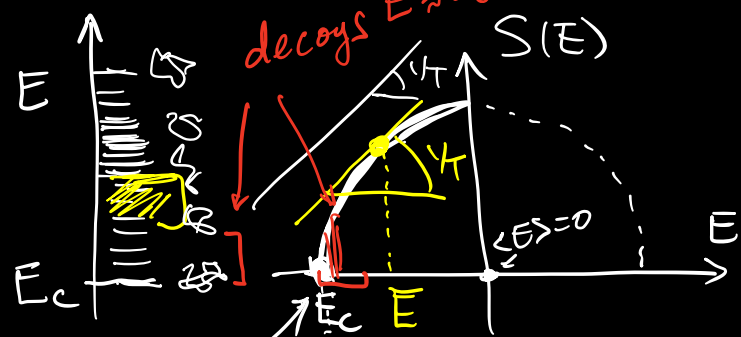
Protein Folding 2 : REM, designed sequences

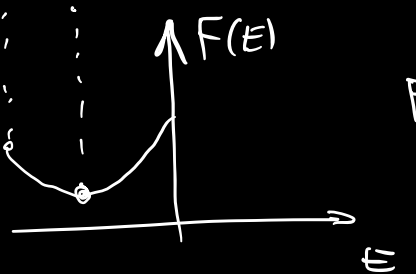
"Energy gap"

① REM : random a.a. sequence folded into all possible conformations



$$\frac{1}{T} = \frac{dS}{dE}$$

$$S(E) = \log M - \frac{E^2}{2\Sigma^2}$$

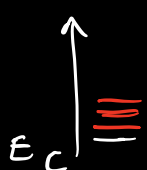


$$E_c = -\Sigma\sqrt{2\log M}$$

$$T_c = \frac{\Sigma}{\sqrt{2\log M}}$$

$$F = E - TS$$

@ T_c : E_c is min of $F(E)$



freezing

folding into E_c state of one of the decoys

Kinetic problems

@ $T \lesssim T_c$: freezing into decoys

But E_c state is stable

@ $T > T_c$: changing states

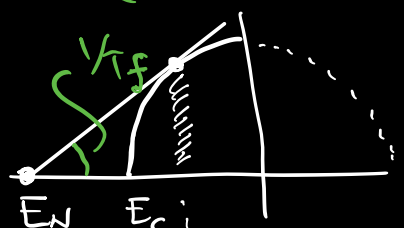
But E_c state is unstable

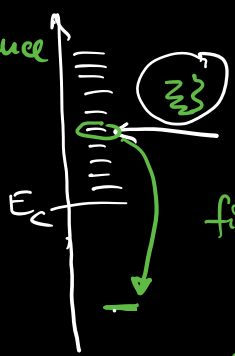
native state

SPEED - STABILITY PARADOX

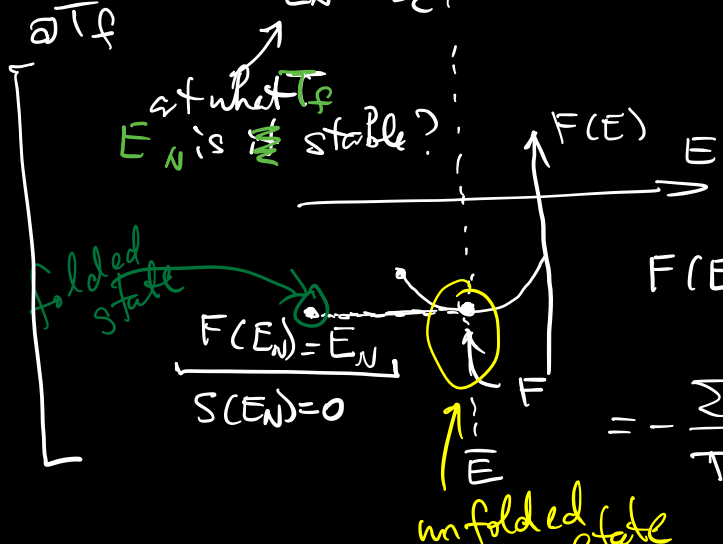
- Is it possible an a.a. sequence such that it has no near-native decoys

Designed REM ⟷ selected sequence





find a.a. sequence such that some selected structure has $E \ll E_c$



$$F(E) = E - T_f S(E)$$

$$= -\frac{\Sigma^2}{T_f} - T_f\left[\ln M - \frac{\Sigma^2/T_f^2}{2\Sigma^2}\right]$$

$$\frac{1}{T} = \frac{dS}{dE}$$

$$S = \ln M - \frac{E^2}{2\Sigma^2}$$

$$\bar{E} = -\frac{\Sigma^2}{T}$$

$$= -\frac{\Sigma^2}{T_f} - T_f\left[\ln M - \frac{\Sigma^2}{2T_f^2}\right] = -\frac{\Sigma^2}{2T_f} - T_f \ln M$$

$$E_N = -\frac{\Sigma^2}{2T_f} - T_f \ln M$$

express in terms of E_c, T_c

$$E_c = -\Sigma\sqrt{2\ln M}$$

$$T_c = \frac{\Sigma}{\sqrt{2\ln M}}$$

$$\Sigma^2 = -T_c E_c$$

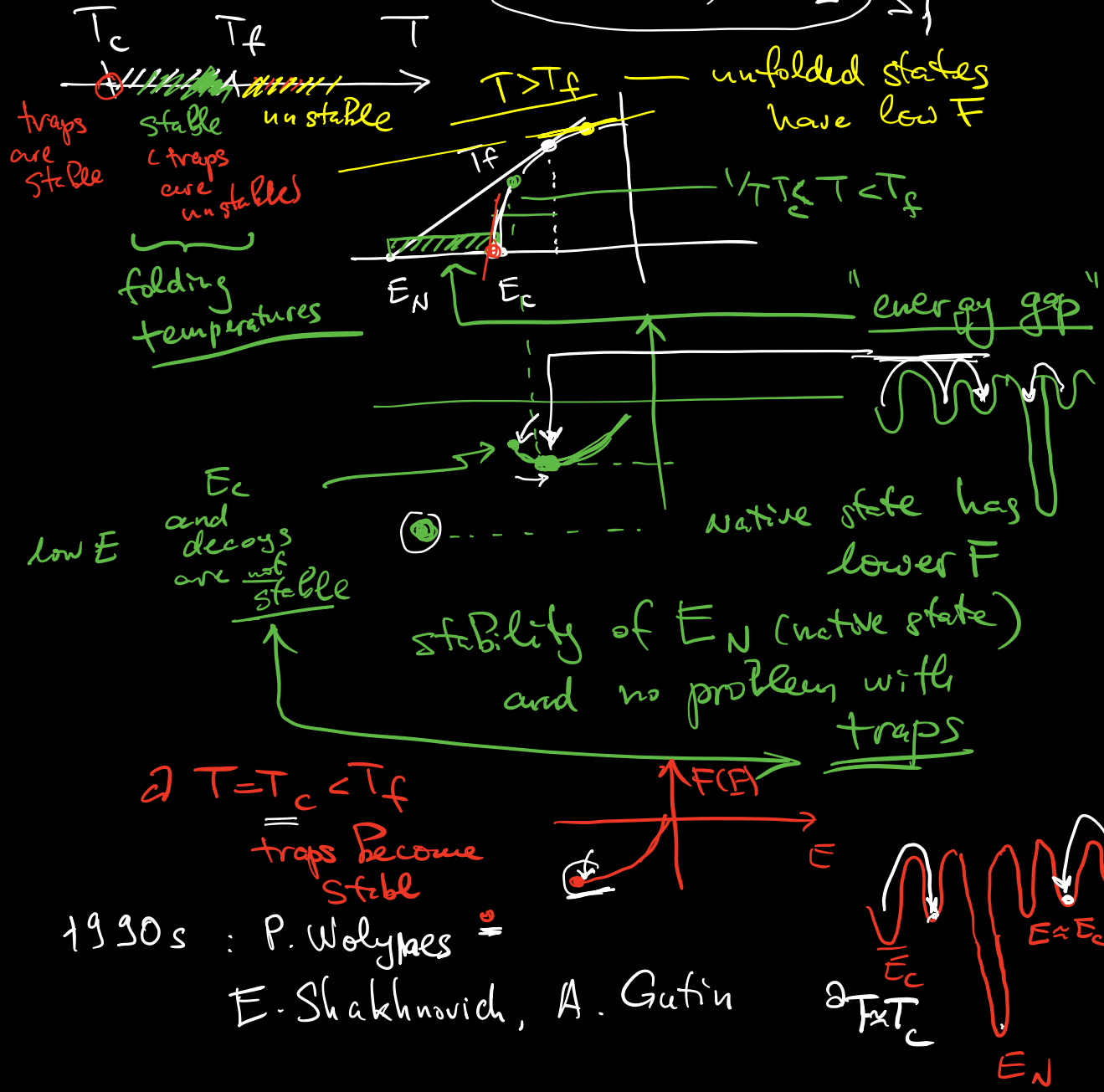
$$\ln M = -\frac{E_c}{T_c}\cdot\frac{1}{2}$$

$$= \frac{T_c E_c}{2T_f} + T_f \cdot \frac{E_c}{2T_c}$$

$$\frac{E_N}{E_c} = \frac{T_c}{2T_f} + \frac{T_f}{2T_c} \times$$

$$2y = x + \frac{1}{x} \qquad x^2 - 2xy + 1 = 0 \;;\; x_{1,2} = y \pm \sqrt{y^2 - 1}$$

$$T_f = T_c \underbrace{\left[\frac{E_N}{E_c} + \sqrt{\left(\frac{E_N}{E_c}\right)^2 - 1} \right]}_{>1}$$



⋆ Kinetic theory (Gutin et al 1998)

(1) Folding proceeds through a transition state ≠ folding nucleus

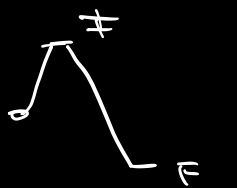
$$F^{\neq} = E^{\neq} - T \log M^{\neq}$$

↑ # of conformations in the transition state

(2) Reaction (folding) proceeds down-hill from the folding nucleus.

Folding time is determined by going into the transition state (nucleus)

$$t_{fold} = \tau e^{\frac{F^{\neq} - F}{T}}$$



↑ for different T ; T_{opt} - ?

⋆ Hypothesis: – random sequences don't fold

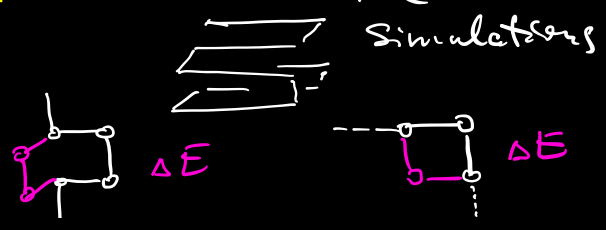
– selecting a sequence with large energy gap ($\frac{E_N}{E_c}$) provides a foldable protein

Lattice model

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

Δ_{ij} = 1 if $r_{ij} < R_{cutoff}$, 0 otherwise

U: 20×20 matrix of interaction energy — fix

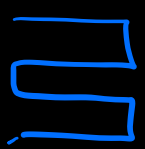


① Test random sequences

accept if $\Delta E < 0$
or otherwise $p = e^{-\Delta E/T}$

② "Design" sequences ⇒ simulate evolution

native structure



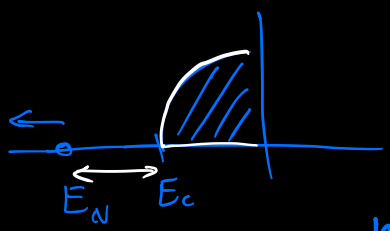
...AVLTSTDA...

...AVLVSTDA...

$\Delta E = E(\Delta_N, a') - E(\Delta_N, a)$

accept if $\Delta E < 0$
or if $p = \exp(-\frac{\Delta E}{T})$

T = evolution temp.



E_N / E_c

minimize E_N

1/population size