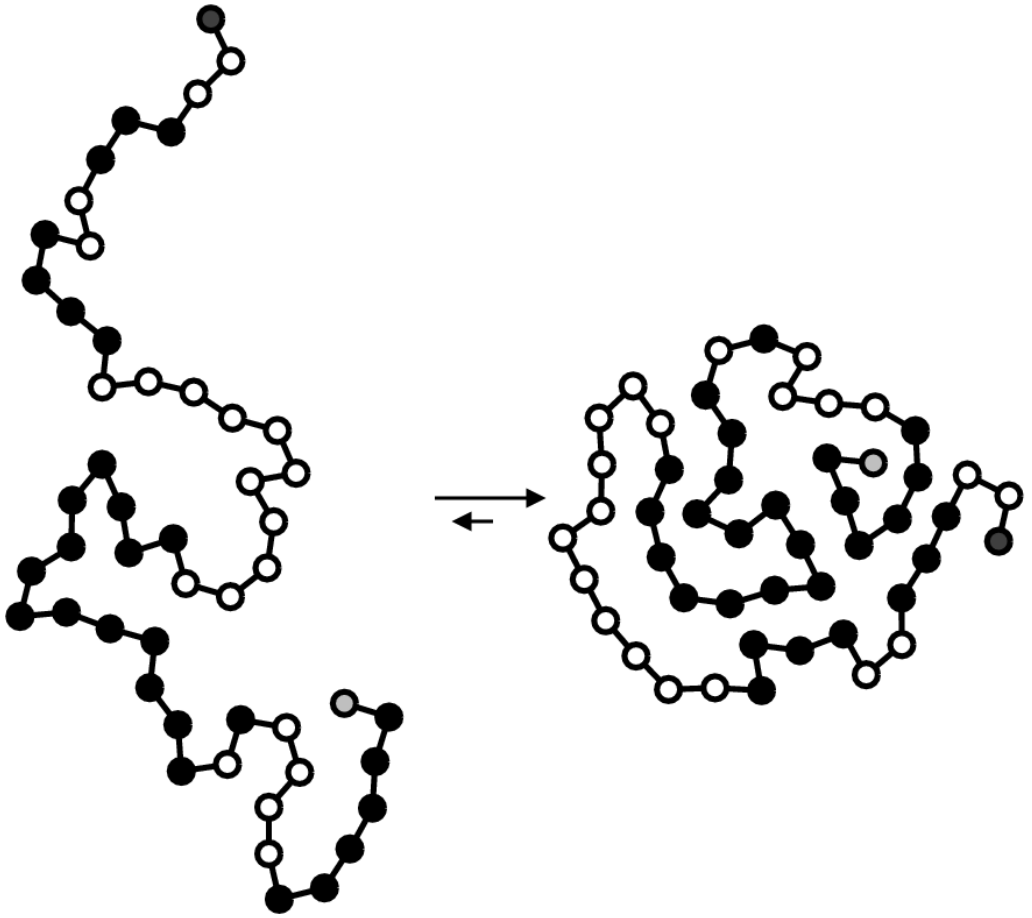
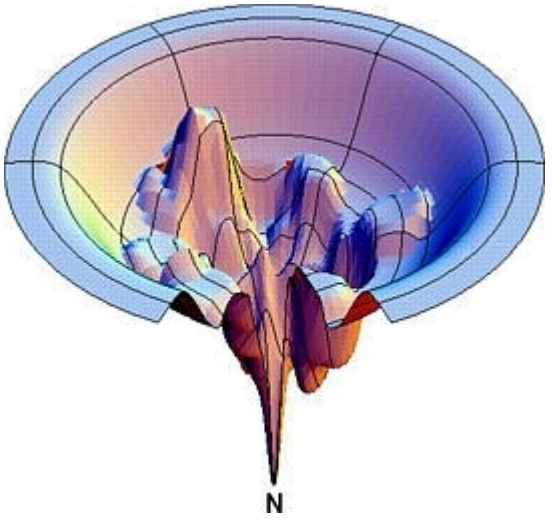
Wintersemester 2016/2017
Biomolecular Engineering/Nanobiophysics
Module

LECTION 4: PROTEIN FOLDING



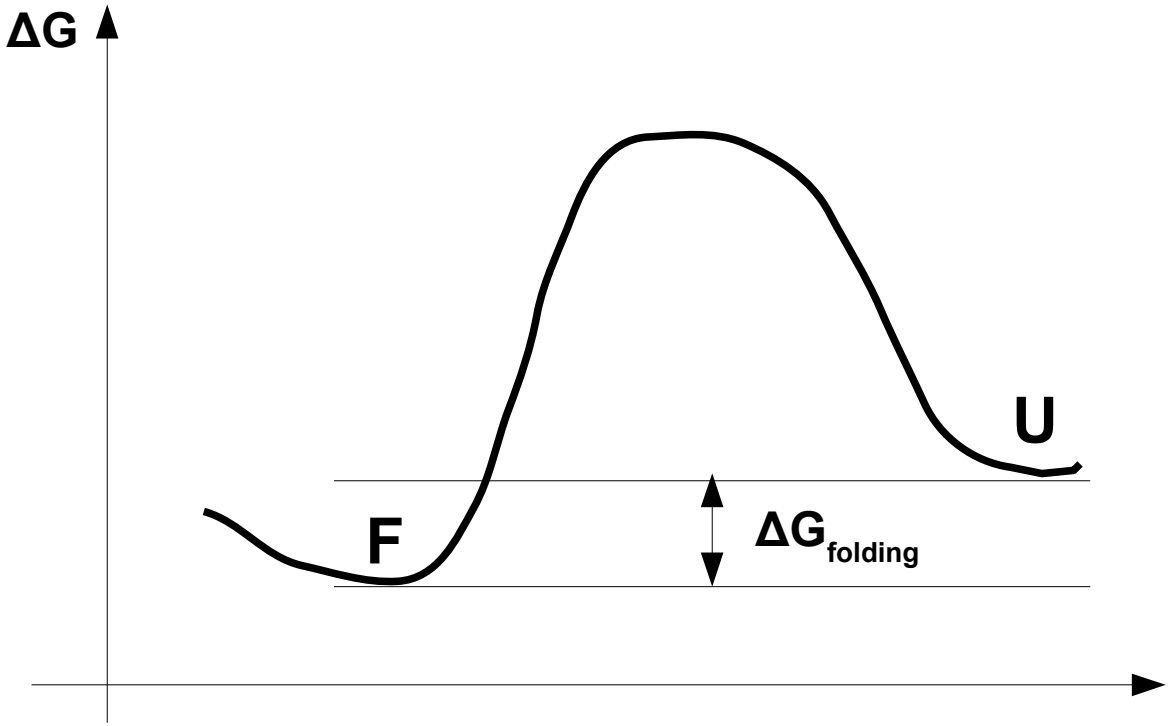


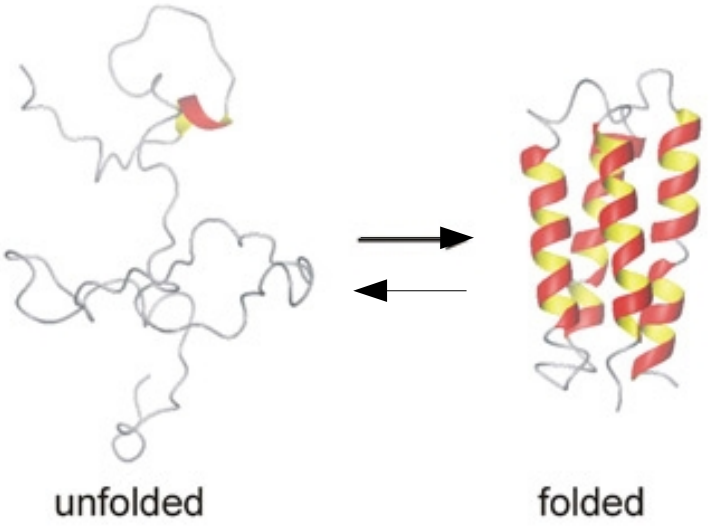
LECTION 4: PROTEIN FOLDING

- Protein folding and denaturation
- Anfinsen's dogma
- Folding of α- and β-secondary structures
- Hydrophobic effect
- Levinthal's paradox
- Protein *vs* other polymers
- Thermodynamics and kinetics of protein folding
- Protein folding in the cell
- Computational approaches to study folding
- CASP
- Case study: folding of an interdomain linker

FOLDING AND DENATURATION

- Protein folding is the process by which a protein structure assumes its native (functional) conformation.
- Protein denaturation is the process by which a protein loses its native conformation.

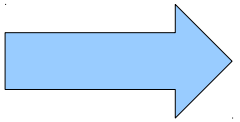




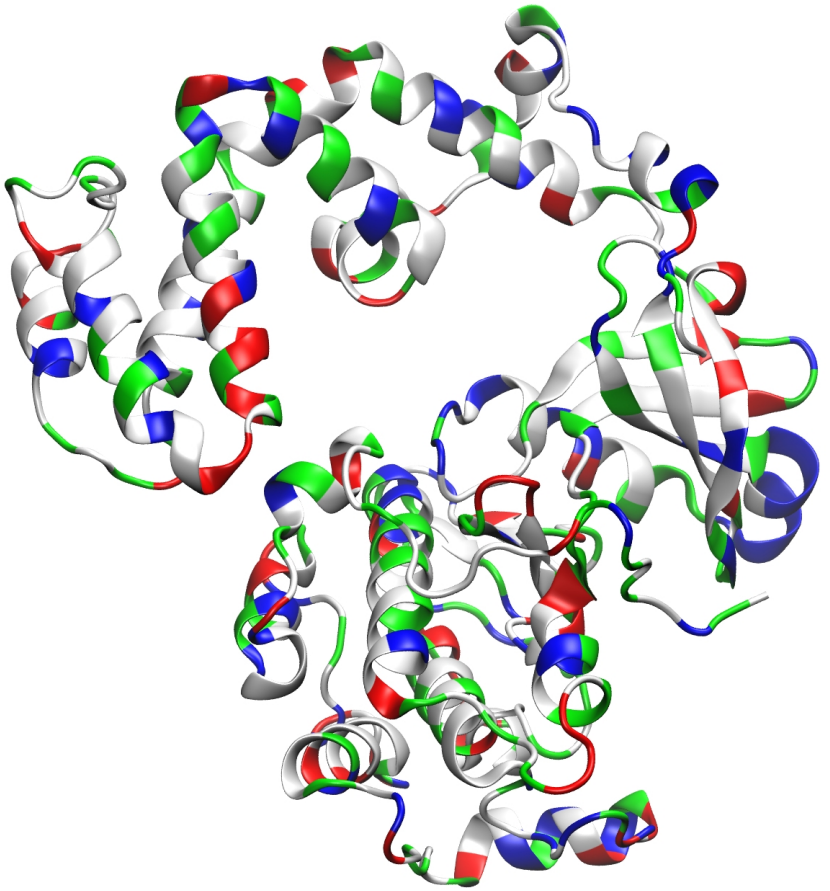
FOLDING: 1D=>3D

AA sequence

```
>FASTER_FORMAT_SEQUENCE
MDFGSLETVVANSAFIAARGSFDASSGPASRDRKYLARLKLPPLSKCEALR
ESLDLGFEGMCLEQPIGKRLFQQFLRTHEQHGPALQLWKDIEDYDTADDAL
RPQKAQALRAAYLEPQAQLFCSFLDAETVARARAGAGDGLFQPLLRAVLAH
LGQAPFQEFLDSLYFLRFLQWKWLEAQPMGEDWFLDFRVLGRGGFGEVFAC
QMKATGKLYACKKLNKKRLKKRKGYQGAMVEKKILAKVHSRFIVSLAYAFE
TKTDLCLVMTIMNGGDIRYHIYNVDEDNPGFQEPRAIFYTAQIVSGLEHLH
QRNIIYRDLKPENVLLDDDGNVRISDLGLAVELKAGQTKTKGYAGTPGFMA
PELLLGEEYDFSVDYFALGVTLYEMIAARGPFRARGEKVENKELKQRVLEQ
AVTYPDKFSPASKDFCEALLQKDPEKRLGFRDGSCDGLRTHPLFRDISWRQ
LEAGMLTPPFVPDSRTVYAKNIQDVGAFSTVKGVAFEKADTEFFQEFASGT
CPIPWQEEMIETGVFGDLNVWRPDGHHHHHH
```



Protein structure

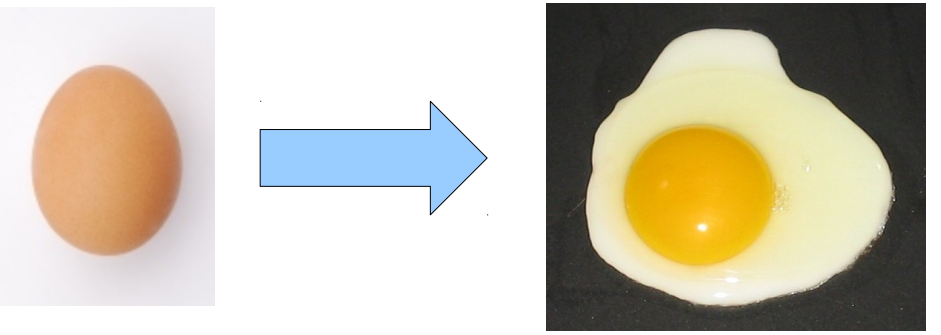


- In general — NOW — computationally unfeasible task

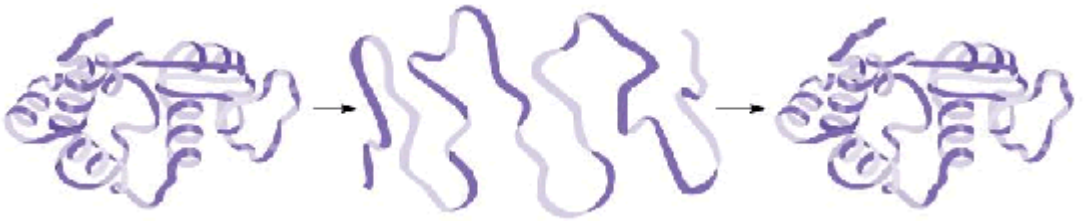
DENATURATION

- **Caused by:**
  - temperature (high and low)
  - pH
  - ionic strength
  - pressure

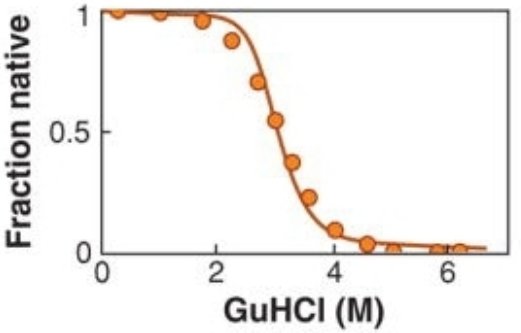
$$\Delta G = \Delta H - T\Delta S$$

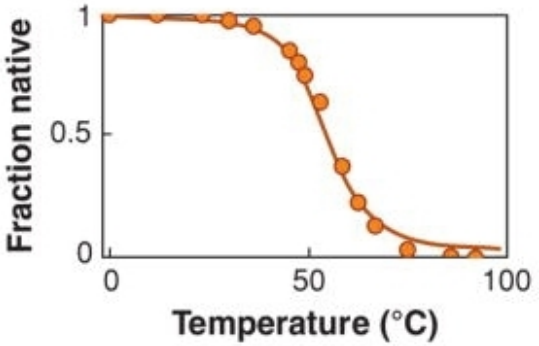


- **Reversibility:**
  - reversible
  - irreversible



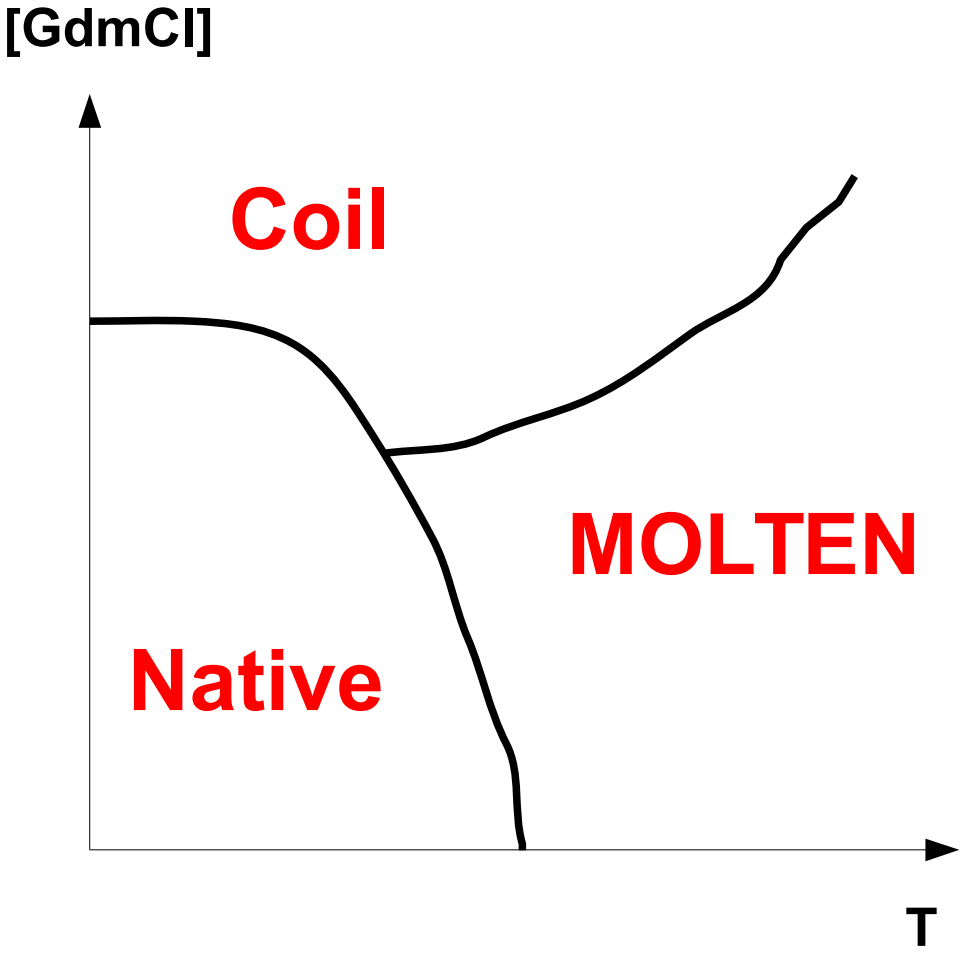
- **Structural levels affected:**
  - Tertiary
  - Secondary
  - * S-S bridges
- **Structural levels not affected:**
  - Covalent bonds
  - Primary structure





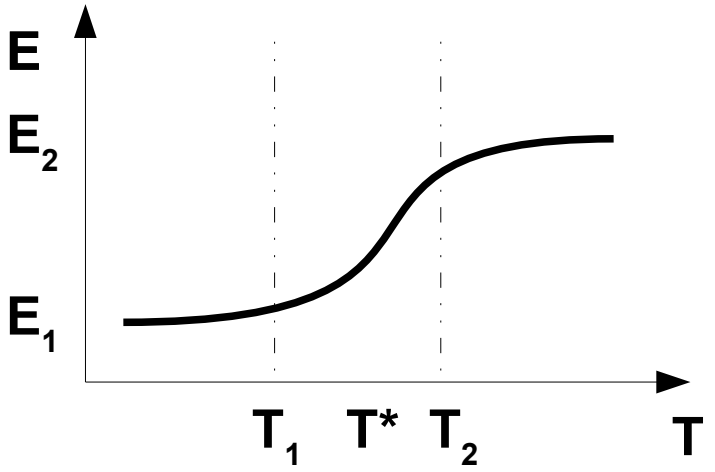
PROTEIN STATES

- **Native:**
  - native function
  - accurately defined contacts/residus packing (CD, fluorescence)
  - fixed H-exchange rate (NMR)
- **Molten globule:**
  - still compact but no unique packing
  - hydrophobic core fluctuates
  - secondary structure, S-S bonds
  - increased mobility of side-chains
  - partial solvent accessibility of Trp
  - absence of some «remote» contacts
- **Coil:**
  - function is lost
  - most of contacts are lost
  - secondary structure partly affected

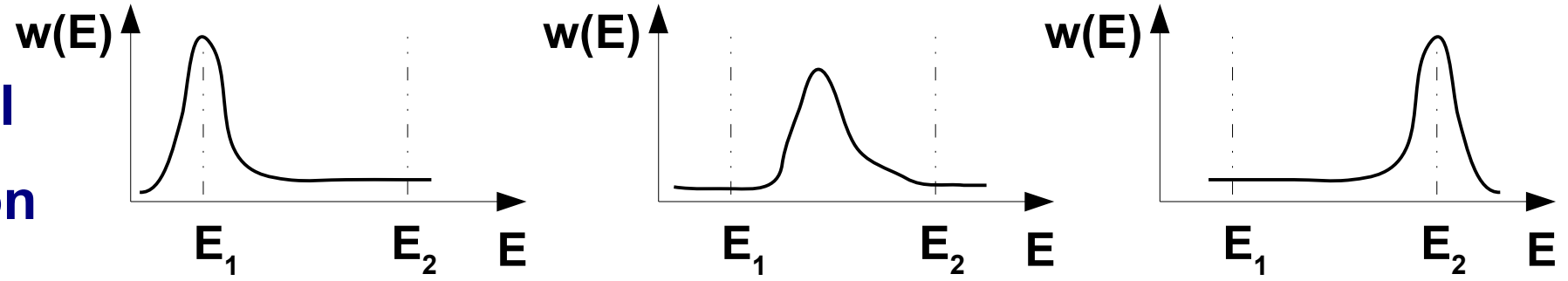


TRANSITION: «ALL OR NOTHING»

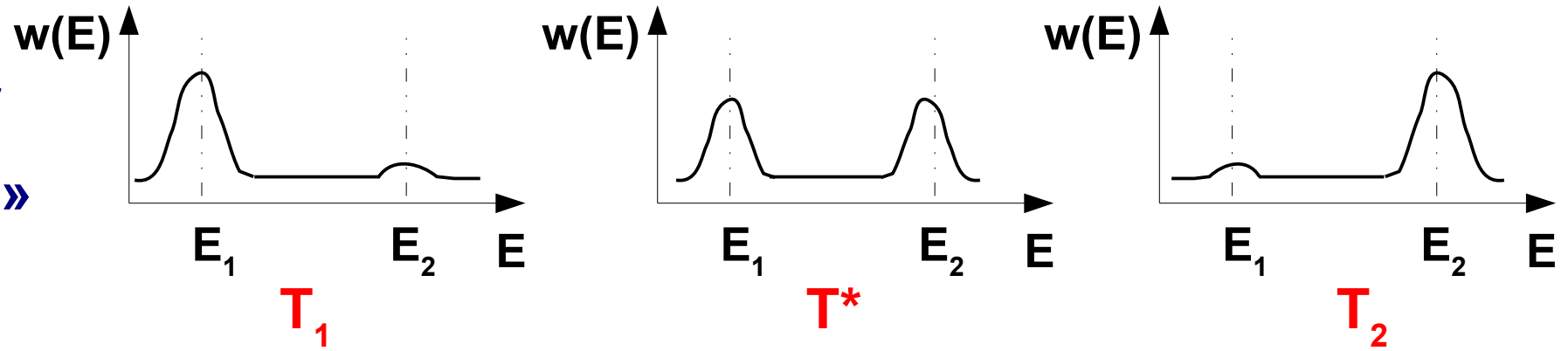
- Protein changes from not native to native state by dramatic change of properties: heat capacity, single molecule energy distribution etc.



Gradual transition

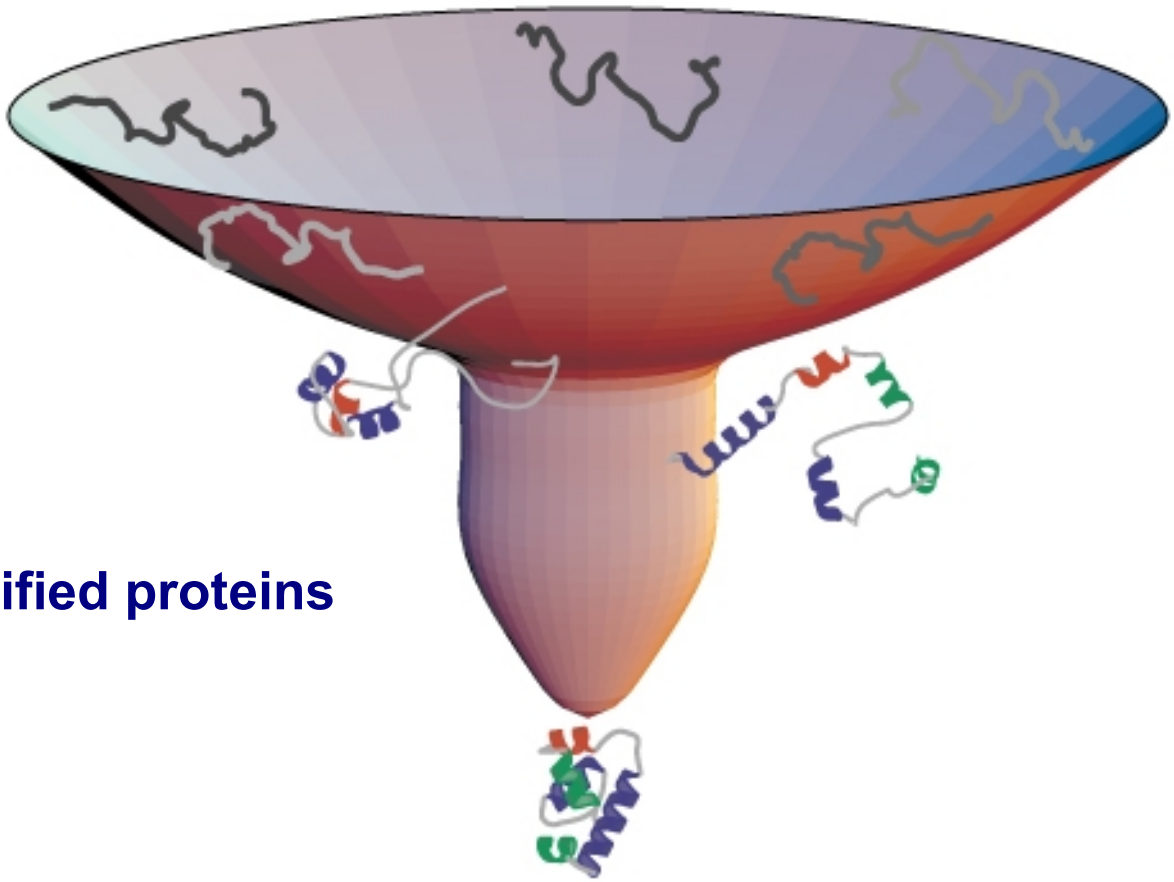


«All or nothing»



ANFINSEN'S DOGMA

- Protein's native structure is determined by its aa sequence and it is:
  - unique
  - stable
  - kinetically accessible



➢ Anfinsen's experiment (1957):
- Bovine ribonuclease A
- Denaturation
- Works for not covalently modified proteins
- $\Delta G_{folding}$~1-10 kcal/mol

LEVINTHAL'S PARADOX

- How much time needs protein to be folded?

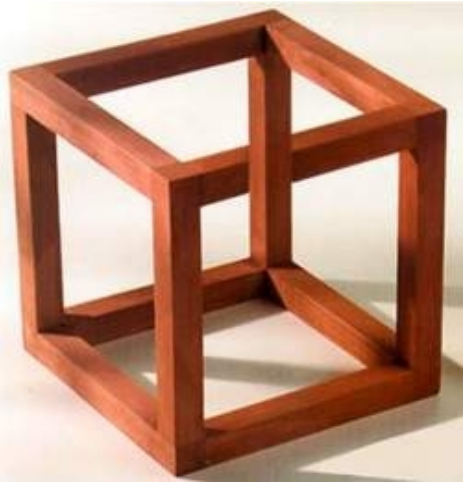
  - Every amino acid ~ 10 conformations

  - 100-aa polypeptide ~ 10^{100} conformations

  - Time of 1 conformational change ~ 10^{-13} s

  - 10^{80} years needed, Universe life ~ 10^{10}

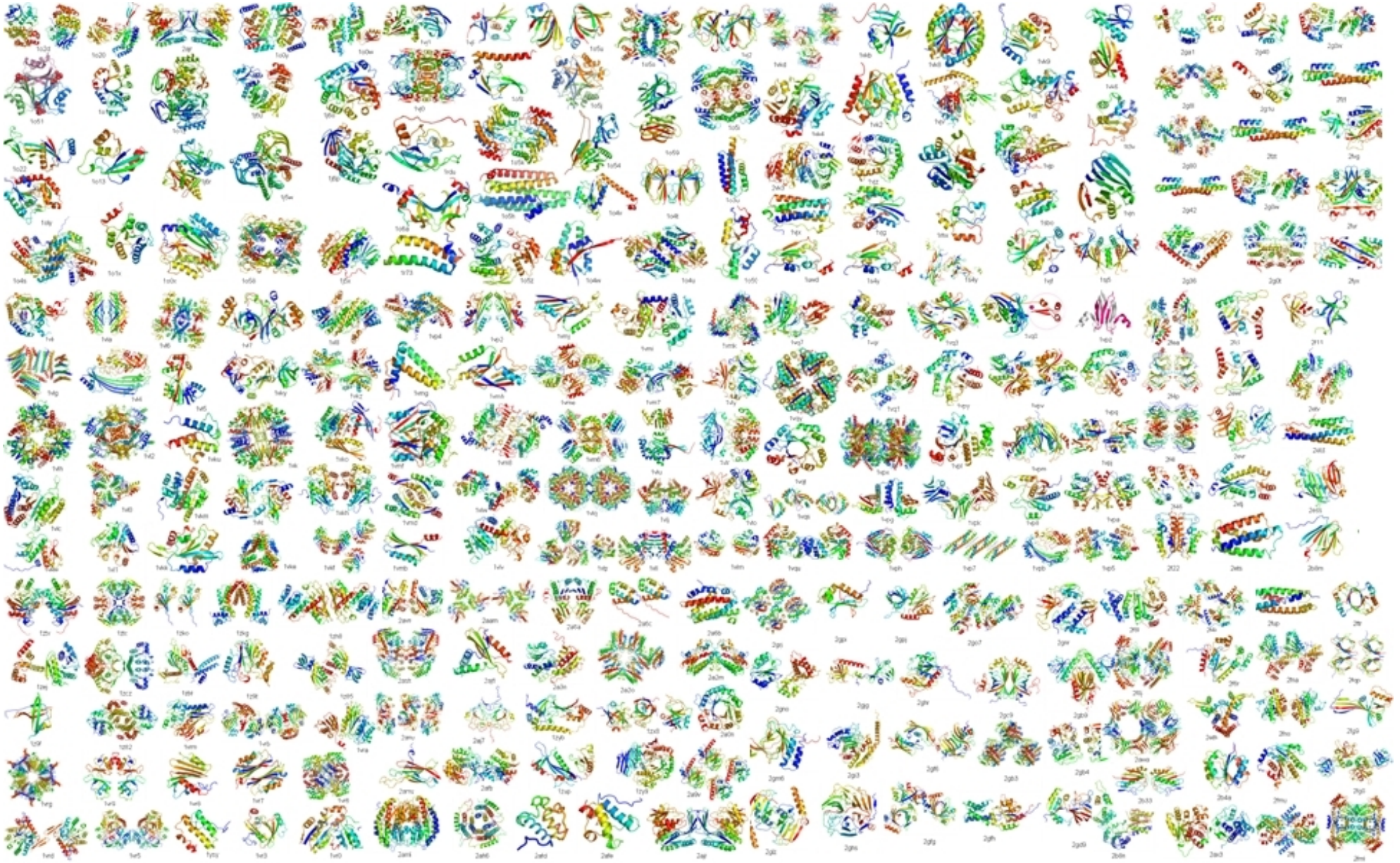
  - In reality: folding time ~ minutes-days



- Is the native conformation the most stable???

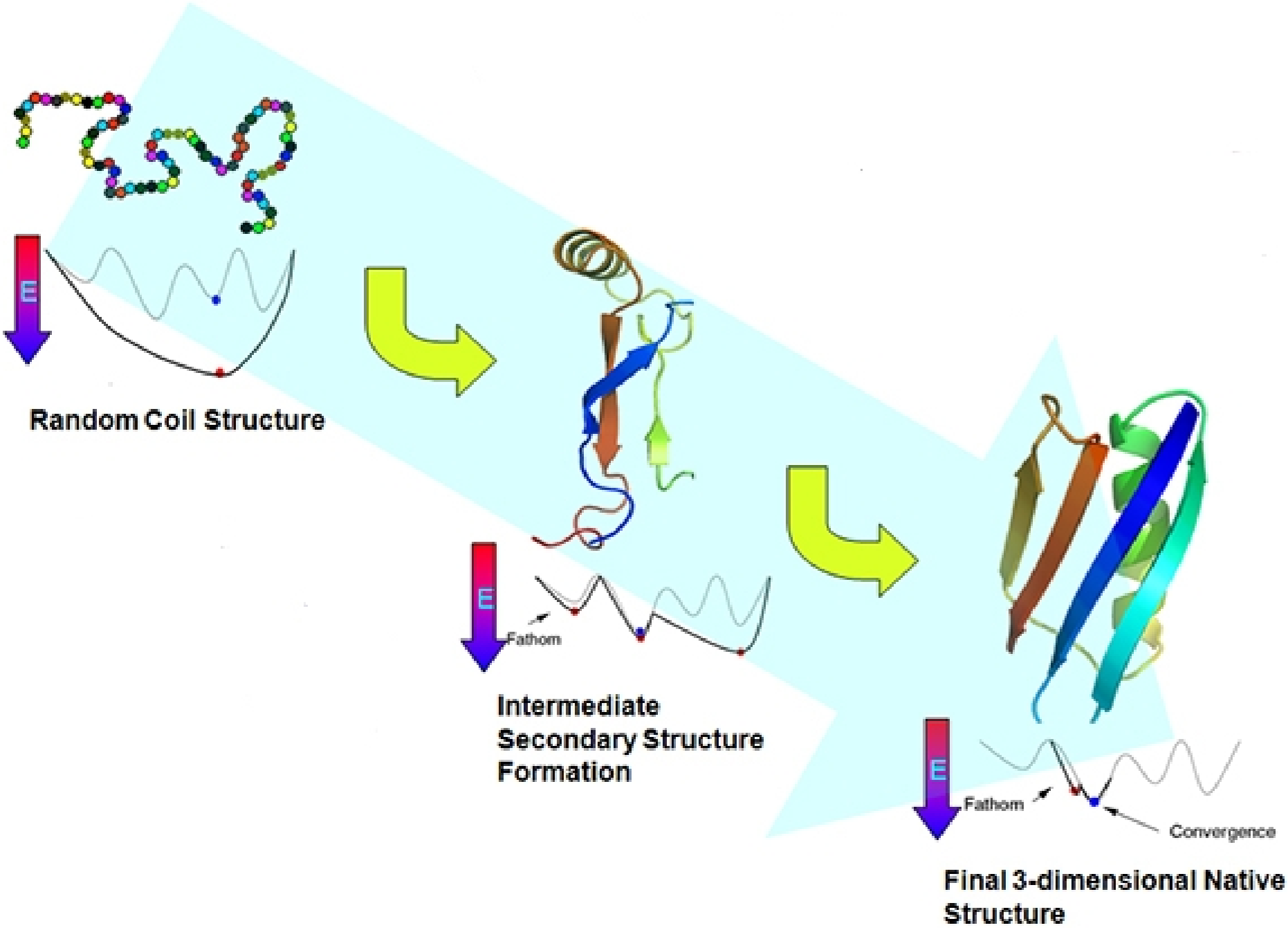
- Kinetics or thermodynamics???

MANY SEQUENCES/FEW FOLDS



SCOP (2008): 1393 folds

FRAMEWORK MODEL



α-HELIX ORGANIZATION

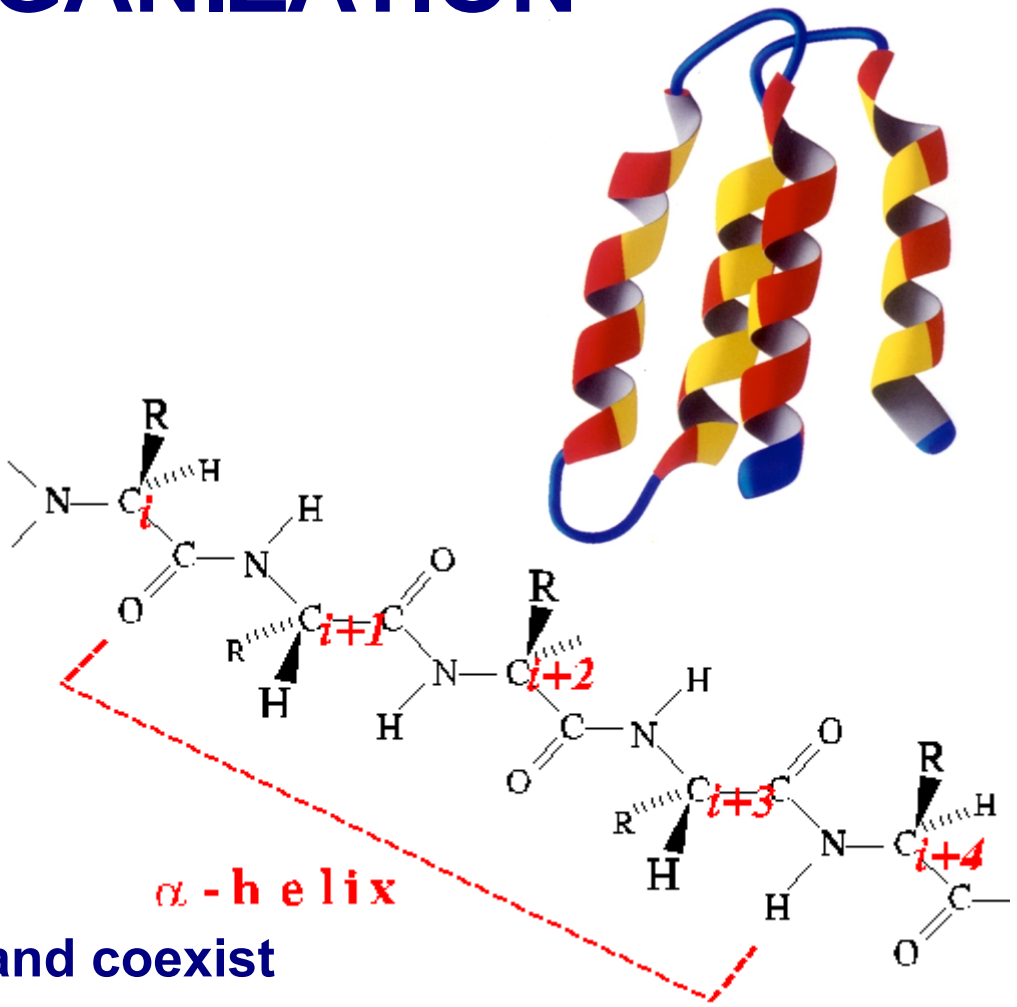
$$\delta G_\alpha = G_\alpha - G_{coil} = (n-2) f_{HB} - nTS_\alpha$$

$$g_{init} = -2g_{HB} \qquad g_{elong} = g_{HB} - TS_\alpha$$

$$\delta G_\alpha = g_{init} + n\, g_{elong} \Rightarrow n_0$$

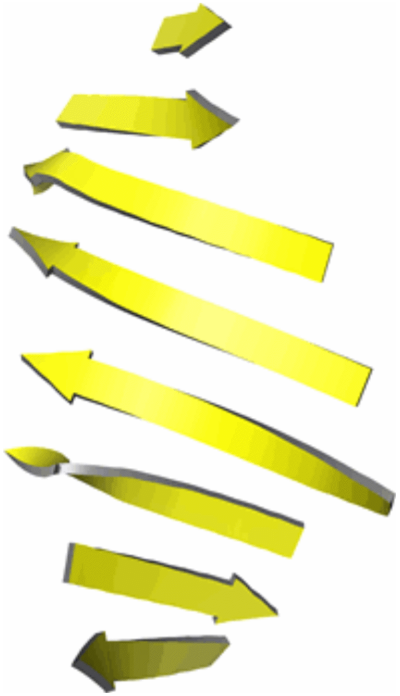
$$K = e^{\frac{-\delta G_\alpha}{kT}} = e^{\left(\frac{-g_{init}}{kT}\right)} \left(e^{\left(\frac{-g_{elong}}{kT}\right)}\right)^n = \sigma S^n$$

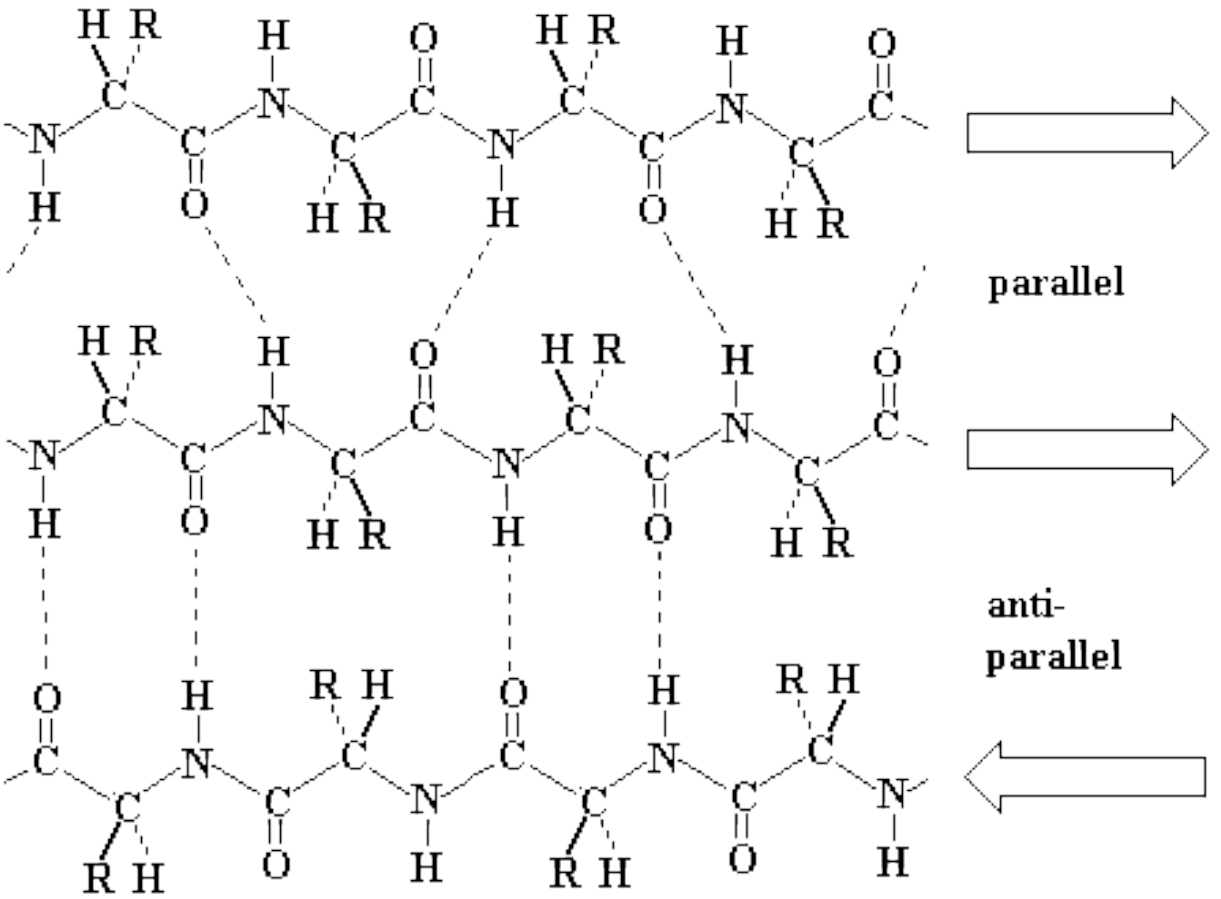
$$\sigma \ll 1$$



- 1D transition, both phases are stable and coexist
- n_0 = f(aa) ≈ 30; g_{init} ≈ 4 kcal/mol; σ ≈ 0.001; g_{HB} ≈ TS_α ≈ -2 kcal/mol
- ↑T => ↓ helix content (helicity)
- $Time_{initiation}$ ≈ $Time_{full\ elongation}$
- Rate ≈ 1 aa/ns

β-SHEET ORGANIZATION

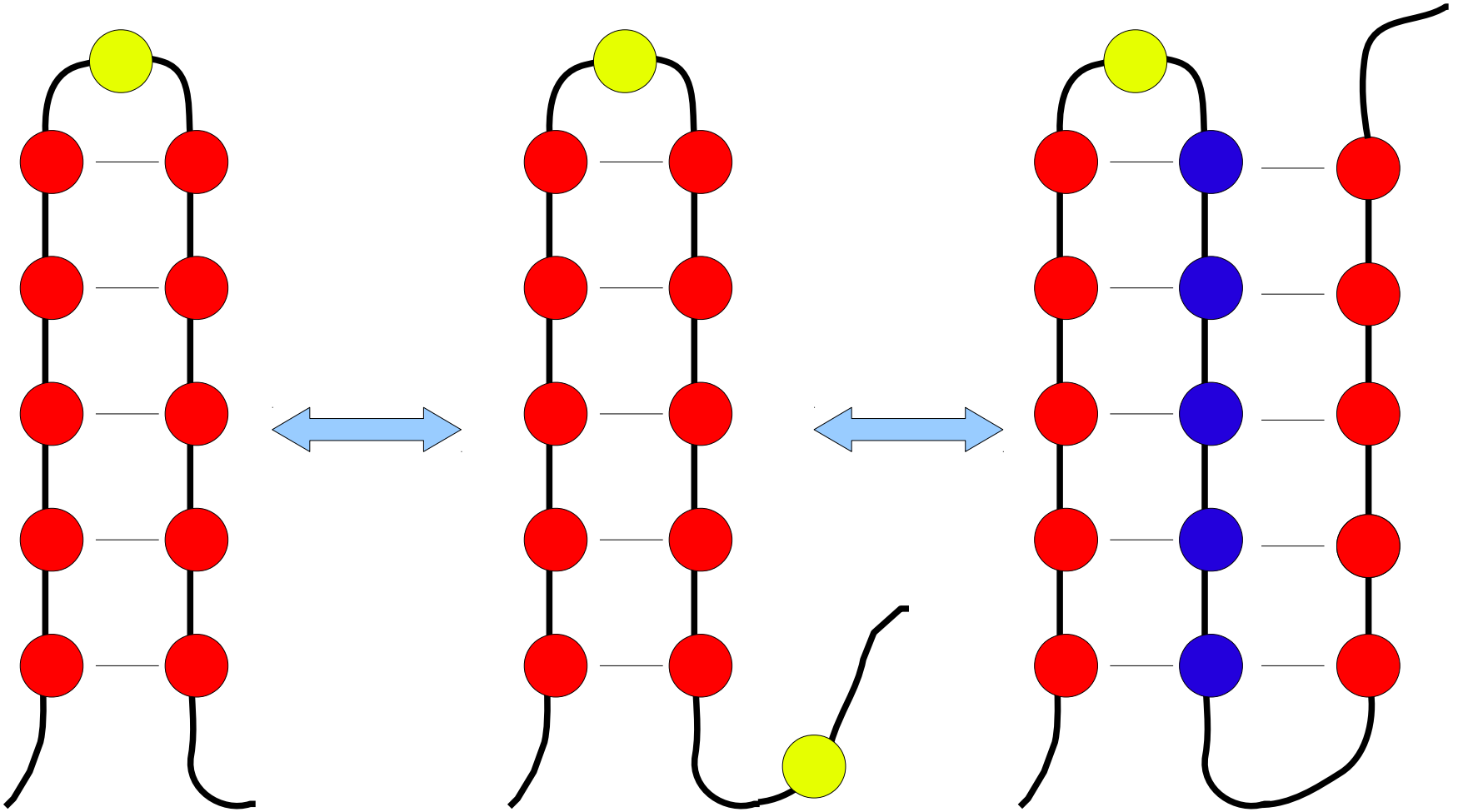




1D ⟶ 2D

- β-sheet assembles much slower than α-helix (~ms-hours or weeks).

β-SHEET ORGANIZATION



Unstable

- External β-layers have higher energies than internal.
- II type of transition.
- High energetical barrier.

β-SHEET ORGANIZATION

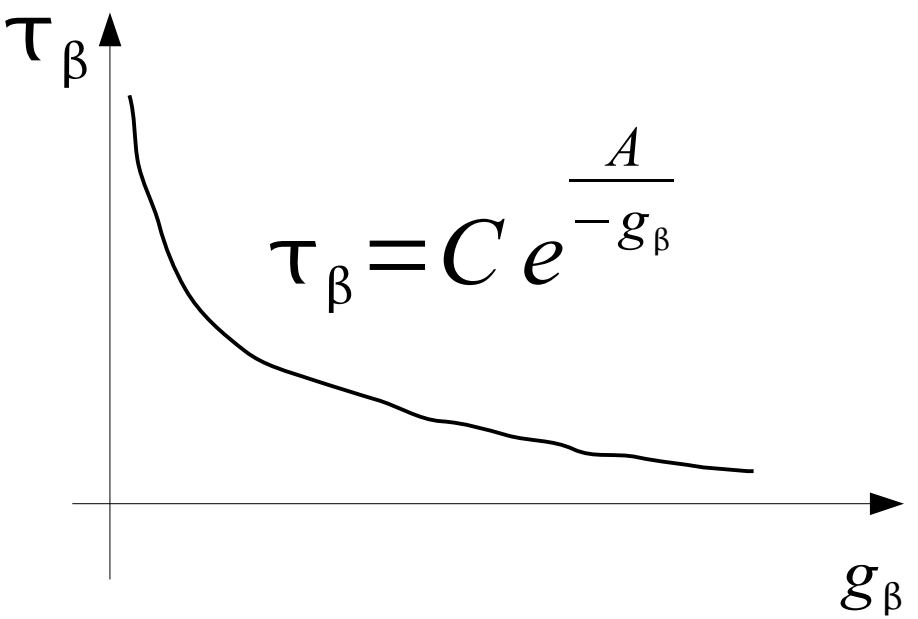
$g_\beta + \delta g_\beta$ – *energy of a residue within* β – *sheet* + *at the border*

$g_\beta + \delta g_\beta < 0$ – β – *hairpin is stable*

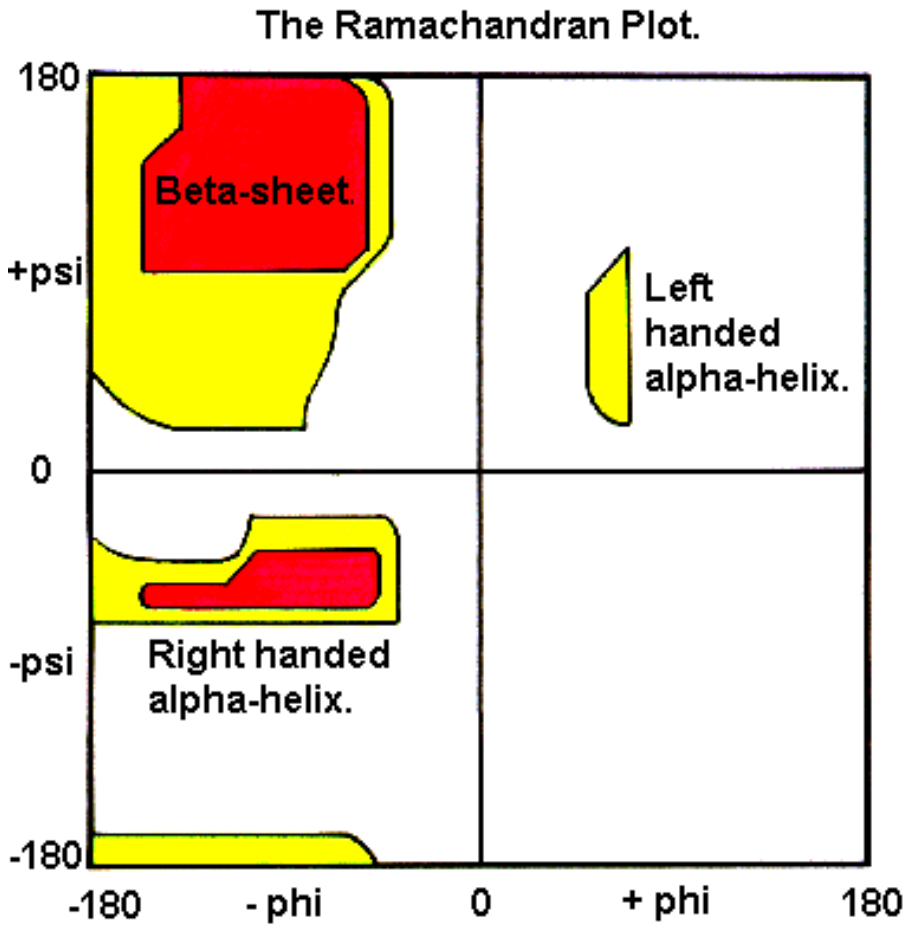
$g_\beta + \delta g_\beta > 0$ – β – *hairpin is unstable*

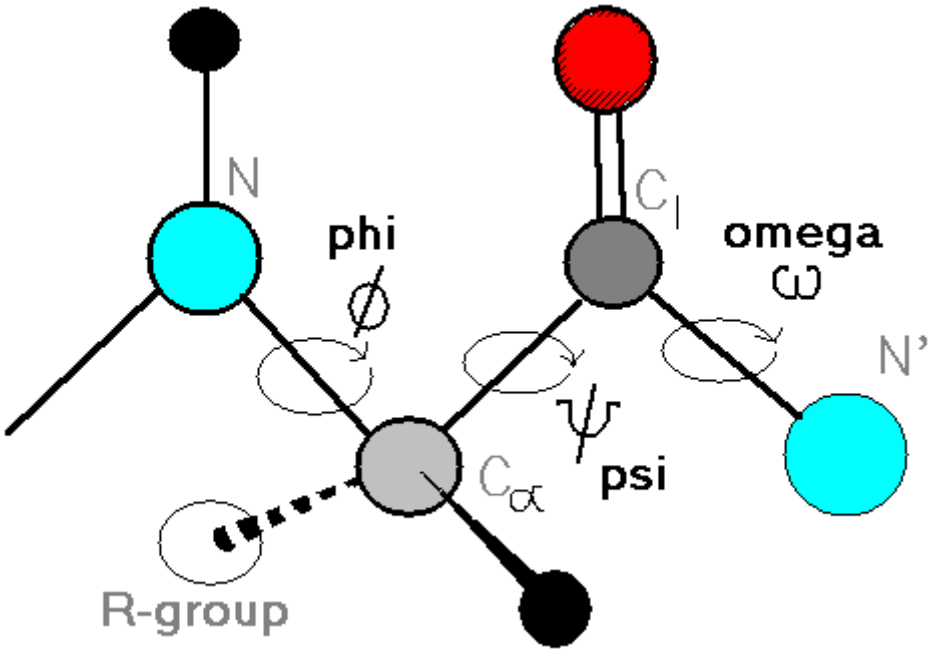
$N_{min} = \dfrac{g_\beta + \delta g_\beta}{-g_\beta}$ – *Minimum number of residues required*

- **No equilibrium with unfolded structure.**
- **Initiation is a limiting stage.**
- **Stable hairpins form as fast as α-helices.**
- **$V \sim N^{3/2}$**
- **$\langle n(\alpha) \rangle = 11$; $\langle n(\beta) \rangle = 6$**



SECONDARY STRUCTURE PROPENSITIES



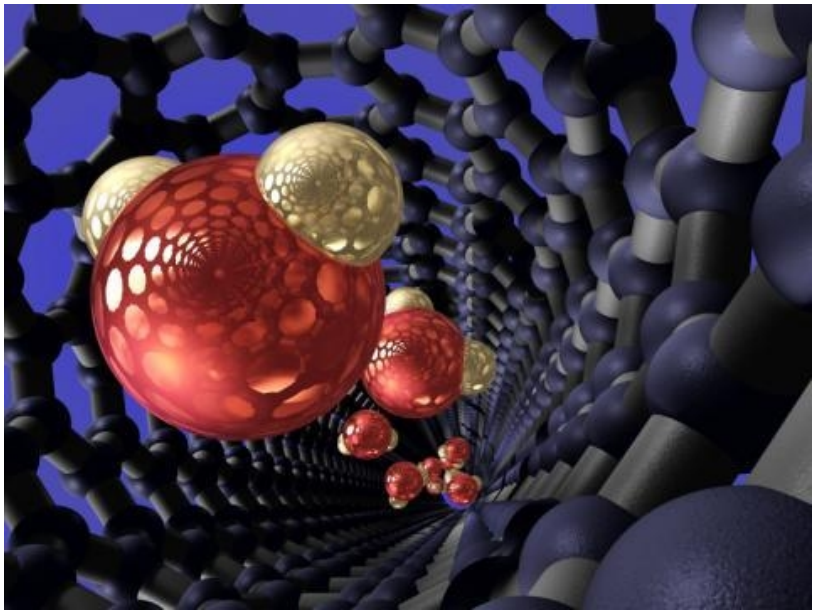


- α-forming: Met, Ala, Leu, Glu, Gln, Lys.
- β-forming: Thr, Ile, Val, Phe, Tyr, Trp.
- Disorder-supporting: Gly, Ser, Pro, Asp, Asn.
- Indifferent to secondary structure: His, Arg, Cys.

FOLDING DRIVING FORCES

ΔG=ΔH-TΔS

- **Enthalpy**
- **Entropy**



Protein

+

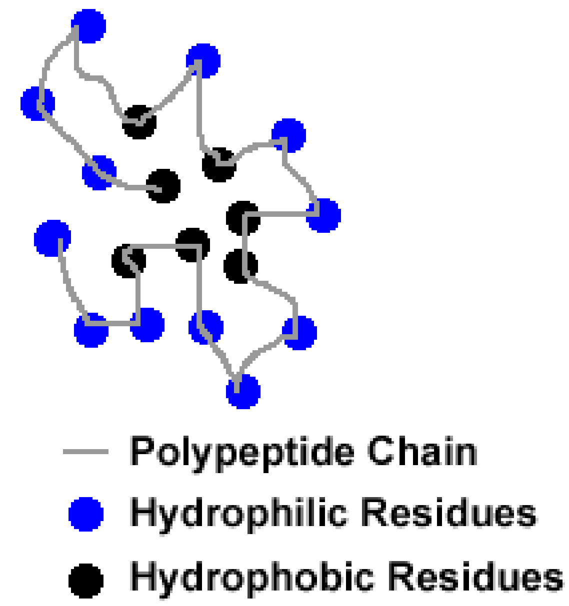
Solvent

$$V(\vec{r}) = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) + \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + \sum_{i<j}^{atoms} \frac{q_i q_j}{\epsilon R_{ij}}$$

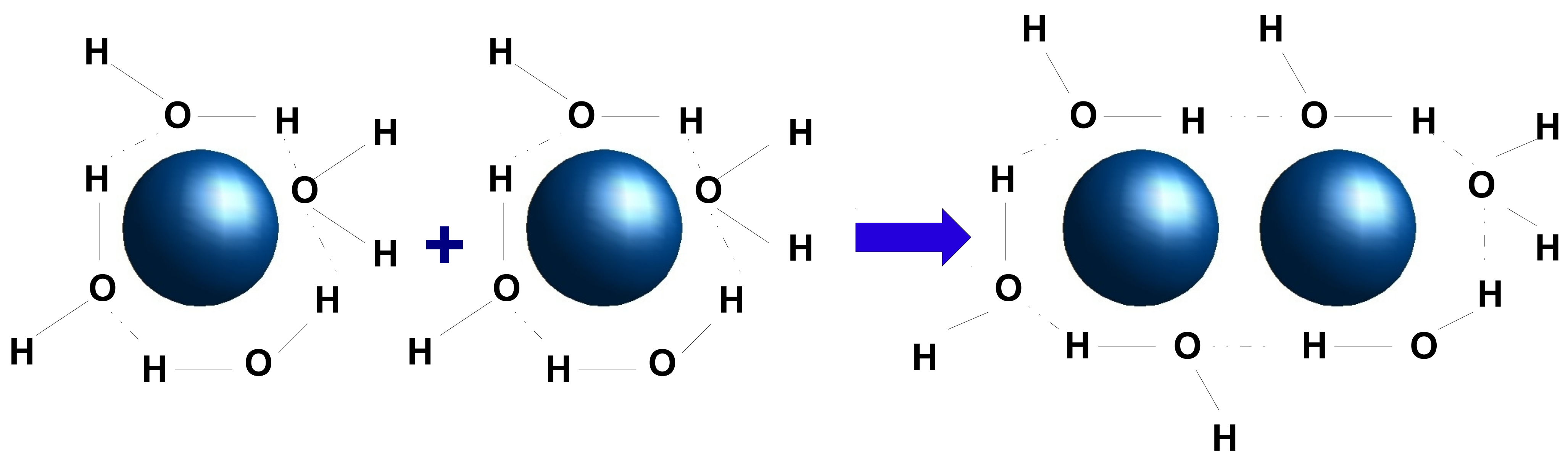
- **Electrostatic potential:**
  - **H-bonds**
  - **Salt bridges**
- **Lennard-Jones potential**
  - **vdW contacts**

HYDROPHOBIC EFFECT

- [CH_4] above water surface = 10 [CH_4] in water
- 90% of work spent on protein folding
- Hydrophobic molecules:
  - disturb H-bonds
  - do no create H-bonds themselves



Entropic nature

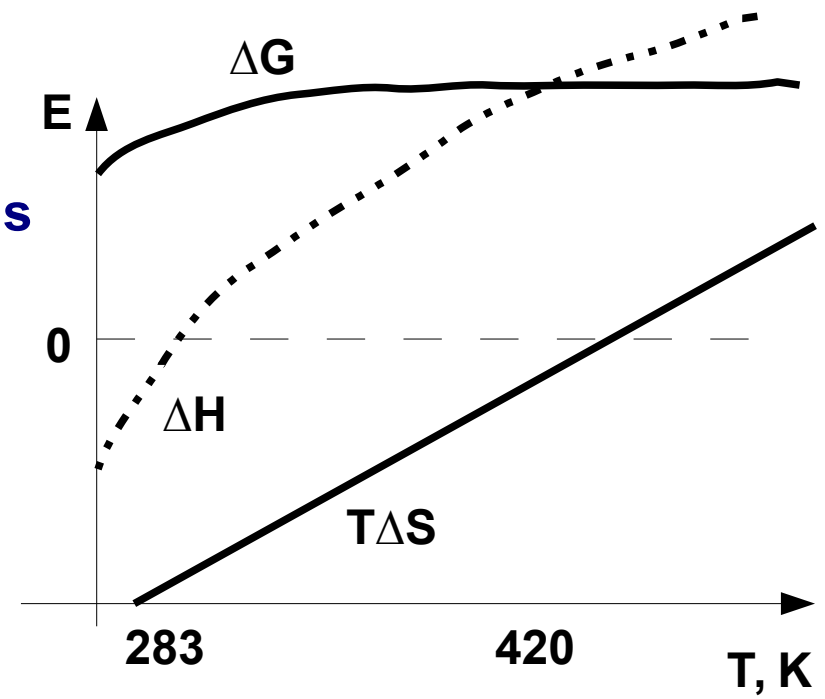


HYDROPHOBIC EFFECT

- Hydrophobicity = function (T)
  - T↑ up to 420K, HΦ↑, melting surfacial H-bonds
  - T↑ (>420K), HΦ↓

- $\Delta G = k\Delta ASA$; $k \sim 10^{-2}$ kcal/(mol·Å)

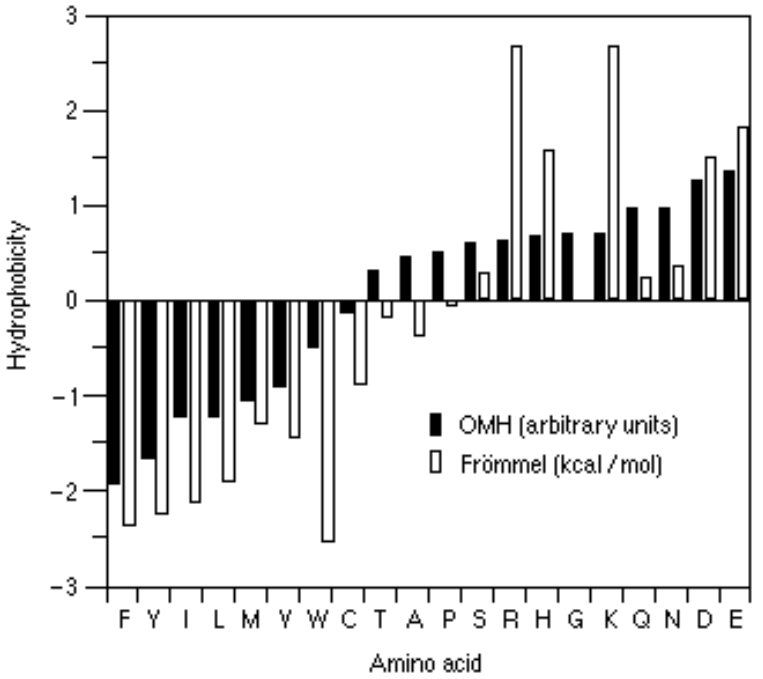
  ΔG(100 aa)~kcal/mol



- Index of hydrophobicity for amino acids:

  equilibrium constant between water

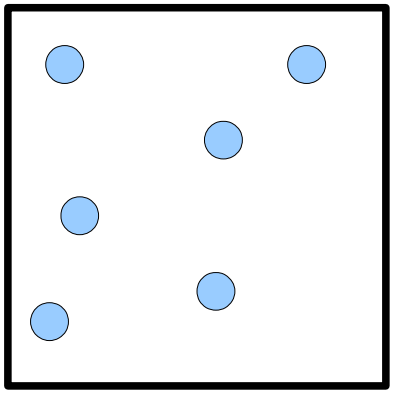
  and nonpolar solvent



POLYMER *vs* GAS

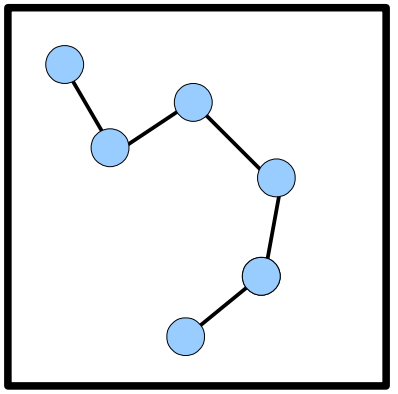
Gas:

V, N, ω



Polymer:

V, N, ω



$$\rho = \frac{N\omega}{V}$$

$$V_{access} = \frac{(V - N\omega)}{N} = \left(\frac{V}{N}\right)(1-\rho) = \frac{\omega/\rho}{(1-\rho)}$$

$$\delta S = k \ln(V_{access}) = k \ln((\omega/\rho)(1-\rho))$$

Could be two ΔG_{min} => two phases

$$V_{access} = A(1-\rho)$$

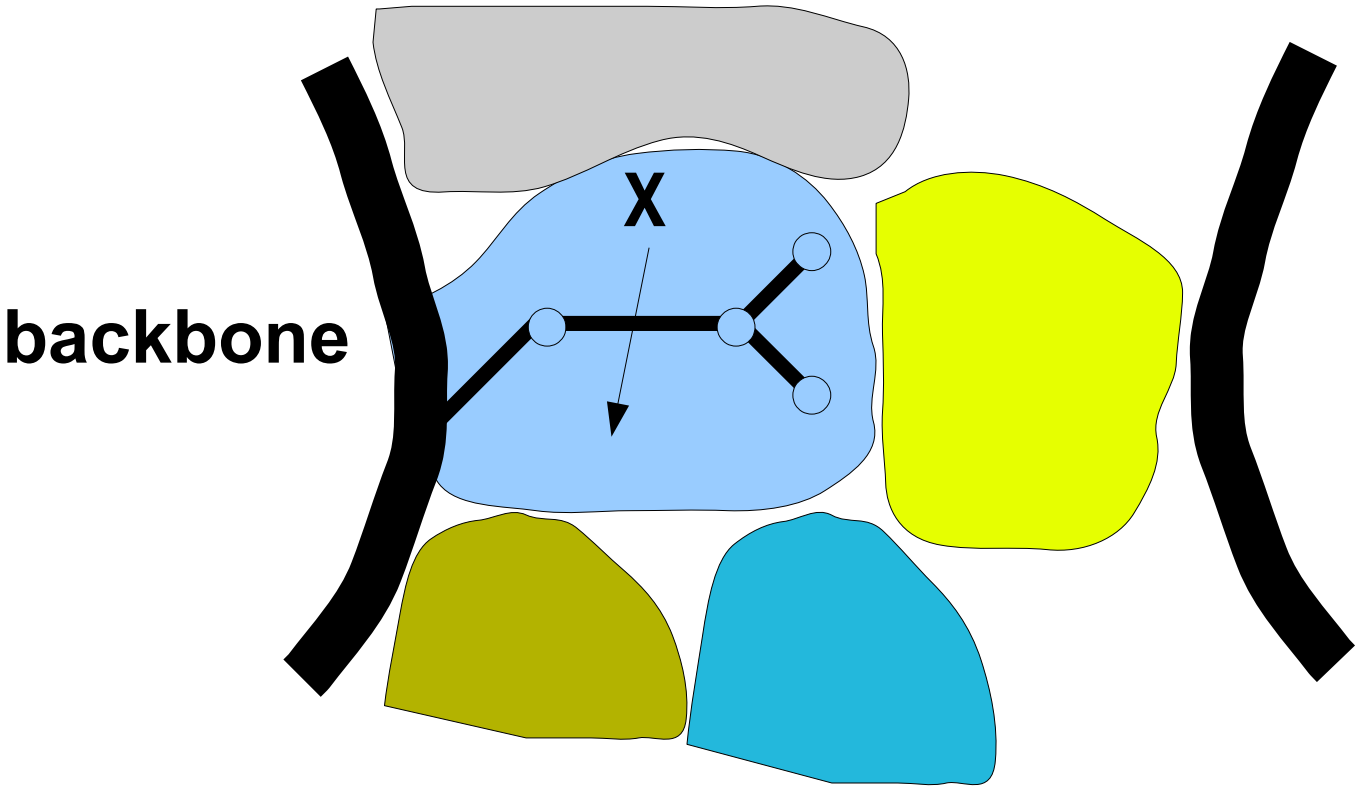
$$\delta S = k \ln(V_{access}) = k \ln(A(1-\rho))$$

Only one ΔG_{min} => gradual transition

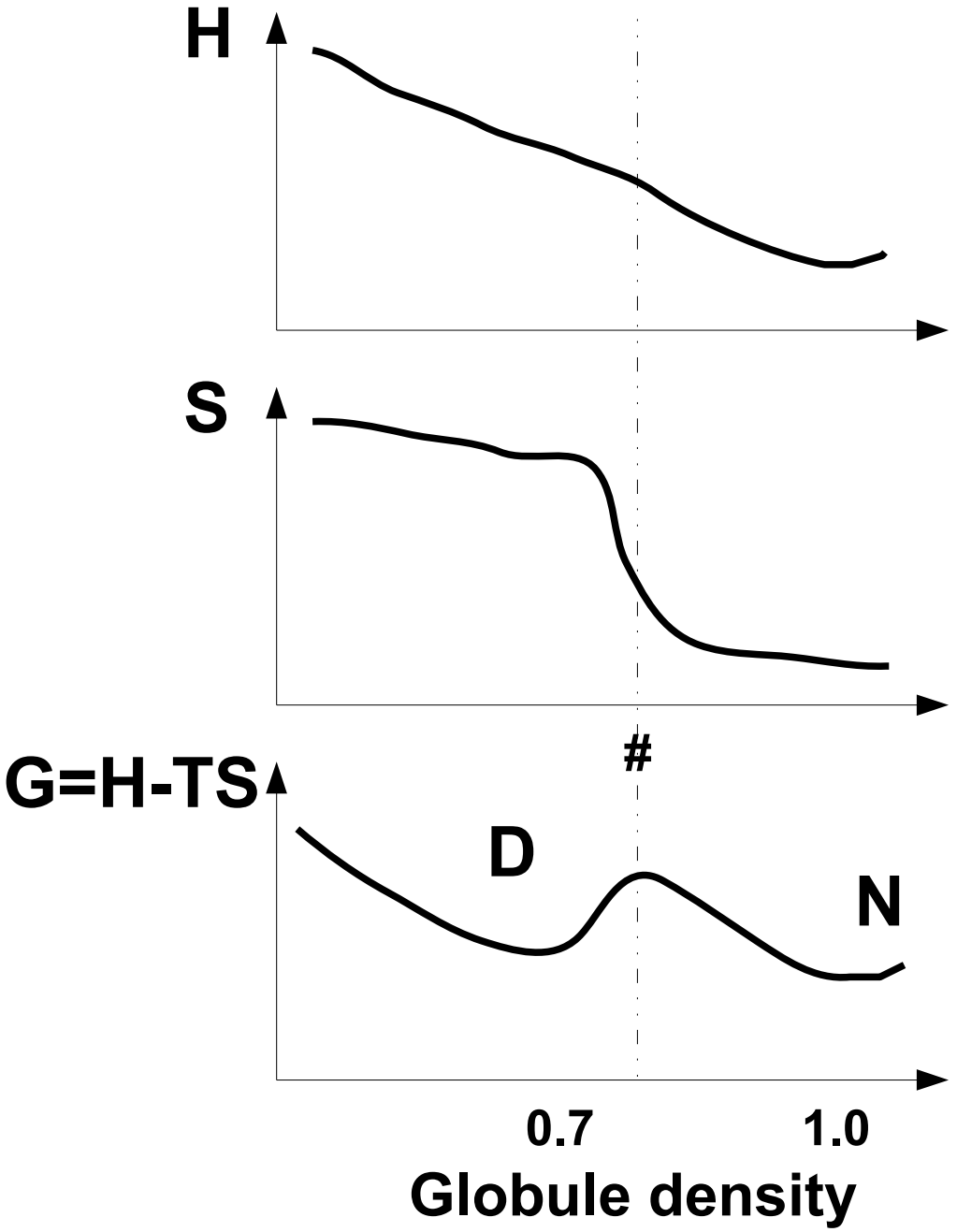
- **Why protein does not fold like a normal polymer?**

PHYSICS OF «ALL OR NOTHING»

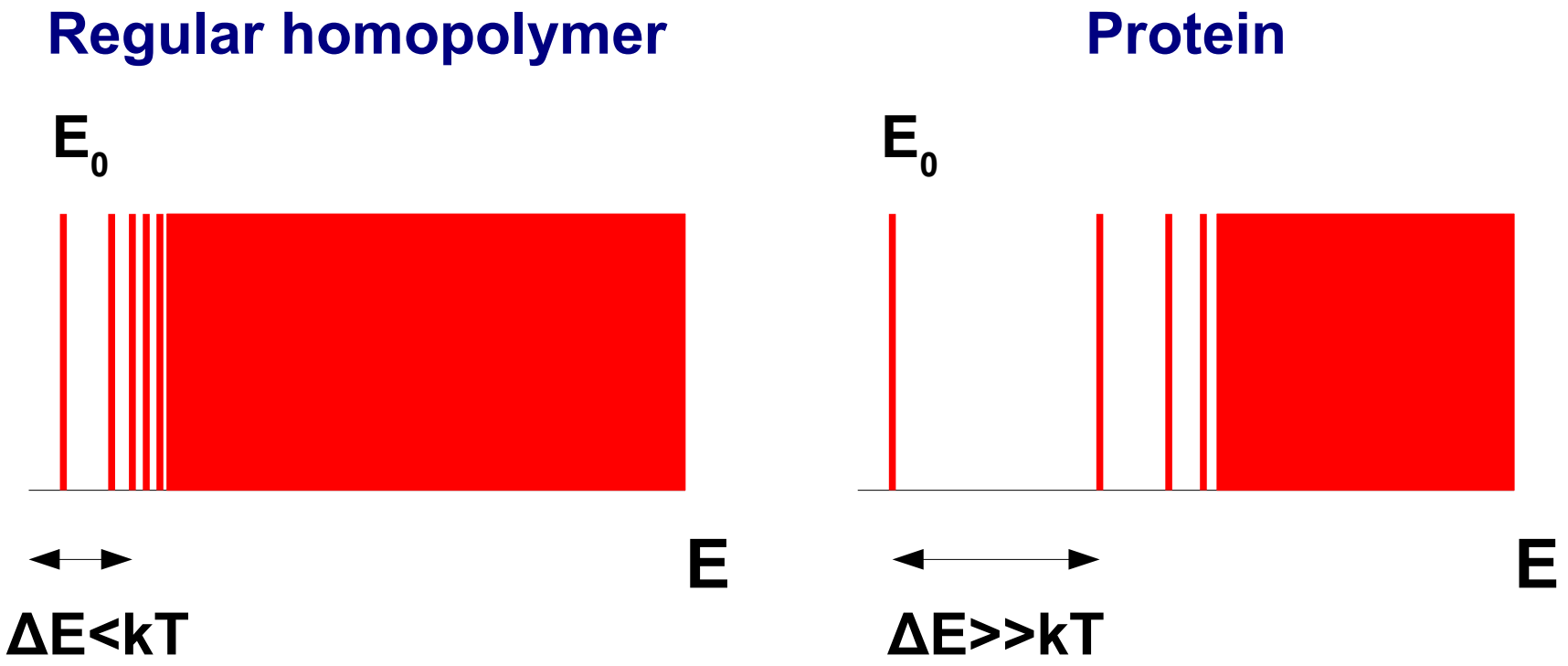
- Protein is not a standard polymer: high heterogeneity.
- Rigid backbone and more flexible side-chains.
- V_{vdW} occupy 70-80% of $V_{protein}$. Roughly equal distribution of free V.
- Rotamers associated entropy jump occurs in unfolding + action of solvent (minimal frustration concept).
- Density remains high.



THERMODYNAMICS



POLYMER *vs* PROTEIN: SPECTRUM

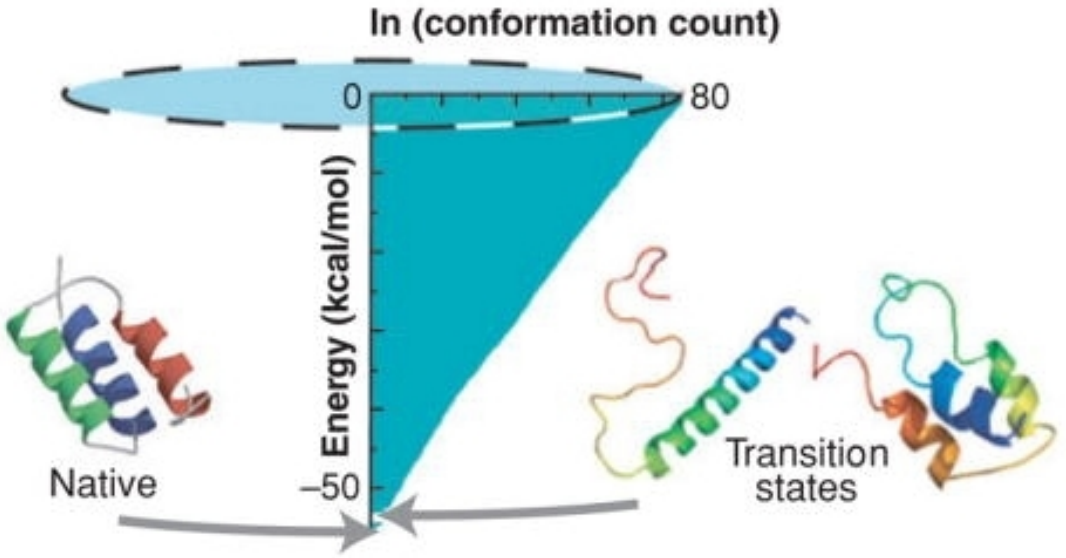


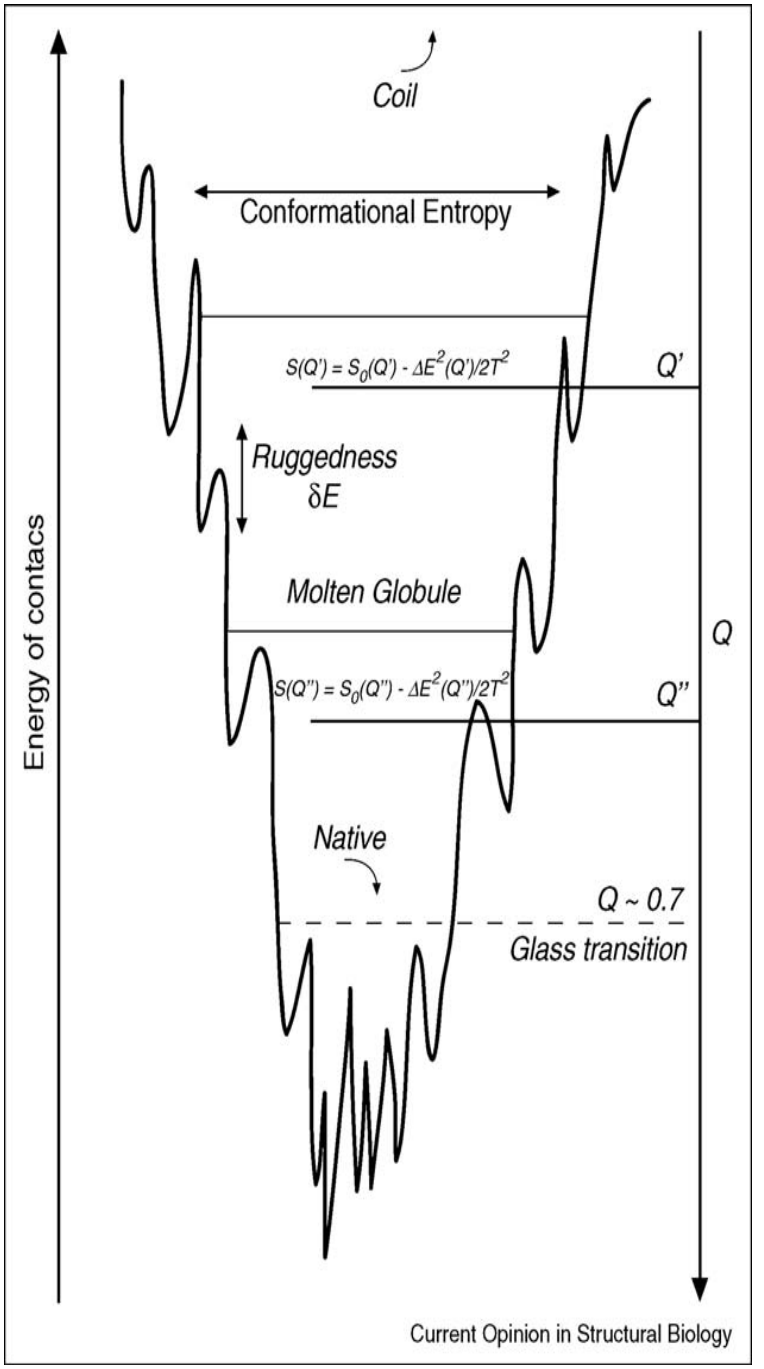
- Protein has anomal stable energetic state (native).
- ΔE~10 kcal/mol => $p=\exp(-\Delta E/kT)\sim10^{-8}$
- If there are > 1 stable structure $p=\exp(-2\Delta E/kT)\sim10^{-16}$

  - polyLys, prions, β-amyloids, serpins

THERMODYNAMICS *vs* KINETICS

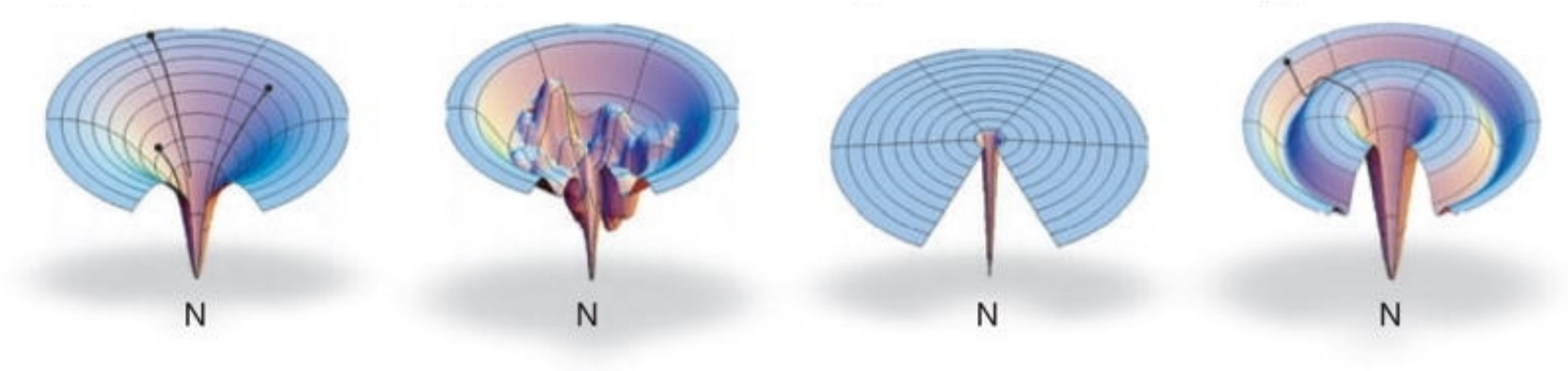
- **Protein folding is not only determined by a native state thermodynamical properties but also by a kinetic path:**
  - **fast**
  - **selected during evolution**
  - **leads to native state**
- **Energy landscapes have hierarchical organization**





FUNNELS

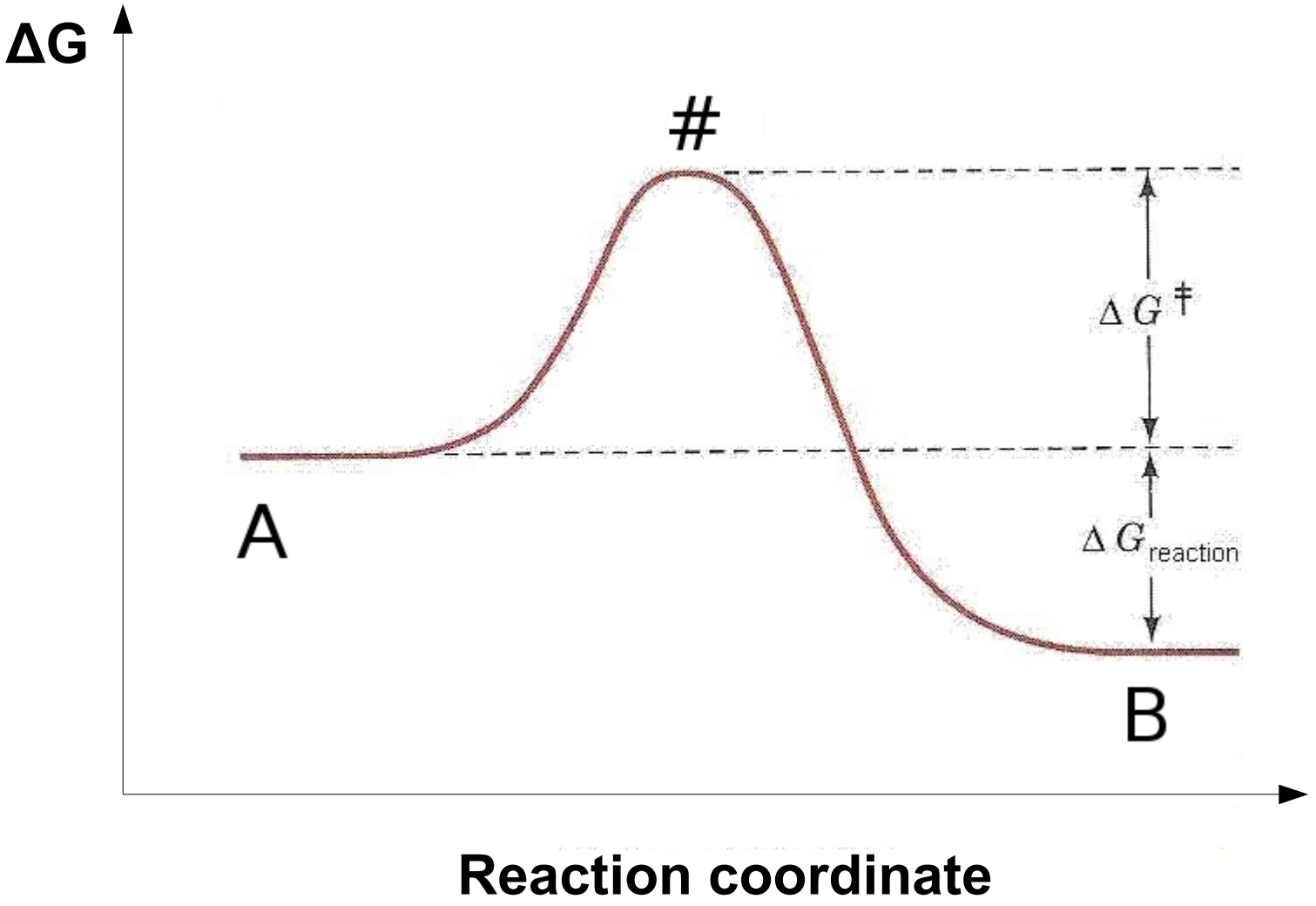
- The folding funnel hypothesis is a the energy landscape theory of protein folding assuming that a protein's native state corresponds to its free energy minimum.



- One path or several:
  - thermodynamic view
  - statistical physics view

TRANSITIONAL STATE

- For small proteins (one-domain proteins) folding process includes only one energetical barrier.



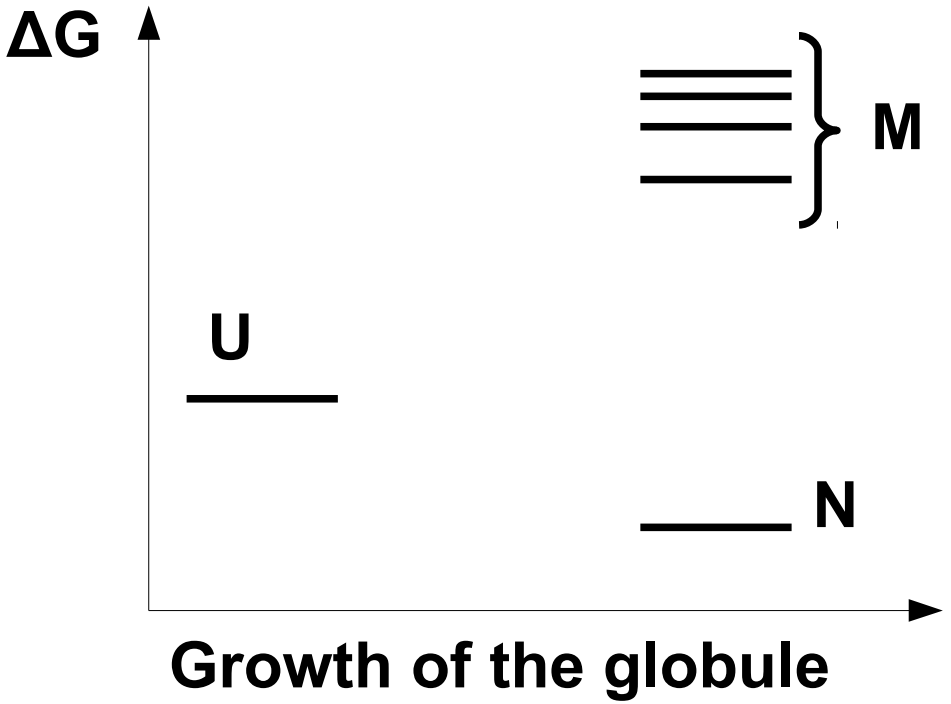
$$k_{A\to B} = k_0 e^{\frac{-(G^{\#} - G_A)}{RT}}$$

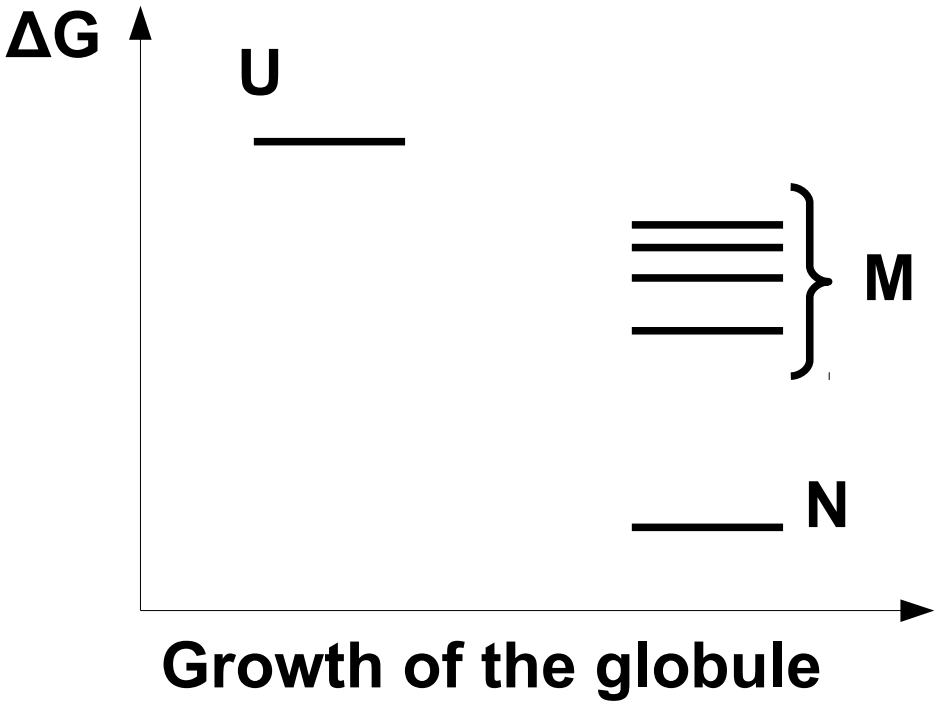
$$k_{B\to A} = k_0 e^{\frac{-(G^{\#} - G_B)}{RT}}$$

$$t_{A\to B} = \frac{1}{k_{A\to B}} \sim e^{G^{\#}}$$

ESTIMATION OF $G^{\#}$ BARRIER

- During folding:
  - $\Delta H = An + Bn^{2/3}$, decreases
  - $\Delta S = Cn + Dn^{2/3}$, decreases
  - $\Delta G = \Delta H - T\Delta S$ => equilibrium $F \leftrightarrow U$ if there were no surfacial effects!
- $G^{\#} \sim n^{2/3}$
- Folding time ~ exp $[(1 \pm 0.5)N^{2/3}]$ 10ns, defined by amount of native contacts.





CHEVRON FIGURES

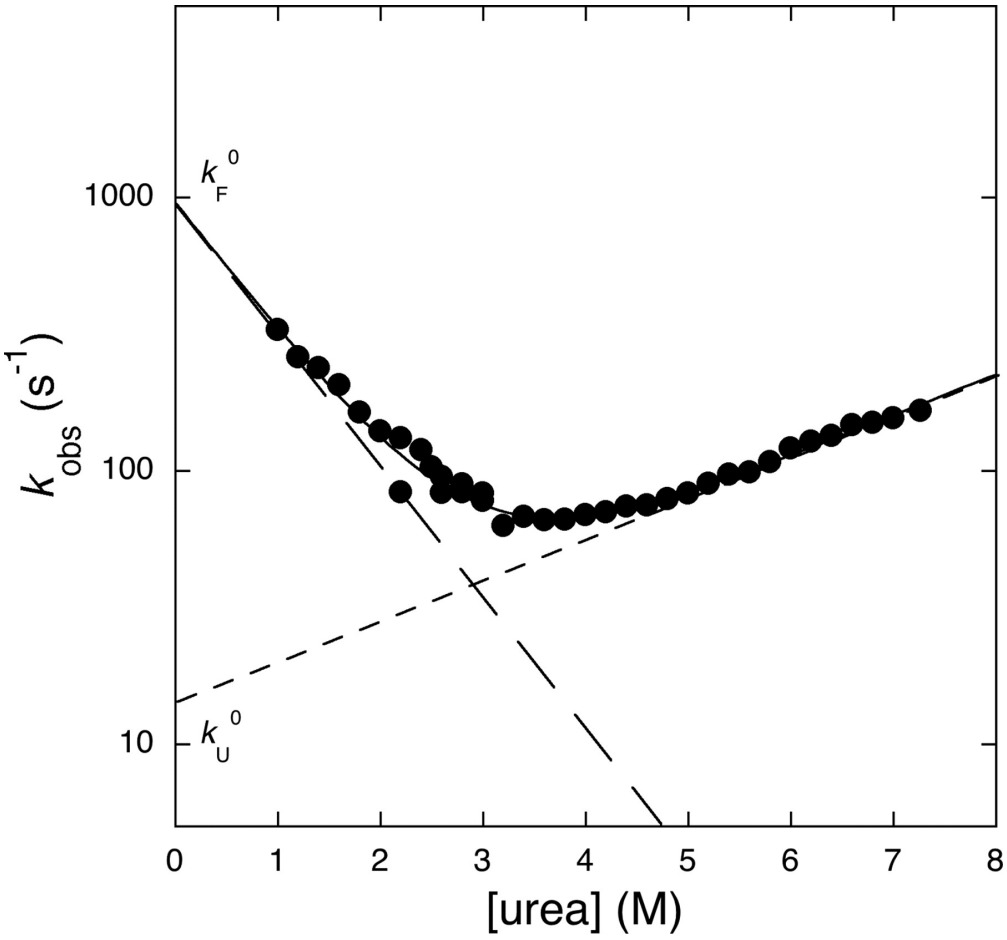
$$K_{B:A} = \frac{n_B^{\infty}}{n_A^{\infty}} = \frac{k_{A\to B}}{k_{B\to A}}$$

$$\frac{dn_A}{dt} = -k_{A\to B} n_A + k_{B\to A} n_B$$

$$n_A + n_B = n_0$$

$$n_A(t) = (n_A(0) - n_A^{\infty})\, e^{-(k_{A\to B} + k_{B\to A})} + n_A^{\infty}$$

$$k_{obs} = k_{A\to B} + k_{B\to A}$$

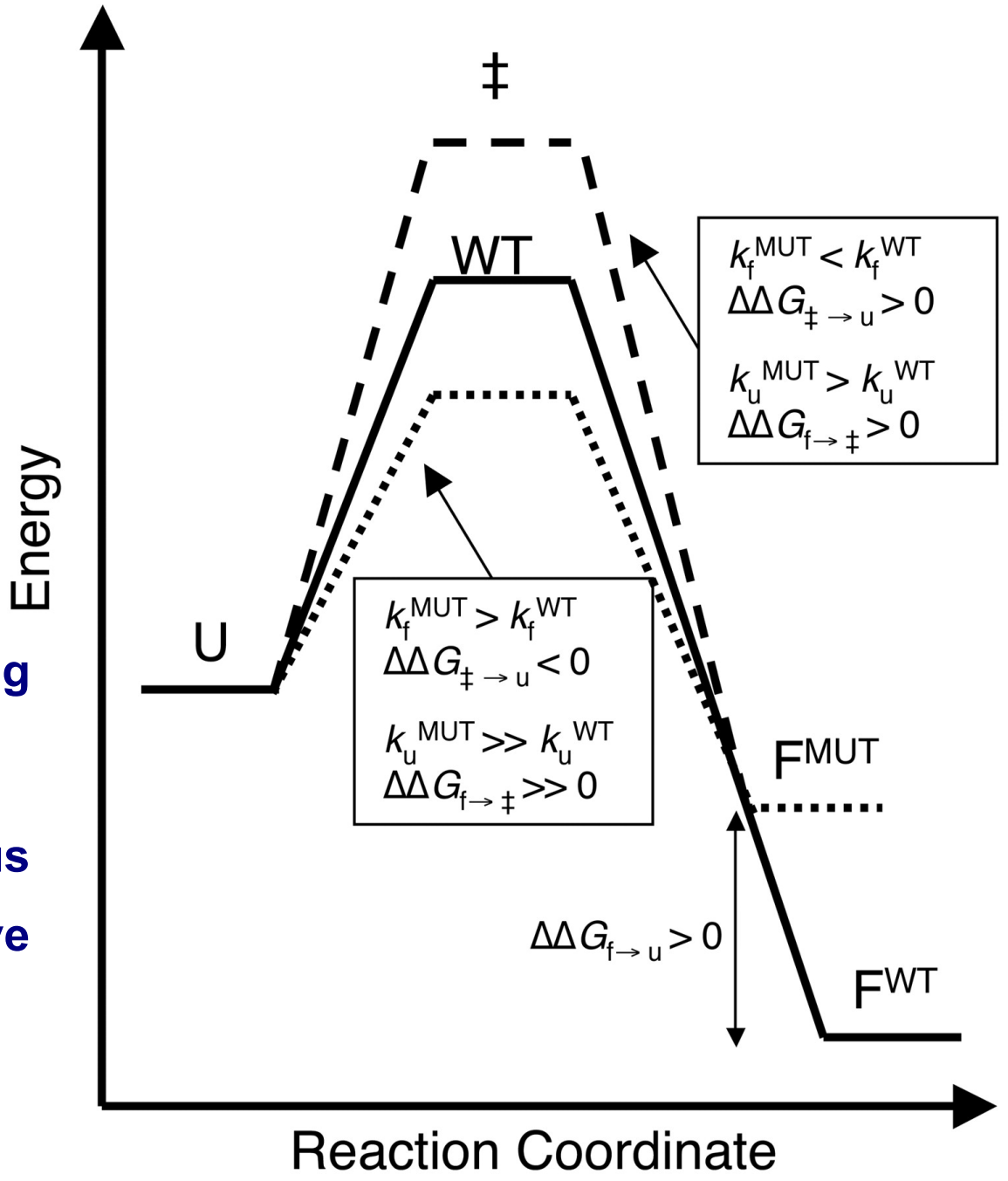


MUTAGENESIS IN FOLDING STUDIES

- **Mutation affects:**
  - **folding rate**
  - **stability of the native state**
- **Nucleation mechanism**

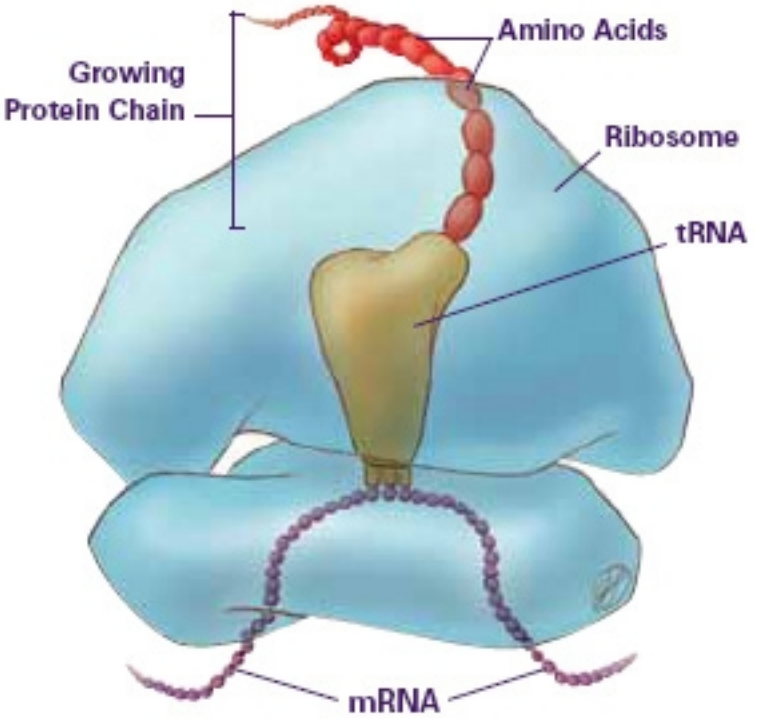
$$\varphi_f = \frac{\delta\left(G^{\ddagger} - G_U\right)}{\delta\left(G_F - G_U\right)}$$

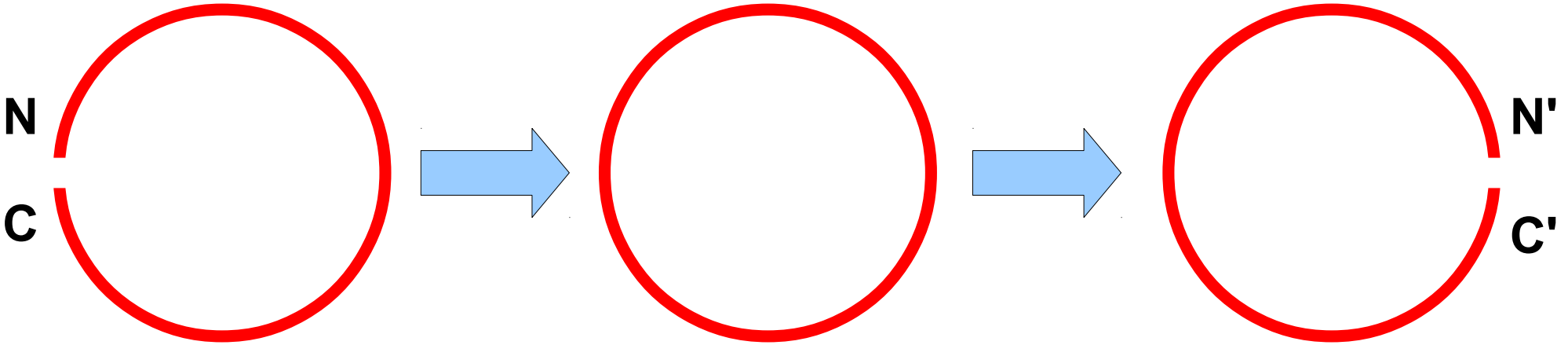
- **$\Phi_f \approx 1$, residue is in folding nucleus; $\Phi_f \approx 0$, not.**
- **Most of mutations in nucleus affect stability of the native structure.**



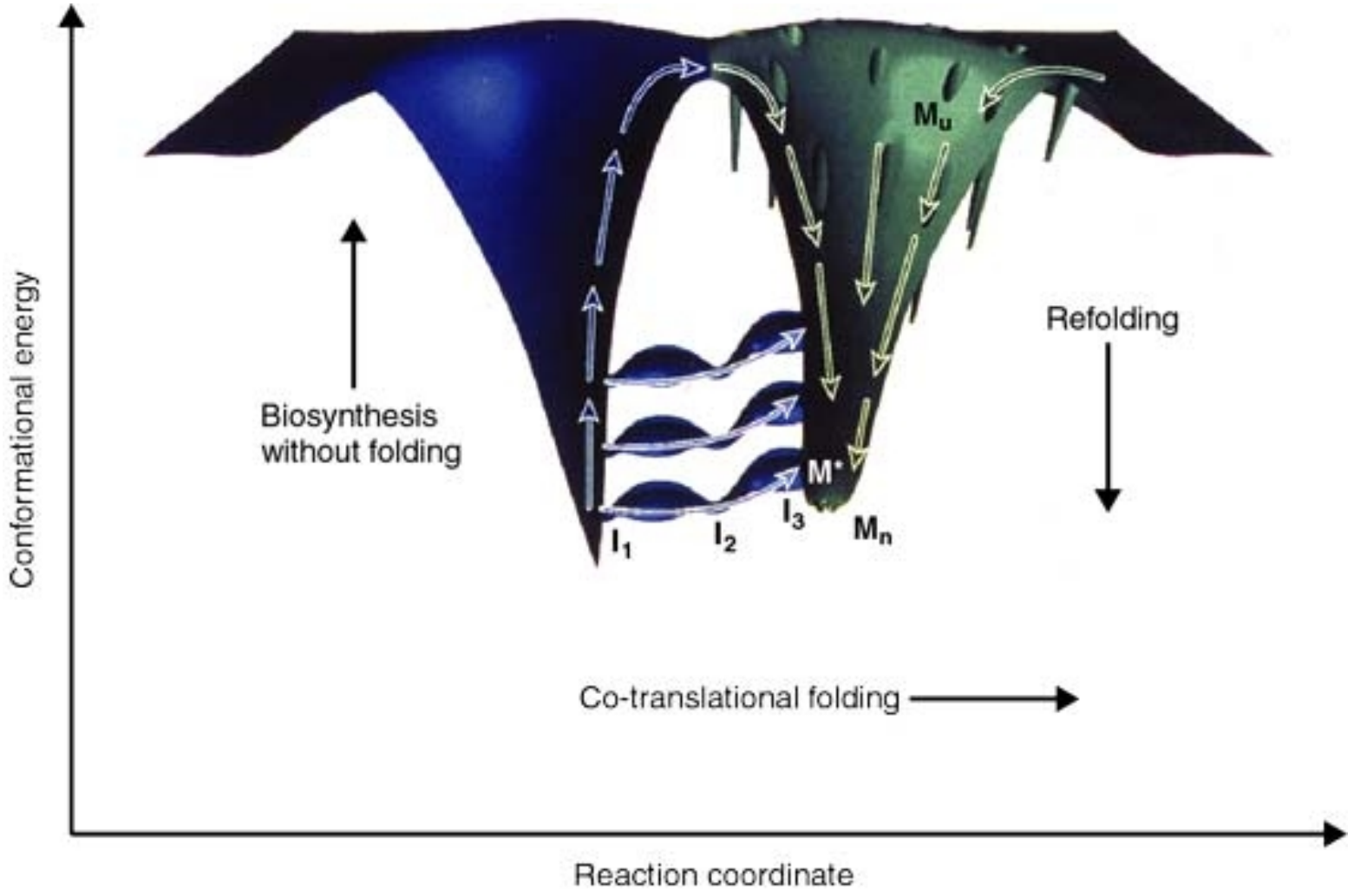
FOLDING IN THE CELL

- Ribosome synthesis ~ 1 minute.
- Synthesis speed is not constant.
- For multidomain proteins N-terminal domains are folded before the synthesis is completed.
- Domain is a unit of folding (*in vitro*; globin).
- Cotranslational folding (luciferase).
- Self-organization experiments: chemical synthesis and cyclic proteins.

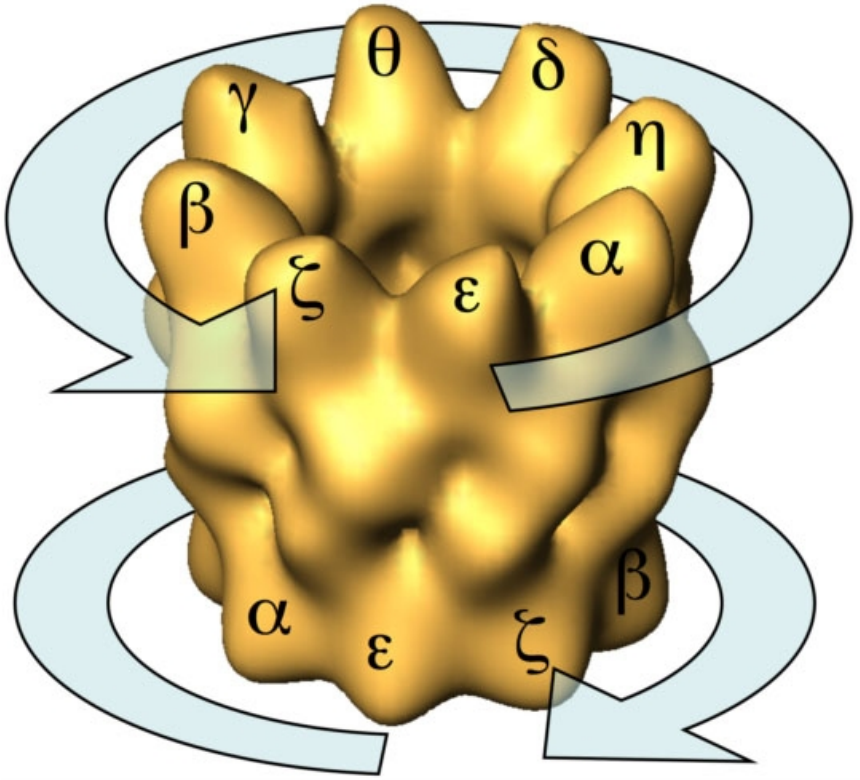


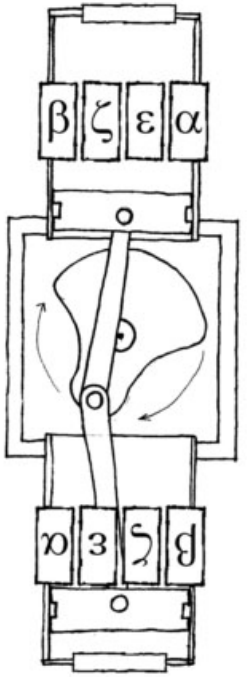


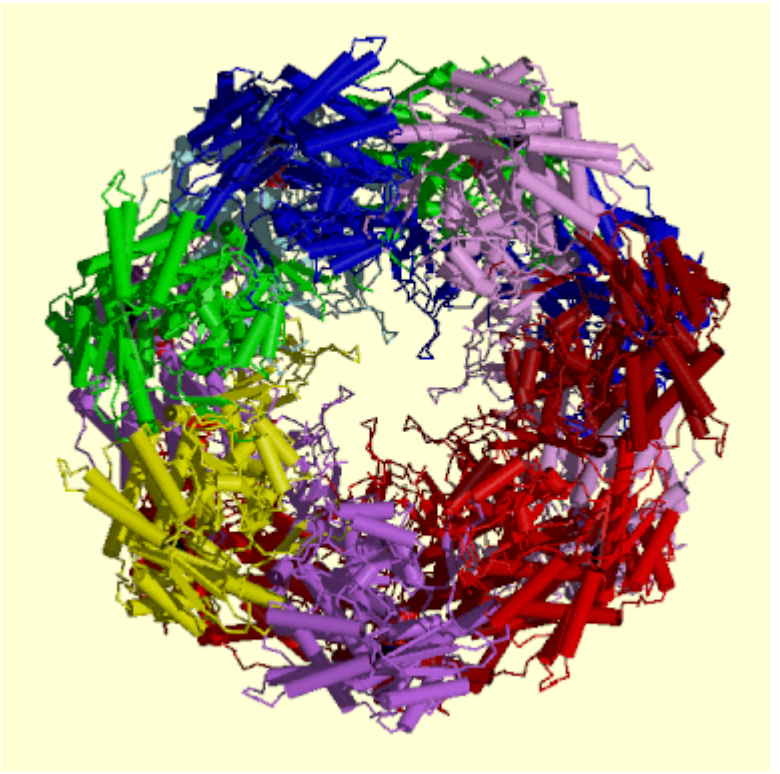
ENERGETICS OF COTRANSLATIONAL FOLDING



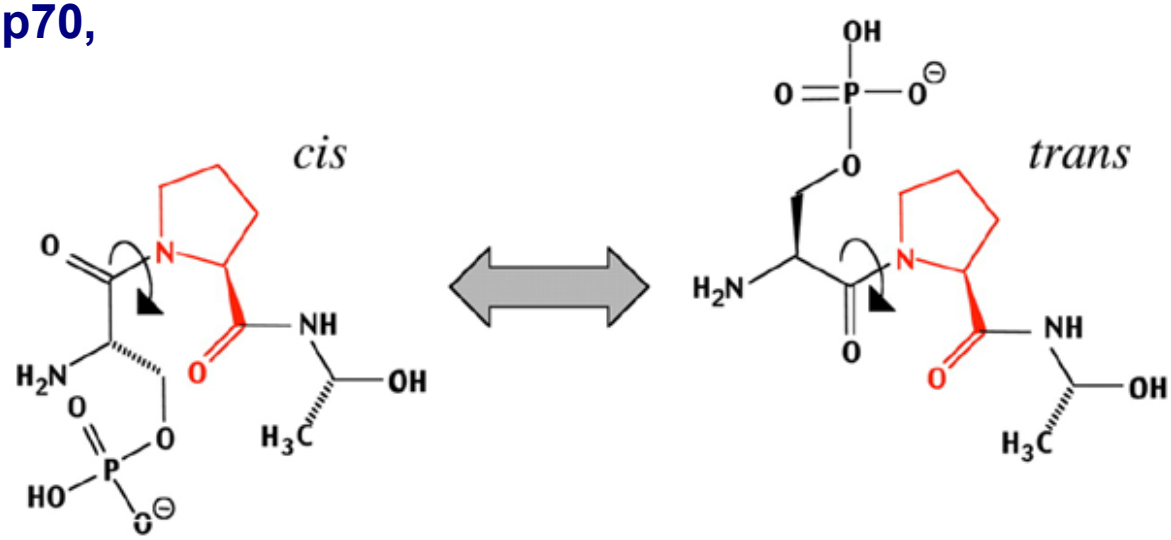
CHAPERONS







- Hsp(Heat Shock Proteins): Hsp60, Hsp70, Hsp100 etc. Regulated by T, pH etc.
- Use ATP hydrolysis
- Decrease the hydrophobic ΔASA
- Ribosomes selves
- Prolyl- and disulfide-isomerase



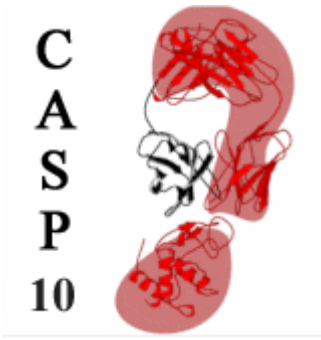
COMPUTATIONAL APPROACHES

- MD
  - replica exchange
  - Principal Component Analysis (PCA)
  - Clustering by contacts/RMSD/radius of gyration/H-bonds
- Monte Carlo
- «Zip and assemble» approach
- Up to 100 aa proteins could be tractable:
  - 36-residue villin, RMSD = 4.5 Å (Duan, Kollman, 1998)
  - 20-residue Trp-cage peptide, RMSD = 1 Å (Simmerling, 2002)
  - 47-residue albumin-binding domain, RMSD = 2 Å (Lei, 2007)
  - β-hairpins up to 20-residues

CASP COMPETITION

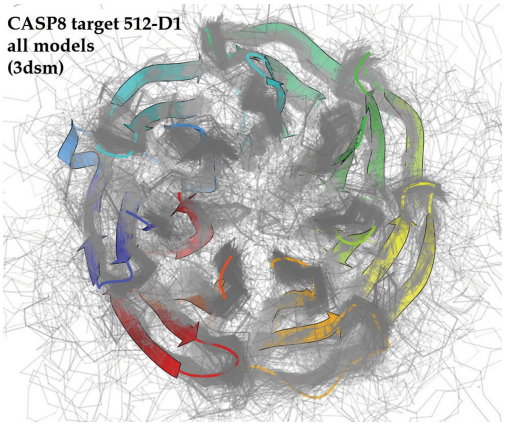
- *C*ritical *A*ssessment of Techniques for Protein *S*tructure *P*rediction
- predictioncenter.com
- Goal: to obtain an in-depth and objective assessment of current abilities in the area of protein structure prediction
- Prediction of 'soon known structures', no postpredictions
- CASP questions:
  - Models similarity to the corresponding experimental structures
  - Mapping of the target sequence onto the proposed structure
  - Model usefulness for similar structures
  - Model accuracy *vs* best template use
  - Has there been progress from the earlier CASPs?
  - What methods are most effective?
  - Where can future effort be most productively focused?

CASP Experiments
CASP_ROLL
CASP10 (2012)
CASP9 (2010)
CASP8 (2008)
CASP7 (2006)
CASP6 (2004)
CASP5 (2002)
CASP4 (2000)
CASP3 (1998)
CASP2 (1996)
CASP1 (1994)



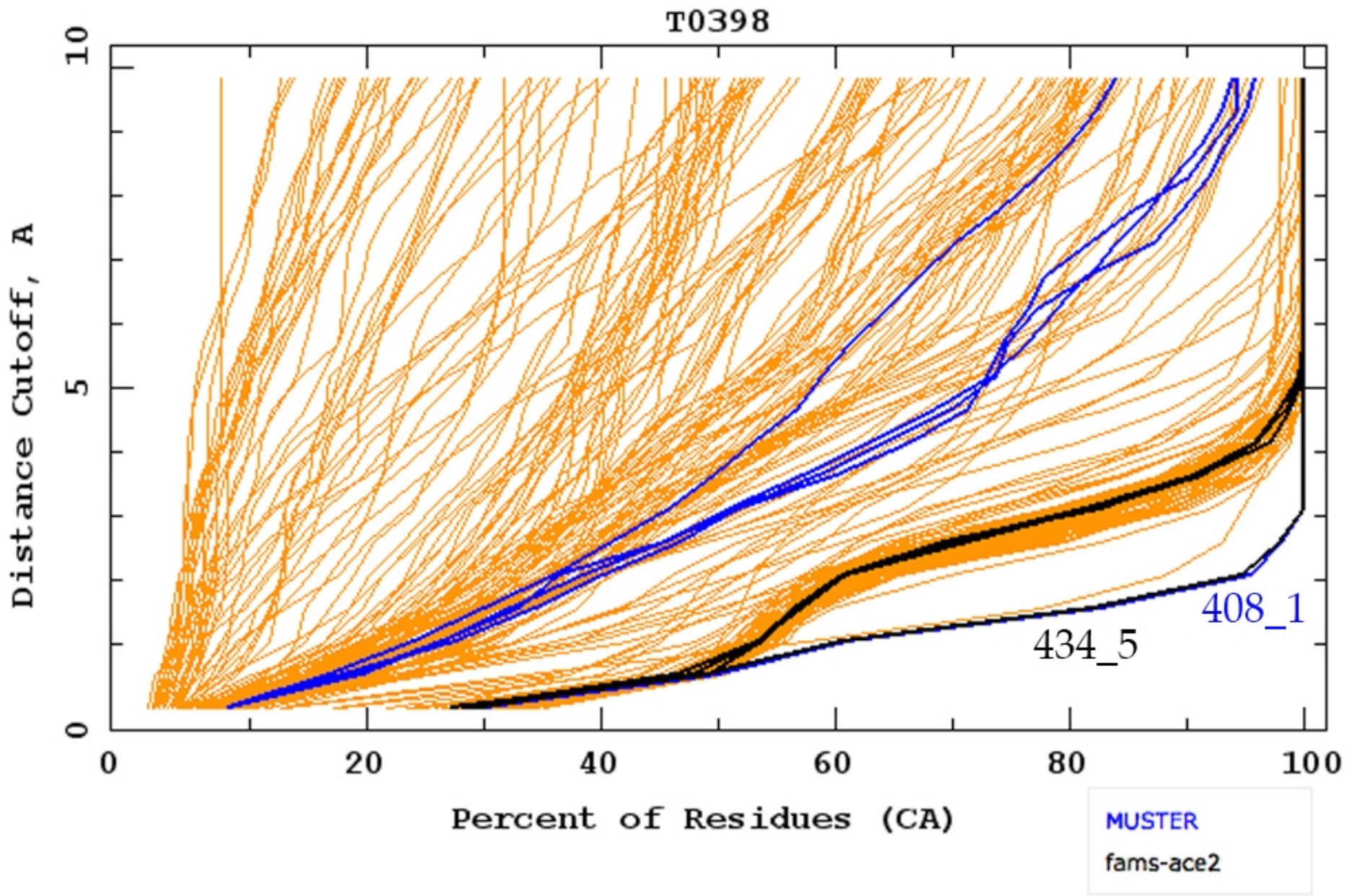
SCOPE OF CASP

- **Tertiary structure prediction:**
  - **The 'Template based modelling'**
  - **The 'Template free modelling'**
  - **Detailed analysis of the side chains, loops, and active sites for those structure models where the backbone is sufficiently accurate.**
  - **Success in refining models beyond the quality obtained by simply copying from a single template will be analyzed.**

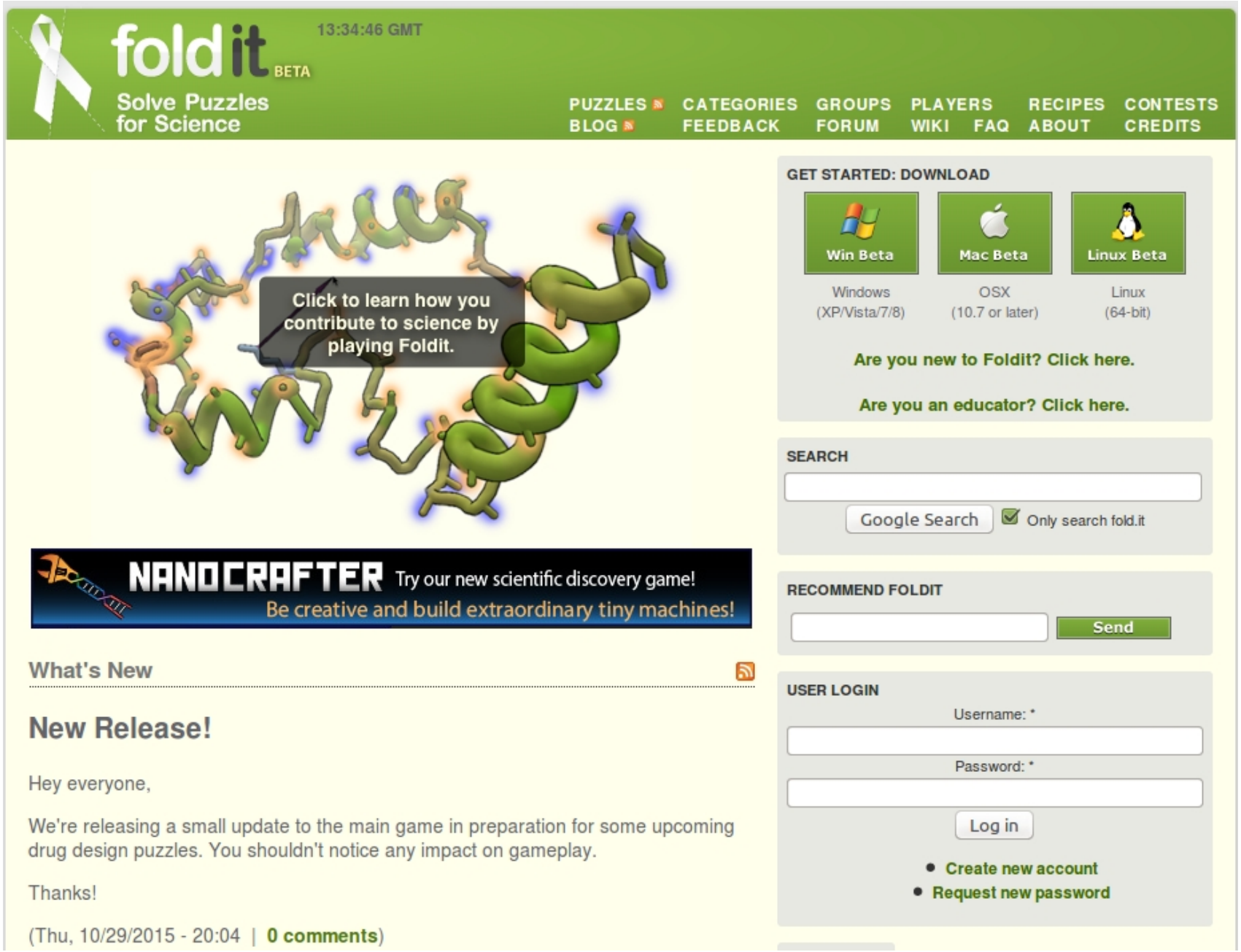


- **Other prediction cathegories:**
  - **Detecting residue-residue contacts in proteins.**
  - **Identifying disordered regions in target proteins.**
  - **Function prediction (prediction of binding sites).**
  - **Quality assessment of models in general and the reliability of predicting certain residues in particular.**

CASP EXAMPLE



FOLDIT



http://fold.it/

FOLDING@HOME

DOWNLOAD Home Page

WHAT IF...

PLAY VIDEO

you could help find a cure?

Help Stanford University scientists studying Alzheimer's, Huntington's, Parkinson's, and many cancers by simply running a piece of software on your computer.

The problems we are trying to solve require so many calculations, we ask people to donate their unused computer power to crunch some of the numbers.

FOLDING@HOME

In just 5 minutes ...

Add your computer to over **163,000** others around the world outputting **38,000** teraflops of computing power to form the world's largest distributed supercomputer.

Step 1. Download protein folding simulation software called Folding@home.

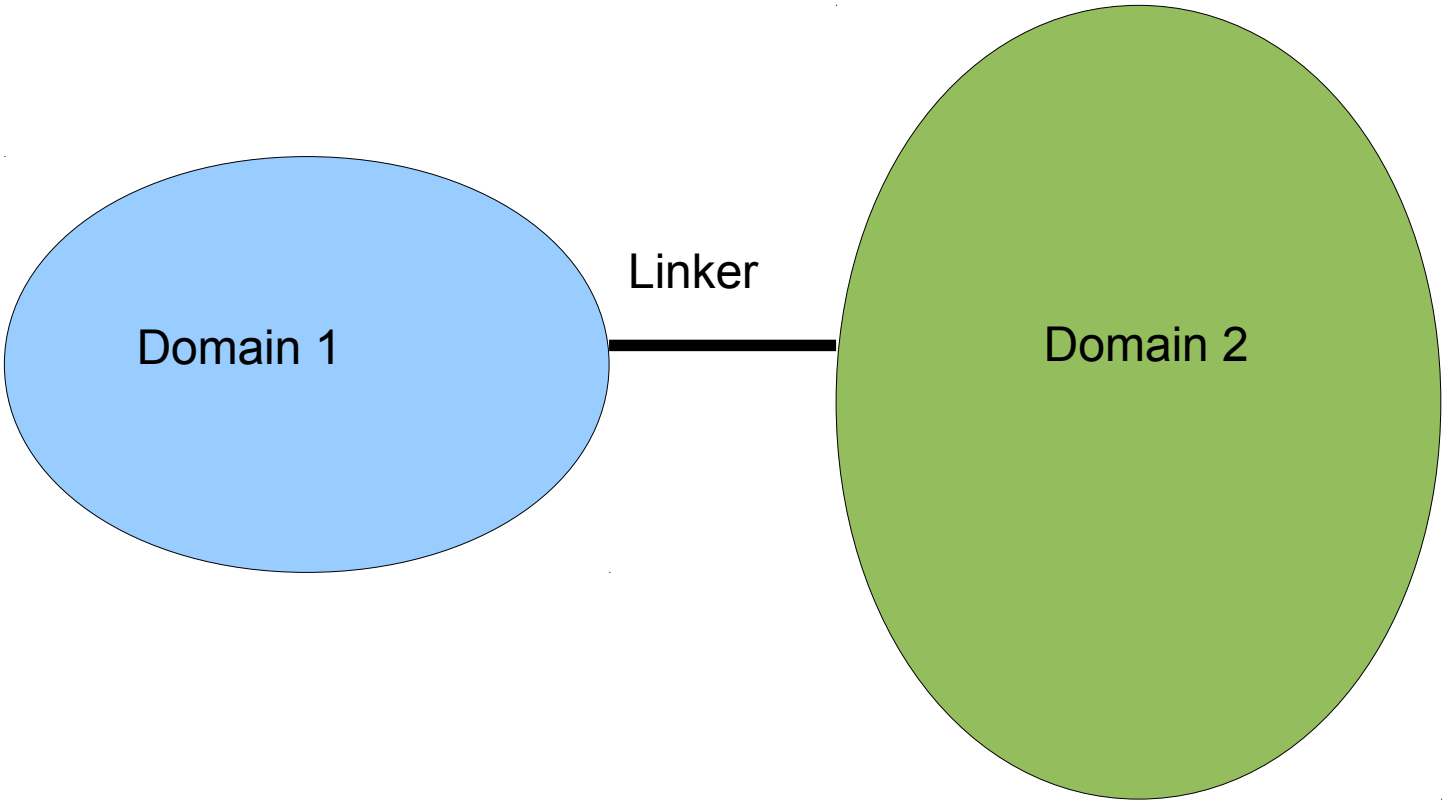
Step 2. Run the installation. The software will automatically start up and open a web browser with your control panel.

Step 3. Follow the instructions to Start Folding.

Stanford University will send your computer a folding problem to solve. When your first job is completed, your computer will swap the results for a new job.

START FOLDING

CASE STUDY: FOLDING AN INTERDOMAIN LINKER



- **No structure of KDR motif in the linker is available**

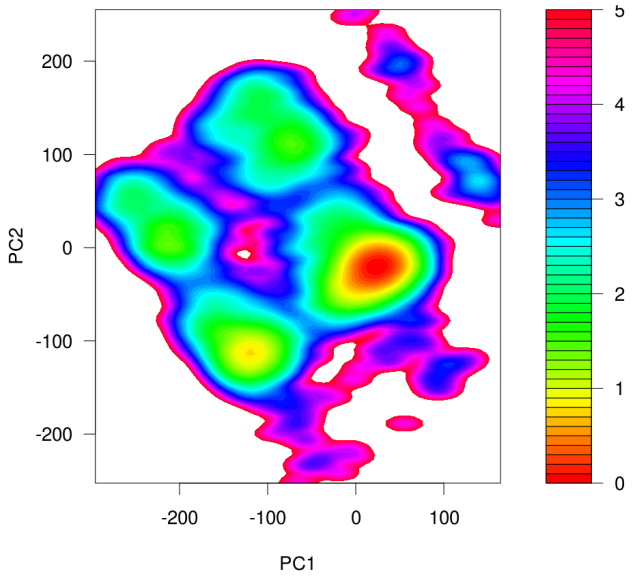
Aim: to calculate the conformation of the linker

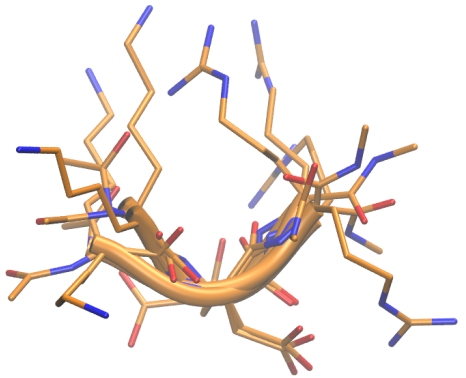
CASE STUDY: FOLDING AN INTERDOMAIN LINKER

- **Metrhodology: REMD**
- **Peptides: KDR, KKDRA, PKKDRAR, RPKKDRARQ**

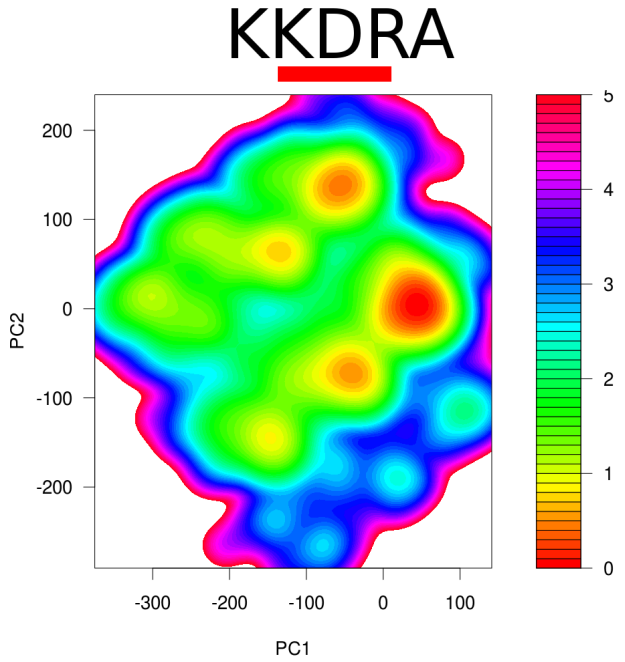
CASE STUDY: FOLDING INTERDOMAIN LINKER

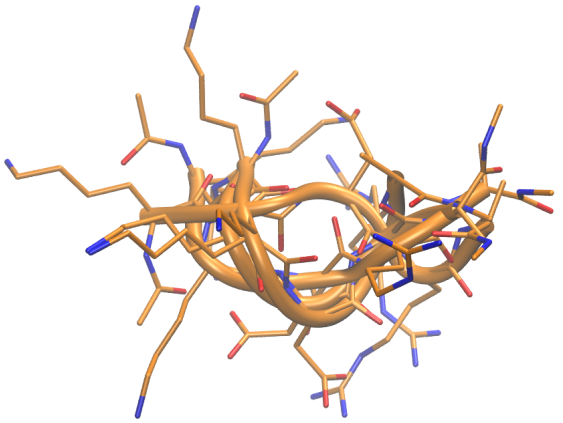
KDR



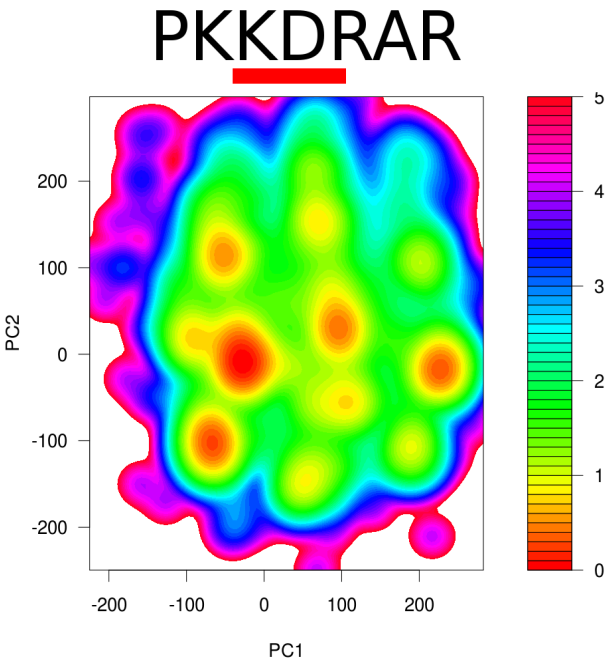


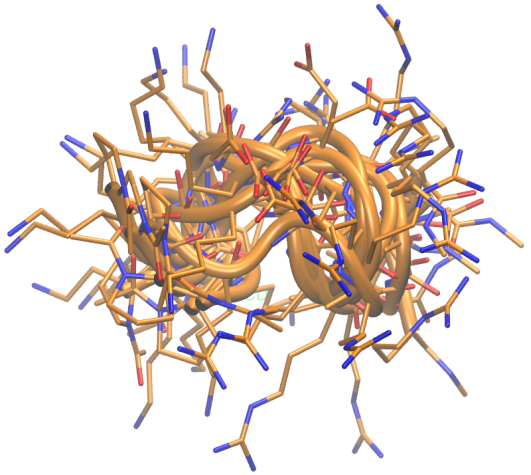
CASE STUDY: FOLDING AN INTERDOMAIN LINKER



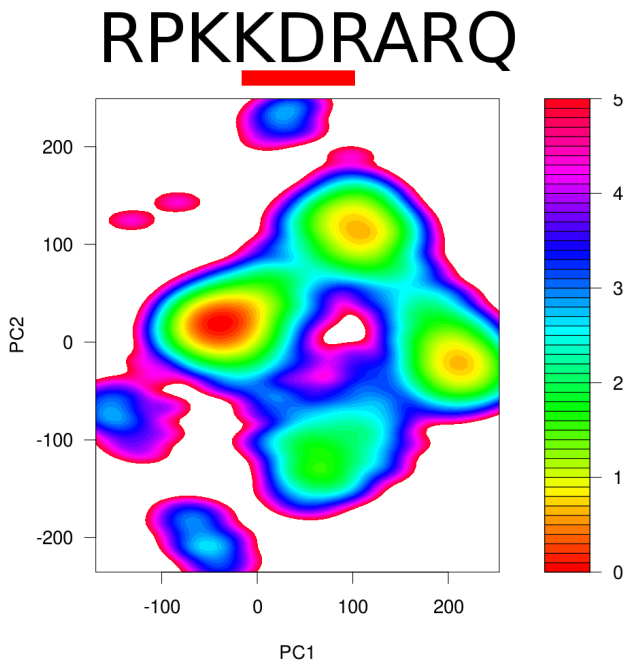


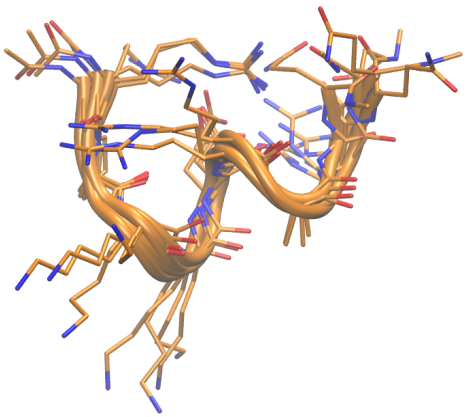
CASE STUDY: FOLDING AN INTERDOMAIN LINKER





CASE STUDY: FOLDING AN INTERDOMAIN LINKER





CASE STUDY: FOLDING AN INTERDOMAIN LINKER

The studied interdomain linker folds into α-helix

LECTION 4: PROTEIN FOLDING

- Protein folding and denaturation: «all or nothing»
- Anfinsen's dogma
- Folding of α- and β-secondary structures
- Hydrophobic effect
- Levinthal's paradox
- Protein *vs* other polymers
- Thermodynamics and kinetics of protein folding
- Protein folding in the cell
- Computational approaches to study folding
- CASP
- Case study: folding of an interdomain linker