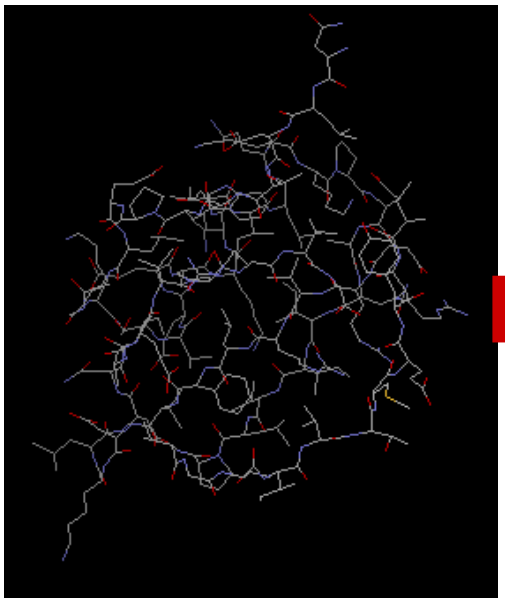
Multiple minima problems in biopolymers

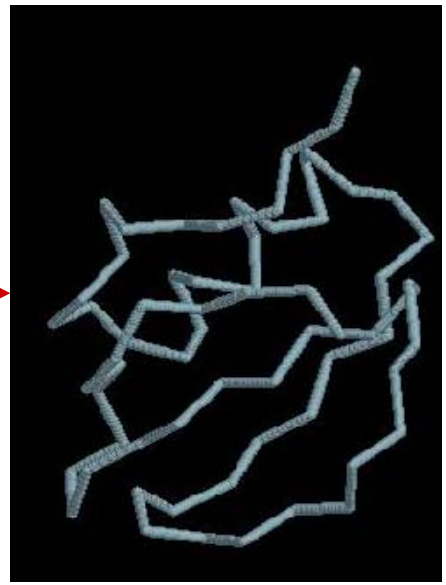
Steven Plotkin

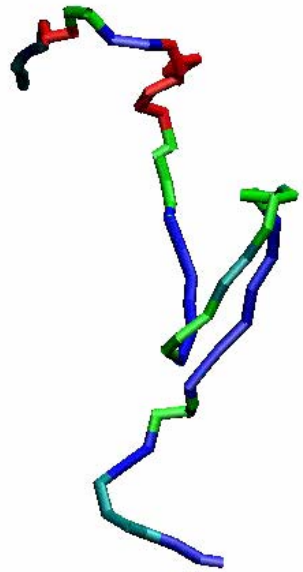
There are at present fundamental problems in theoretical physics awaiting solution, e.g. the relativistic formulation of quantum mechanics and the nature of atomic nuclei (to be followed by more difficult ones such as the problem of life), the solution of which problems will presumably require a more drastic revision of our fundamental concepts than any that have gone before.[1]

-P.A.M. Dirac

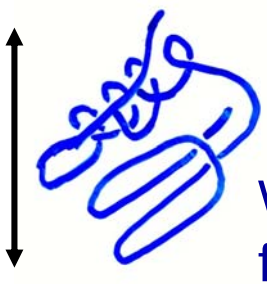
Consider a protein such as this one:







2ci2
Proteinase inhibitor
~30 nm end-to-end,
~70 aa (small)
M=10,000 Au
Size ~ 30 Å



$\Delta E \sim 10^2\ k_B T$ (~10^2 H-bonds, H-ϕ) ,
$\Delta S \sim 10^2\ k_B$,
$\Delta F = \Delta E - T\,\Delta S \sim 10\ k_B T \sim 0.1$ kT/aa
Proteins are only marginally stable
(Facilitates their degradation)

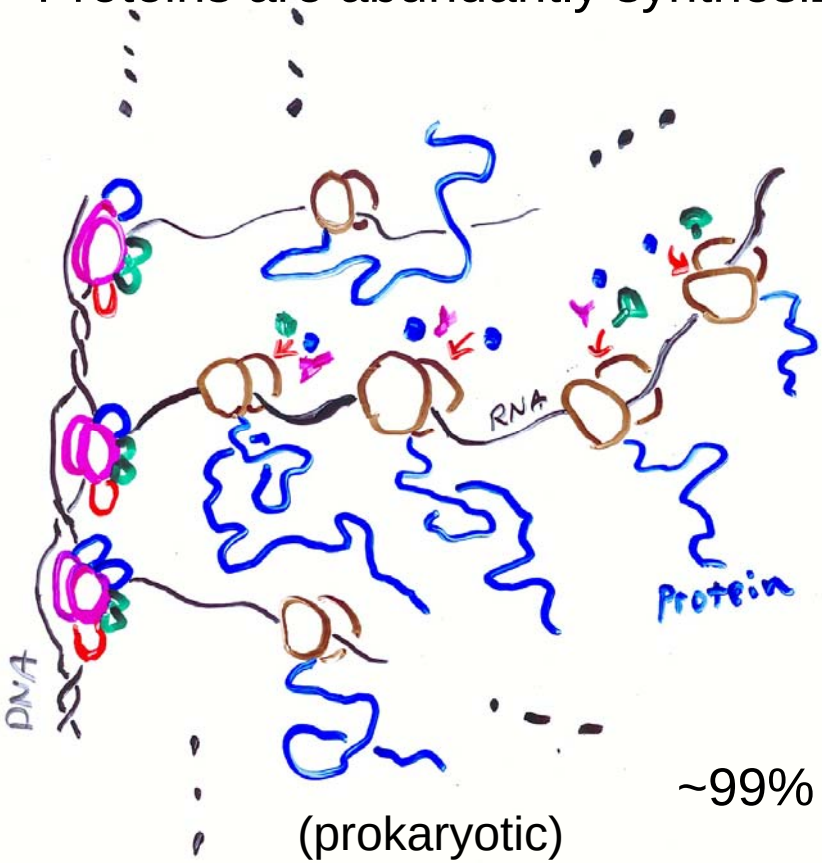
We can coarse-grain it so we can study its folding process on a computer –(must be careful!)

CI2 is 1 member of 1 class of proteins

How many proteins are there in humans?

There are betw 30,000-40,000 genes in humans

Proteins are abundantly synthesized



(prokaryotic)

How many proteins are there in humans?

It has been hard enough counting genes. Proteins can be spliced in different ways and decorated with numerous functional groups, all of which makes counting their numbers impossible for now.

SCIENCE VOL 309 1 JULY 2005

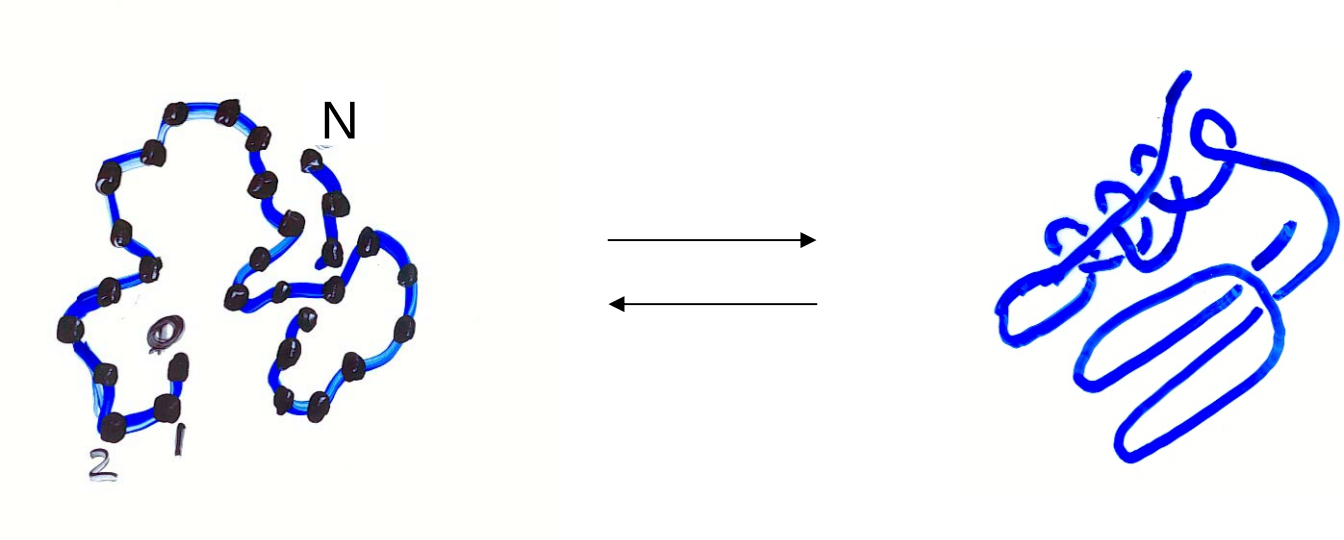
Published by AAAS

From translation rate,
ribosomes/cell,
avg protein length:
~10,000 proteins $\cdot s^{-1} \cdot cell^{-1}$

>> metabolic requirements
~ **100 proteins $\cdot s^{-1} \cdot cell^{-1}$**

~99% of synthesized proteins have been recycled

For proteins the energy landscape must be largely downhill



$\Omega \sim 2^N$

$\nu \sim 10$

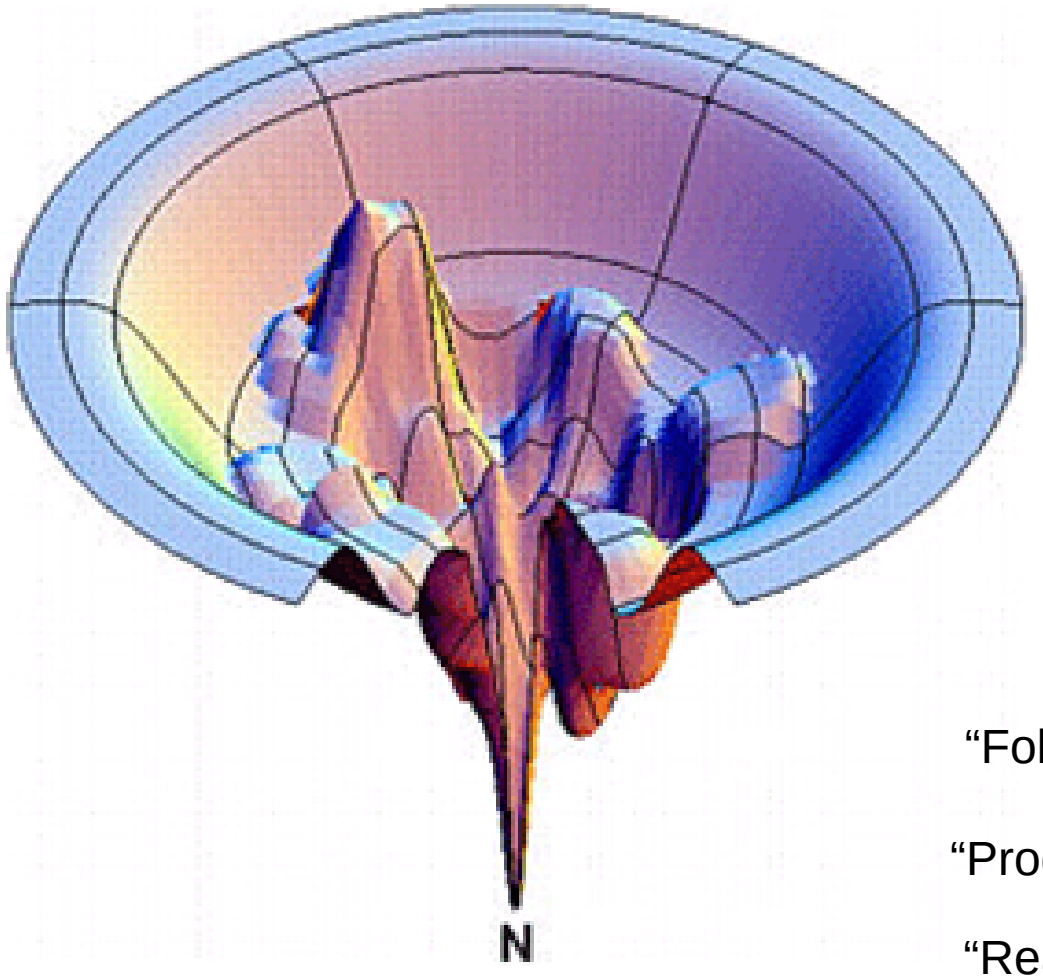
$N \sim 70$ (Cl2)

$\tau_F^{(rand)} \sim \tau_0 \, 10^{70}$

$\sim (10^{-11} s) \, 10^{70}$

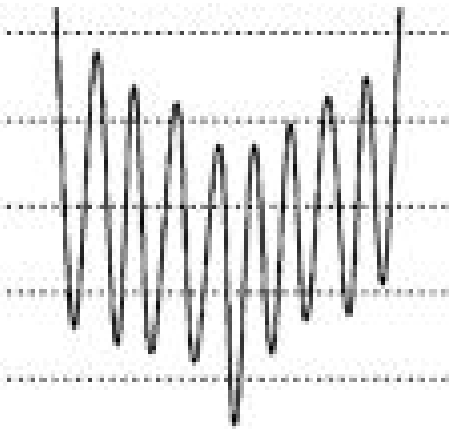
$\sim 10^{59}$ s.

Paradigm: Protein dynamics occurs on a funneled energy landscape

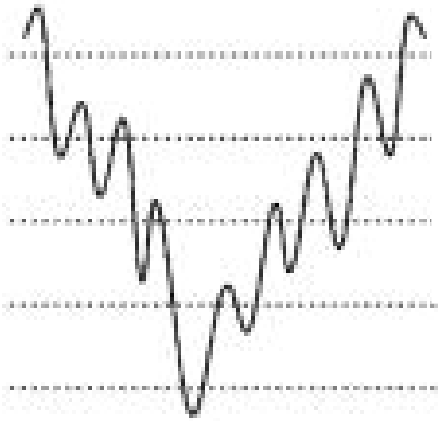


Wolynes, Onuchic, Dill…

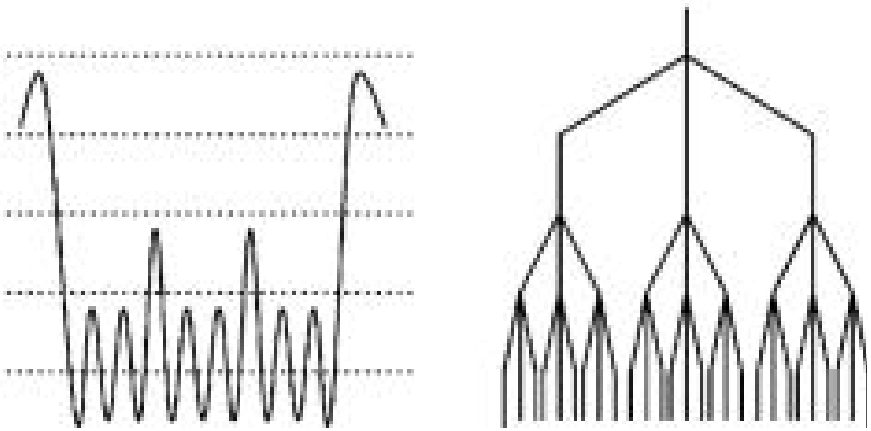
Disconnectivity trees



Funneled, high barriers



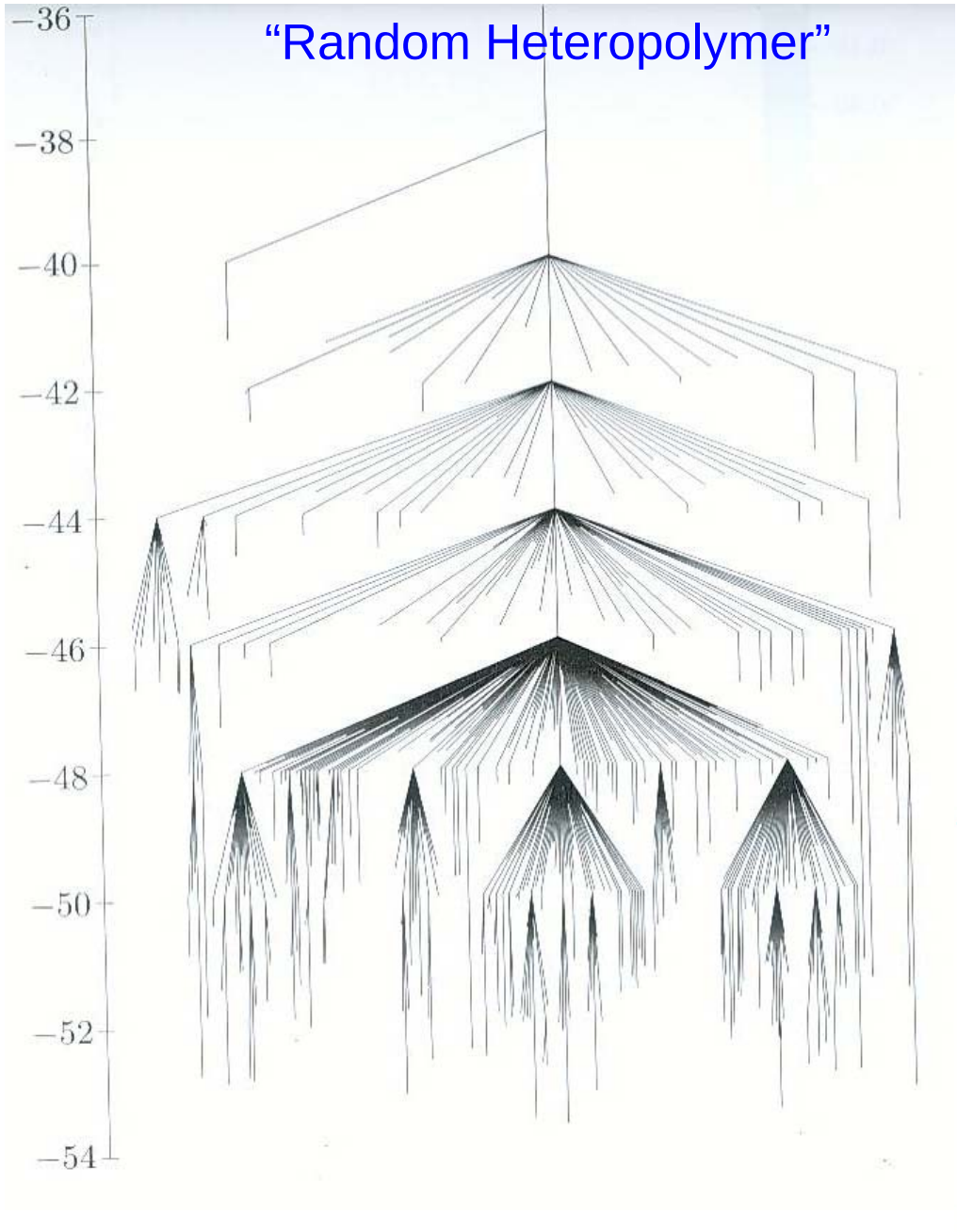
Funneled, low barriers

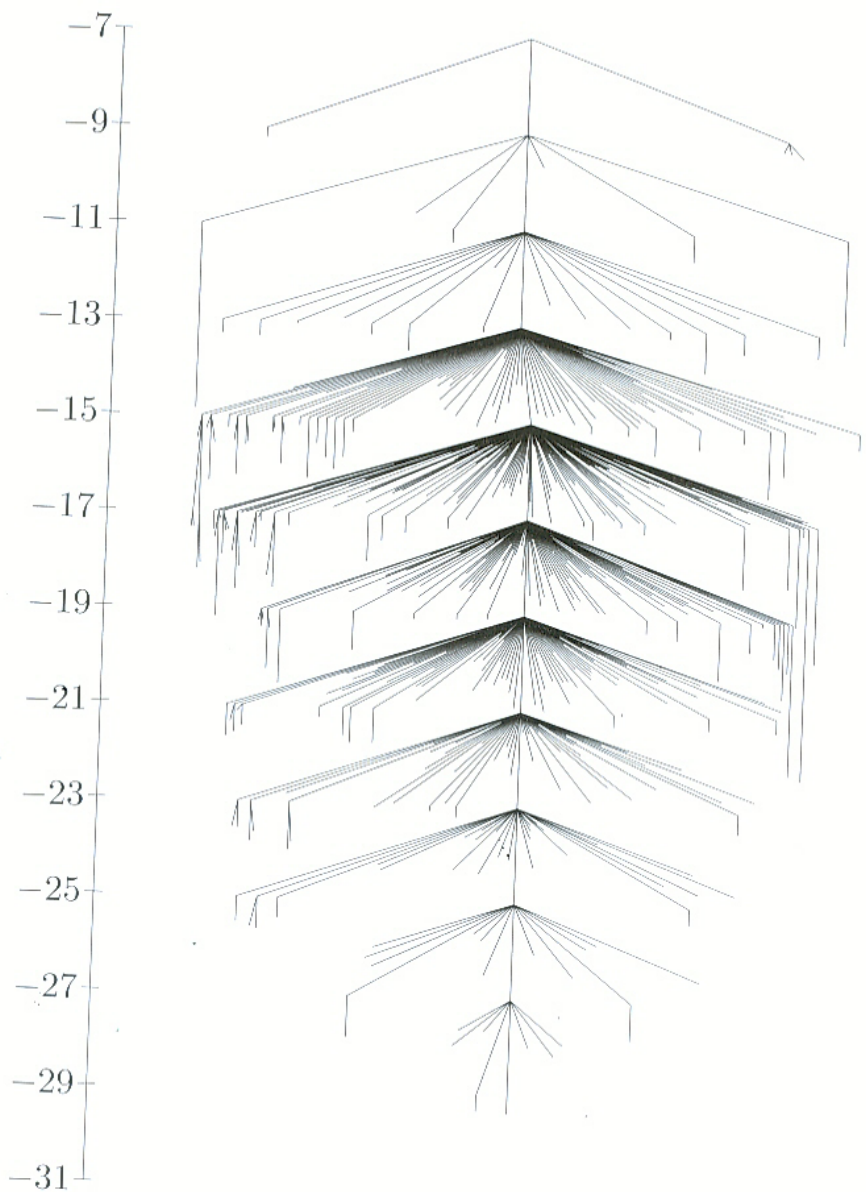


Frustrated/Glassy

Clusters together minima w barriers no higher than a given threshold

DJ Wales “Energy Landscapes” (Cambridge 2003)



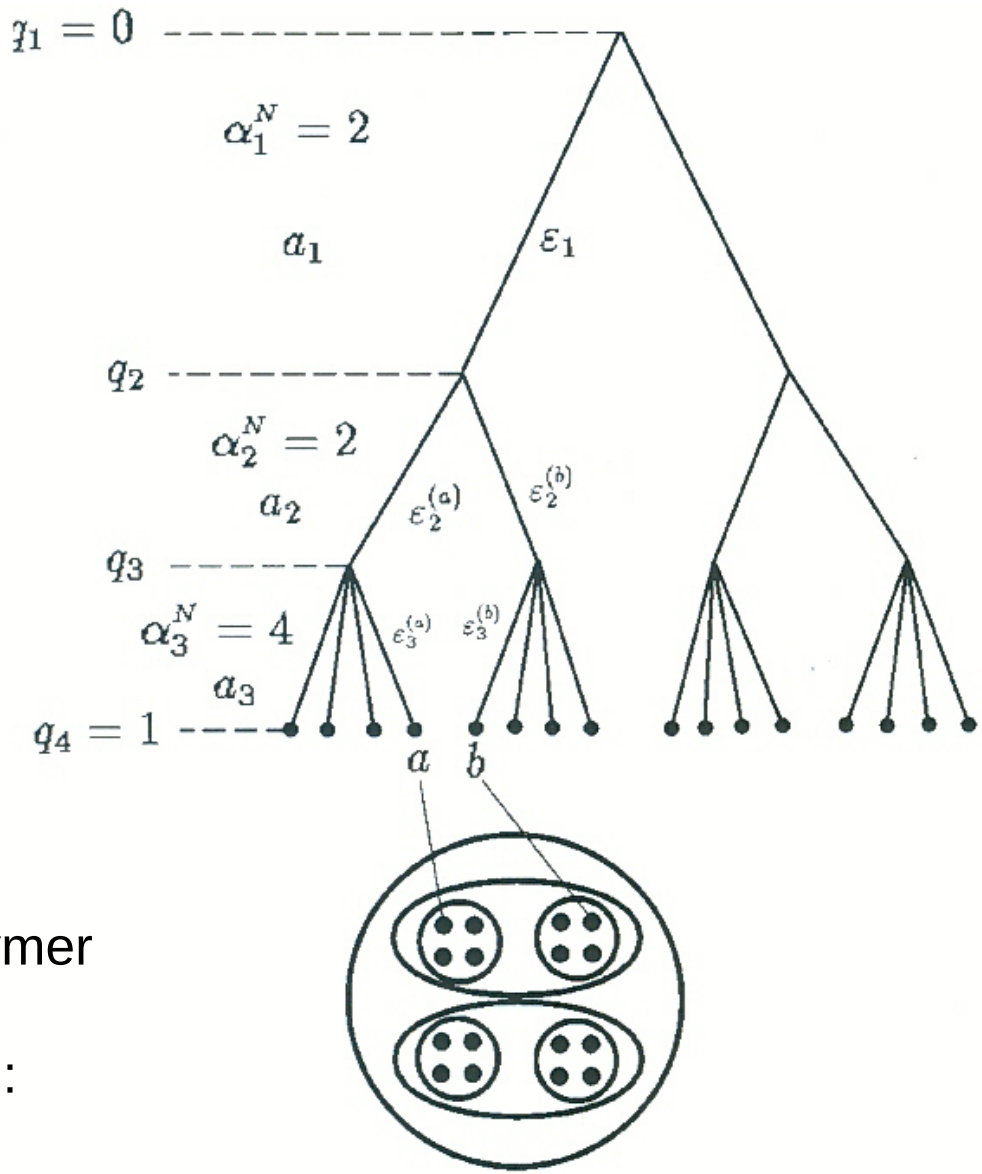


DJ Wales “Energy Landscapes” (Cambridge 2003)

Replica symmetry breaking and ultrametricity

The glass transition represented through the sequence avgd tree that describes how replica states are clustered as T<Tg

Map a model of RSB accurate up to pair correlations (Derrida's GREM) to the random heteropolymer

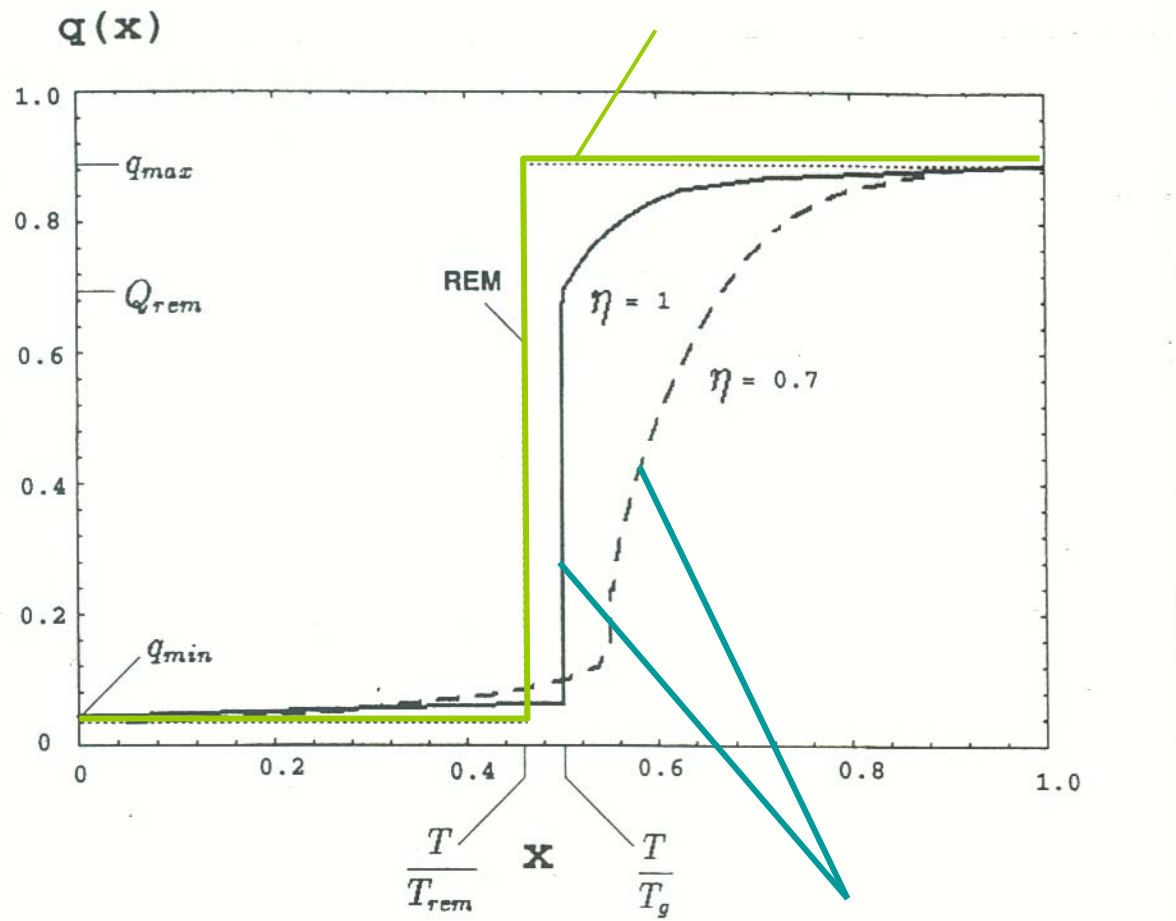


2 quantities in the model:

$$P_q(E_1, E_2) \ , \quad S(q)$$

RSB order parameter

Replica calculations for Virial expansion
(H Orland ‘88, Shakhnovich ‘89)

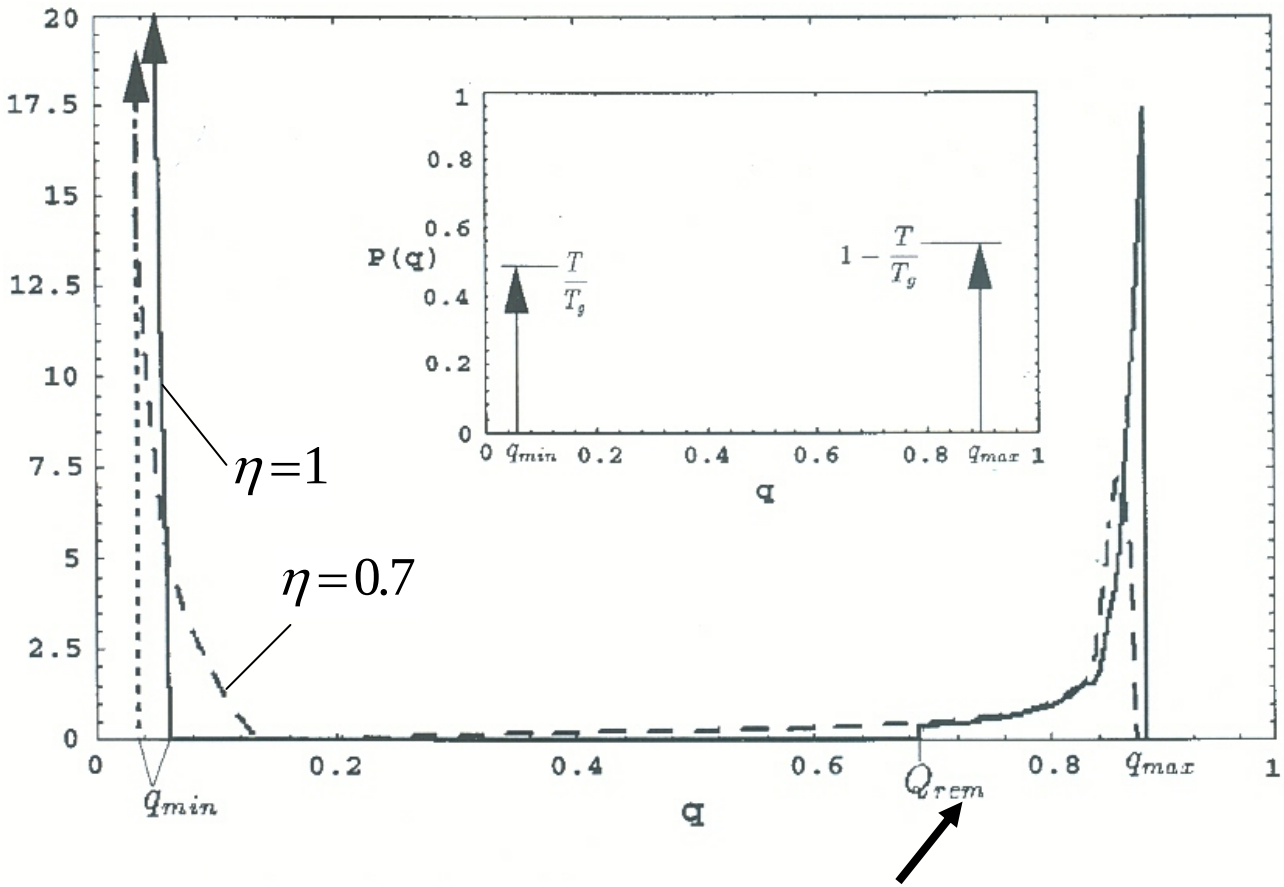


Including correlations, calc
of residual polymer entropy

Plotkin, Wang, Wolynes (1996)

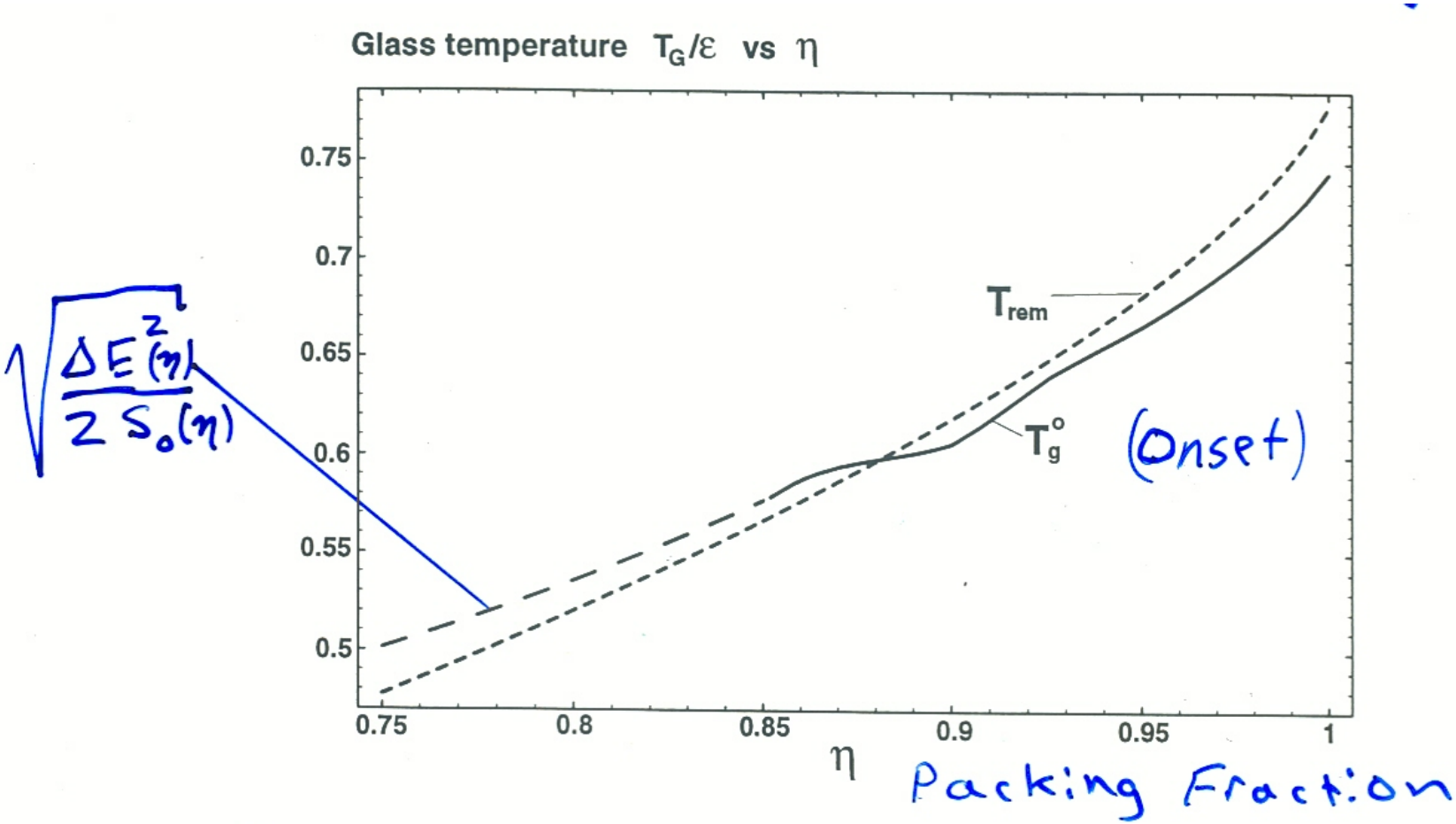
Probability distribution of similarity q, for $T \approx \frac{1}{2} T_G$

$P(q) = dx(q)/dq$



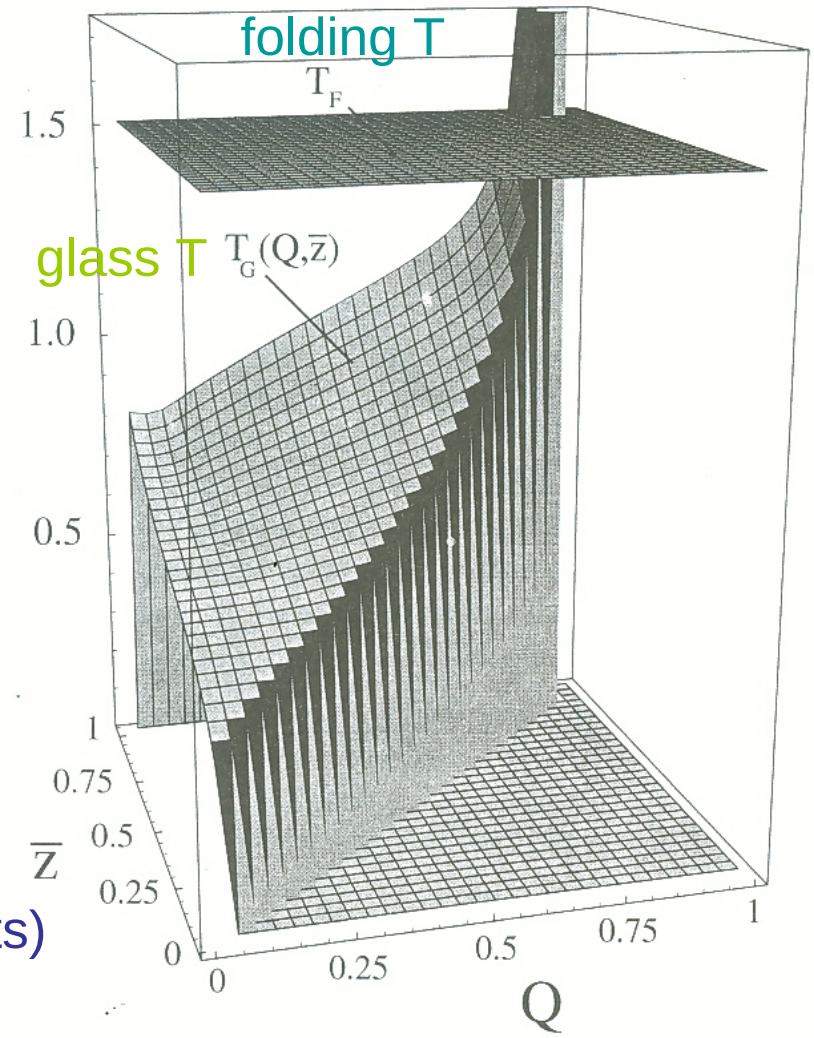
At T< Tg, system freezes into a basin of states rather than a single one
If the system is less dense, basin be large with much entropy.

The glass temperatures for the REM and GREM are very close



So in practice the REM glass temperature may be used for the onset of non-self averaging behavior

The glass temperature is below the folding temperature over most of the range of the landscape.

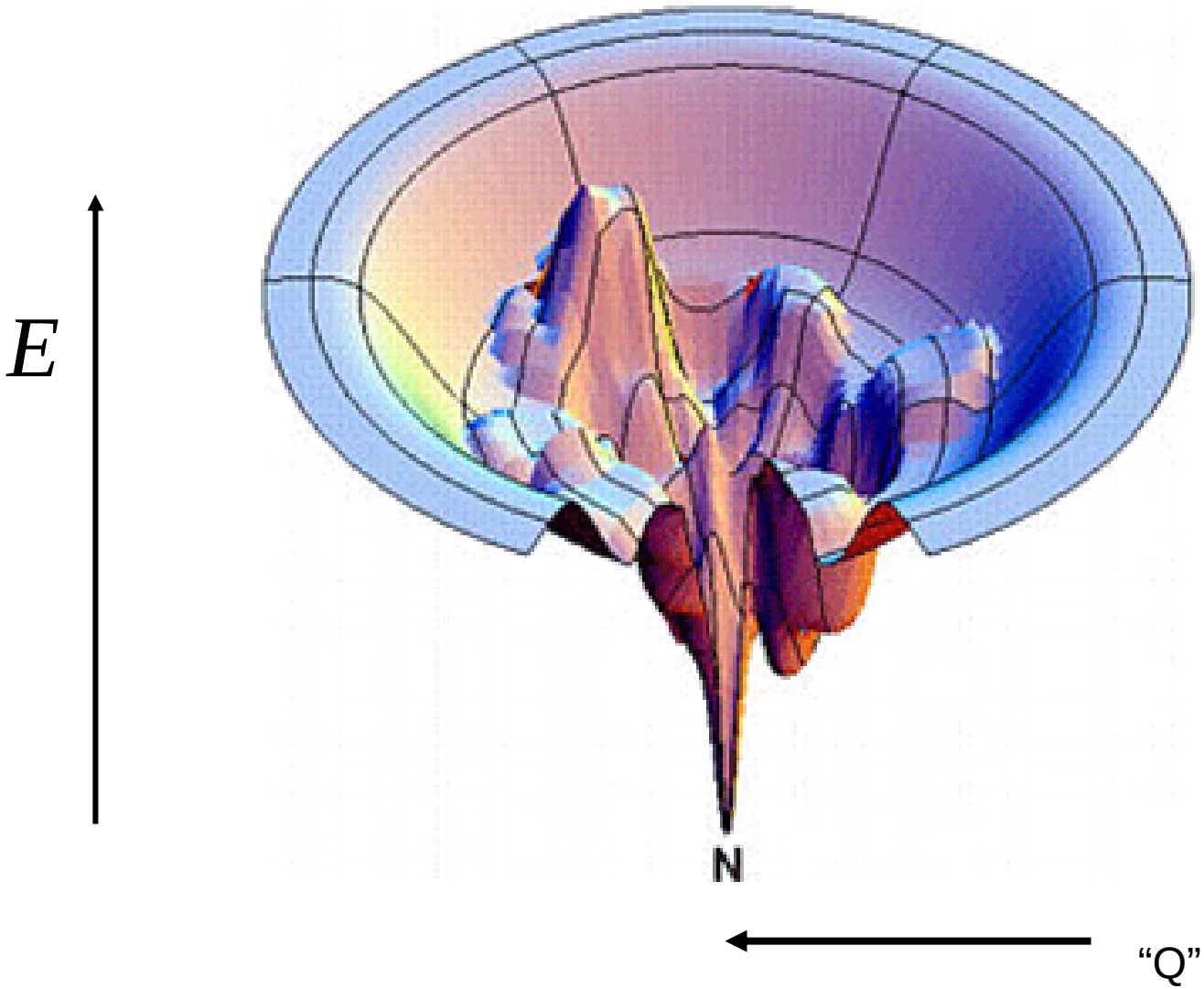


"compactness" (fract of total possible contacts)

"foldedness" (fract of native contacts)

- Diffusion coefficient decreases as the protein folds.
- Free energy landscape is self-averaging.
- Folding dynamics are non-glassy and fast.

A funneled landscape avoids the glassy physics that characterizes random sequences.



Is this the only way?

One can imagine other landscapes that satisfy the thermodynamic and kinetic requirements of folding

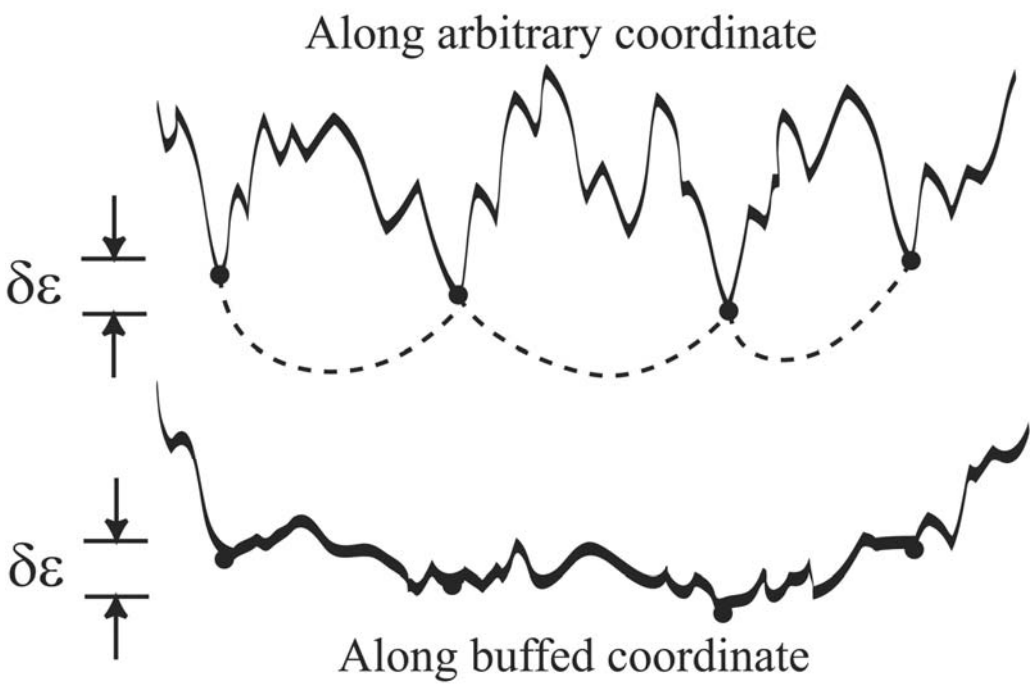
Density of states could still be RHP-like

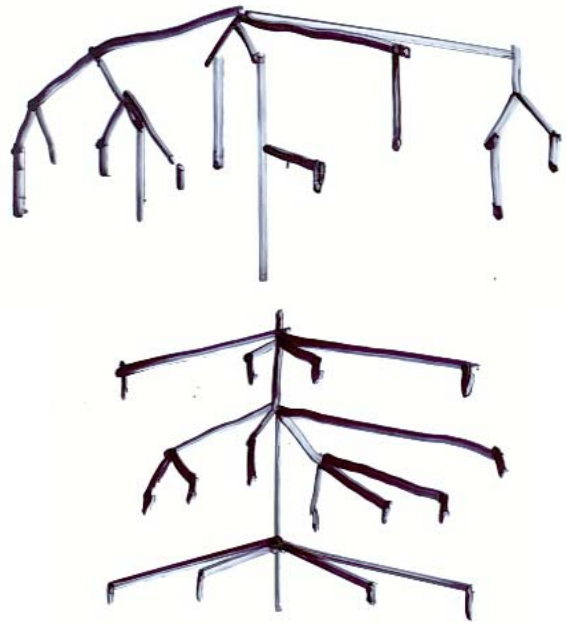
But barriers that make folding slow could be unusually small along a particular “buffed” coordinate.

These landscapes must be rugged enough to be “glassy” at T_{bio}

Only weakly gapped/funneled in the conventional sense

$$P_{GND} > 0.5$$





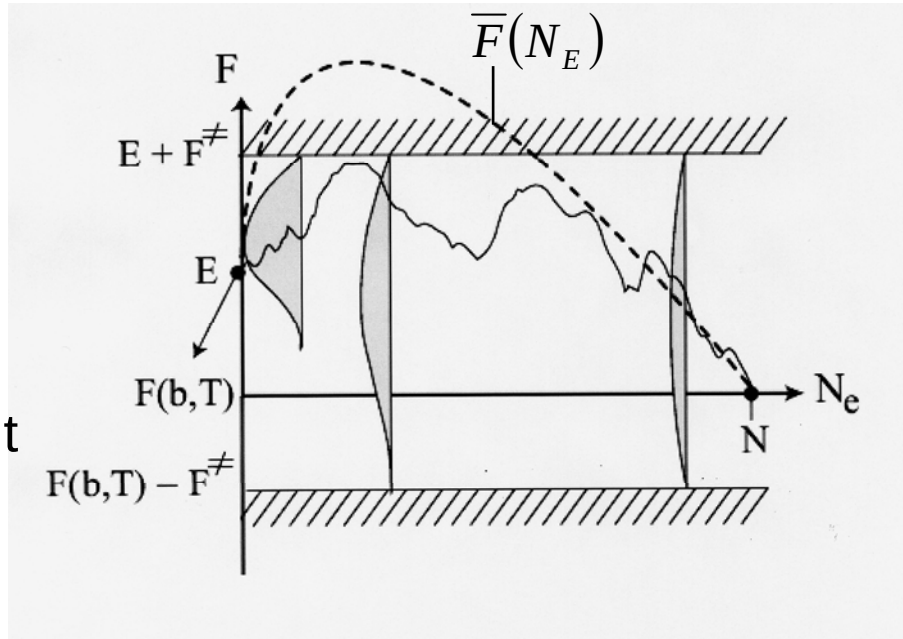
Plotkin, Wolynes PNAS (2003)

The buffing mechanism for folding

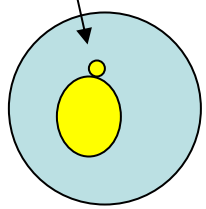
What are the fraction of paths below $F^{\ddagger}$?

Diff seqs have different F(Ne) to escape from local traps

These are distributed about the mean F(Ne)



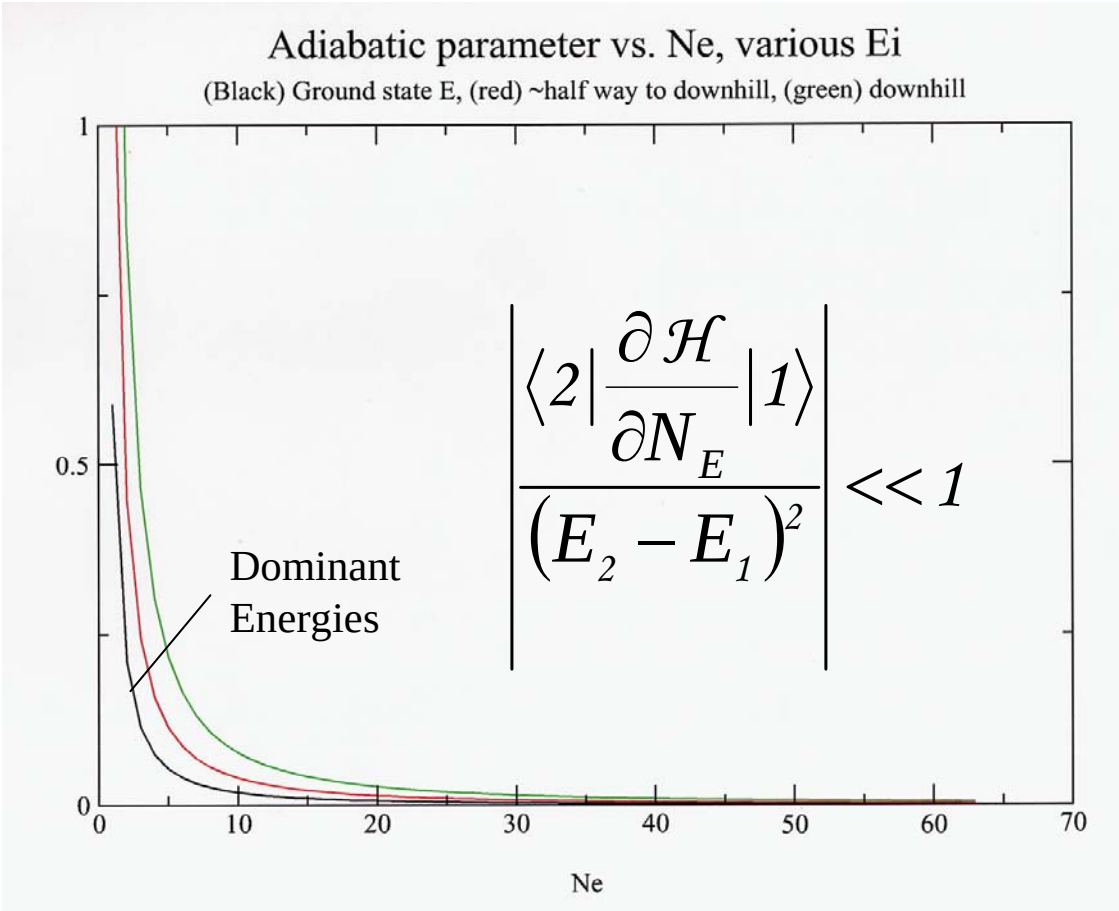
Gaussianly distributed increment



$$p\left(F^{\neq}\,|\,E\right)=\frac{\int \mathcal{D}F\, \mathrm{e}^{-\int_0^N dN_{\mathrm{E}}\left[m^*\left(\frac{\partial F}{\partial N_{\mathrm{E}}}-\frac{\partial \overline{F}}{\partial N_{\mathrm{E}}}\right)^2-\mathcal{V}_{\mathrm{SQ}}\right]}\delta_{\mathrm{i}}\,\delta_{\mathrm{F}}}{\int \mathcal{D}F\, \mathrm{e}^{-\int_0^N dN_{\mathrm{E}}\left[m^*\left(\frac{\partial F}{\partial N_{\mathrm{E}}}-\frac{\partial \overline{F}}{\partial N_{\mathrm{E}}}\right)^2\right]}\delta_{\mathrm{i}}\,\delta_{\mathrm{F}}}=\frac{\mathcal{K}_{SQ}\left(F_N,N\,|\,E_i,0\right)}{\mathcal{K}_U\left(F_N,N\,|\,E_i,0\right)}$$

$$\frac{\partial \mathrm{K}}{\partial N_{\mathrm{E}}} - \frac{1}{4m^*}\frac{\partial^2 \mathrm{K}}{\partial F^2} + \left(\frac{d\overline{F}}{dN_{\mathrm{E}}}\right)\frac{\partial \mathrm{K}}{\partial F} + \mathrm{V_{SQ}} \quad \mathrm{K} = \delta\left(F - E_{\mathrm{i}}\right)\delta\left(N_{\mathrm{E}}\right)$$

Slowly varying



The adiabatic approximation with ground state dominance gives an exponentially rare fraction of buffed sequences

$$f\left(F^{\neq}\right) \approx \mathrm{e}^{-\frac{\pi^2 T^2 \ln\nu}{8F^{\neq 2}}N \; -\frac{\left(b_G^2 - \left\langle b^2 \right\rangle\right)^2}{2\sigma_b^2}N \; - \text{"surface tension"} \times N^{1/2}}$$

Sequence fluctuations

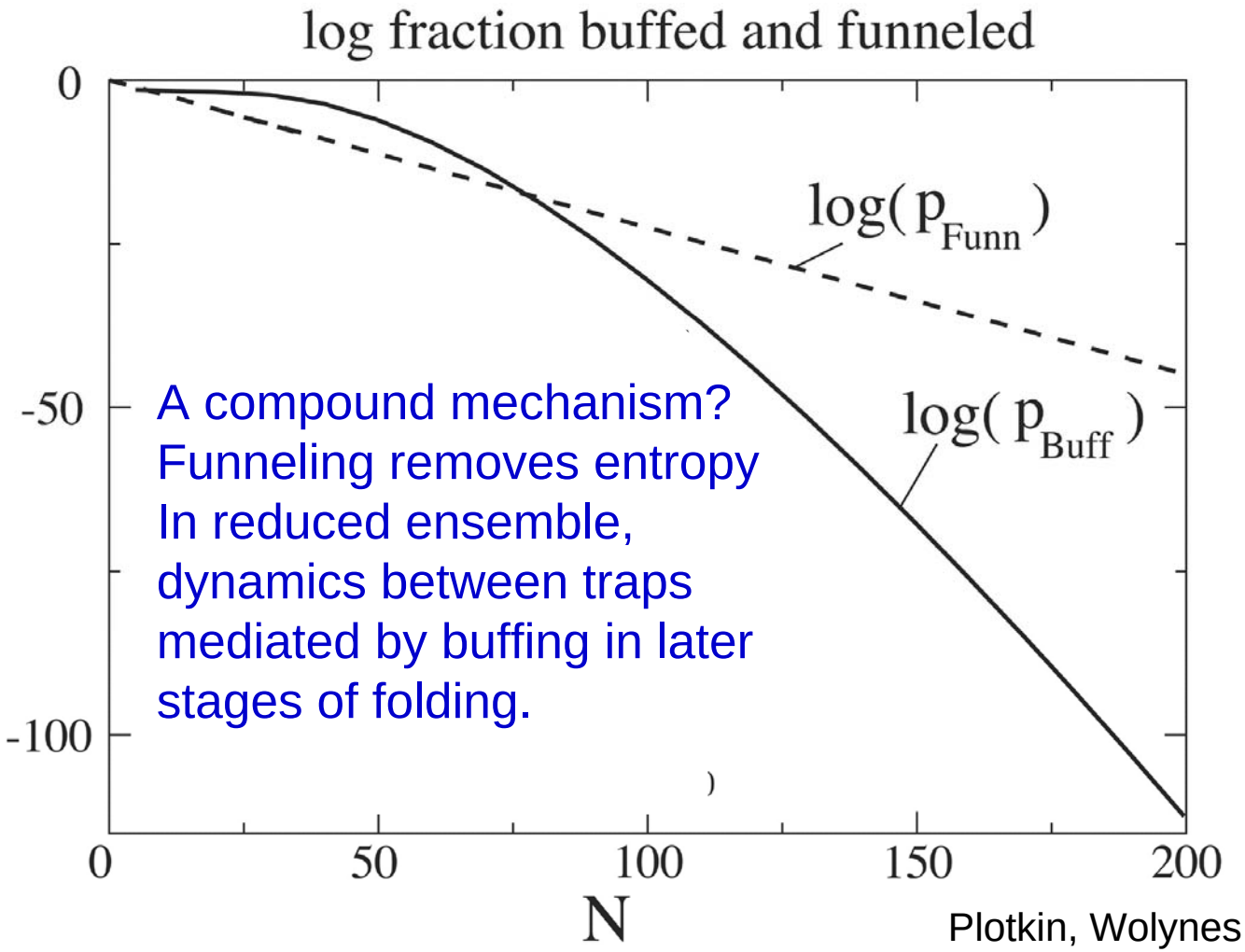
Ruggedness /stability

Mean-field behavior

Funneled sequences are also exponentially rare in N:

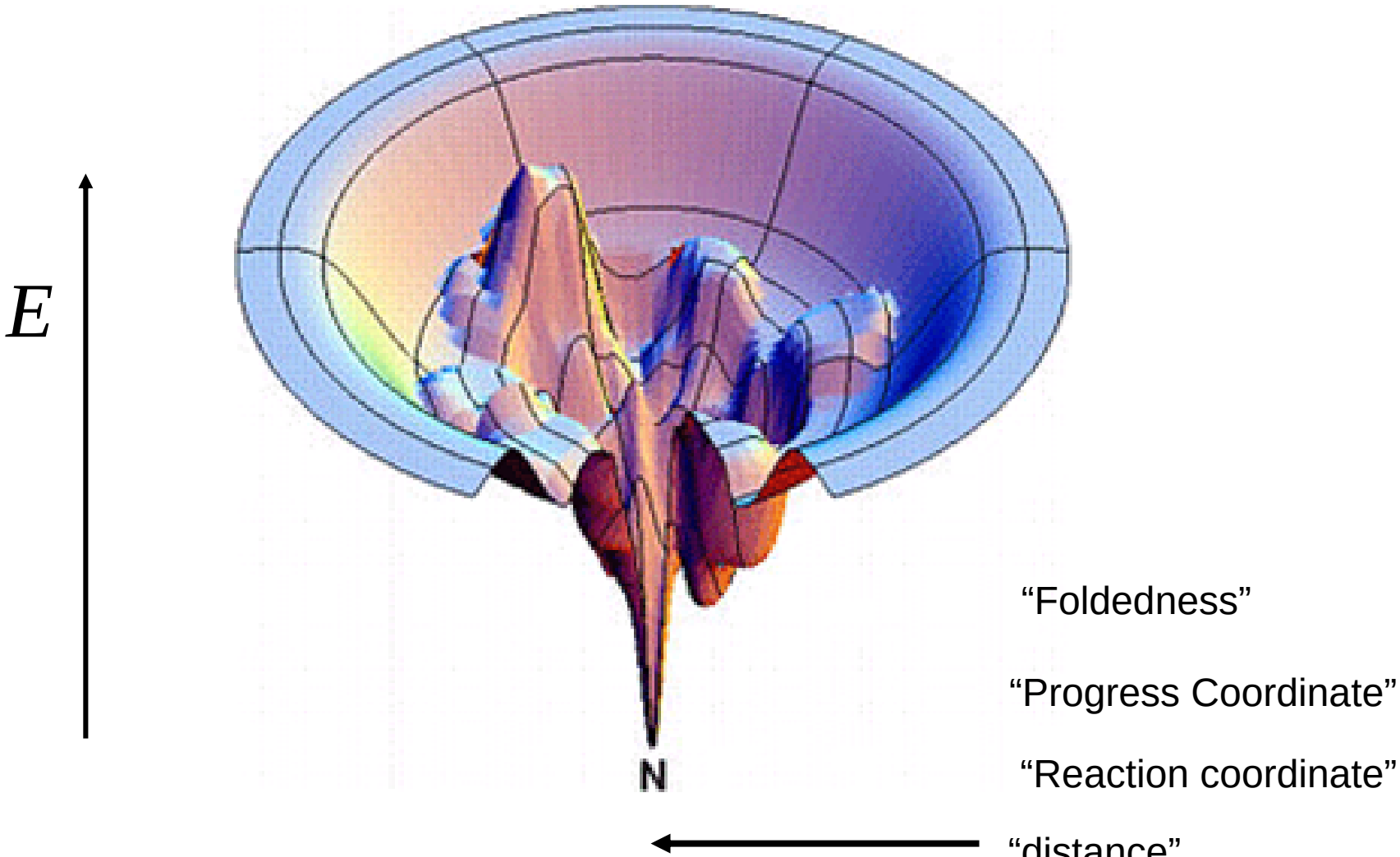
$$f\left(\Delta\right) \approx \mathrm{e}^{-\Delta/T_G} = \mathrm{e}^{-N\,\varepsilon/T_G}$$

How rare are “buffed seqs” compared to funneled seqs?
Quantitatively address this by finding the probability pathways are buffed to say 4 kT, vs. the probability a sequence is funneled enough s.t. $\Delta G_{U\neq} = 4\ kT$
Plot of p_{fun} and p_{buff} vs. chain length:
Funneling becomes dominant for long chains
But there is a crossover

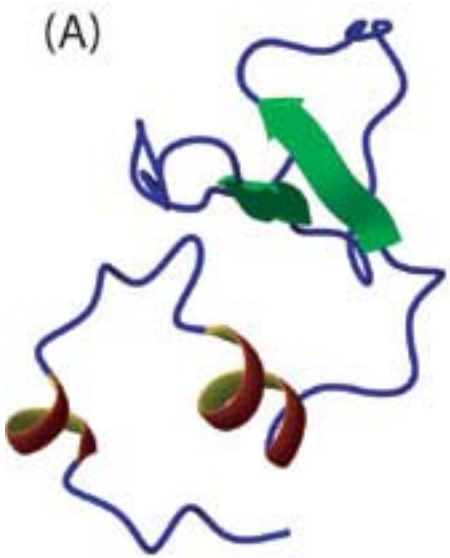


A compound mechanism?
Funneling removes entropy
In reduced ensemble, dynamics between traps mediated by buffing in later stages of folding.

Plotkin, Wolynes PNAS (2003)

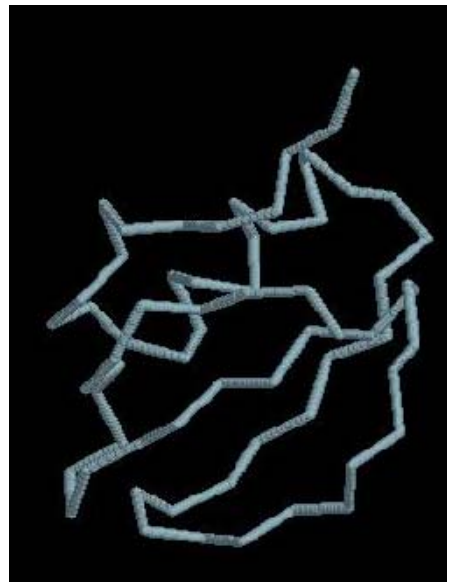

E
N
“Foldedness”
“Progress Coordinate”
“Reaction coordinate”
“distance”

What quantity can tell us how far a structure is from folded?



D_{AN}

(N)

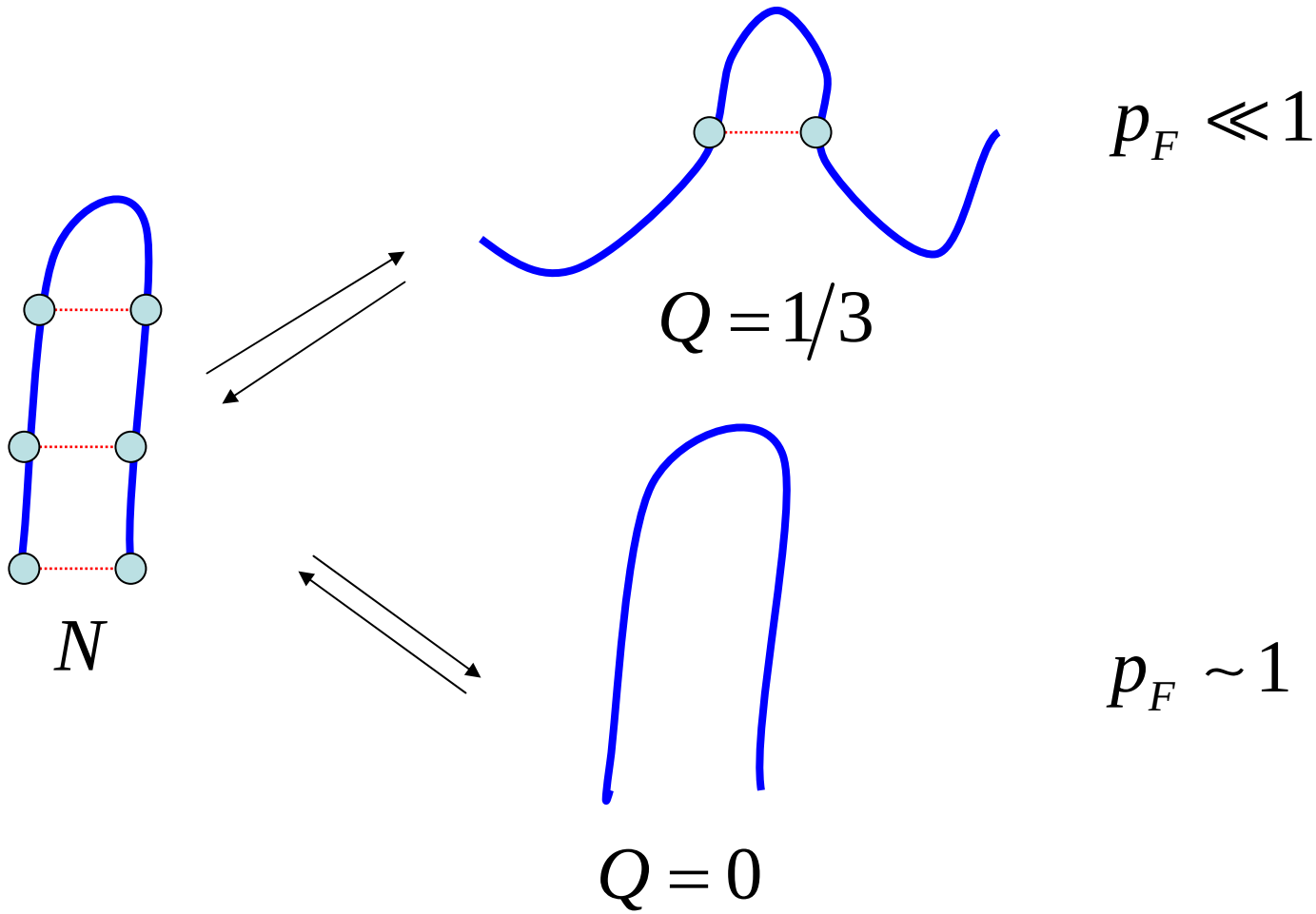


D_{BN}

Should be a geometrical quantity depending only on the 2 structures considered.

Some heuristic distance measures widely used:

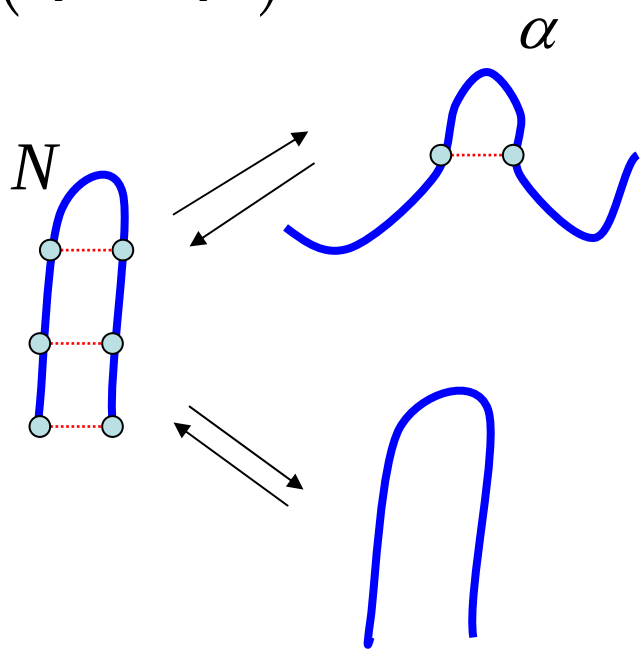
Q = fraction of native ‘contacts’ present



Another heuristic distance measure widely used is

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i^{\alpha} - \mathbf{r}_i^{N}\right)^2}$$

works better for these structures:



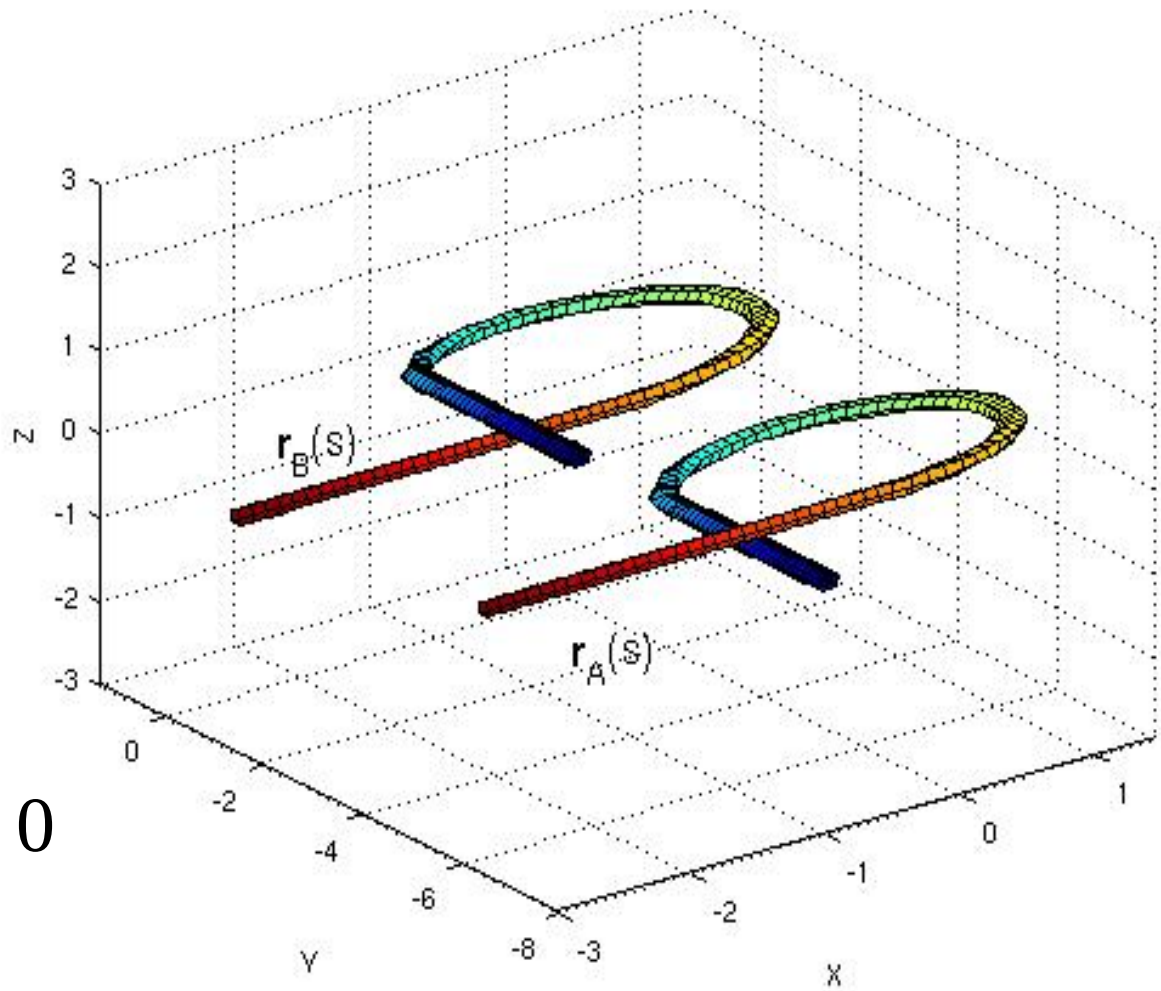
However it is not a true “distance”

Real (minimal) distance beads could move would be:

$$MRSD = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(\mathbf{r}_i^{\alpha} - \mathbf{r}_i^{N}\right)^2} \qquad \leq RMSD$$

Both RMSD, MRSD fail to give the proper distance a partly folded protein needs to move if chain non-crossing is important.

e.g.



$MRSD \approx 0$

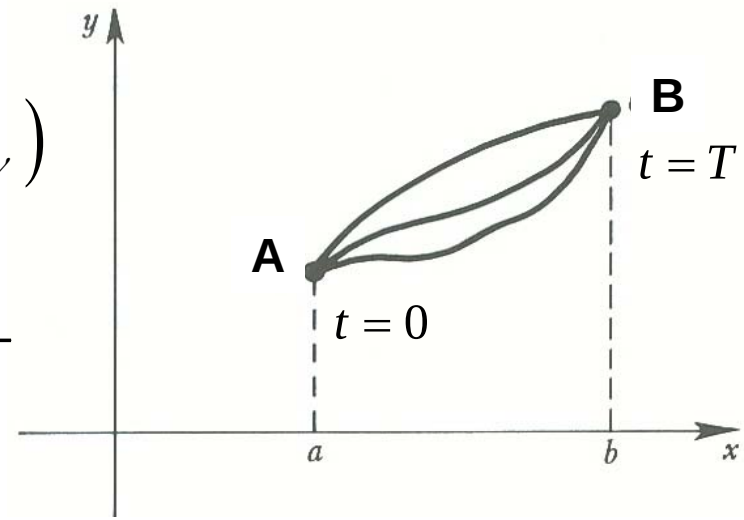
Can we generalize the notion of distance between 2 points to calculate the distance between 2 curves? **(Plotkin PNAS (2007) in press).**

The distance between 2 points is obtained from the variational minimization of the functional:

$$D\left[\mathbf{r}(t)\right] = \int_A^B \left|\Delta \mathbf{r}\right| \qquad \left(g_{\mu\nu} = \delta_{\mu\nu}\right)$$

$$= \int_A^B \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

$$= \int_0^T dt \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \qquad \left(\dot{x} = \frac{dx}{dt}\right)$$



Usually this action is written non-parametrically:

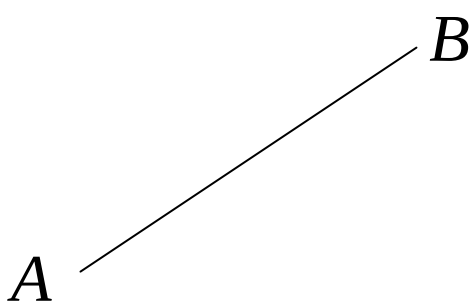
$$D = \int_{x_A}^{x_B} dx \sqrt{1 + y'^2 + z'^2} \qquad y' = \frac{dy}{dx}$$

But for the general problem we have no guarantee y,z are simple functions of x.

Minimization $\delta D = \delta \int_0^T dt \sqrt{\dot{\mathbf{r}}^2} = 0$ **gives the distance**

$D^* = D\left[\mathbf{r}^*(t)\right]$ **and** $\mathbf{r}^*(t)$ **, the minimal transformation**

a 1D object.



Specifically $\delta D = \delta \int_0^T dt \sqrt{\dot{\mathbf{r}}^2} = 0$ **gives**

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{r}}}\right) = 0 \qquad L = \sqrt{\dot{\mathbf{r}}^2}$$

$$\frac{d}{dt}\left(\frac{\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|}\right) = \frac{d\hat{\mathbf{v}}}{dt} = 0$$

$\therefore \quad \hat{\mathbf{v}} =$ const in direction $\hat{\mathbf{e}}_v$

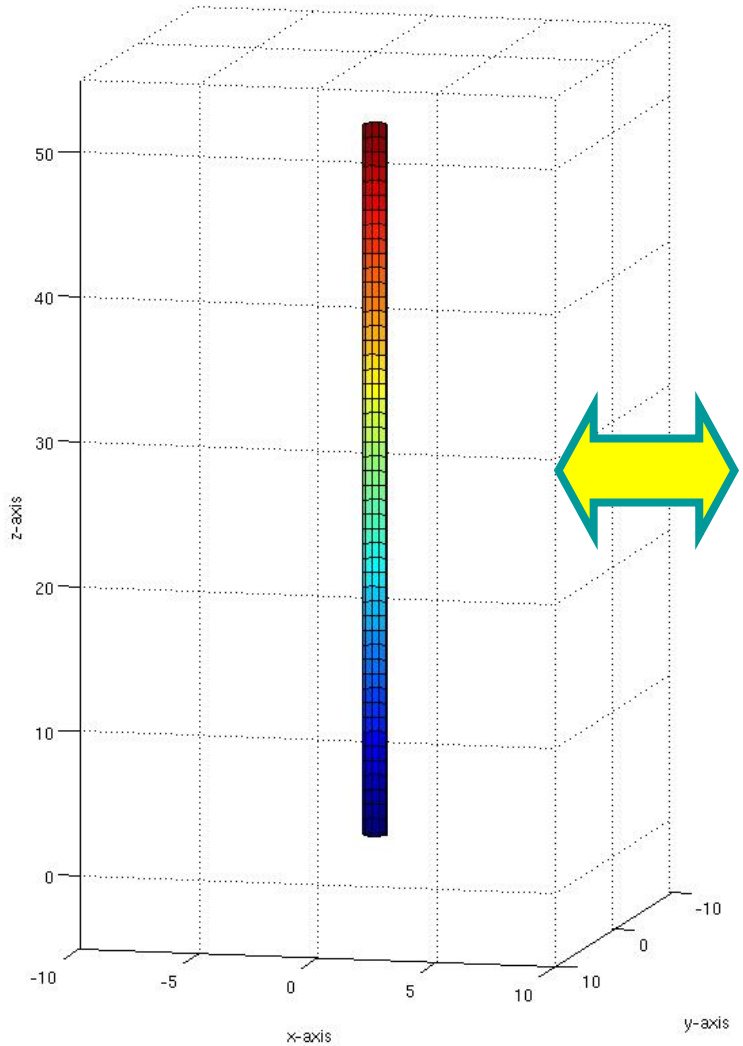
Straight-line motion results

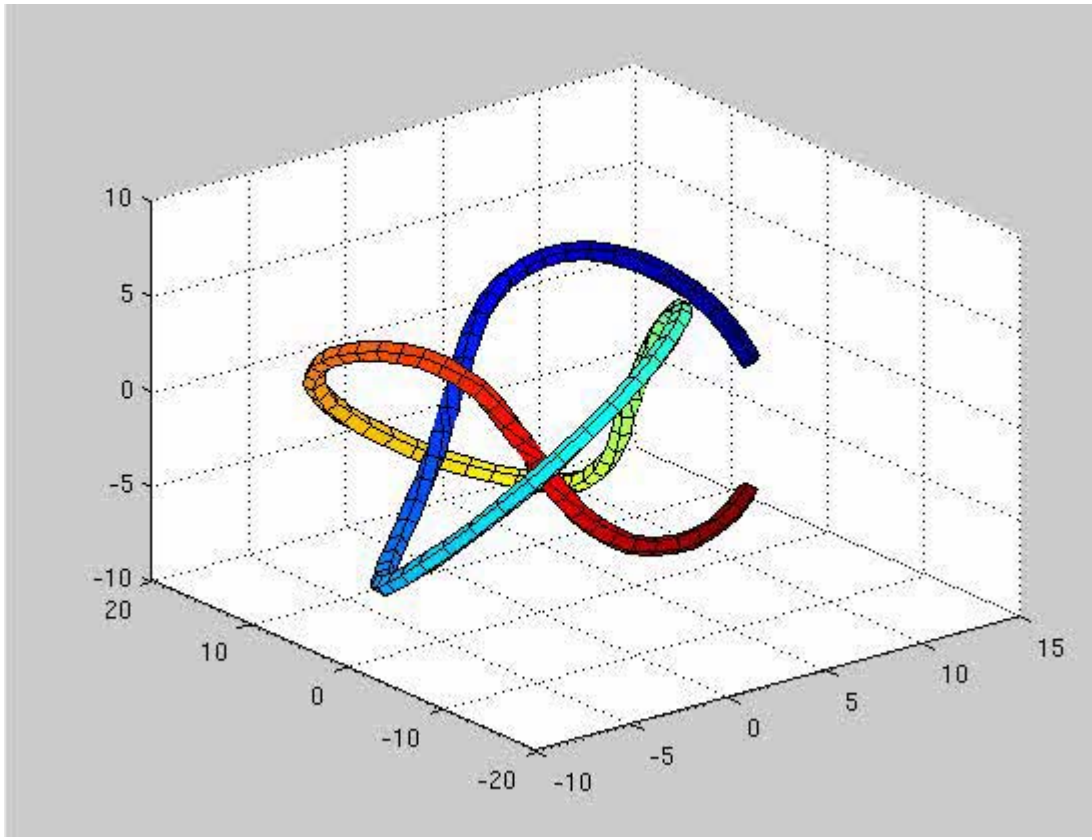
However any $\mathbf{v}(t) = v_0(t)\hat{\mathbf{v}}$ **s.t** $\mathbf{r}(t)$ **satisfies B.C.'s** ***A, B*** **is a solution.**

$$\mathbf{r}(t) = \mathbf{A} + \hat{\mathbf{v}}\int_0^t v_0(t)$$

We could have fixed a gauge from the outset, or we can fix it at the end (easier).

We want to generalize this treatment to find ***the*** distance between any two space curves $\mathbf{r}_A(s),\ \mathbf{r}_B(s)$



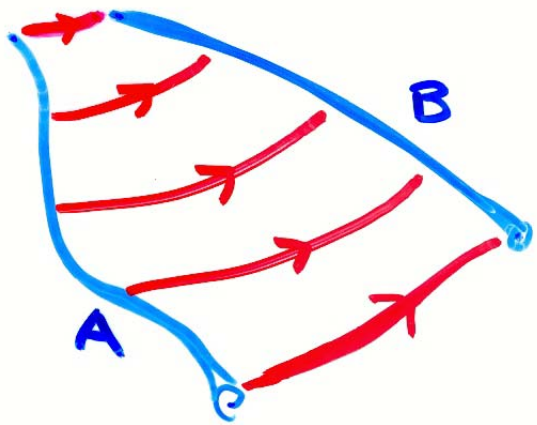


First, write the distance functional *in the same way as before:*

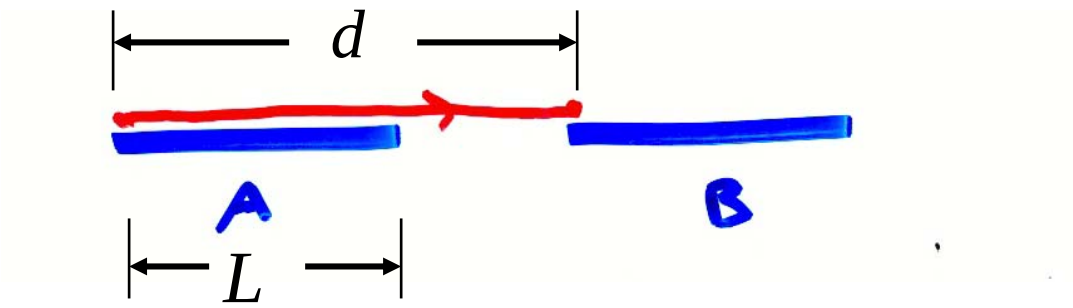
$$D[\mathbf{r}(s,t)] = \int_0^L ds \int_0^T dt \sqrt{\dot{\mathbf{r}}^2}$$

The distance between 2 space curves $\mathbf{r}_A(s),\ \mathbf{r}_B(s)$

is the accumulated 'area' of the minimal transformation $\mathbf{r}^*(s,t)$,a 2-D object, with B.C.s $\mathbf{r}^*(s,0) = \mathbf{r}_A(s),\ \mathbf{r}^*(s,T) = \mathbf{r}_B(s)$



It is not a 3-D “soap-film” area:



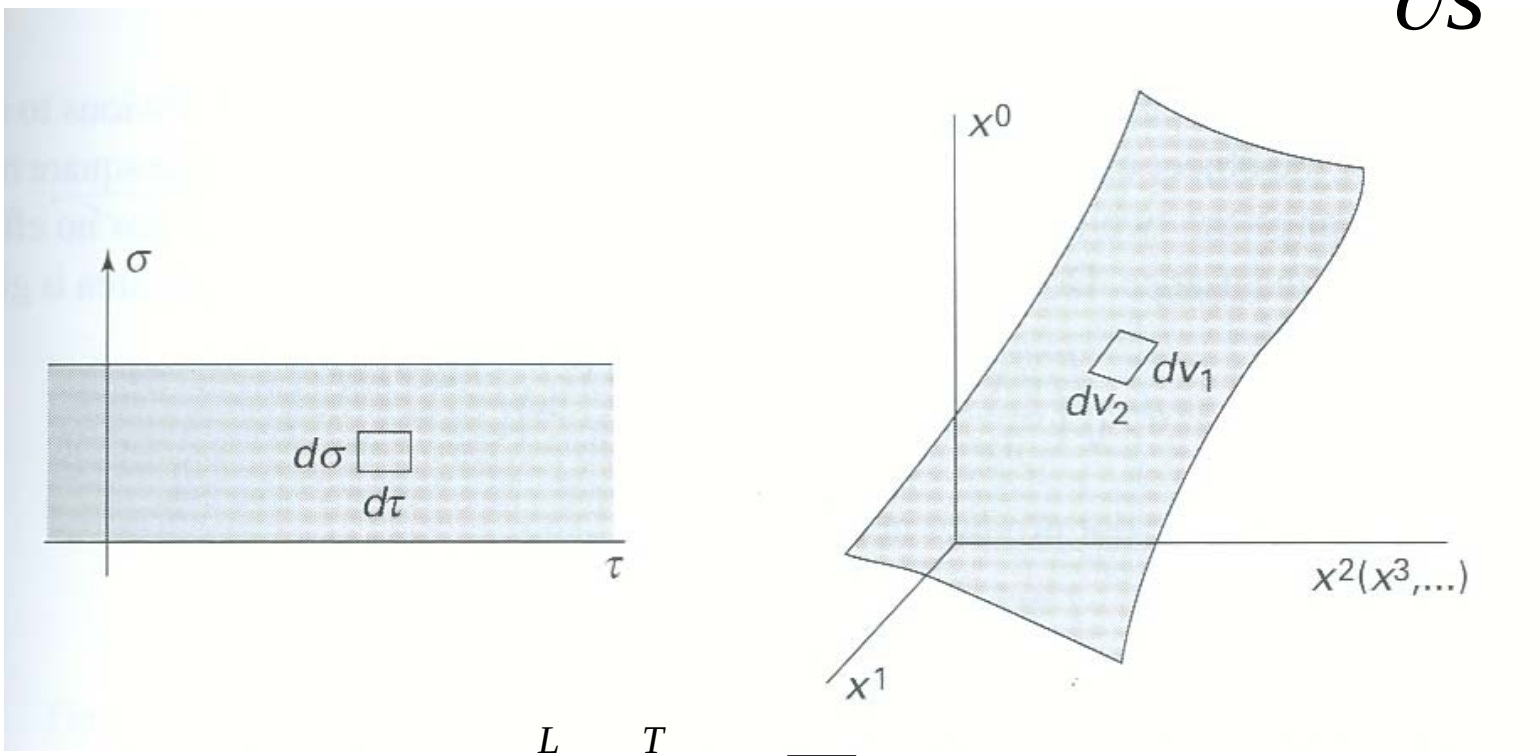
$$A_{soap} = 0$$

$$D_{AB} = Ld$$

Is it a 4-D space-time area, as in classical relativistic string theory?

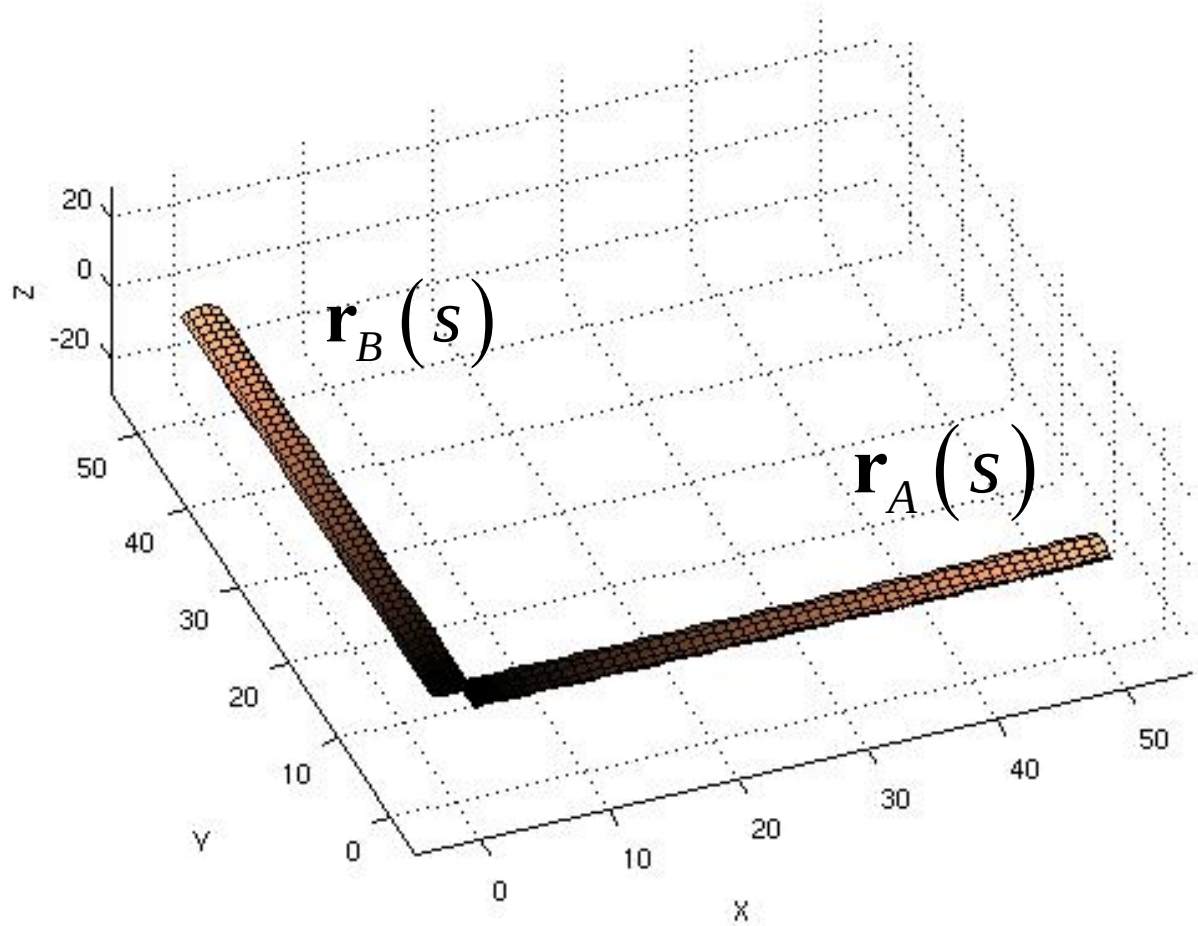
$$S_{Nambu-Goto}[\mathbf{r}(s,t)] = \int_0^L ds \int_0^T dt \sqrt{\dot{\mathbf{r}}^2 \mathbf{r}'^2 - (\dot{\mathbf{r}} \cdot \mathbf{r}')^2}$$

$$\mathbf{r}' \equiv \frac{\partial \mathbf{r}}{\partial s}$$

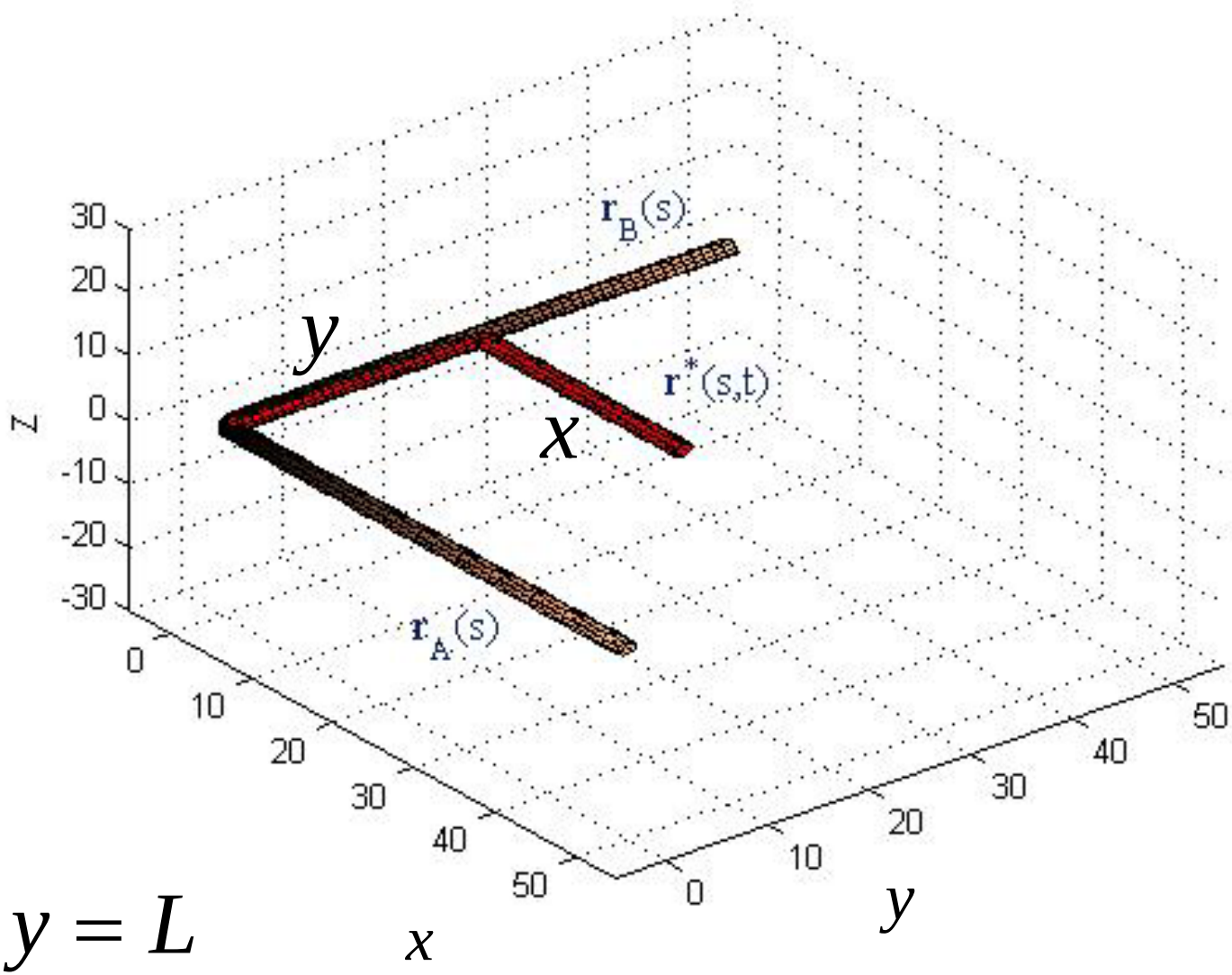


our $$D[\mathbf{r}(s,t)] = \int_0^L ds \int_0^T dt \sqrt{\dot{\mathbf{r}}^2}$$

The “simplest” example:



The minimal transformation is **not** a simple rotation!

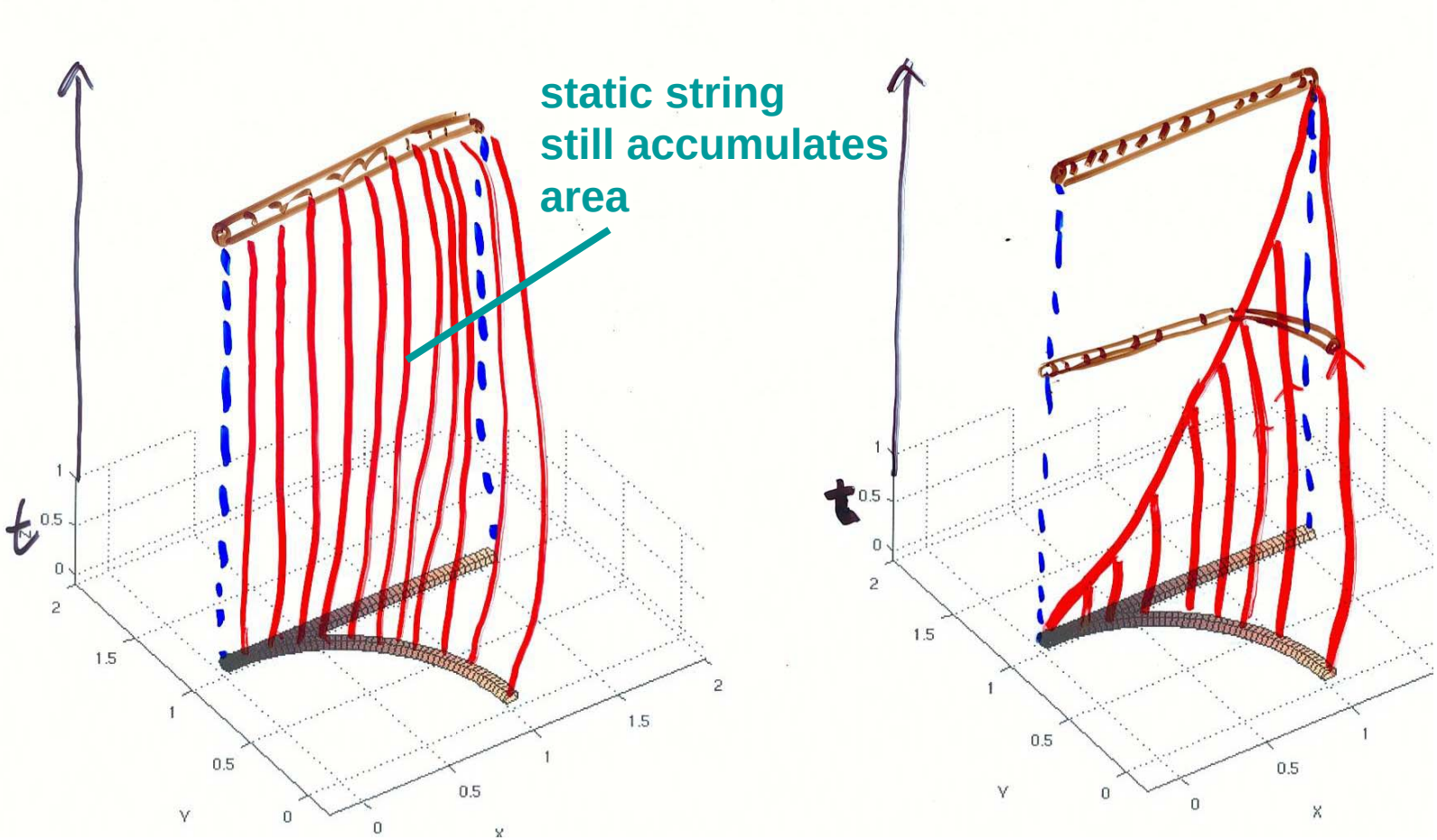


$$x + y = L$$

$$y = -x + L$$

$$D = L^2 \Big/ \sqrt{2}$$

Our action does not map to the world-sheet area of the classical relativistic string.

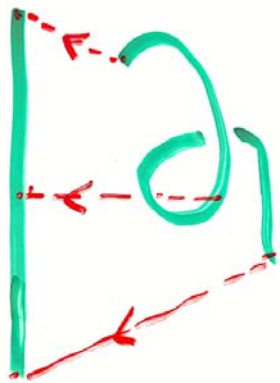


Minimizing the action: $$D[\mathbf{r}(s,t)] = \int_0^L ds \int_0^T dt \sqrt{\dot{\mathbf{r}}^2}$$

results trivially in straight line motion- the MRSD.

Inextensible strings must include the constraint:

$$\sqrt{\mathbf{r}'^2(s,t)} = 1 \qquad \mathbf{r}' = \frac{\partial \mathbf{r}}{\partial s}$$



"rubber band"

Minimize:

$$D[\mathbf{r}(s,t)] = \int_0^L ds \int_0^T dt \left(\sqrt{\dot{\mathbf{r}}^2} - \lambda(s,t) \sqrt{\mathbf{r}'^2} \right)$$

The extremum $\dfrac{\delta L\left[\mathbf{r}(t)\right]}{\delta \mathbf{r}(\tau)} = 0$ gives the Euler-Lagrange equation:

$$\left(\dot{\mathbf{r}}^2\right)\ddot{\mathbf{r}} - \left(\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}\right)\dot{\mathbf{r}} = \left|\dot{\mathbf{r}}\right|^3\left(\lambda\mathbf{r}'' + \lambda'\mathbf{r}'\right)$$

Solve for $\mathbf{r}^*(s,t)$, the minimal transformation.

recast EL eqn:

$$\dot{\hat{\mathbf{v}}} = \lambda\boldsymbol{\kappa} + \lambda'\hat{\mathbf{t}}$$

unit velocity — curvature — tangent

whether an extremal transformation is a minimum can be determined from the second variation:

$$\delta^2 D = \frac{1}{2}\int_0^L ds \int_0^T dt\left(\delta\dot{\mathbf{r}}\cdot\mathbf{I}\cdot\delta\dot{\mathbf{r}} + \delta\mathbf{r}'\cdot\boldsymbol{\Lambda}\cdot\delta\mathbf{r}'\right) > 0$$

B.C.s at t=0,T are given by $\mathbf{r}^*(s,0)=\mathbf{r}_A(s),\quad \mathbf{r}^*(s,T)=\mathbf{r}_B(s)$

String ends are free at s=0,L, so conjugate momenta must vanish:

$$\therefore \mathbf{p}_s = \left.\frac{\partial L}{\partial \mathbf{r}'}\right|_{s=0,L} = 0$$

but for our $$L=\sqrt{\dot{\mathbf{r}}^2}-\lambda(s,t)\sqrt{\mathbf{r}'^2}$$

$$\mathbf{p}_s = \lambda\hat{\mathbf{t}} = 0$$

$$\therefore \quad \lambda(0,t)=\lambda(L,t)=0$$ (string tension vanishes)

Boundary conditions

But then EL eqn $\dot{\hat{\mathbf{v}}} = \lambda\boldsymbol{\kappa} + \lambda'\hat{\mathbf{t}}\Big|_{\text{end pts}}$ simplifies to $\dot{\hat{\mathbf{v}}} = \lambda'\hat{\mathbf{t}}$

And since $\dot{\hat{\mathbf{v}}}$ is $\perp$ $\hat{\mathbf{v}}$

$\therefore \quad \lambda'\hat{\mathbf{v}}\cdot\hat{\mathbf{t}} = 0$

so either $\mathbf{v}\cdot\hat{\mathbf{t}} = 0$ -> pure **rotation**

or $\lambda' = 0$ and therefore $\dot{\hat{\mathbf{v}}} = 0$ -> straight line motion

Also trivial soln $\mathbf{v} = 0$

For numerical solutions, discretize the chain (method of lines)
EL PDE becomes a set of coupled ODEs

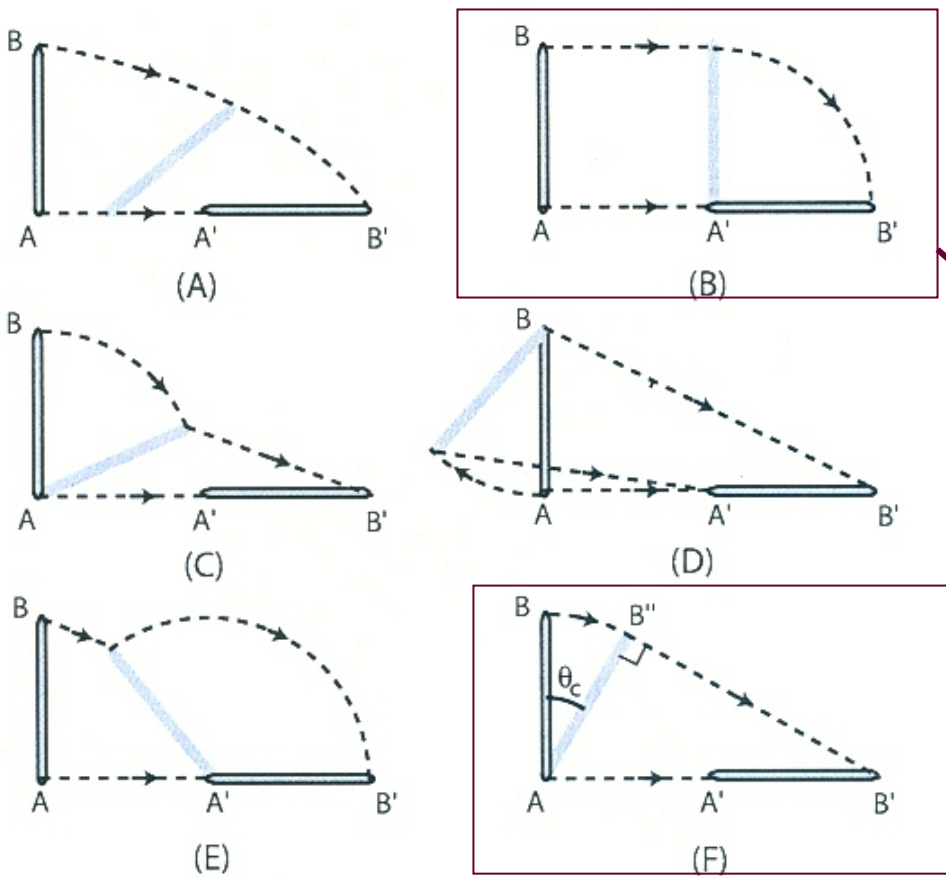
Similar to the analysis of B.C.s, the ODEs for particle i have solutions:

-> pure **rotation**

-> straight line motion

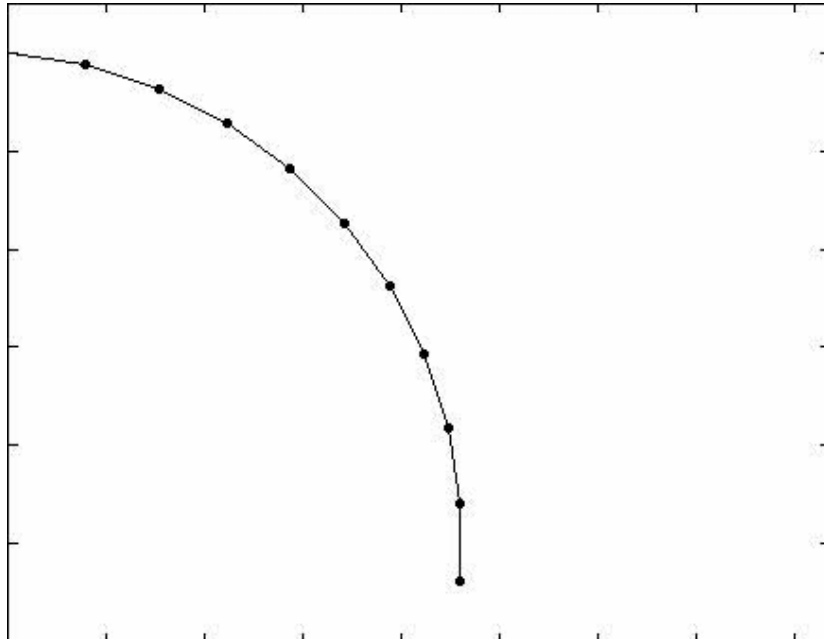
-> trivial soln $\mathbf{v} = 0$

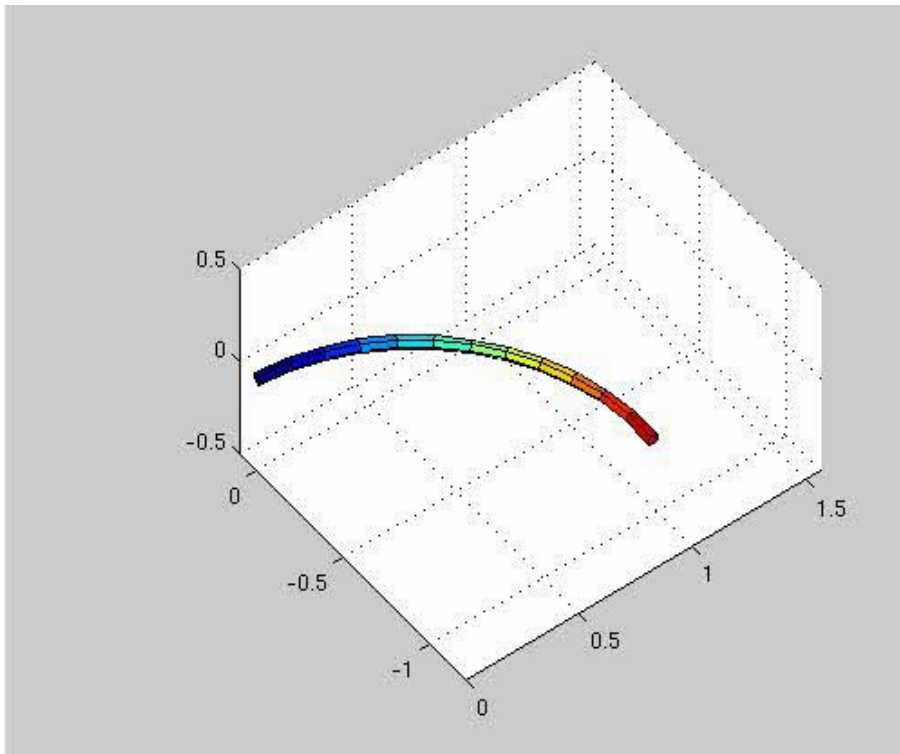
Simplest case:



Weierstrass-Erdmann corner conditions

Mohazzab, Plotkin




0.5
0
-0.5
0
-0.5
-1
0
0.5
1
1.5

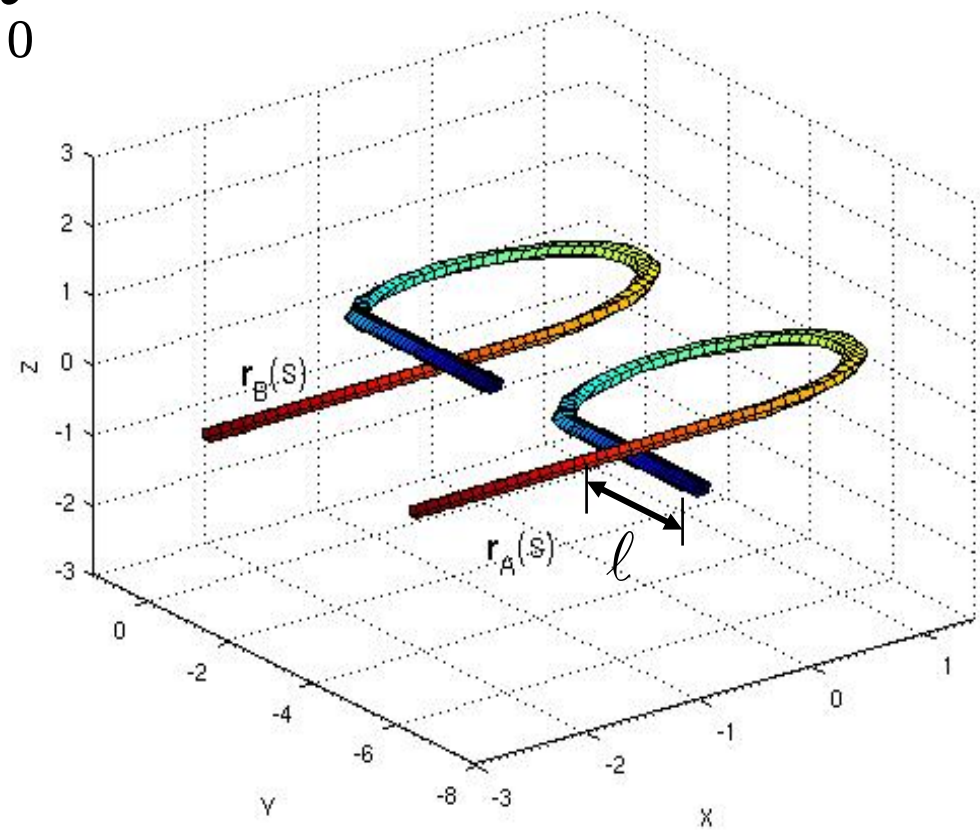
To apply the theory, additional terms may be included in the Lagrangian:

- Curvature constraints – persistence length
- Non-crossing - Edwards constraint:

$$V_{NC}[\mathbf{r}(s,t)] = \int_0^L ds_1 \int_0^L ds_2 \; \delta\left(\mathbf{r}(s_1,t) - \mathbf{r}(s_2,t)\right)$$

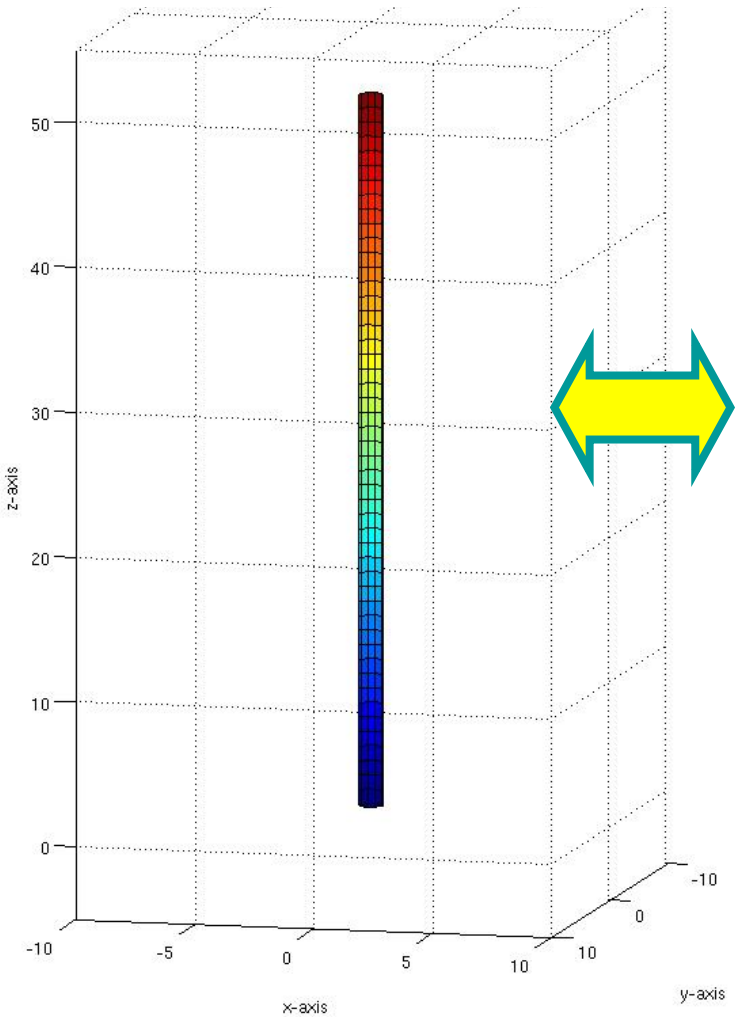
$$D_{RMSD} \approx 0$$

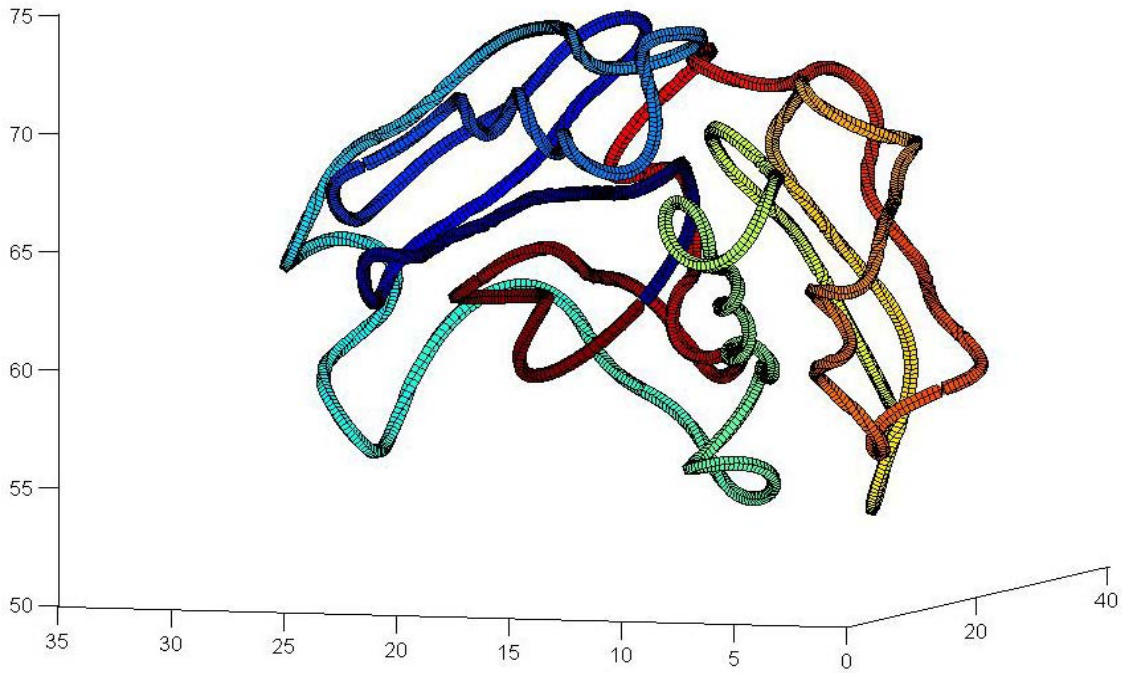
$$D = \ell^2$$



Applications to proteins.
Structures with D_{MIN} , D_{MAX}

Does $\langle Dk_F \rangle$?
Does $\langle D_{iN} p_F^{(i)} \rangle$?





Research opportunities!

in the Plotkin Group

1 PDF

1 Grad student