The Energy Landscape of Biomolecules

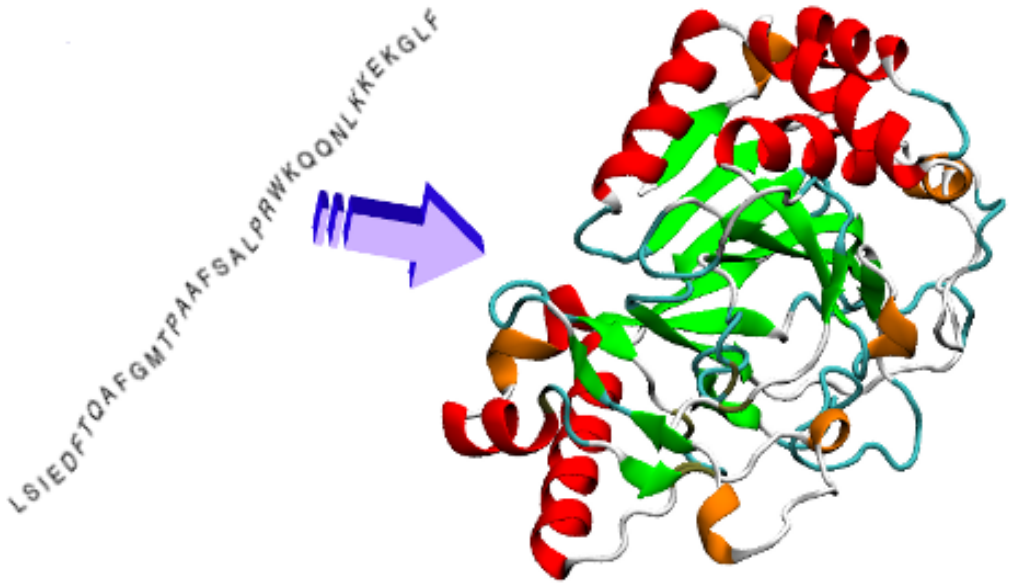
Protein Folding

Charlotte Voigt

10.1.2017

1. Proteins and folding
2. Levinthal's Paradox
   - A simple theoretical model
3. Computational Modelling
4. Energy landscapes
   - Levinthal's Paradox in Energy landscape terms
   - Energy landscapes are funnel-like
   - Dimensionality reduction
5. Summary

$$\text{sequence} \xrightarrow[\textit{folding}]{} \text{structure} \xrightarrow[\textit{binding}]{} \text{function}$$



Alessandro Laio, SISSA

Levinthal 1968: By random search, it takes very long for the protein to discover all possible structures

For example:

$$\left.\begin{array}{l} N = 101 \text{ amino acids} \\ 3 \text{ possible states of each bond} \end{array}\right\} 3^{100} \approx 5 \cdot 10^{47} \text{ configurations}$$

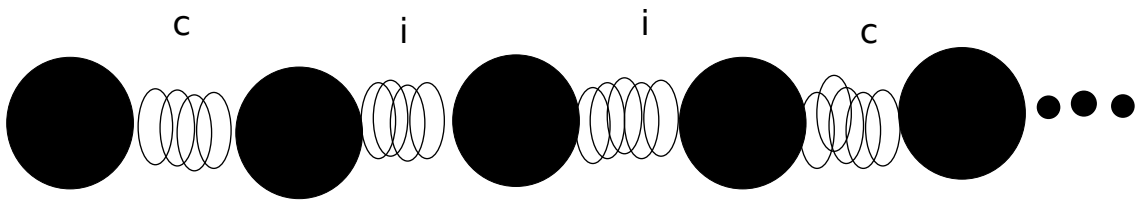
Even with a sampling rate of $10^{13} \frac{1}{s}$, it would take 10^{27} years to take all configurations!

This effect can be expressed in an easy mathematical model [Zwanzig 1992]:
$N + 1$ amino acids, N bonds
States:

$$\text{c} \leftrightarrow \text{correct bond}$$

$$\text{i} \leftrightarrow \text{incorrect bond}$$

Start with a random sequence (c, i, i, c, ...)



and change bonds with rates

$$k_0 \text{ for c} \rightarrow \text{i}$$

$$k_1 \text{ for i} \rightarrow \text{c}$$

Number of incorrect bonds: S
Number of correct bonds: $N - S$
Probability of S incorrect bonds a time t: $P(S, t|S_i, t_0) = P(S, t)$
Master or gain-loss equation:

$$\frac{d}{dt}P(S,t) = \overbrace{(N-S+1)k_0}^{\text{rate}(S-1\to S)} P(S-1,t) + \overbrace{(S+1)k_1}^{\text{rate}(S+1\to S)} P(S+1,t) - \underbrace{(N-S)k_0}_{\text{rate}(S\to S+1)} P(S,t) - \underbrace{Sk_1}_{\text{rate}(S\to S-1)} P(S,t)$$

Using absorbing boundary conditions at the final state S_f, define the survival probability $s(t, S_f|S_i)$ by

$$s(t, S_f) = \sum_{S \neq S_f} P(S, t) = 1 - \int_0^t F(t'; S_f) dt'.$$

The mean first-passage time

$$\tau(S_f|S_i) = \int_0^\infty tF(t; S_f) dt$$

is the mean time for a protein to reach state S_f from starting state S_i. For large N and not too small k_0, τ is independent of S_i

$$\tau(S_f = 0|S_i) \approx \frac{1}{Nk_0} \left(1 + \frac{k_0}{k_1}\right)^N$$

Introduce an energy penalty:
A correct bond has degeneracy 1 and energy ϵ_c.
An incorrect one has degeneracy ν and energy $\epsilon_i = \epsilon_c + U$.

$$\frac{d}{dt}[c] = -k_0[c] + k_1[i], \qquad [c] + [i] = 1$$

is the kinetic sheme for a single bond. In equilibrium we have

$$\frac{[i]_{eq}}{[c]_{eq}} = \frac{k_0}{k_1}$$

And from statistical mechanics:

$$K = \frac{k_0}{k_1} = \frac{\nu e^{-(\epsilon_c + U)/k_b T}}{e^{-\epsilon_c/k_B T}} = \nu e^{-U/k_B T}.$$

Mean first-time passage τ with energy penalty U:

$$\tau \approx \frac{1}{Nk_0}(1+K)^N$$

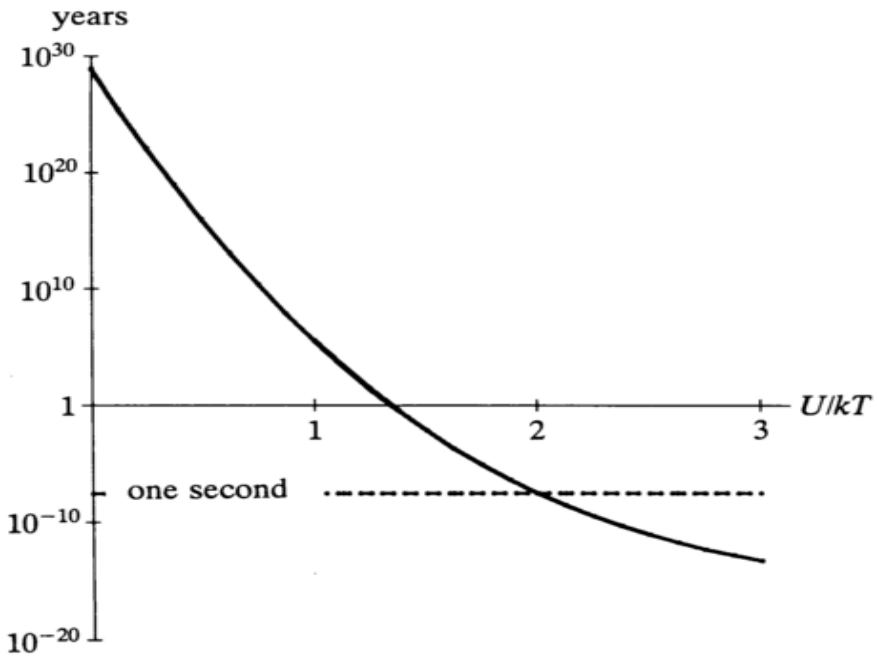
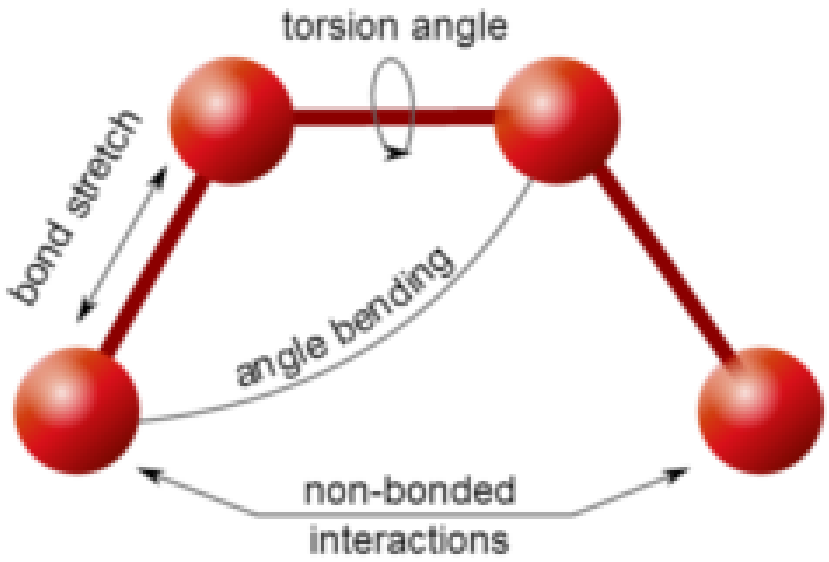


Figure : Mean first-passage time τ in years. With $N = 100$, $\nu = 2$, $S_i = 66$, $k_1 = 10^9 s^{-1}$ and $k_0 = 2\exp(-U/kT) \cdot 10^9 s^{-1}$. [Zwanzig, PNAS, 89:20-22, 1992]

A realistic force field for computational modelling

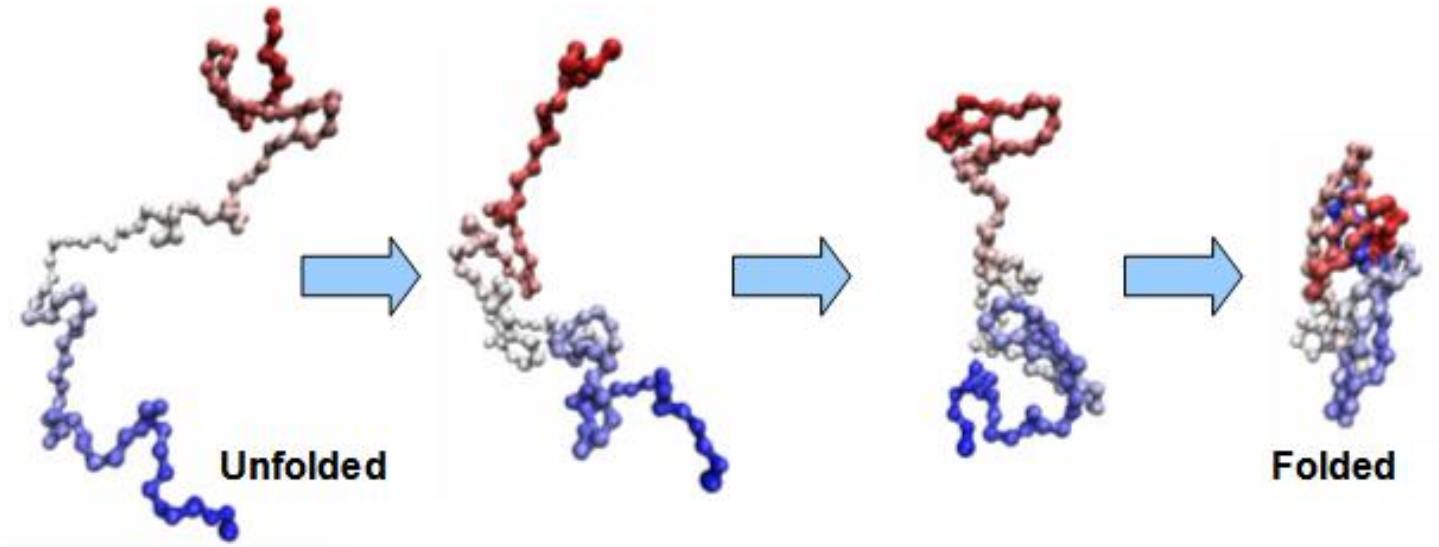


mmb.irbbarcelona.org

$$U = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} \frac{V_n}{2}(1 + cos(n\phi - \delta_n)) + \sum_{i<j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right)$$

Molecular Dynamic (MD) simulations: Use force fields, solve Newtons equations of motion.



inst.eecs.berkeley.edu

Energy landscapes

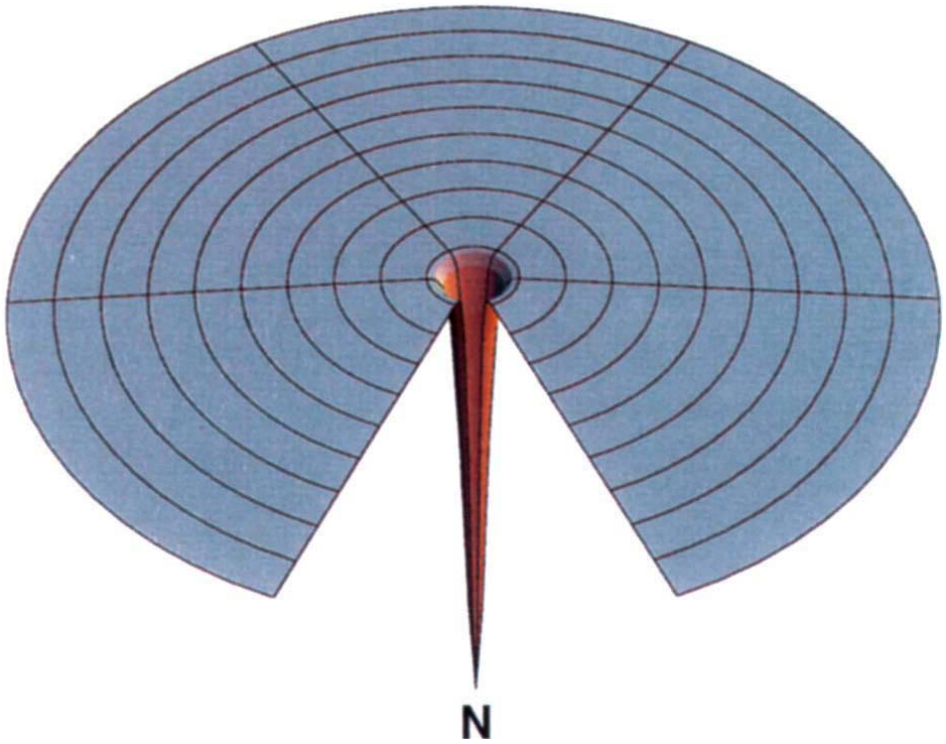
Plot the Free Energy of a protein over its degrees of freedom.

$$\Delta G(x) = -kT \ln P(x)$$

The movie shown in the talk can be found at
https://www.youtube.com/watch?v=YANAso8Jxrk

Dill group

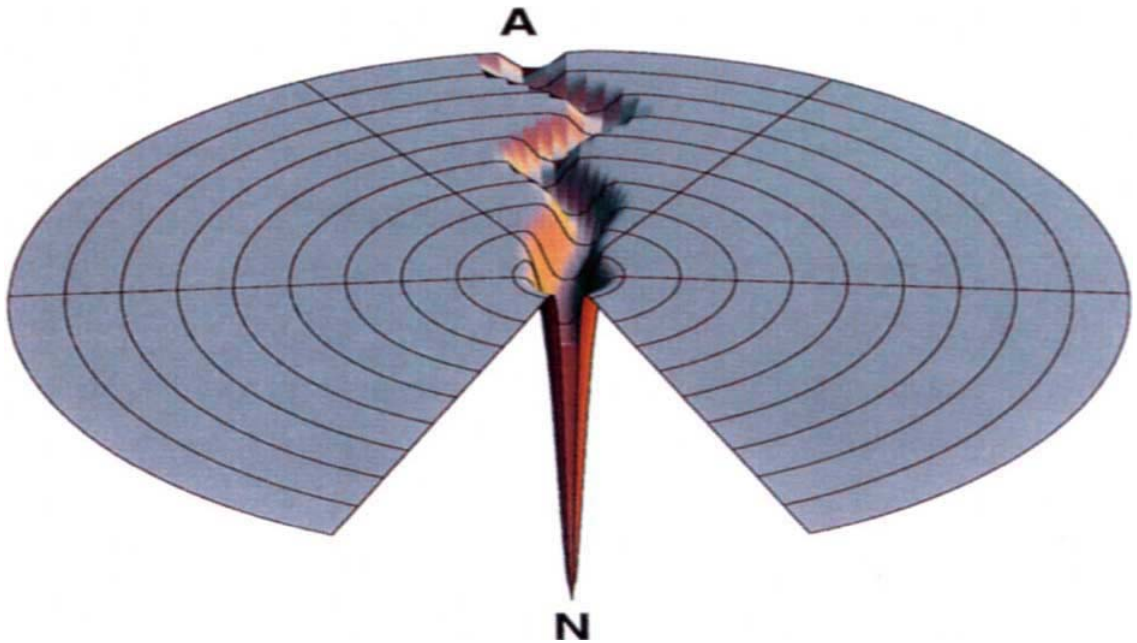
Illustration of Levinthal's paradox:



Dill 1997

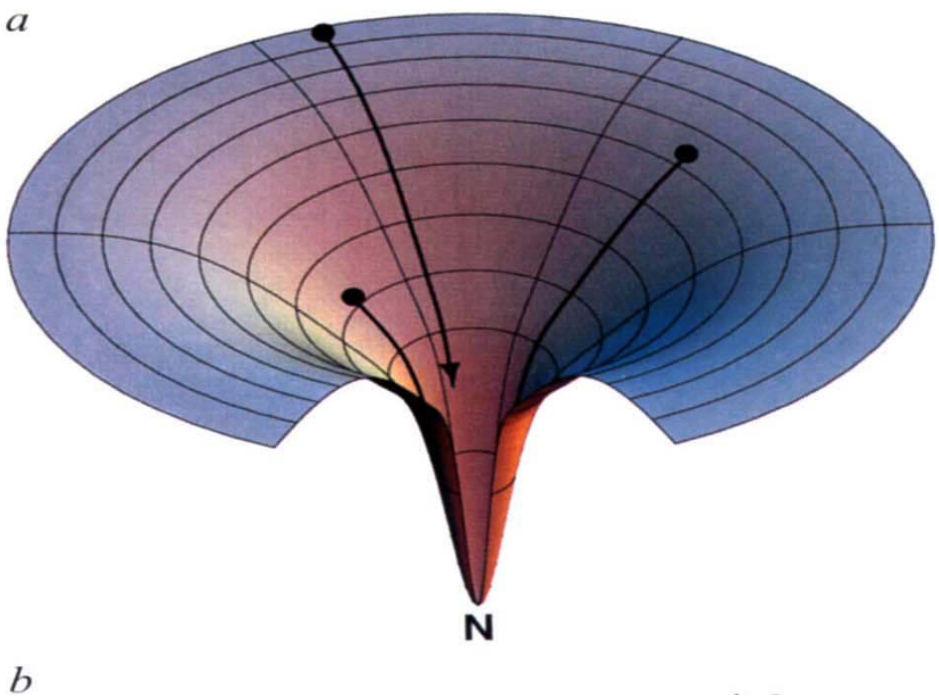
Problem: wide, flat space of conformations, only one native state.

People also thought about pathways like this:



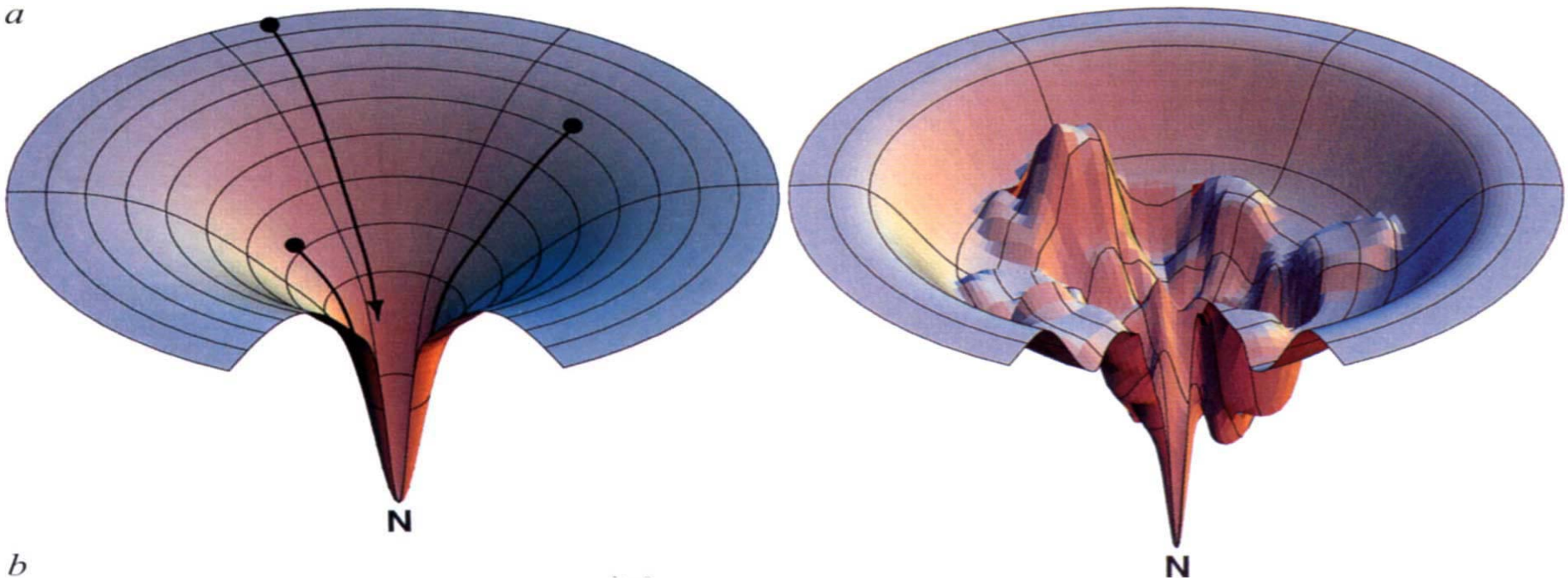
Dill 1997

Illustration of the solution of Levinthal's Paradox:



Dill 1997

Illustration of the solution of Levinthal's Paradox:

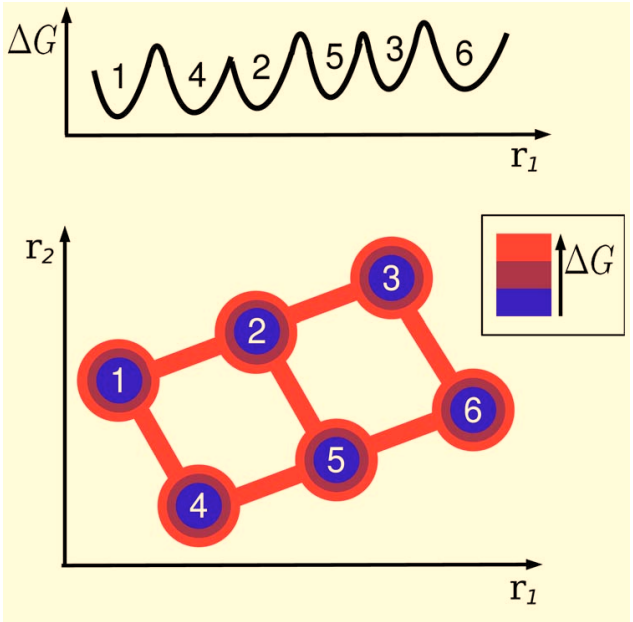


Dill 1997

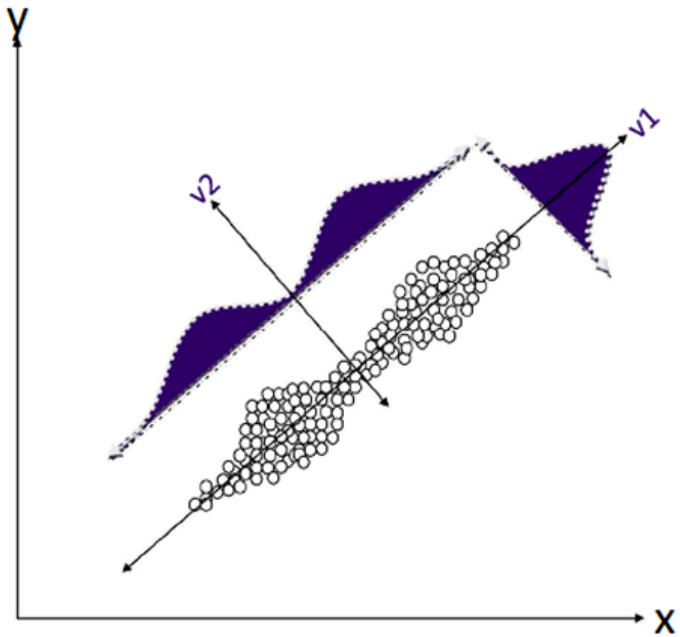
Constructing Energy Landscapes

Important question in illustrating Energy Landscapes: How to choose adequate coordinates for the axes?

- Natural reaction coordinates in the simple Zwanzig model could be the number of correct bonds.
- Be careful not to lose connectivities.



Altis, J. Chem. Phys., 128:245102, 2008



Remember Principal Component analysis (PCA):
Do a basis transformation via diagonalization of the covariance matrix σ_{ij}.
Project data into the direction of the greatest variance.

Illustrative example: Folding of villin headpiece

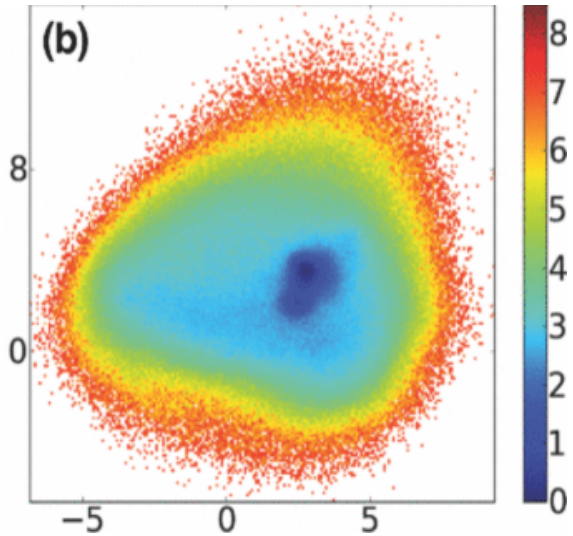


Figure : Energy landscape of the folding of the villin headpiece HP35 [Sittel, J. Chem. Phys., 141:014111, 2014]

In a cartesian coordinate PCA, one needs to separate overall and internal motion.

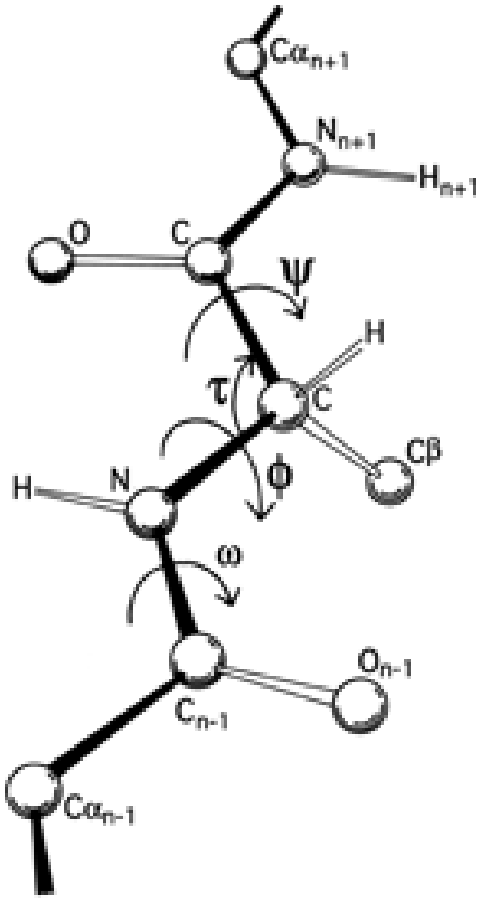
- Set center of coordinate system to the proteins center of mass $\sum_i m_i \vec{r}_i = 0$.
- Decompose atomic velocities $\vec{v}_i = \vec{u}_i + \vec{\omega} \times \vec{r}_i$ into a vibrational and an angular part. This leads to the kinetic energy

$$\begin{aligned} T &= \frac{1}{2} \sum_{i=1}^{N} m_i \left(\vec{u}_i^2 + (\vec{\omega} \times \vec{r}_i)^2 + \vec{\omega}(\vec{r}_i \times \vec{u}_i) \right) \\ &= T_V + T_R + T_{RV} \end{aligned}$$

- Remove the overall motion via $\vec{r'}_i = R\vec{r}_i$, by minimizing $\sum_{i=1}^{N} m_i (\vec{r'}_i - \vec{\bar{r}}_i)^2$ with reference structure $\vec{\bar{r}}$.

$\rightarrow$This works well only for small, rigid structures!

Alternative: Use internal coordinates such as dihedral angles Φ_i, Θ_i.
Transform those circular coordinates to linear ones:



David C. and Jane S. Richardson, 2000-2007

$$q_{2n-1} = cos\phi_n$$
$$q_{2n} = sin\phi_n$$

This figure shows clearly seperated free energy minima in a) which are not resolved in b).

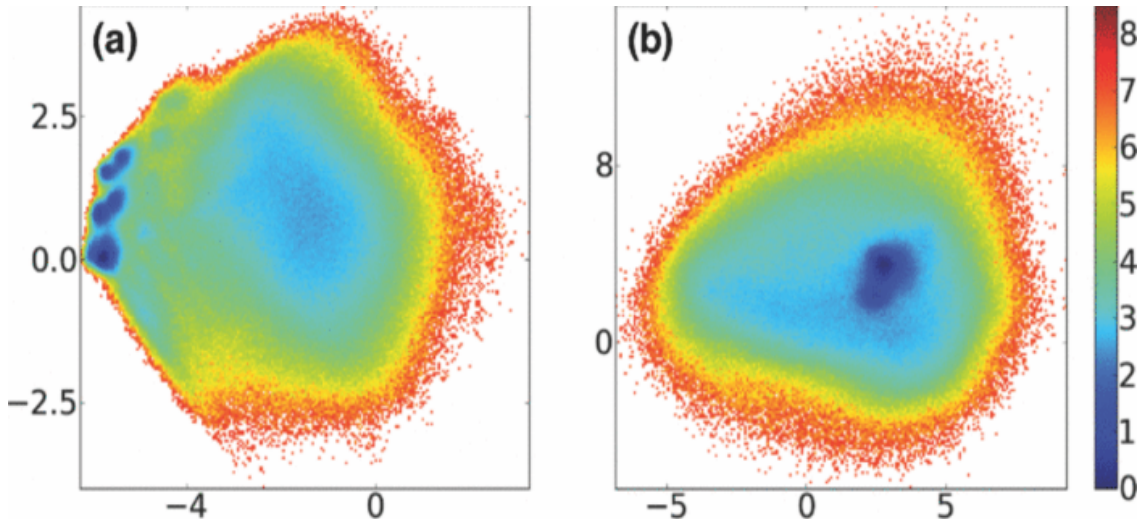


Figure : Free Energy landscape of the folding of HP35 obtained from a PCA using a) internal coordinates and b) cartesian coordinates [Sittel, J. Chem. Phys., 141:014111, 2014]

1. Levinthal's paradox isn't so paradox after all.
2. Energy landscapes are a good way to illustrate the folding process of a protein.
3. Energy landscapes can have different qualities and it can be hard to display the key features.


K. A. Dill.

Polymer principles and protein folding.

Protein Science, 8:1166–1180, 1999.


K. A. Dill and H.S. Chan.

From levinthal to pathways to funnels: The "new view" of protein folding kinetics.

Nature Structural Biology, 4:10–19, 1999.


K. A. Dill and J. L. MacCallum.

The protein folding problem, 50 years on.

Science, 338:1042–1046, 2012.


A. Jain F. Sittel and G. Stock.

Principal component analysis of molecular dynamics: On the use of cartesian vs. internal coordinates.

J. Chem. Phys., 141:014111, 2014.


A. Szabo R. Zwanzig and B. Bagchi.

Levinthal's paradox.

Proceedings of the National Academy of Sciences of the United States of America, 89:20–22, 1992.