and therefore folded variants from combinatorial libraries of heavily simplified SH3 domains displayed on phage.
- Reduced alphabet: I, K, E, A, G.

Schematic of Fully Simplified SH3 Domain

Red--I, K, E, A, G Blue--residues in the peptide binding interface



Simplified SH3 Domain



The stability of library variants is less than wt

But the folding rates are sometimes faster.

Thus, folding rates probably not optimized by evolutionary selection

В



II. Investigating protein folding using simple models and wealth of published data on folding of small proteins

Two-State Folding Proteins:

Cytochrome B ₅₆₂	160 000 s ⁻¹	Gray and Winkler, Pers Com
Cytochrome-C	6 350 s ⁻¹	Mines et al., 1996
λ-repressor	3 640 s ⁻¹	Huang and Oas, 1995
Ubiquitin	1 530 s ⁻¹	Khorasanizadeh et al., 1995
CspB	1 075 s ⁻¹	Schindler and Schmid, 1996
ADA2h	890 s ⁻¹	Villegas et al., 1995
Amino Domain of L-9	720 s ⁻¹	Khulna and Raleigh, Pers Com
ACBP	690 s ⁻¹	Kraglund et al., 1996
Protein G	490 s ⁻¹	Park et al., 1997
¹⁰ FNIII	155 s ⁻¹	Plaxco et al., 1997
Fyn SH3	95 s ⁻¹	Plaxco et al., In Press
Protein L	66 s ⁻¹	Scalley et al., 1997
CI-2	48 s ⁻¹	Jackson and Fersht, 1991
HPr	15 s ⁻¹	van Nuland et al., 1998
Acyl Phosphatase	$0.2 \mathrm{s}^{-1}$	F. Chiti, Pers Com

Simple Model for Folding

Free energy landscape dominated by tradeoff between formation of attractive native interactions and loss of chain configurational entropy (non-native interactions ignored)

All contacts equal in energy

Free energy of ordering segments of same length equal

Contact order: average sequence separation between residues in contact in native structure

LOW CONTACT ORDER

HIGH CONTACT ORDER



Proteins with low contact order fold faster



III. Is there a hierarchy to structure formation in the folding of small (< 100aa) proteins?

Method:

- Kinetic analysis of side chain truncation mutants using simple transition state theory (pioneered by Alan Fersht's group)
- Determine the fraction of the interactions made by each residue in the native state that are also made in the folding transition state
- Experimentally, determine the ratio $\Delta\Delta G_{U-TS} / \Delta\Delta G_{U-N}$. This is often called the Φ value in the literature.

TS theory for protein folding

• Simple transition state theory is often used to treat protein folding



• rate of folding/unfolding proportional to fraction of population at the transition state



Fig. 1. Free energy diagrams illustrating the possible effects of altered helix stability on folding and unfolding rates. Activation energies for folding and unfolding of the wild-type sequence (solid lines) and variant (dashed lines) are indicated by the height of the arrows. A decrease in equilibrium stability (helical propensity) may be expressed kinetically as a decrease in folding rate (top, $\Phi_{\rm f}$ = 1), an increase in unfolding rate (middle, $\Phi_{\rm f}$ = 0), or a combination of both (bottom, $0 < \Phi_{\rm f} < 1$) depending on whether helix is present, absent, or partially formed in the transition state at the site of the substitution. The energetic contribution of helical structure to the transition state can be quantified by the $\Phi_{\rm f}$ parameter (Eq. 4) given by $\Delta\Delta G_{\rm f}$ +, the difference in height of the folding-side arrows, divided by $\Delta\Delta G^{\circ}$, the change in equilibrium stability.





Protein topology is a major determinant of protein folding mechanism

The time it takes a protein to fold, and the rate limiting step in folding depend on the topology of the native structure:

- Insensitivity of folding rate to large scale sequence changes
- Similarity between src and spectrin SH3 domain transition states
- Correlation between contact order and folding rate

Good news for simple models of folding: simple geometric considerations may capture bulk of physical chemistry

IV: Computational remodeling of protein folding pathways

- Topology appears to be major determinant of protein folding mechanisms
- However, in proteins with symmetry, factors such as the relative stabilities of local structural elements can have large roles
- Can folding pathways be altered by modulating stabilities of local structural elements?



Search the PDB for Alternate Hairpin Conformations



- ~ 300 hairpins found for protein G and protein L from PDB database
- redesign the hairpin sequences in the context of the whole protein

Switched folding mechanism of protein G



Conclusions

- Protein folding rates are not optimized by natural selection.
- Correlation between contact order and folding rate, and similarity of folding transition states of homologous proteins suggests *folding mechanisms are determined in part by topology of native state*
- Protein folding transition state can be mapped by investigating effects of point mutations on folding/unfolding rates
- Folding transition states can be understood using computational methods
- Protein folding transition *states/pathways can be switched* using computational design methods.

References

There is a huge body of literature on folding....

In the books recommended for the class:

Creighton, Ch 7, paragraph 7.5 covers folding kinetics Fersht, Ch 18 and 19 have a more thorough description