TRANSITION STATES AND THE MEANING OF PHI-VALUES IN PROTEIN FOLDING KINETICS

Fast folding proteins:
Many proteins fold and unfold rapidly, following single-exponential (two-state) kinetics. Rate processes that involve a single-exponential relaxation in both forward and reverse directions have been traditionally interpreted in terms of rate limiting steps, called **transition states**.

But, how does classical rate theory relate to protein folding? It appears that protein folding is fundamentally different from processes studied by the classical rate theory. There is no single microscopic reaction coordinate that every chain follows. Also, there may not be identifiable barriers such as traditional transition states studied by classical rate theory. The landscapes that explain protein folding might be funnel-shaped.

Main questions:
1. What chain conformations are responsible for the observable relaxation rate?
2. How do they cause this rate?
3. What is the experimental evidence?

The main approach so far has been performing mutational studies of folding rates and equilibrium constants. This was developed as the phi-value analysis. It has been widely applied.

Phi-value analysis:
A particular amino acid is mutated. If the mutation changes the stability of the protein by an amount of $\Delta\Delta G$, and the folding barrier by an amount of $\Delta\Delta G^{++}$, the phi value is:

$$\text{phi} = \Delta\Delta G^{++} / \Delta\Delta G$$

For protein folding, phi is usually regarded as a «kinetic ruler»: how native the transition-like conformation of a monomer is. If phi = 0, the transition state is denatured-like, if phi = 1, the transition state is native-like, in other words, the structure is folded at the site of mutation as much as in the native state. Phi can take any values between 0 and 1.

The nativeness of the transition state is defined by the coordinate ξ:
$\xi = 0 \rightarrow$ denatured conformations
$\xi = 1 \rightarrow$ native conformations

The kinetic ruler hypothesis is deeply rooted in classical rate theories, implies that phi is monotonically related to ξ (this does NOT mean that the relationship is linear).

In other words, the kinetic ruler implies that all protein conformations can be lined up along a single axis of nativeness, and somewhere along this axis will be the transition state conformations ξ that are responsible for the single exponential time constant. That is, there cannot be non-classical phi values, because a negative phi-value would mean that the transition state is more denatured than the denatured state, and phi values greater than 1 would mean that the transition state is more native than the native state, which makes no sense.

So, what happens when phi takes non-classical values? 10-20% of the measured phi-values are outside the 0-1 range. Negative values have been observed in computer simulations. These values appear in funnel-shaped energy landscapes.

- When a mutation destabilizes a slow flow channel and causes a backflow from a slow channel into a fast one, we obtain **negative** phi-values.
- Conversely, phi-values **>1** are caused by a backflow from a fast channel into a slow one.

In the article, they propose a new and more accurate meaning for phi:
Phi correlates with the acceleration/deceleration of folding induced by mutations, rather than with the degree of nativeness of a transition state.

Classical phi-values are restricted to systems having a single reaction coordinate, while non-classical values can arise from parallel, coupled flows. For example : funnel.shaped energy landscapes.

In the paper:
- **Conformation** (=microconformation): single arrangement of chain monomers in space.
- **Microroute**: one particular trajectory between two conformers.
- **Macroconformations**: ensemble of all conformations having a particular specified set of contacts.
- **Macroroute** (between two conformations): collection of microtoutes from one conformation to another.

The velocities can be very heterogeneous when we start from a partially folded conformation. Sometimes, we come to the native state more rapidly from a 2-contact fold than from a six-contact fold. Why is that? Because those six contacts are not located on a macroroute that brings the protein to the native state, so the system has to surmount some barriersin the energy landscape.
Folding nuclei: some contacts can make the protein gain kinetic access to the native state. Folding events are like zippers: the most local contacts happen earlier, then secondary structures form non local contacts.

The phi-value is not a kinetic ruler, but it is correlated to the acceleration or deceleration of the process induced by a mutation. Gatekeeper contacts are the ones that strongly influence the folding speed (those that have big absolute values of phi).

<u>Computer-based redesing of a protein folding pathway</u>
- IgG binding domains of protein G and protein L are very similar: a single alpha-helix packed against a four-stranded beta-sheet formed by two oposed beta-hairpins.
- They have no sequence similarity
- At the rate-limiting step:
  - In protein G: the first beta-turn is disrupted and the second is formed.
  - In protein L: the first beta-turn is formed and the second is disrupted.
  - Hypothesis: the B-hairpin with the lowest energy is the one that is formed in the rate limiting step of the protein folding pathway. To test this, they tried to switch the folding pathway of G to that of L.
    - Increase the stability of the first hairpin
    - Decrease the stability of the second hairpin

The first hairpin was changed by redesigning the conformation of an 8-residue segment. Positive phi-angles of the beta strands were removed and a more favorable beta-turn was included (there different types of beta-turns, which are defined by particular patterns of phi and psi angles).

- They created redesigned G proteins, where only amino acids 6-16 were changed:

- NuG2: the lowest energy sequence-structure combination using type I turn.
- NuG1: the lowest energy sequence-structure combination using type I turn.
- The final structure of these proteins and the wild type G protein is the same, except for the small redesigned part. However, the folding pathway is different, it mimicks that of protein L.

The new proteins had an increased stability. First time that protein stability has been increased by computer-based beta-hairpin redesign. They both fold 100x faster than the wild type G protein.

Then they mutated one residue in the two redesigned proteins. Asp → Ala. This mutation removed a hydrogen bond that was necessary to stabilize the second hairpin. → The first hairpin folds now before the second.

Mutations in the first turn speed up the unfolding rate of wt protein G, and slow down the folding rate of the mutated proteins. These changes can be seen in the phi-values of the various mutations.

Transition state contact orders correlate with protein folding rates

They used MD simulations restrained by experimental phi values. They determined the structures of the transition state ensembles of ten proteins that fold with two-state kinetics.

For each of the protein, they calculated the average contact order in the transition state ensemble. They compared it to the experimental folding rate. → they are correlated.

Two state: no partially folded intermediates ccumulate during folding.

We can use phi values al restrains in a MD simulation if we interpret phi values as the “ratio of the number of native interactions in the transition state to that in the native state”.

Native state contact order: it is a measure of the topological complexity of the protein fold. It correlates with logarithm of the rates of folding. Since the rate of folding depends on the free eneergy difference between the unfolded state and the transition state, the existence of this correlation suggests that the native state contact order reflects this free energy difference.

Tow factors to explain this correlation:

- The folding of small proteins are entropy-dominated (the loss of entropy when moving from unfolded to transition state is crucial for the variations in the free energy barrier to folding.
- Transition state topology must resemble that of the native state → the contact order and the entropy loss in the native state are related to those in the transition state.