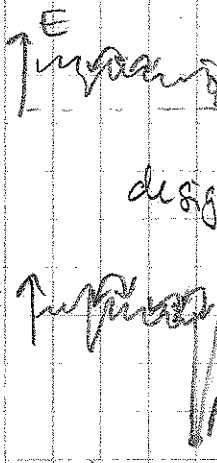
* REM and protein folding

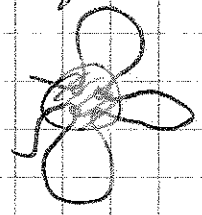
non-designed REM = random a.a. sequence
- ground state at E_c
- stable at T_c
- slow kinetics due to high barriers around E_c

designed REM = evolved a.a. sequence
- ground state at $E_J \ll E_c$
- stable at $T_f > T_c$
- fast kinetics at $\approx T_f$ as "traps" with $E \approx E_c$ are unstable at $T_f > T_c$



* Kinetic model (Gutin et.al. 1998)

1. Folding proceeds through a transition state = folding nucleus



$$F^{\ddagger} = E^{\ddagger} - T \ln \mathcal{N}^{\ddagger}$$

$E^{\ddagger}$: energy of the nucleus; $\mathcal{N}^{\ddagger}$: # of conformations

2. Folding time is determined by time to overcome the (activation) barrier, i.e.

$$t = t_0 e^{\frac{F - F^{\ddagger}}{T}}$$

3. Rapid equilibration, i.e. $F = F(E)$ the minimum of free energy at any T

$\Rightarrow$ calculate $t(T)$

PSF { T_{opt}, show that $T_{opt} > T_c$
{ compare folding time at T_{opt} and T_c

* Lattice model

Structure = chain on a cubic lattice

Energy for any conformation $E = E(\vec{a}, \Delta, U)$

$\vec{a} = \{ W, T, A, A \ldots \}$ protein sequence

$\Delta_{ij} = \begin{cases} 1 & \text{if a.a. } i \text{ and } j \text{ are in contact} \\ 0 & \text{otherwise} \end{cases}$

$\sum_{i>j} \Delta_{ij} = n \leftarrow$ # of contacts in the structure

$U(x,y)$ is a 20×20 matrix of interactions

$$E = \sum_{i>j} \Delta_{ij} U(a_i, a_j)$$

- For a random sequence E is random
  - with $\langle E \rangle_{\substack{\uparrow \\ \text{seq} \\ \text{over} \\ \text{sequences}}} = n \cdot \sum_{x,y} U(x,y) \rho(x) \rho(y)$

where $\rho(x)$ is the frequency of a.a. x in random sequences)

Folding is simulated by Monte Carlo

- steps:  or
- probability to accept a move $= \begin{cases} e^{-\Delta E / T} & \Delta E > 0 \\ 1 & \Delta E < 0 \end{cases}$

Sequence design

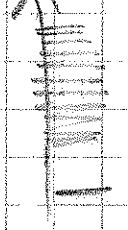
By Monte Carlo (EVOLUTION)

in the sequence space, i.e.

- choose a structure, to be the native structure

Δ^N_{ij}

- minimize energy $E(a, \Delta^N)$ while keeping $\langle E \rangle_{\substack{\uparrow \\ \text{str.} \\ \text{structures}}} = \text{const}$



Or by maximizing $Z = \frac{E - \langle E \rangle_{str}}{\sigma_{str}(E)}$

Monte Carlo: steps = mutations

$\vec{a}$ AVSTAWYH
$\vec{a}'$ AVSTAWYH $\Rightarrow \Delta E = E(\vec{a}', \Delta^N) - E(\vec{a}, \Delta^N)$

Prob to accept $= \begin{cases} 1 & \text{if } \Delta E \le 0 \\ \exp(-\Delta E / T_{seq}) & \text{if } \Delta E > 0 \end{cases}$

T_{seq} = temperature in seq. space.

— How to compute $\langle E \rangle_{str}$ and $\sigma_{str}(E)$ without costly enumeration or Monte Carlo?

Model of random structure

Random structure = random N contacts in Δ_{ij}

$\langle \Delta_{ij} \rangle = \frac{1}{n}$

$\langle E \rangle = \sum_{i,j} \langle \Delta_{ij} \rangle U(a_i, a_j) = \frac{1}{n} \sum_{i,j} U(a_i, a_j)$

$= \frac{1}{n} \sum_{x,y} U(x,y) \cdot p(x) p(y)$

p = freq. of amino acid x in the sequence

$\sigma^2(E) = \langle E^2 \rangle - \langle E \rangle^2 \approx n \cdot \sigma^2(U)$

$\sigma(E) = \sqrt{n}\, \sigma(U)$

Both σ and $\langle E \rangle$ depend on the a.a. composition of the sequence $\Rightarrow$ either $Z \to$ max or $E \to$ min while keeping a.a. composition constant

* Results of lattice models

Designed REM $\Leftrightarrow$ fast folding

Unnecessary & sufficient!

* Go-model

$E = \sum_{i<j} B_{ij} \Delta_{ij}$ $\quad B_{ij} = -\Delta^N_{ij} \Rightarrow$ unrealistic $\frac{E_U}{E_C}$