Levinthal's paradox:

protein with $N$ amino-acids ($N=100$)

# of configurations $3^N = 3^{100} = 5 \times 10^{47}$ config.

$\tau_0 \approx 10^{-11}$ ns.

time for folding $= 5 \times 10^{47} \times 10^{-9}$ s $= 5 \times 10^8$ s

1 yr $\approx 3 \times 10^7$ s.

$\Rightarrow$ time for folding $\approx 2 \times 10^{31} \gg 10^{30}$ years.

Random searches are not efficient way

of searching for energy space.

Levinthal suggested folding pathways.

What about biased random search?

Dawkins (The blind watchmaker), (1987).

If monkeys had to type the remark

from Hamlet

"He thinks it is like a weasel" 28 characters

 Requires $27^{28}$ key strokes $\times 10^{40}$

If monkeys cannot change a letter or space

that is in the correct place $=$ few thousand

key strokes.
what if there is an energy penalty for disrupting a correct configuration.

protein with N+1 amino acids or N bonds.

$S$ is the number of incorrect bonds in any configuration. Native state of the protein has $S=0$.

what is the mean time, starting from some arbitrary value of $S$, to reach $S=0$ for the first time.

$\langle 2^S(s) \rangle$ is the folding time or the inverse of the folding rate.

$$K = \frac{k_0}{k_1} e^{-\frac{\Delta E}{k_0 T}}$$
\[
\begin{align*}
\text{rate } (S \to S+1) &= (N-S) k_0 . \\
\text{rate } (S \to S-1) &= S k_1 . \\
\end{align*}
\]

\[ P(S, t) \equiv \text{probability that there are } S \text{ incorrect bonds at time } t . \]

\[
\frac{dP(S,t)}{dt} = (N-S+1) k_0 \, P(S-1,t) + (S+1) k_1 \, P(S+1,t) - (N-S) k_0 \, P(S,t) - S k_1 \, P(C,S,t) .
\]

**Result**

\[
Z(S) \approx \frac{1}{N k_0} \left( 1 + \frac{k_0}{k_1} \right)^N . \\
\text{asymptotically correct for base } N .
\]

\[
Z(S) \approx \frac{1}{N k_0} (\gamma + 1)^N
\]

\[
\Delta E \approx 0 \quad \Rightarrow \quad \text{random search } \frac{k_0}{k_1} = \gamma .
\]

Energy bias \( \propto 2k_0 t \) reduces folding time to \( t \)
Folding nucleus.

What is the nature of the transition state 
\( \phi \)-value analysis (Fersht, Alan).

\[
\phi_F = \frac{\Delta \Delta G_{+\rightarrow 0}}{\Delta \Delta G_{0\rightarrow -0}}.
\]

\( \phi_F = 0 \) \( \Rightarrow \) transition state destabilized in the same way as the denatured state 
\( \Rightarrow \) site of mutation denatured in the transition state.

\( \phi_F = 1 \) \( \Rightarrow \) transition state destabilized in the same way as the native state 
\( \Rightarrow \) site of mutation folded in the native state.

Intermediate \( \phi \) value \( \Rightarrow \) partially folded or multiple pathways.
How does the folding rate scale with the length of the protein?

**Theoretical Results**

1) $\ln(k_r) \sim L^{1/2}$

2) $\ln(k_r) \sim \ln(L)$

3) $\ln(k_r) \sim L^{2/3}$

4) Lattice Simulations; $\ln(k_r) \sim L$

**Experimental Results**


Folding rate versus stability, length, topology.

Compared available data from all 2-state folders.

How to measure topology:

\[ CO = \frac{1}{L} \sum_{i=1}^{N} DS_{i,j} \]

$N = \# \text{ of contacts}$

$DS_{i,j} = \# \text{ of residues between two contacts}$
- No correlation with stability.
- Weak correlation with length.

3-state folding proteins

- No correlation with REO.
- Significant correlation with length L.

Finkelstein
Galitski et al. (2003)

\[ \text{Abs - CO} = \frac{C_0 \times L}{N} \leq \sum \Delta S_{i,j} \]

Contributions from both topology & length.

\[ C_0 \times L^p \rightarrow \ln (k_s) \]

Highest correlation for \( p = 1 \).

\[ \text{Abs - CO} \rightarrow \ln (k_s) \]

[\boxed{\text{Abs - CO}}]

chain length as \( L \approx 0.1 \times 10^{12} \).