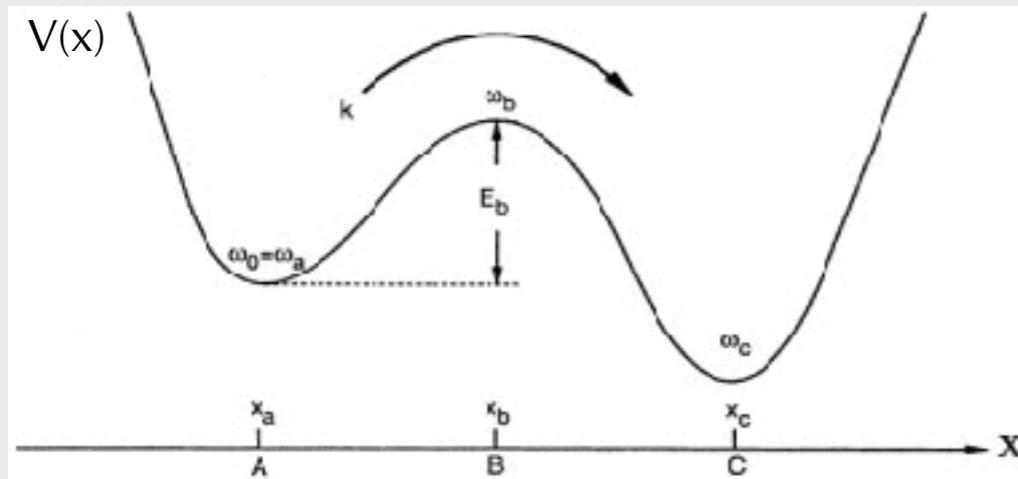# Characterizing energy landscapes:
## a survey

Charles H. Robert

CNRS Laboratoire de Biochimie Théorique

Institut de Biologie Physico-Chimique (IBPC), Paris

# Simple energy landscape

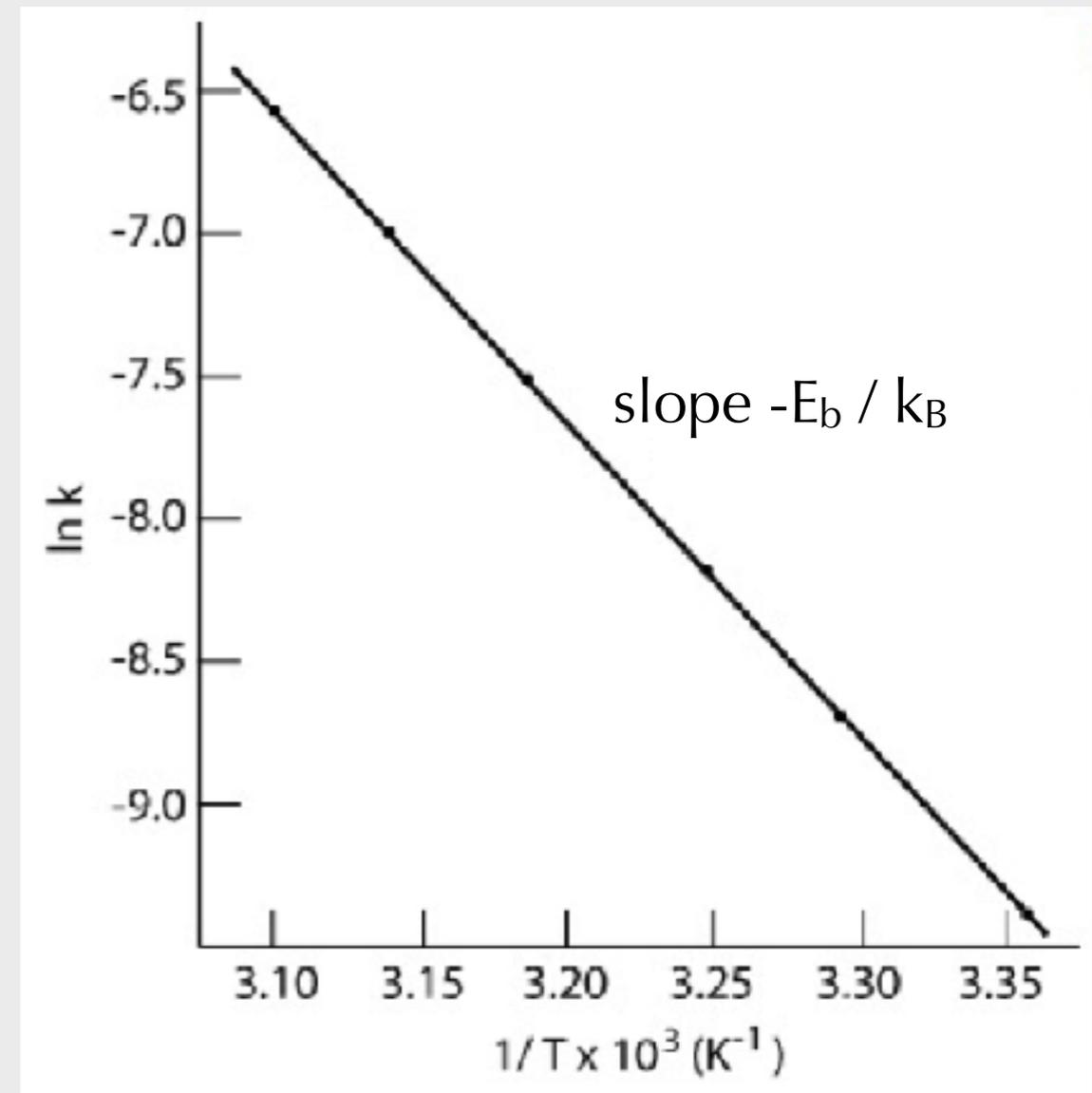## Potential energy minima A and C separated by a transition state B



Rate of forward reaction k, units time$^{-1}$
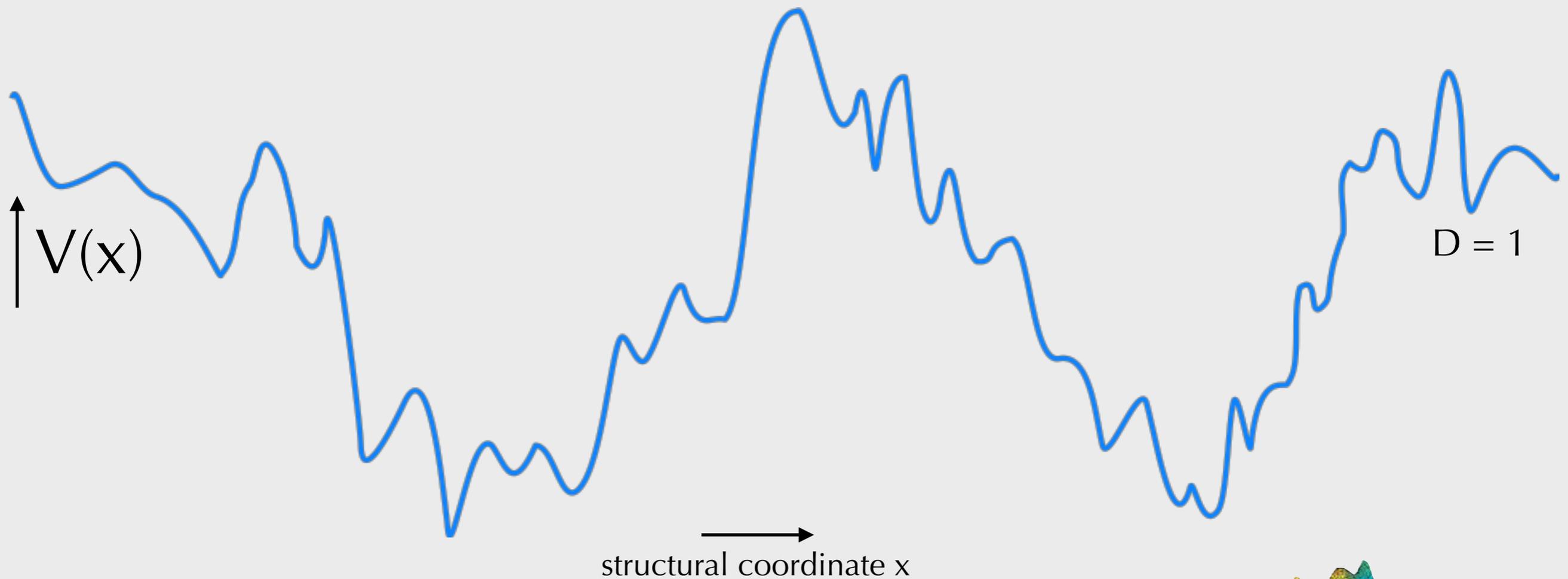
Arrhenius: plotting ln k vs 1/T yields straight line

$$k = \mathcal{A} \exp( - E_b(X) / k_B T )$$

- k : rate constant
- $E_b$ : barrier height (activation energy)
- $\mathcal{A}$ : prefactor

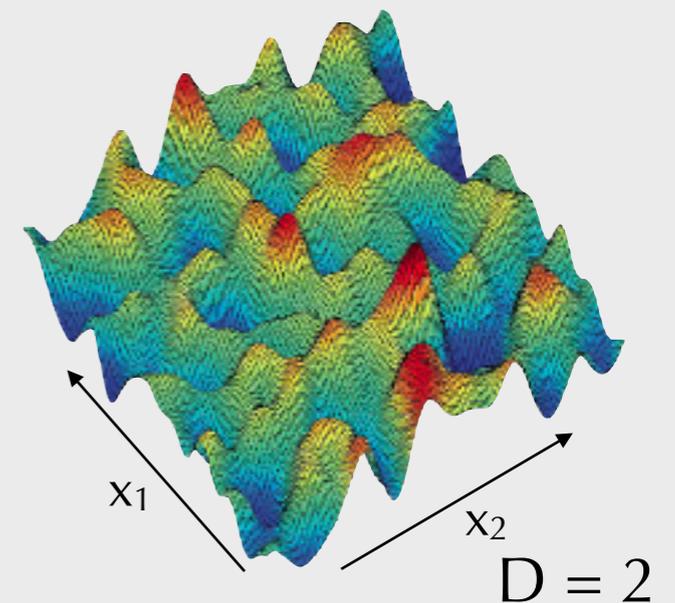Can estimate $\mathcal{A}$ using harmonic transition state theory



slope $-E_b / k_B$

# More complex situations

$V(x)$

D = 1

→ structural coordinate x

Potential energy V as a function of coordinates of N atoms

Defined over D = 3N dimensional space

x₁

x₂

D = 2

# An early energy landscape conundrum: Levinthal's paradox

Christian Anfinsen: Denature an enzyme, then renature it again

Native catalytic activity is regained

→ Native structure must be at a global energy minimum

Cyrus Levinthal: take 100 residue protein, assume 10 backbone states/residue

e.g., ~ phi, psi torsion angles in staggered positions

Number of possible states $\approx 10^{100}$

If 1 ps/state, exhaustive search >> 10 x age of the universe

Yet proteins do find the native state, on µs to s timescale

# Implicit assumption:
# "golf course" energy landscape

- Levinthal's search assumes all non-native states have equal probabilities
- Implies that the potential energy surface is flat except for the native state

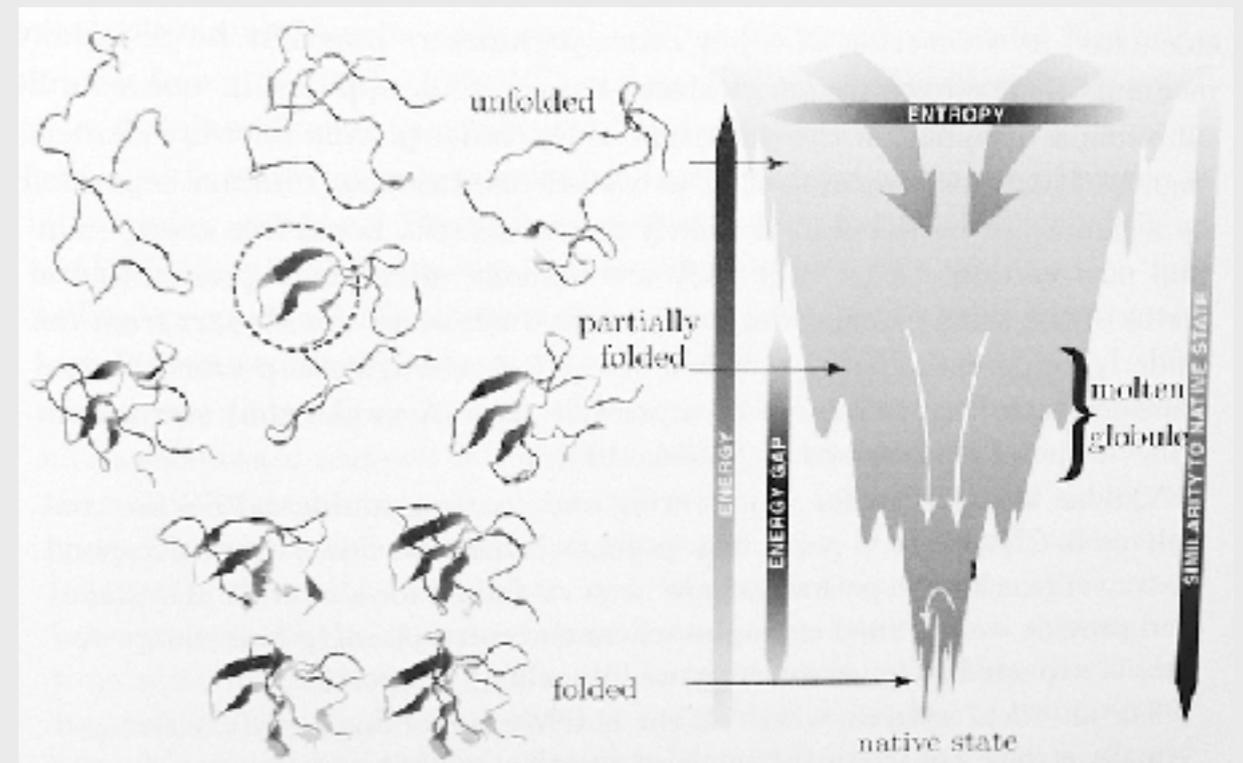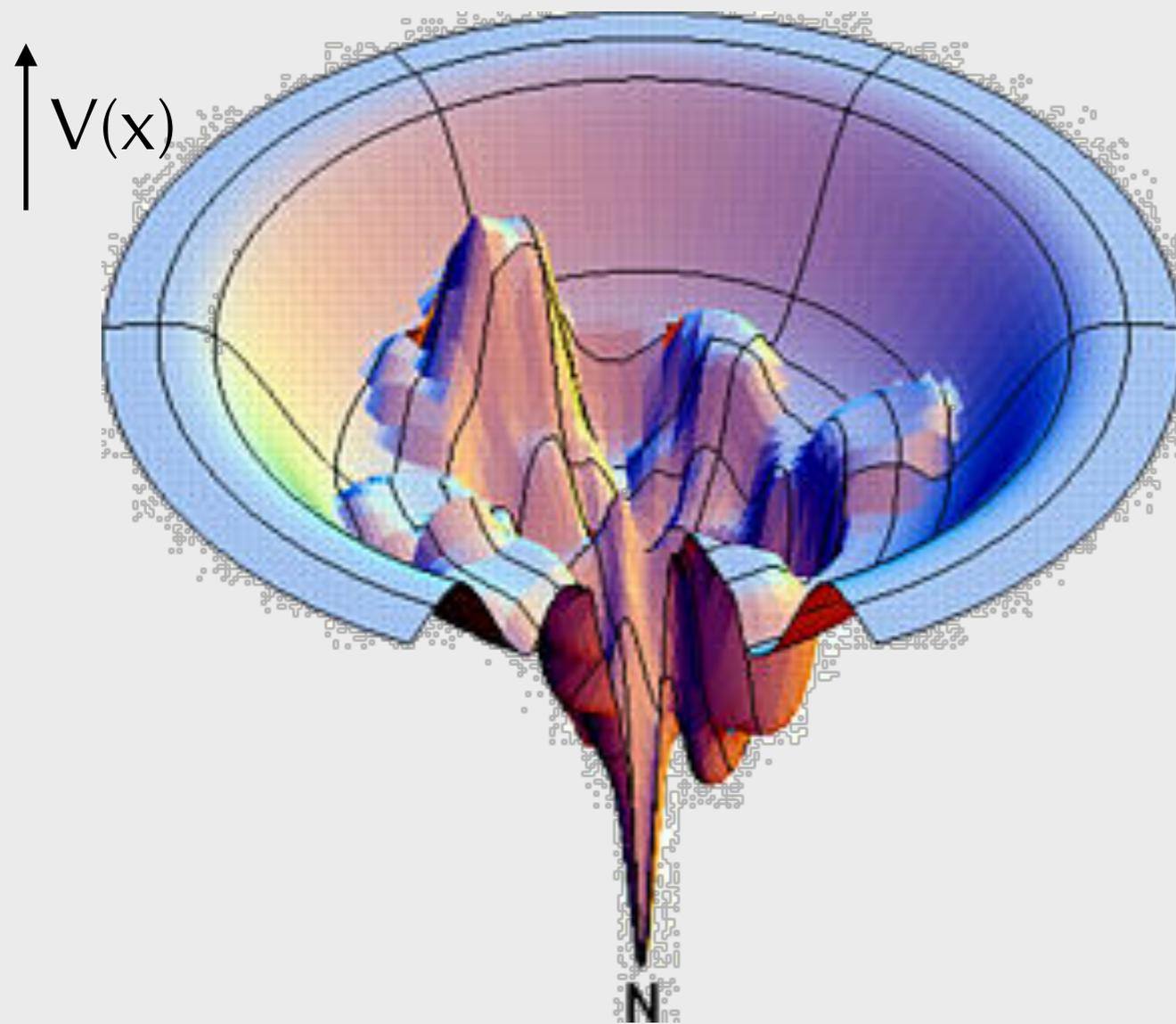Finding native conformation unlikely (like a "hole in one")
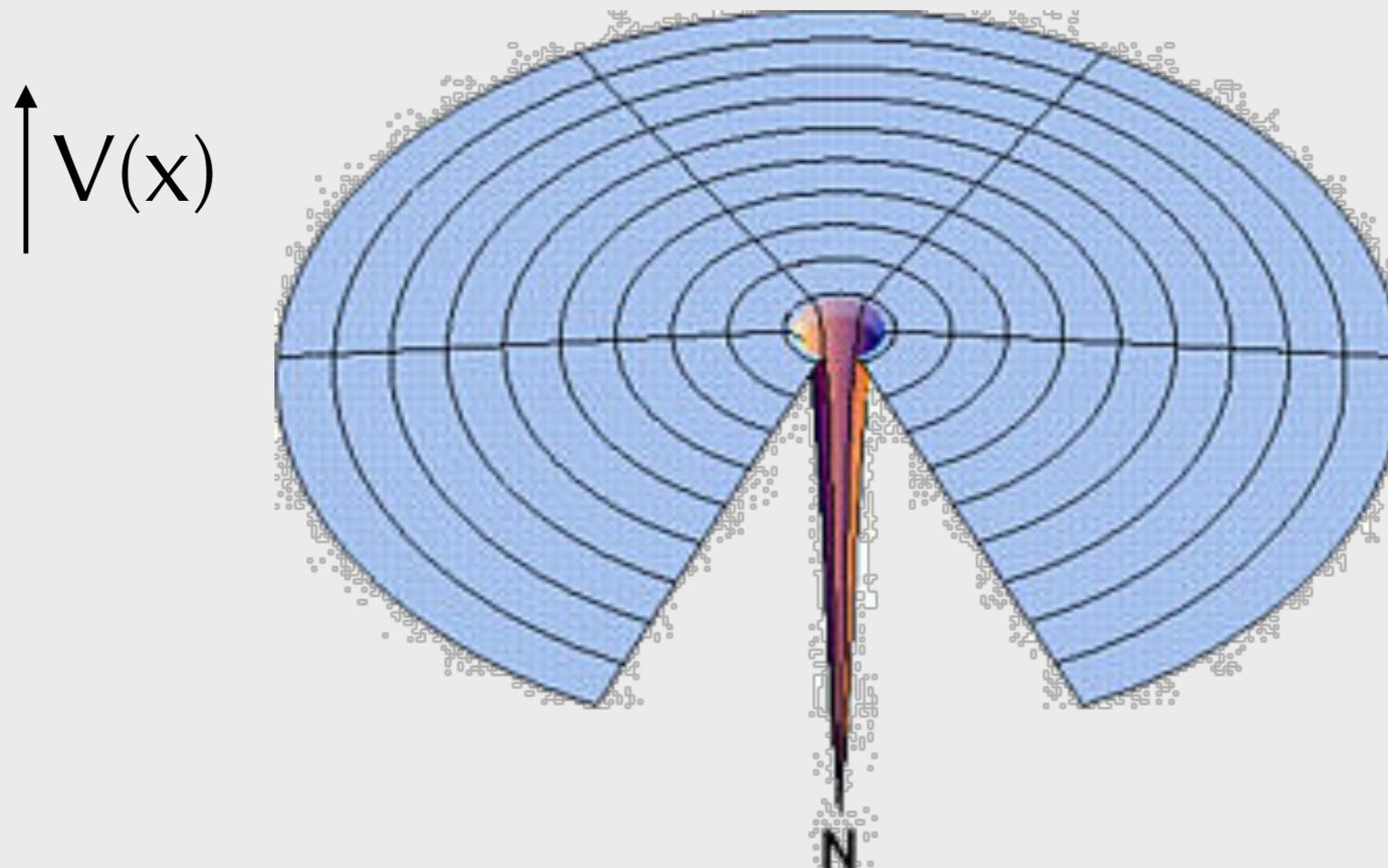
# Resolution of paradox: folding funnel

Local interactions are quickly explored

Native-like local interactions are lower energy than the rest*

Energy decreases as the structure approaches the native form



$V(x)$

N

unfolded

partially folded

folded

native state

ENTROPY

ENERGY

ENERGY GAP

SIMILARITY TO NATIVE STATE

molten globule

* evolution selected them that way

Ken Dill, Peter Wolynes, ...

# A better golf course

# Funnelled ELs are characteristic of self-structuring systems

## Crystals

- Single, ordered state from highly diverse solution

## Many biological systems…

- Folding of globular proteins / RNA
- Specific protein interactions
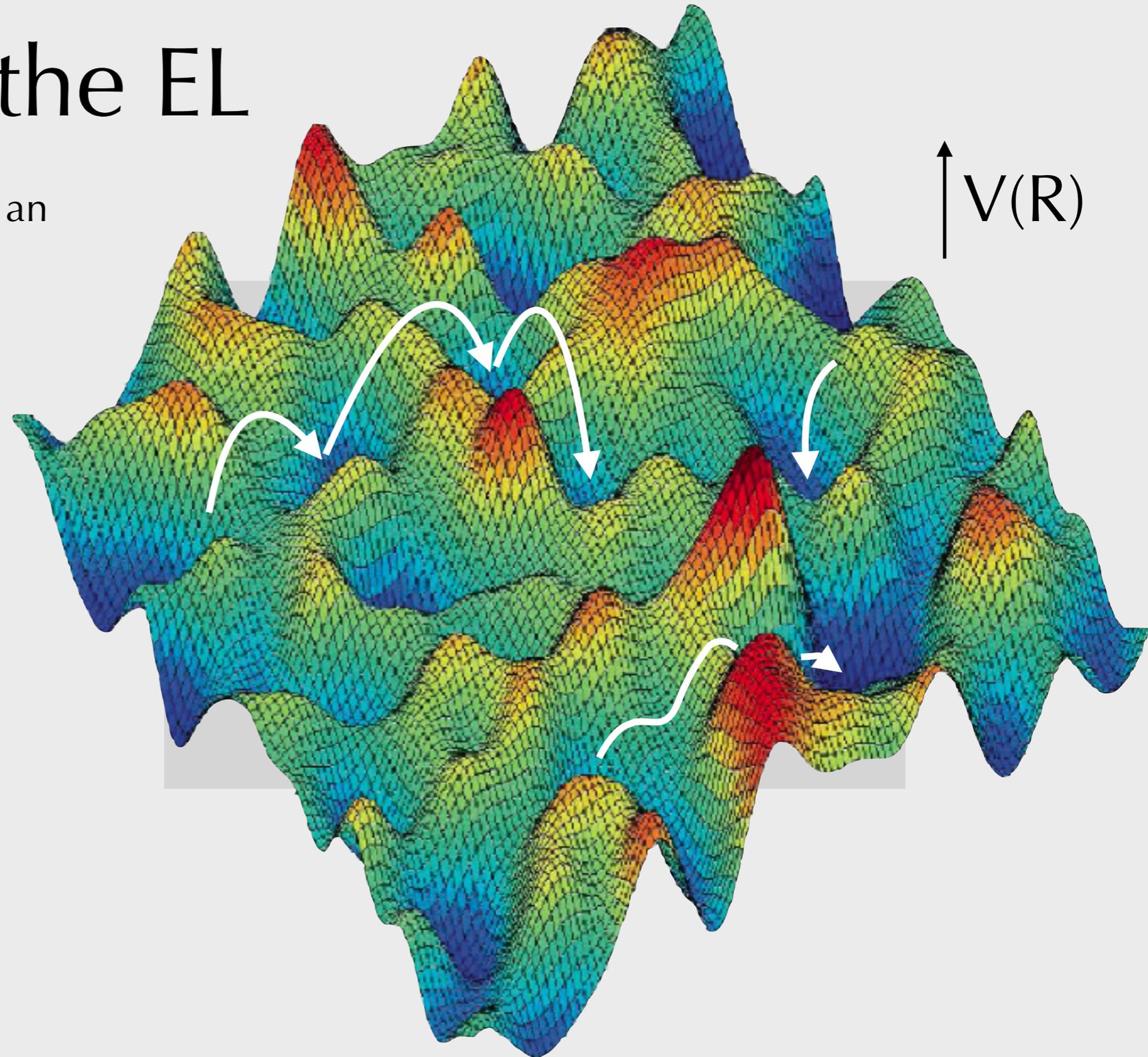- Self-assembly (fibers, virus capsids, …)

## …but not all

- Misfolding/alternative structures (amyloids, …)
- Intrinsically Disordered Proteins/Regions = ID[PR]s (promiscuous interactions, degradation pathways, …)
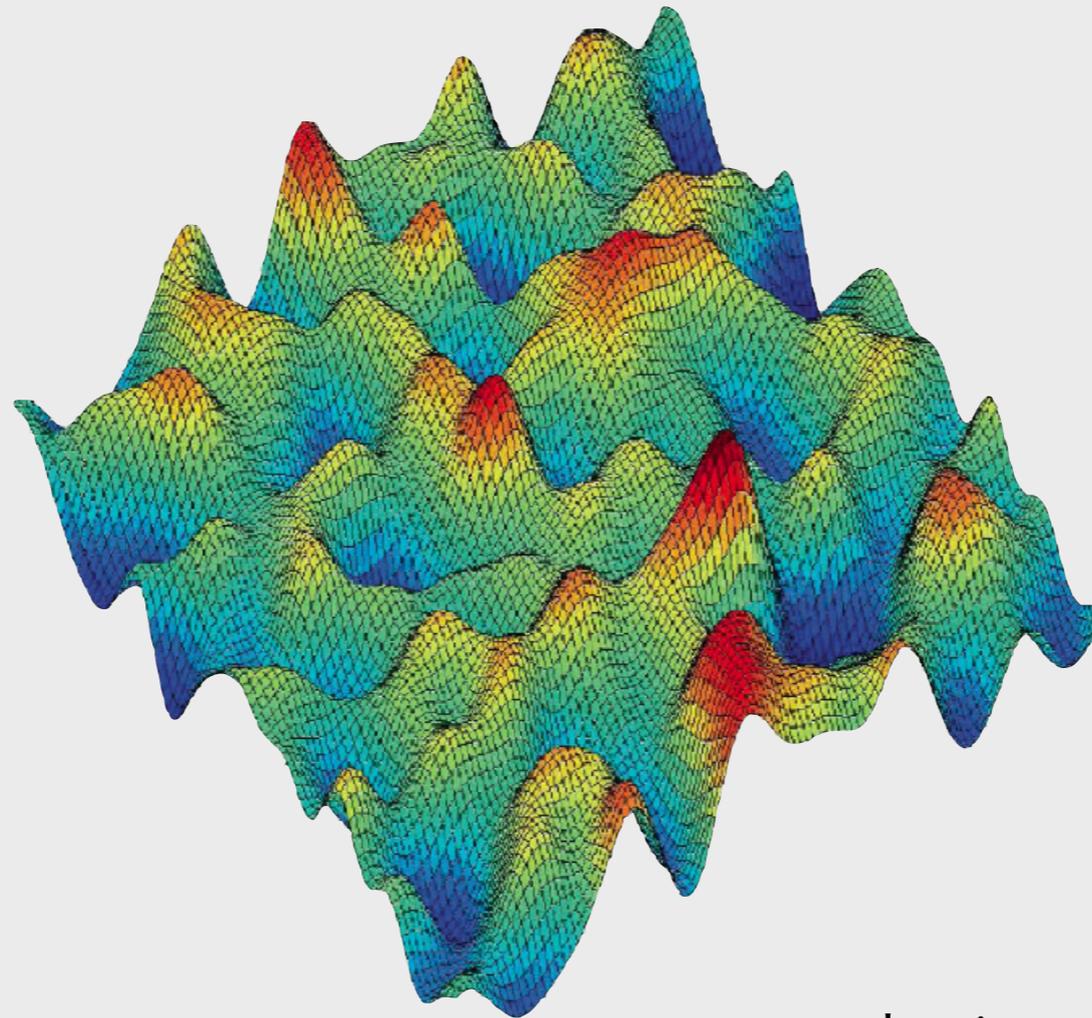
# Exploring the EL

Where does knowledge of an
EL come from?

Optimisation
  minimization, …
Local characterization
Free exploration
Guided exploration



$V(R)$

# Landmarks on an EL



$\uparrow V(R)$

Critical points: gradient = 0
    **minima**
        index 0: no negative eigenvalues in Hessian
    **saddle points**
        index 1, 2, … 3N-1 (number of negative eigenvalues)
    **maxima**
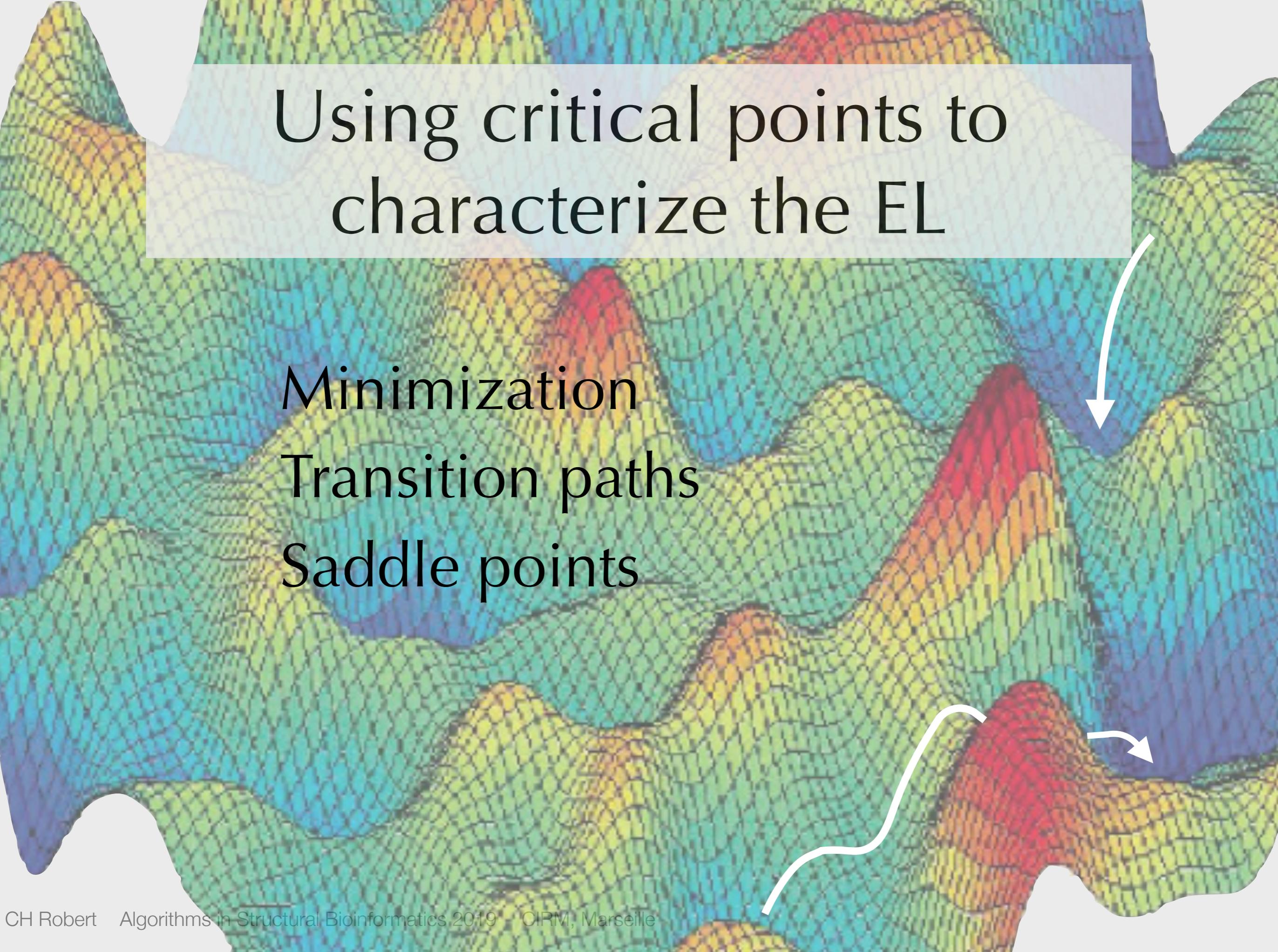        index 3N: negative eigenvalues only

Inflection points: curvature = 0

← number increases exponentially with D

←barriers between minima

…

←may confuse some exploration algorithms

# Using critical points to characterize the EL

Minimization

Transition paths

Saddle points

# Minimization



1st derivative (gradient) approaches
 steepest descent (-gradient of V)
 conjugate gradients (keep list of productive directions)

2nd derivative (curvature of energy surface) approaches
 Approximate curvature matrix (Hessian, also used for NM)
 Would find minimum in one step if surface were quadratic
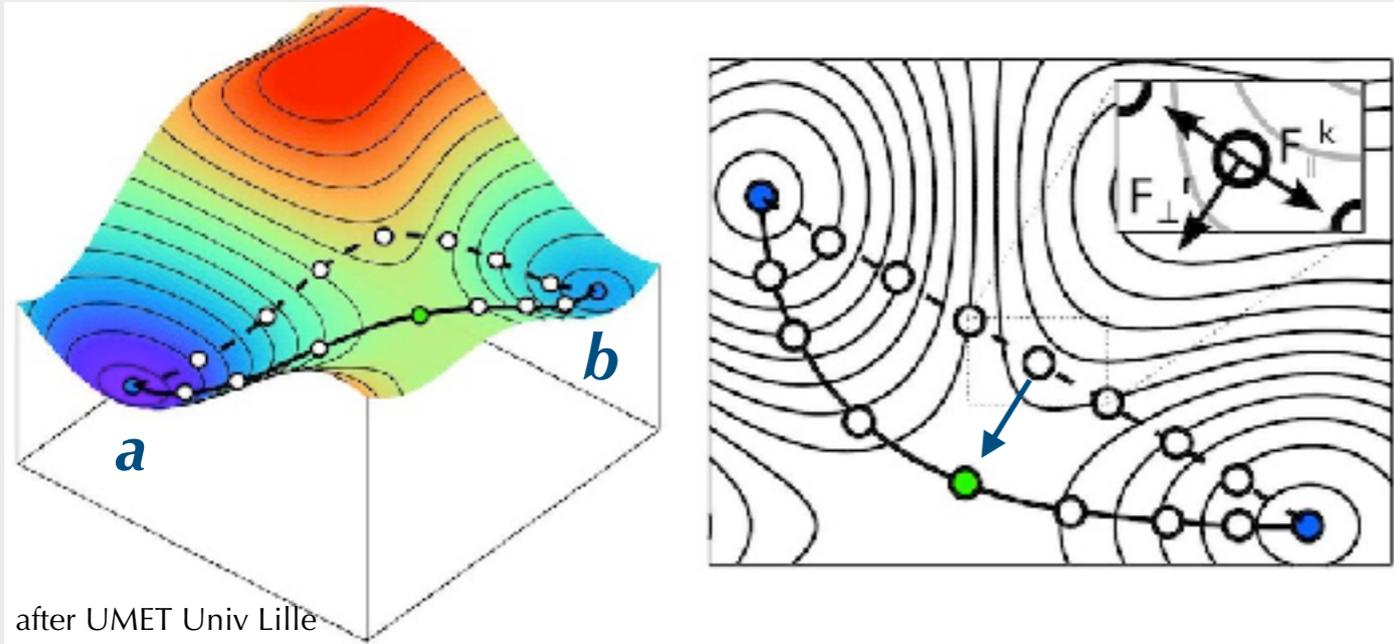   ○ For a real surface, very useful once we are near the minimum
 e.g., BFGS

Quality of minimization judged by magnitude of gradient
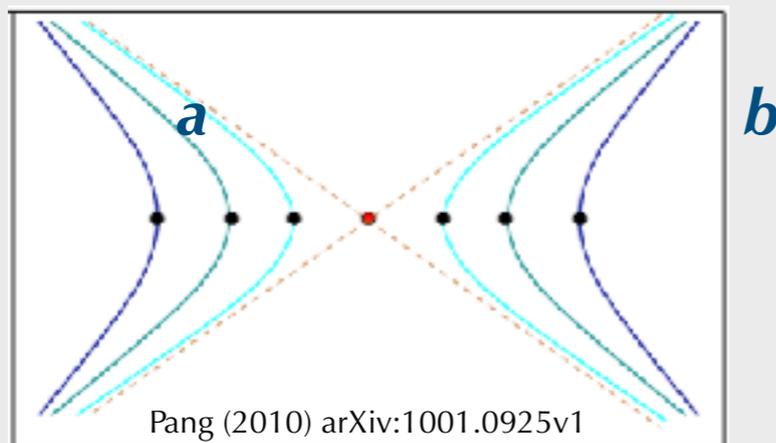 would be exactly zero at true minimum

# Transition paths


after UMET Univ Lille




Pang (2010) arXiv:1001.0925v1
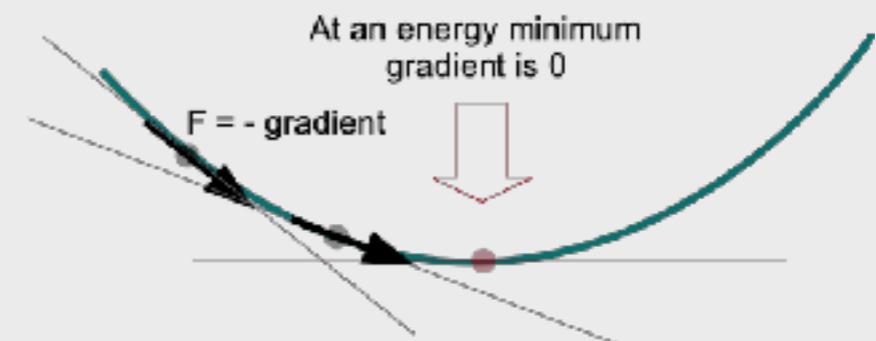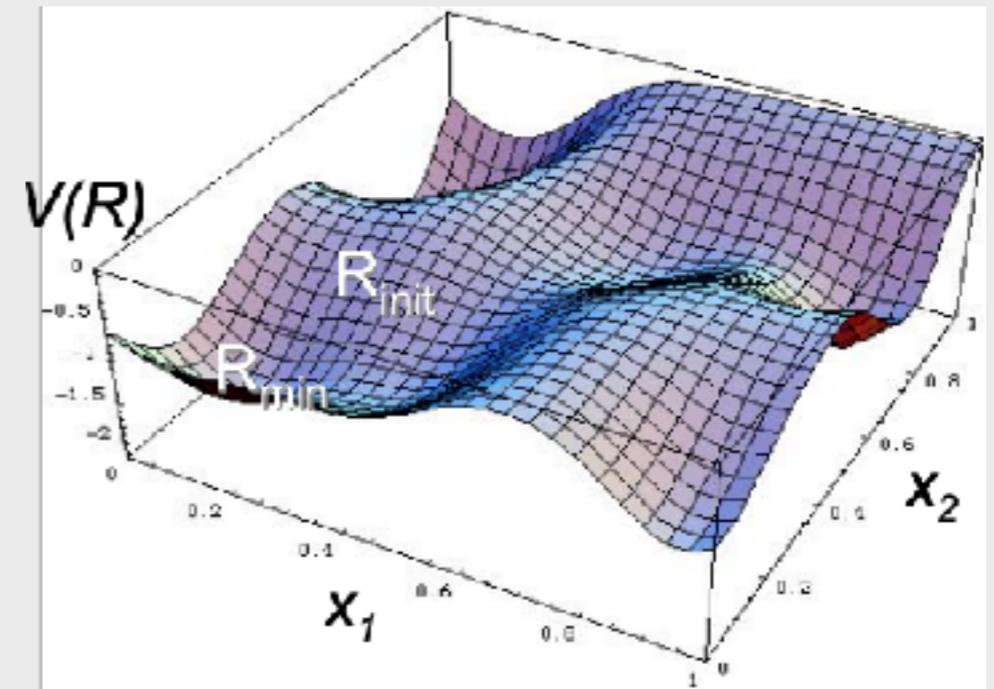
Nudged Elastic Band
- "mountain pass problem" (index 1 saddle points)
- Guess initial points between fixed end-points $a$ and $b$ (e.g., by interpolation)
- Connect points by springs
- Minimize cost function $c(F_\perp, F_{||})$ for the chain

Finding saddle points (an example)
- Use connectivity criteria between $a$ and $b$ in level sets
- Critical value is highest level set for which $a$ and $b$ are not connected
- General Morse index saddles

# Picture the EL through disconnectivity graphs



$G(\Phi)$  $\Phi$

energy "lids"

$V_o + 2\,\partial V$

$V_o + \partial V$

$V_o$
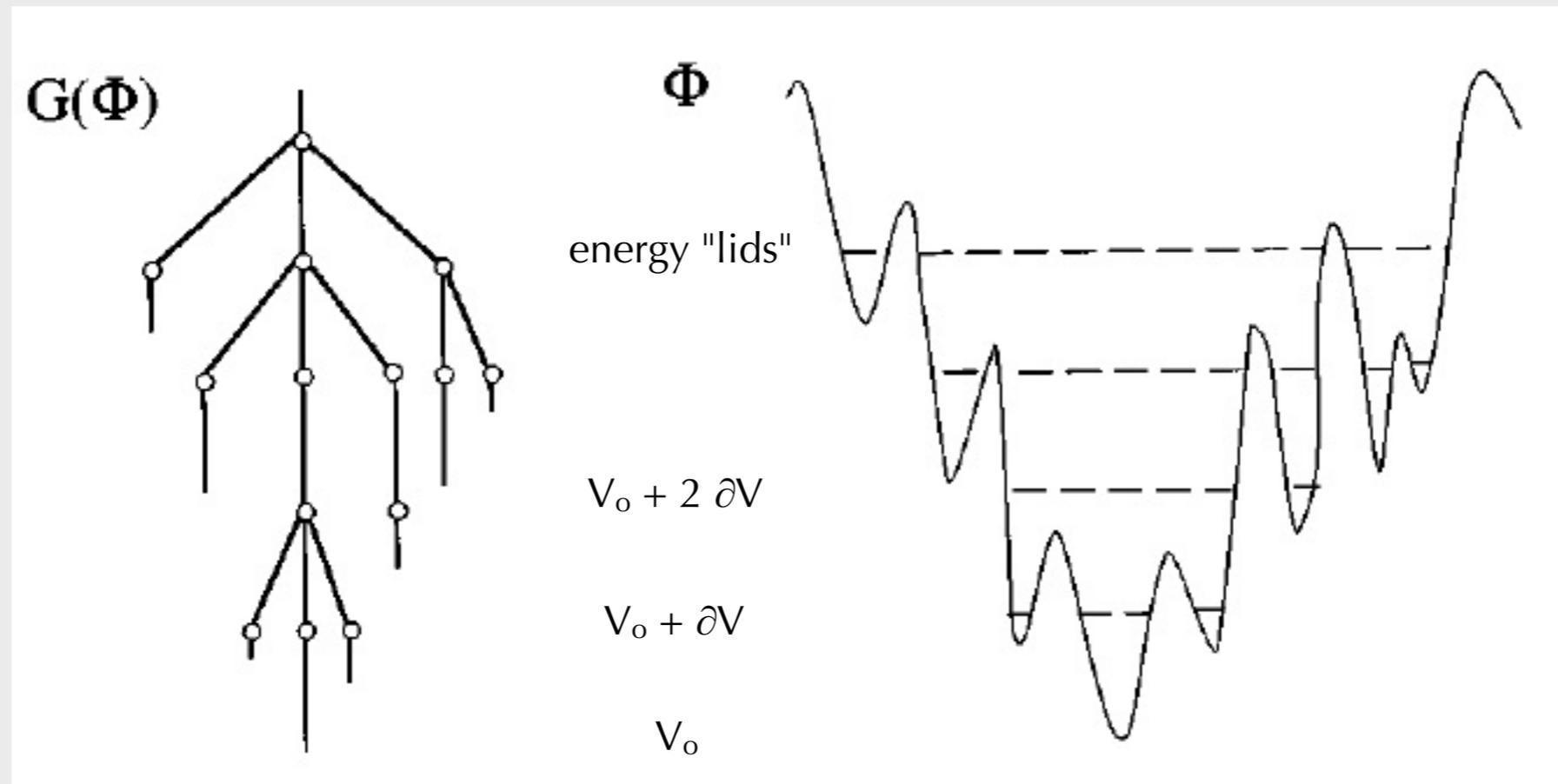
Becker and Karplus (1997) J Chem Phys

Transform connected minima and saddle points into tree structure

Minima separated by the lowest index-1 saddle between $V(j)$ and $V(j+1)$ are connected at level j
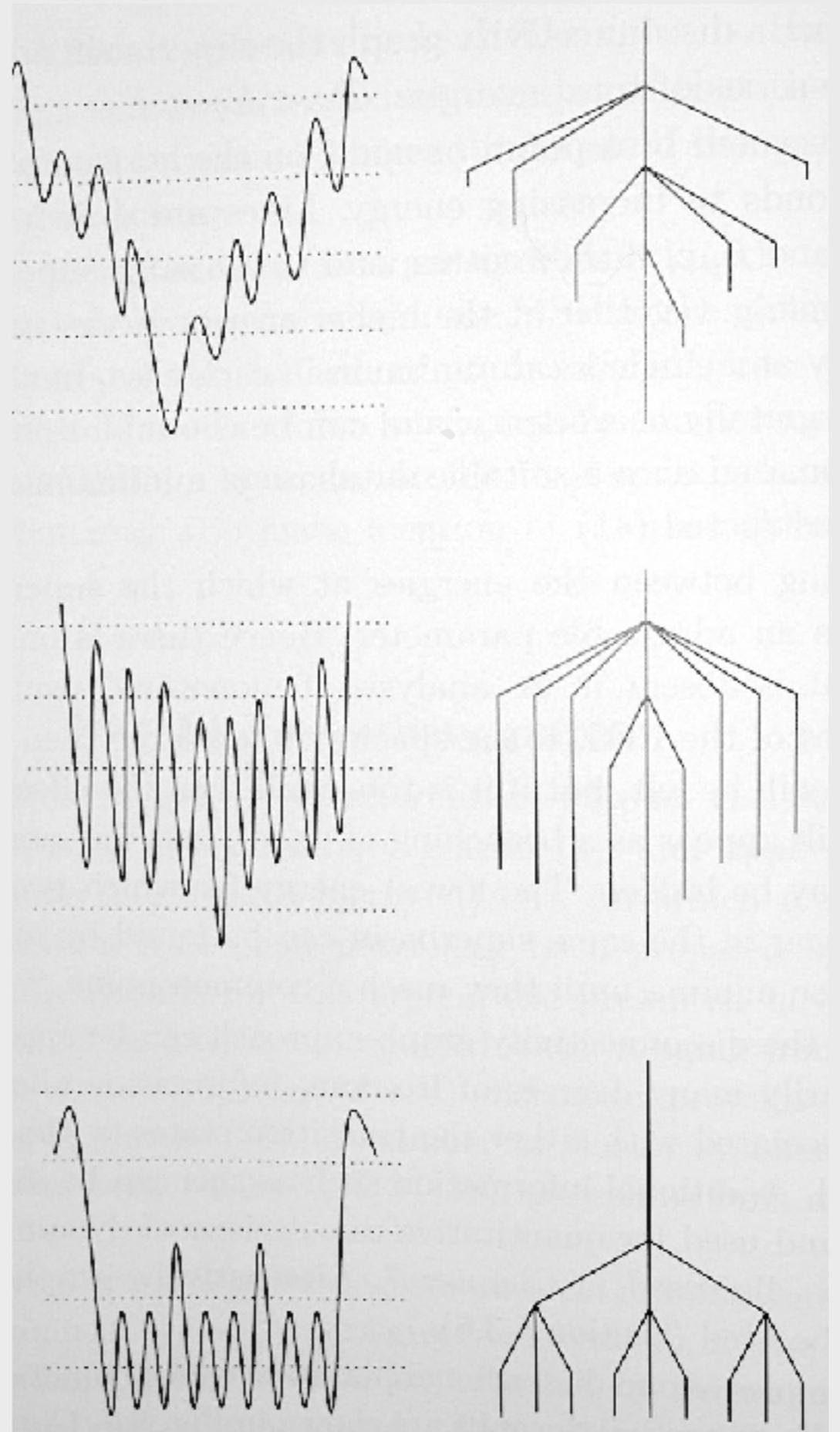
Minima coalesce into connected components (super-basins) at higher and higher energies

# Examples of disconnectivity graphs

Different basin interconnectivities give different classes of landscapes

↓

Different classes of dynamics

Wales (2003) *Energy Landscapes*    15
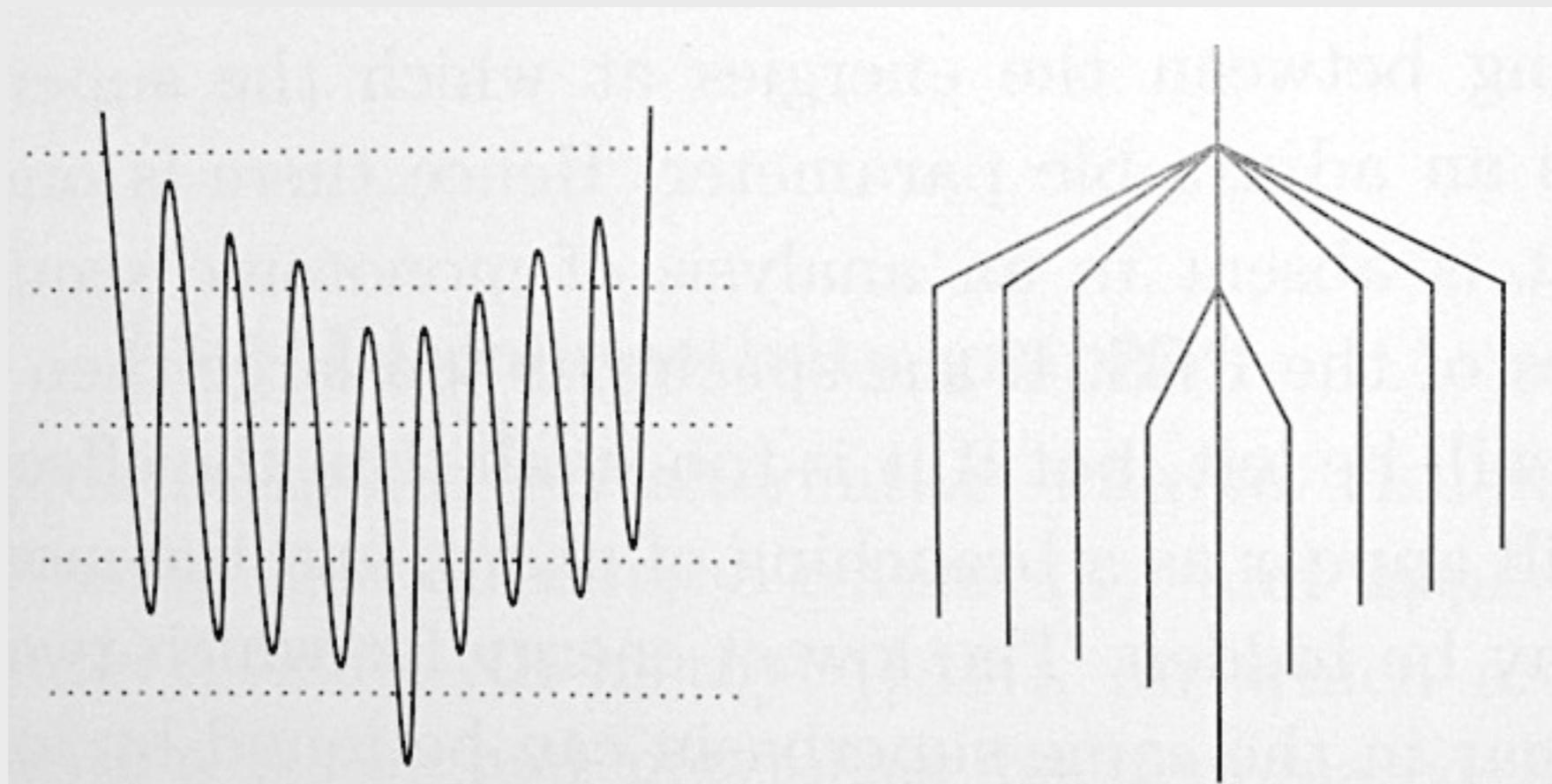
# Non-funneled landscapes: frustration

Energy landscape prevents system from attaining the global energy minimum

- High temperatures— rapid configurational exploration
- Low temperatures— trapping
- → Global energy minimum difficult to reach

# Frustrated systems: glassy behavior

Simple example: cooling of a binary Lennard-Jones system using MD

Plot average energy of quenched snapshots ("inherent structures")

Slower cooling allows accessing lower energy states, but

- amorphous
- no "native" structure
- global minimum not relevant



$<U>$

cooling rate=$1.08 \times 10^{-3}$
cooling rate=$2.70 \times 10^{-4}$
cooling rate=$8.33 \times 10^{-5}$
cooling rate=$3.33 \times 10^{-6}$

slower cooling
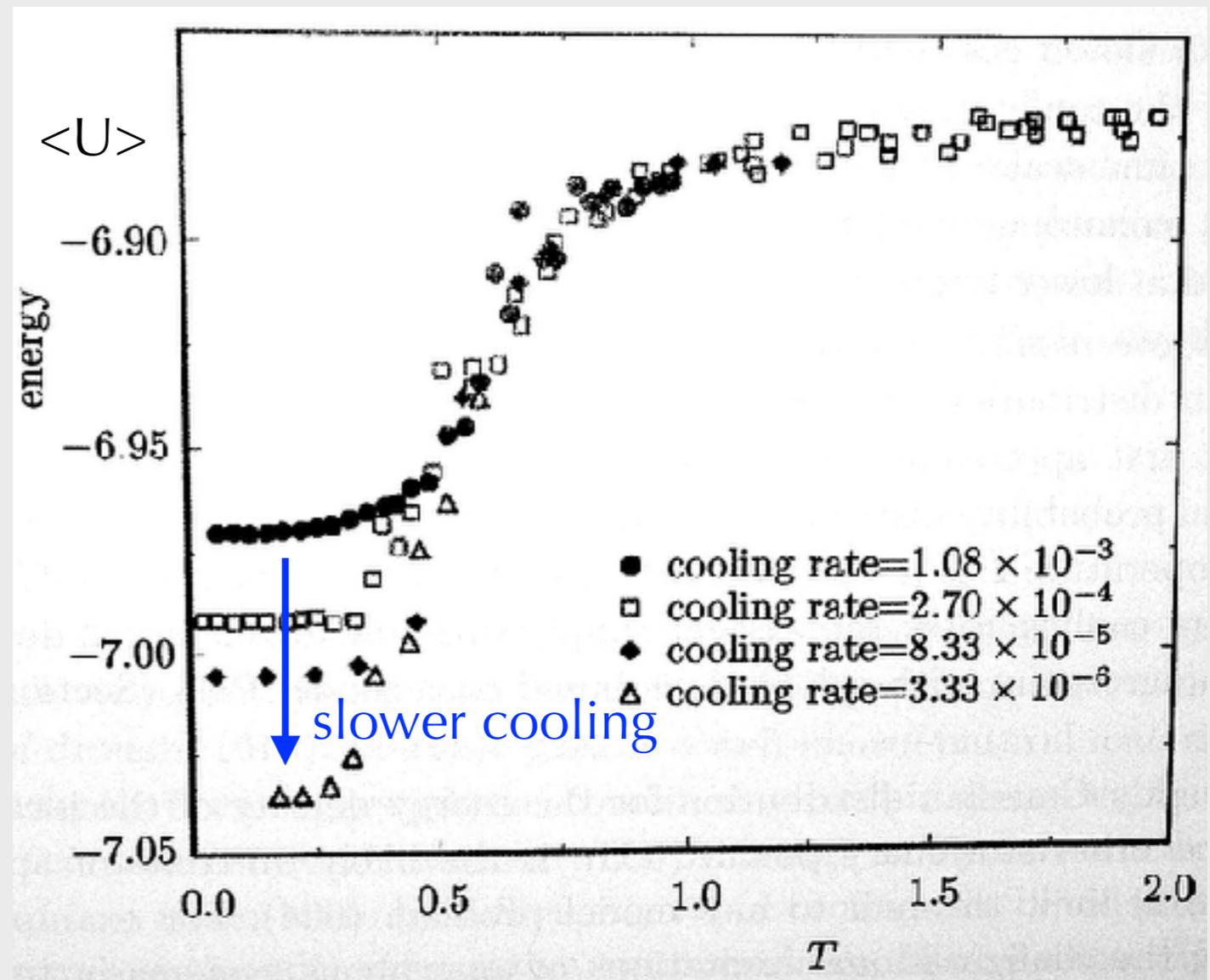
Sastry, Debenedetti, and Stillinger (1998) Nature 393, 554

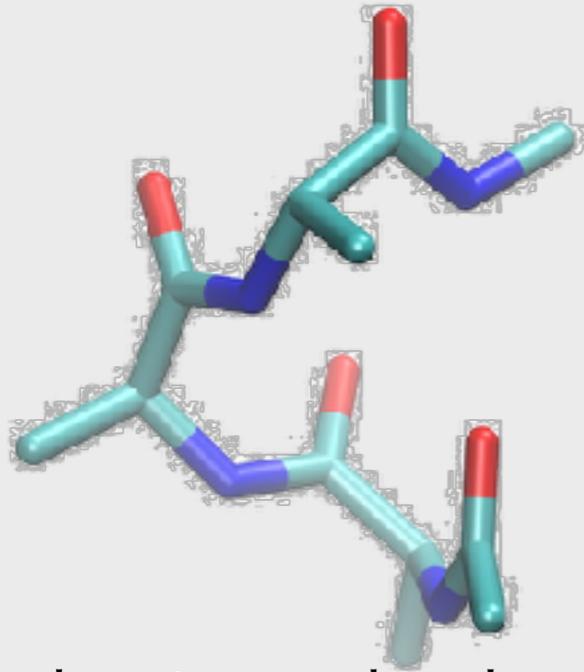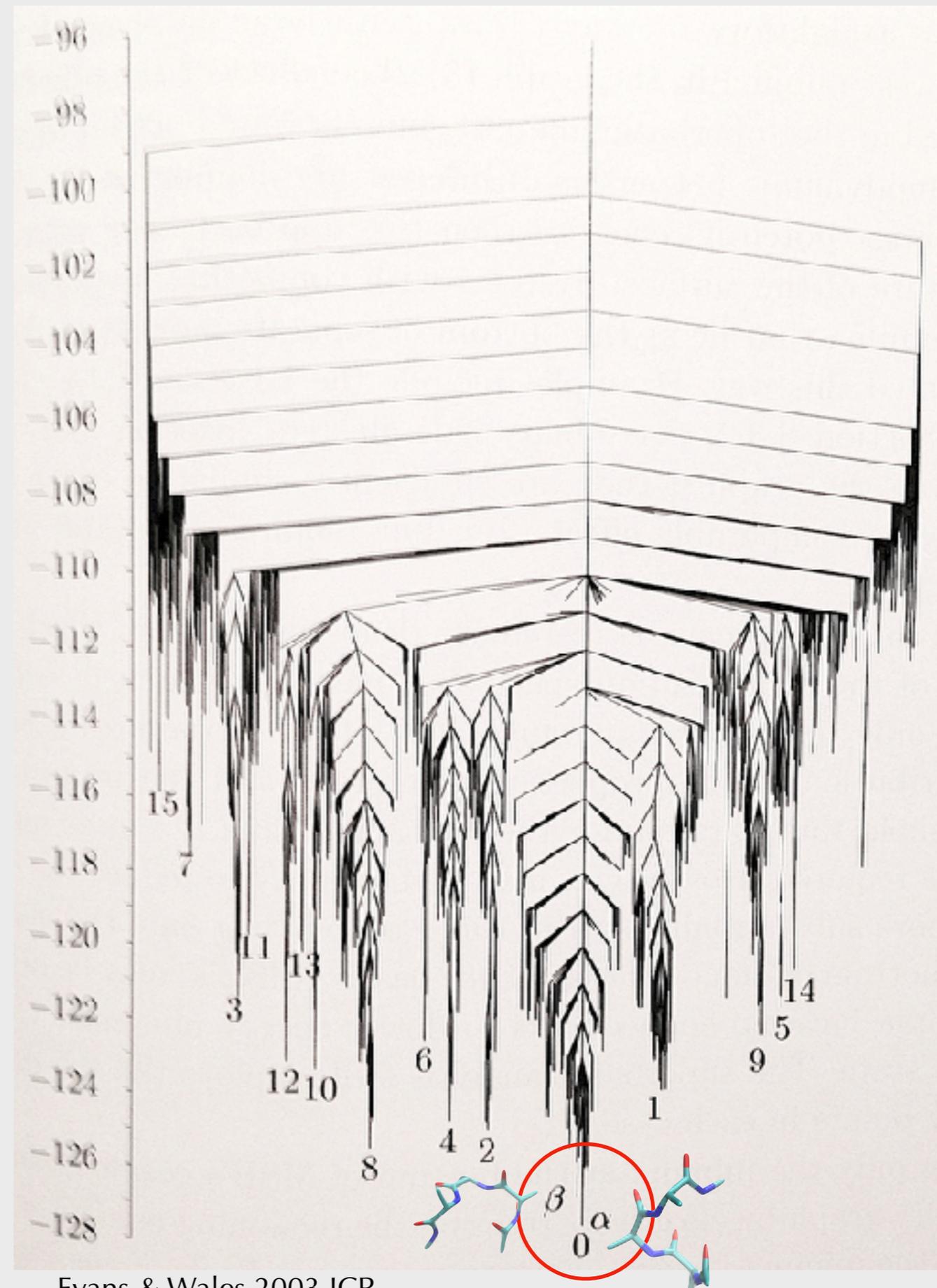# Disconnectivity graph for Alanine tetra-peptide Ac-(Ala)$_3$-NHMe



Blocked amino and carboxy termini: four peptide bonds

Can form one H-bonded turn of an alpha helix

Extended ($\beta$) and helix ($\alpha$) approximately equal potential energy

Evans & Wales 2003 JCP

# Limitations of disconnectivity graphs
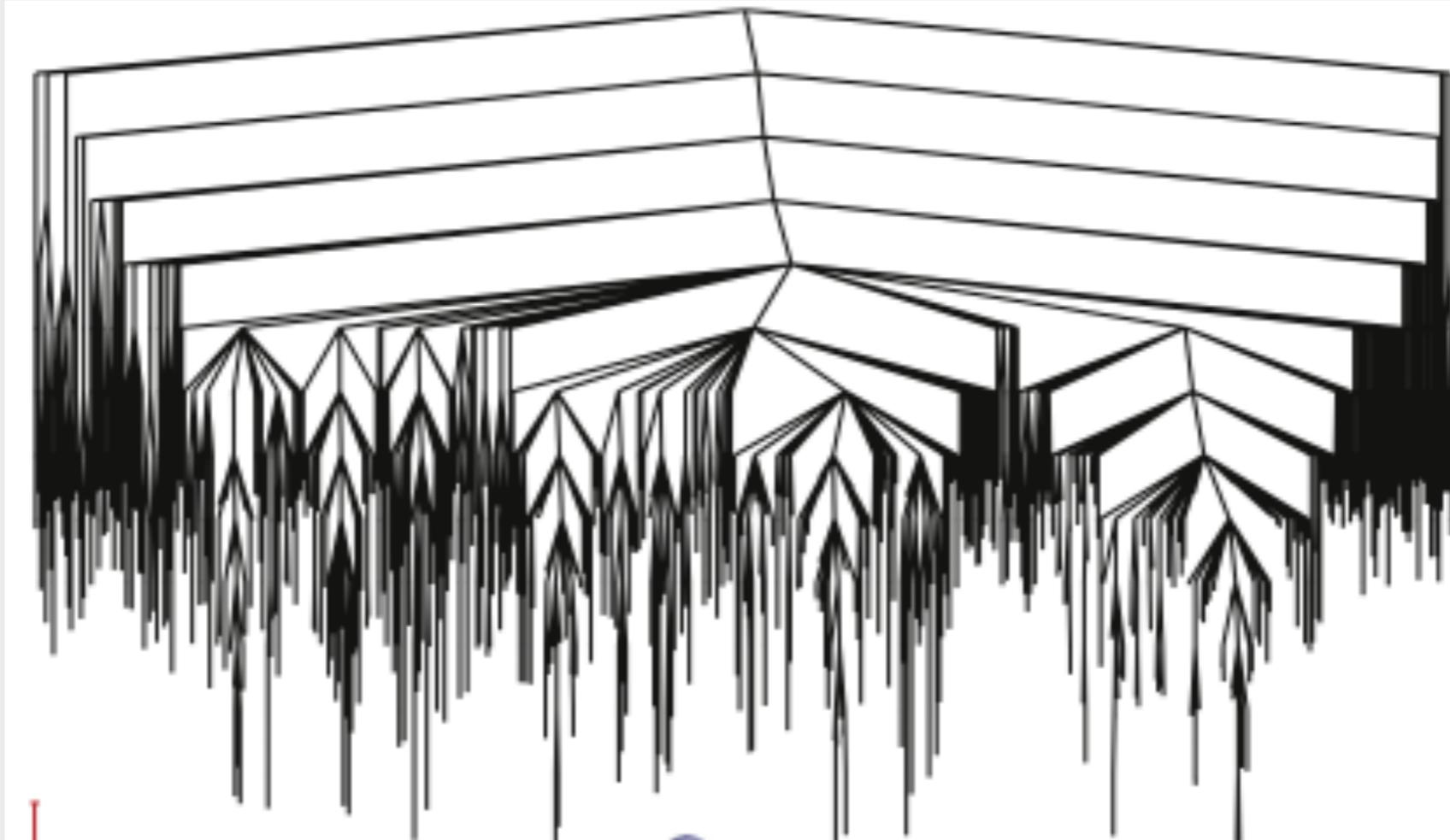


Inconsistency: Minima shown exactly but saddle points only shown within ΔE

Only lowest transition states are seen

Multiple transition states can be hidden in an energy step

  expand ΔE: gain visibility, lose transitions
  reduce ΔE:gain transitions,lose visibility

# Persistence analysis



> ▷ Rationale
> – Sub-level sets: $M_t = f^{-1}(-\infty, t), t \in \mathbb{R}$
> – Birth and death of the
>   connected components of $M_t$
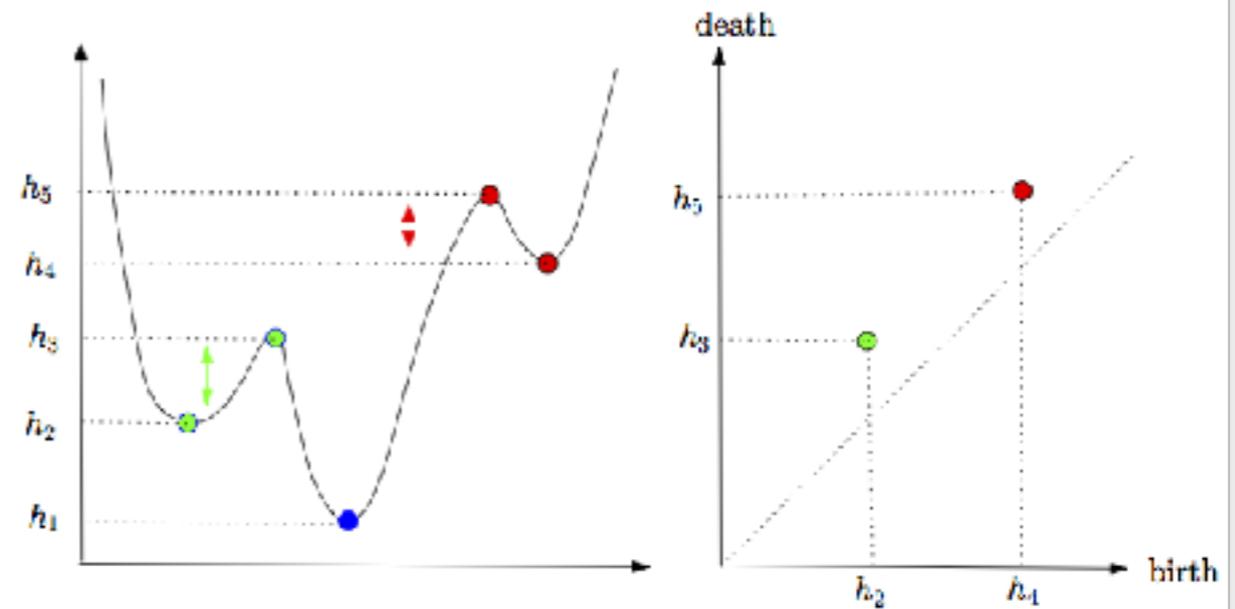
F. Cazals, course slide

Analyze connected components of sub-level sets as function of height (energy)

Plot *birth* of each additional component on *x, death* on *y*

- see the exact ***persistence*** of each component (height of saddle) directly
- natural mechanism to **denoise** landscape (*persistence < threshold*)

in the SBL
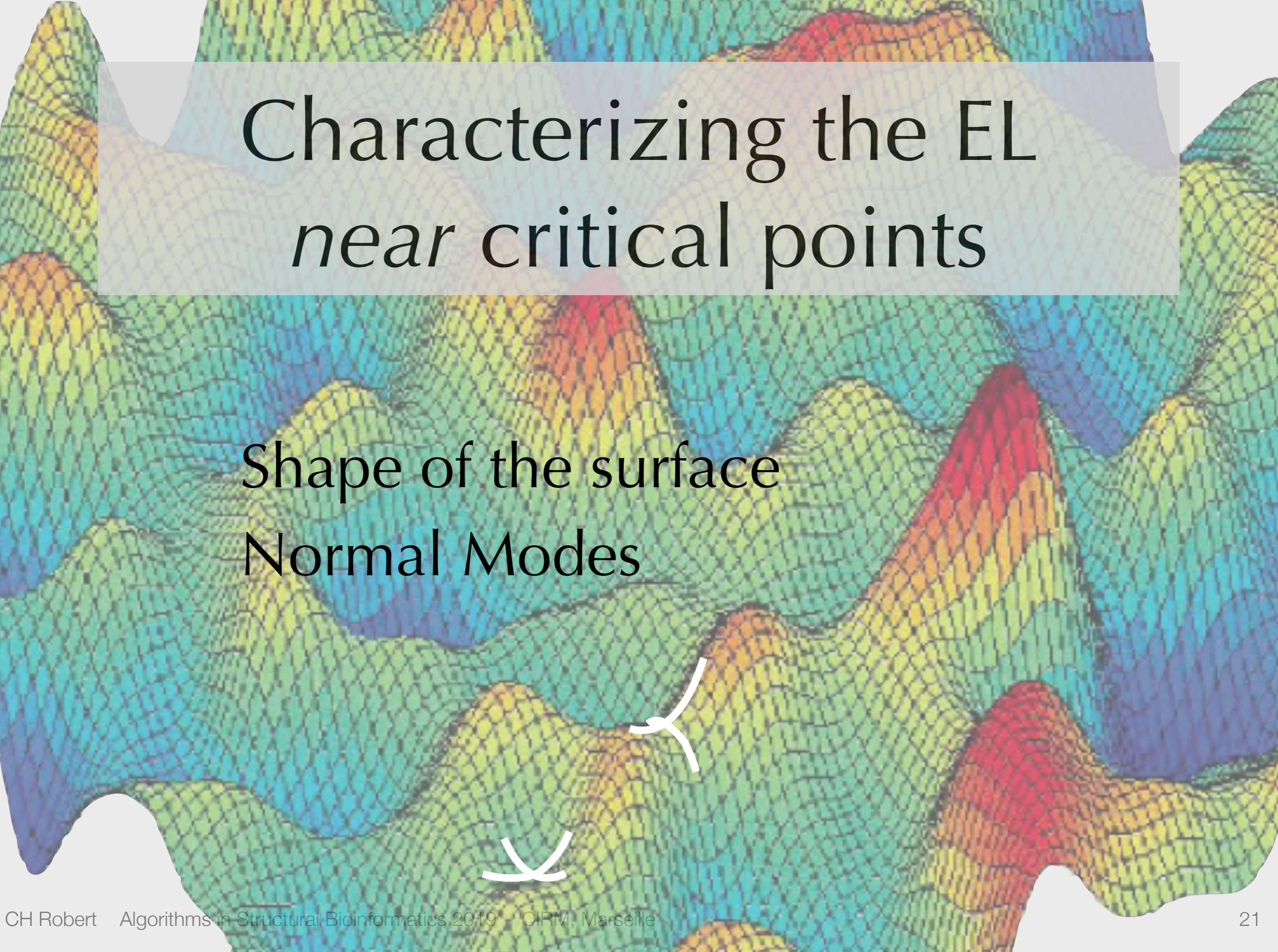
# Characterizing the EL
## *near* critical points

Shape of the surface
Normal Modes

# Idea behind NM



V(R)

Harmonic approximation to potential energy
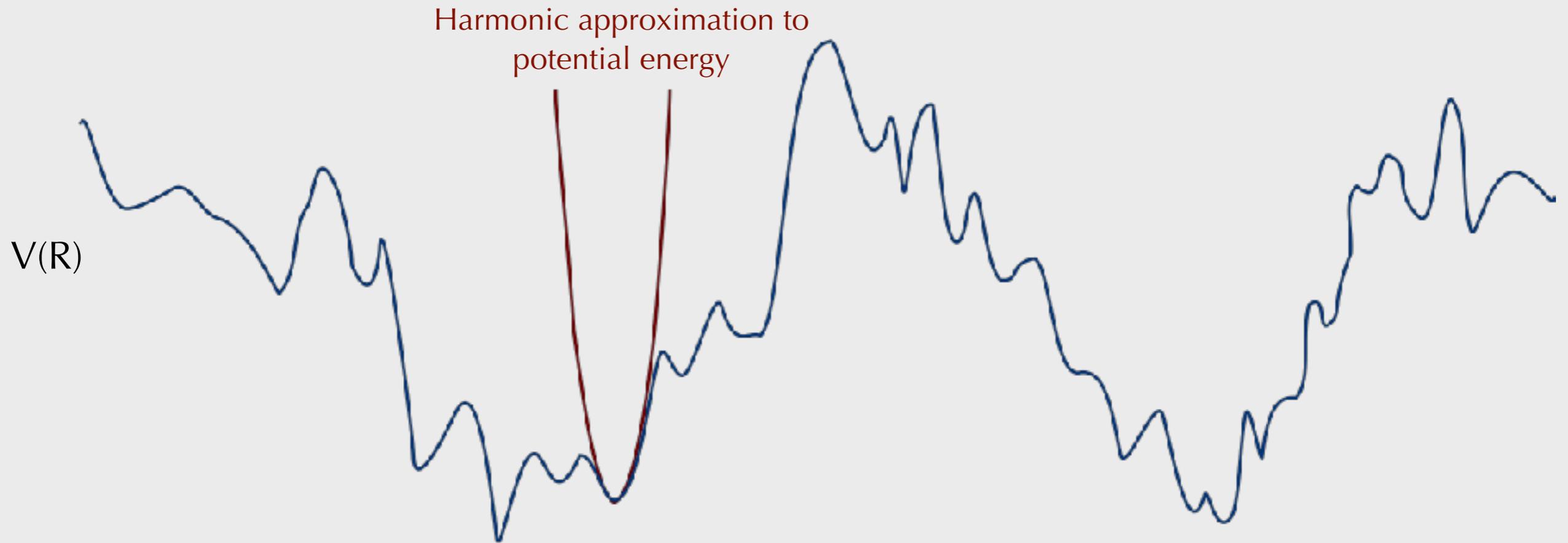
Restrict description to a single harmonic potential centered on a minimum-energy configuration

Vibrational dynamics only

Dynamics depend on the shape of the energy minimum at that point

# Total vibrational energy

Kinetic plus harmonic potential energy

$$E_{vib} = T(\dot{\mathbf{R}}) + V_{harmonic}(\mathbf{R})$$

atom 1    atom 2    ...    atom N

$$\mathbf{R} = [(x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_N, y_N, z_N)]$$

$$x = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad ... \quad x_{3N}]$$

Expand potential energy V about a position $x_o$ at an energy minimum

$$V(\mathbf{x}) = V(\mathbf{x_o}) + \sum_i^{3N} \left(\frac{\partial V}{\partial x_i}\right)_{\mathbf{x_o}} (x_i - x_{io}) + \frac{1}{2!} \sum_{i,j}^{3N} \left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_{\mathbf{x_o}} (x_i - x_{io})(x_j - x_{jo}) + ...$$

**Hessian matrix** of second-order partial derivatives describes the ***curvature at the minimum***

$$E_{vib} = \frac{1}{2} \sum_{atoms, a}^{N} m_a(v_{ax}^2 + v_{ay}^2 + v_{az}^2) + \frac{1}{2} \sum_{i,j}^{3N} \left(\frac{\partial^2 V}{\partial x_i x_j}\right)_{x_o} (x_i - x_{oi})(x_j - x_{oj})$$

Inconvenient form: $T(\dot{R})$ sums over atoms, $V(R)$ over Cartesian coordinates

# Root-mass weighting gives eigensystem

Introduce root-mass weighted coordinate displacements

$$\xi_i = \sqrt{m_i}(x_i - x_{io})$$

$$E_{\text{vib}} = \frac{1}{2}\sum_i^{3N} \dot{\xi}^2 + \frac{1}{2}\sum_{i,j}^{3N}\left(\frac{\partial^2 V}{\partial\xi_i\xi_j}\right)_0 \xi_i\xi_j$$

Symmetric form

$$E_{\text{vib}} = \frac{1}{2}\dot{\xi}^T\dot{\xi} + \frac{1}{2}\xi^T H\xi \qquad \text{(Note: Hessian becomes H, a force-constant matrix)}$$

Can find matrix A that diagonalizes H

$$L = A^T H A \qquad\qquad \boldsymbol{L} = \begin{bmatrix} \omega_1^2 & 0 & 0 & \dots & 0 \\ 0 & \omega_2^2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & \omega_{3N}^2 \end{bmatrix}$$

Eigenvalues $\omega_i^2$

Eigenvectors $\boldsymbol{q_i}$ : columns of $A^T$

# Normal modes: molecular motions can be decomposed into independent oscillators

$$E_{vib} = E_1 + E_2 + ...$$

$$= \underbrace{\left( \frac{1}{2}\dot{q}_1{}^2 + \frac{1}{2}\omega_1^2 q_1^2 \right)}_{\text{E oscillator 1}} + \underbrace{\left( \frac{1}{2}\dot{q}_2{}^2 + \frac{1}{2}\omega_2^2 q_2^2 \right)}_{\text{E oscillator 2}} + ...$$

Because L is diagonal, the total energy is that of **3N independent oscillators** along the q's

Each normal coordinate $q_i$ is an oscillating function of time along eigenvector $q_i$, frequency is $\omega_i$

$$q_i(t) = \sqrt{(2\ E_i)}/\omega_i \ \cos(\omega_i\ t + \varphi_i)$$

Note: angular frequencies (radians/sec) typically converted to wavenumber cm$^{-1}$ by $\bar{\nu} - \omega/2\pi c$

# NM provides info on flexibility



From crystal B-factors
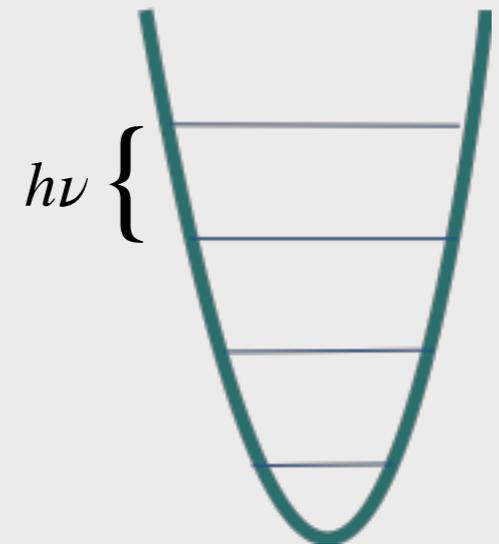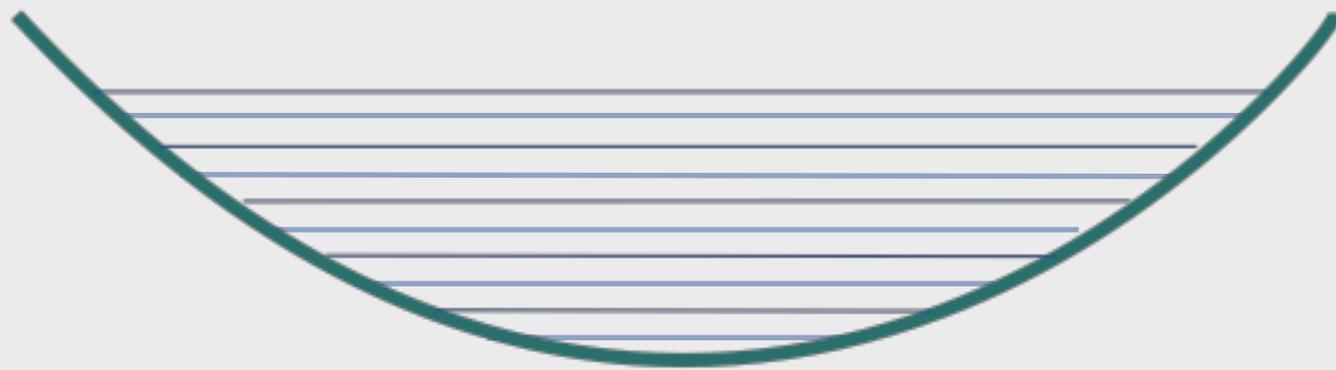
From all-atom NM

Robert et al. (2004) JMB
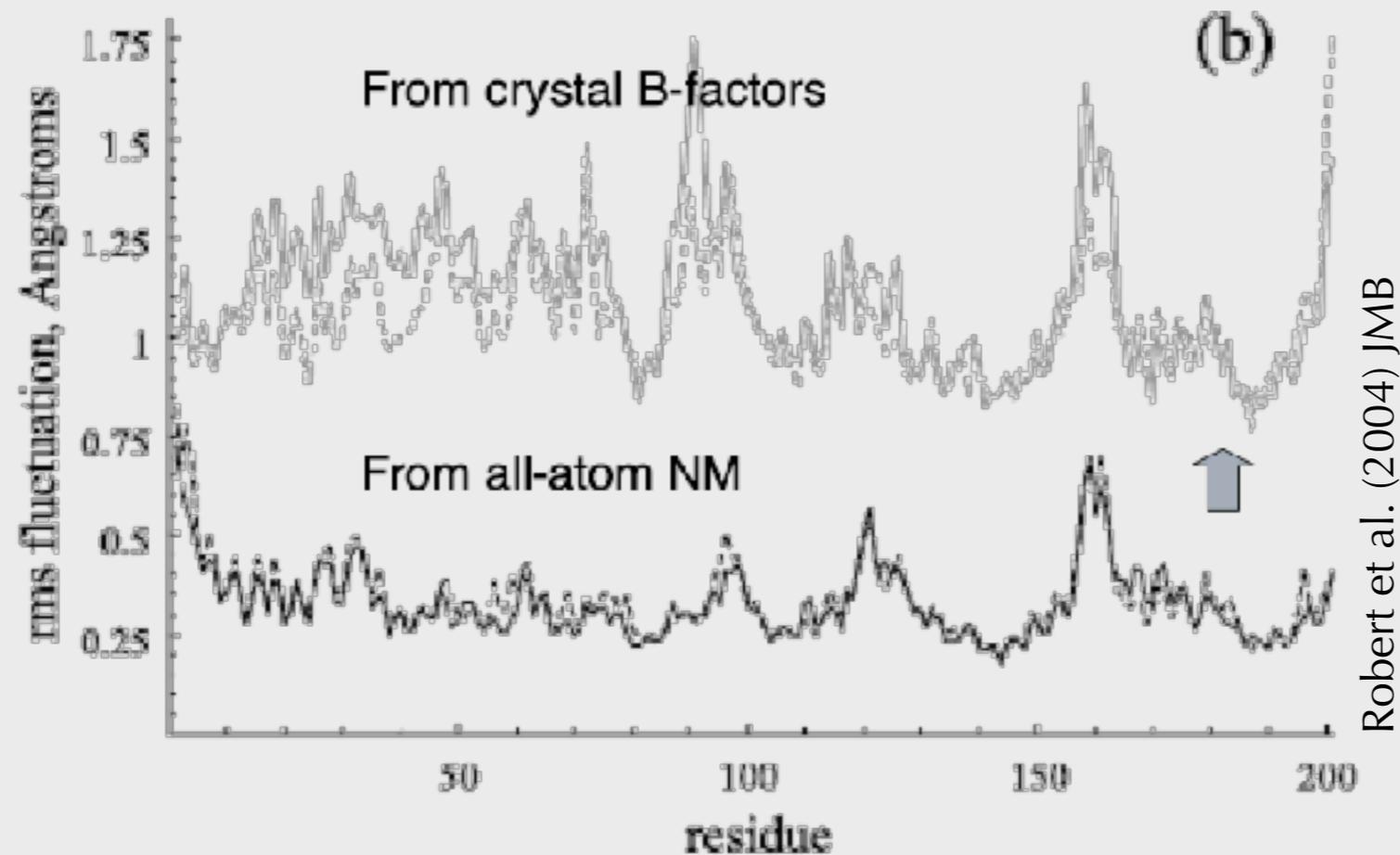
Calculate average squared atom displacements from modes

$$\langle \delta_i^2 \rangle = k_B T \sum_{k=1}^{3N-6} \frac{q_{3i,k}^2 + q_{3i+1,k}^2 + q_{3i+2}^2}{\omega_k^2}$$

(for atom 0…N-1)

Correlate well with temperature factors (B-values) in xtal structures

# Using NM for entropy estimation



Typical vibrational frequencies: $2 \lesssim \bar{\nu} \lesssim 2000$ cm$^{-1}$

○ Energy level spacing is on the order of $k_B T$ (207 cm$^{-1}$): $1/100 \lesssim \dfrac{h\nu}{k_B T} \lesssim 10$

Quantum harmonic oscillator

○ Partition function $q = \sum\limits_{n=0}^{\infty} \exp^{-\epsilon_n/k_B T}$ where $\epsilon_n = h\nu(n + 1/2)$

Vibrational entropy is f(occupation of excited states)

○ $S_{\text{vib}} = k_B \sum\limits_{i}^{3N-6} \{-ln(1 - \exp^{-h\nu_i/k_B T}) + \dfrac{h\nu_i/k_B T}{\exp^{h\nu_i/k_B T} - 1}\}$

Lower frequency NM → smaller spacing → larger entropy contribution

Why use this? e.g., MM-PBSA endpoint calcs of $\Delta G_{\text{assoc}}$

# Ala tetrapeptide **free-energy** landscape



Superposition approximation for basins
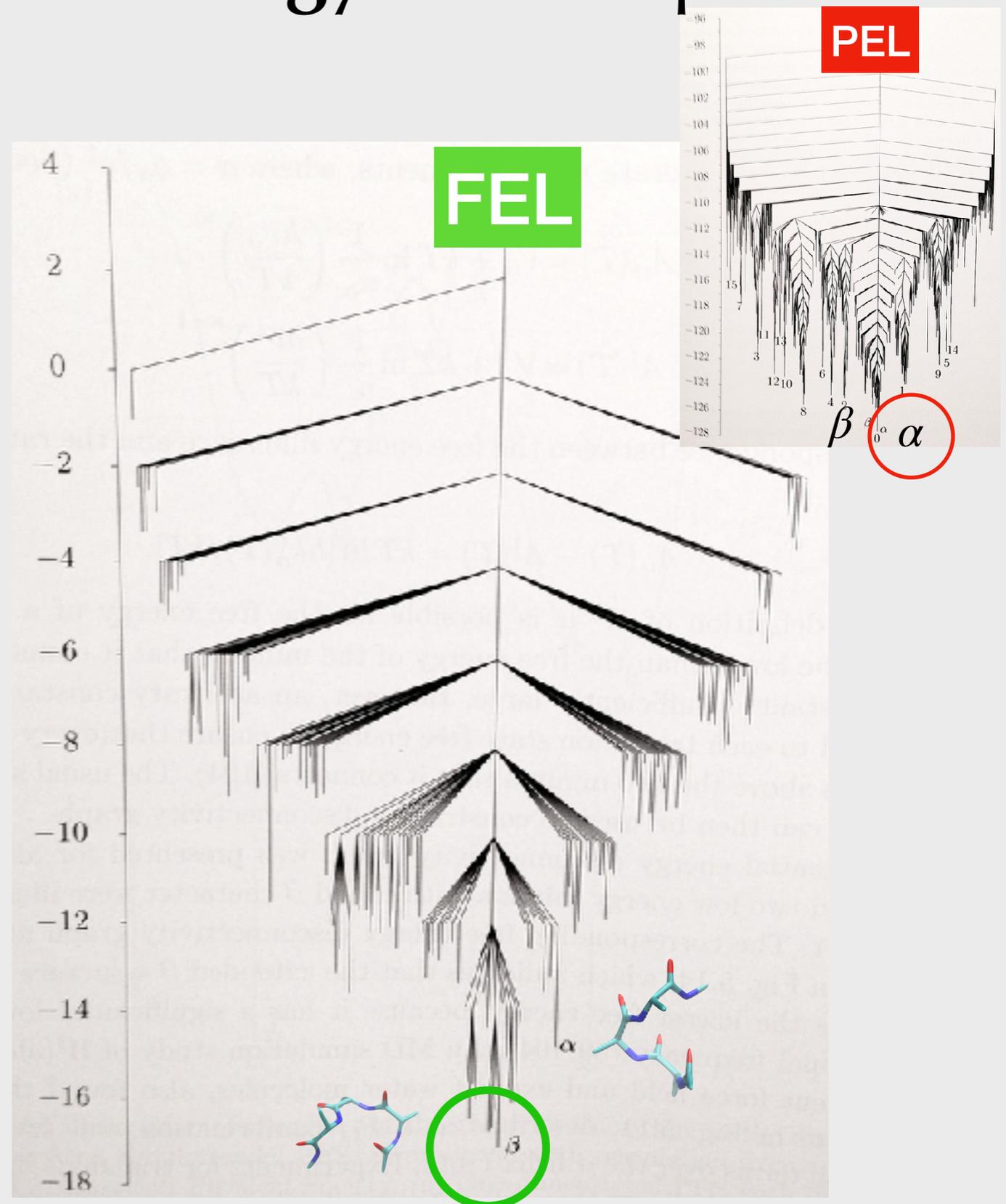
→ Thermodynamics

… for basins and transition points

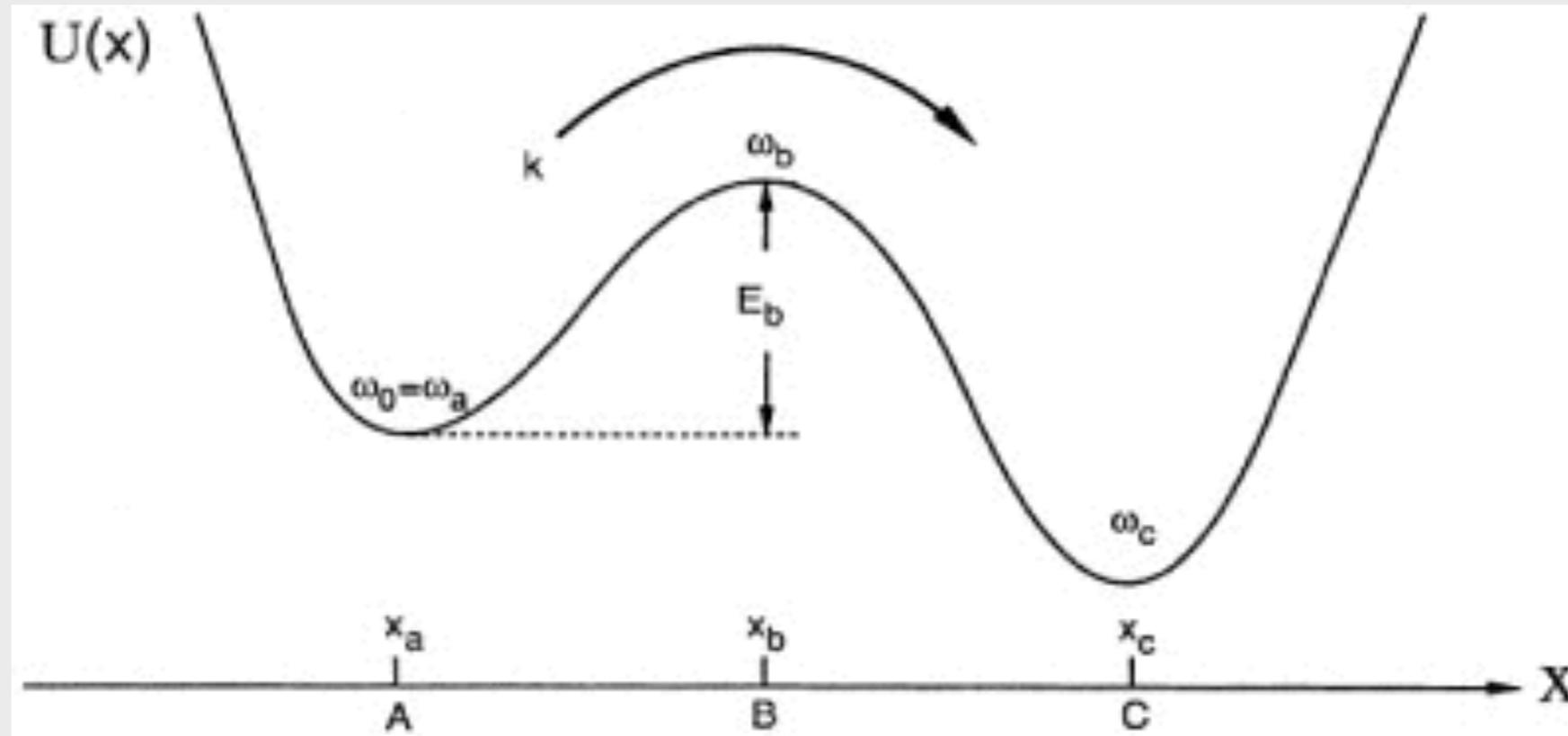→ Free Energy Landscape

- ○ FEL is f(T)

Extended $\beta$ form is more stable at 298K

- ○ lower vibrational mode frequencies than alpha helix → higher entropy

- ○ agrees with MD sampling studies (Krivov&Karplus 2002 PRL)

Evans & Wales 2003 JCP

# NM for kinetics: transition state theories depend on critical points and their shapes



Rate of forward reaction

$k = \mathscr{A} \exp( - E_b(X) / k_B T )$

$E_b$ saddle point energy

saddle point: productive vibrational frequency $\omega_b$

minima: vibrational partition functions $q_A$ and $q_B$ to estimate prefactor $\mathscr{A}$

# What do the NM vectors represent?

Eigenvectors: principal axes of the (hyper) paraboloid describing the minimum
   **"Natural" coordinate axes**

Deforming along an eigenvector $q_1$
   - produces a 1D vibration involving all the atoms
   - has **no effect** on vibration along $q_2$
        ○ no rotational component : each $\mathbf{Lq} = \lambda\mathbf{q}$.
   - Can thus speak of the energy of that mode.

Orthogonal (normal) modes of vibration



**Eigenvectors**

**Cartesian coordinate axes**
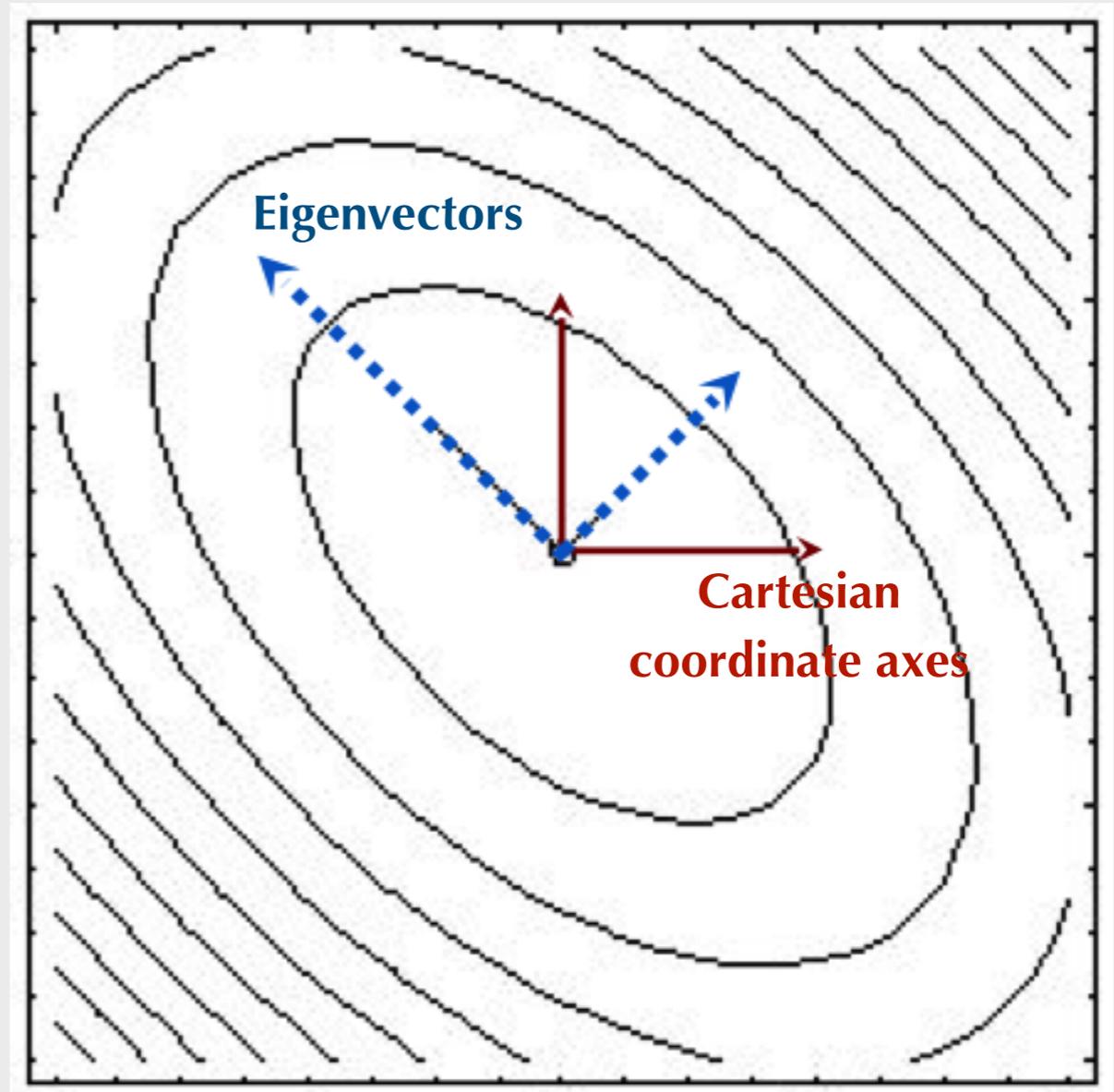
# Normal modes summary

**Advantages**

Analytical

Less computationally demanding than MD

Standard atomic forcefields*

Identify correlated motions

Deviation from crystal struct for 83 proteins



**Disadvantages**

Extensive minimization
- Can be costly
- Structural deviation

Diagonalize large matrices (3Nx3N)
- Memory/time

Dependence on initial structure

Solvent effects poorly incorporated
- Single solvent configuration if any
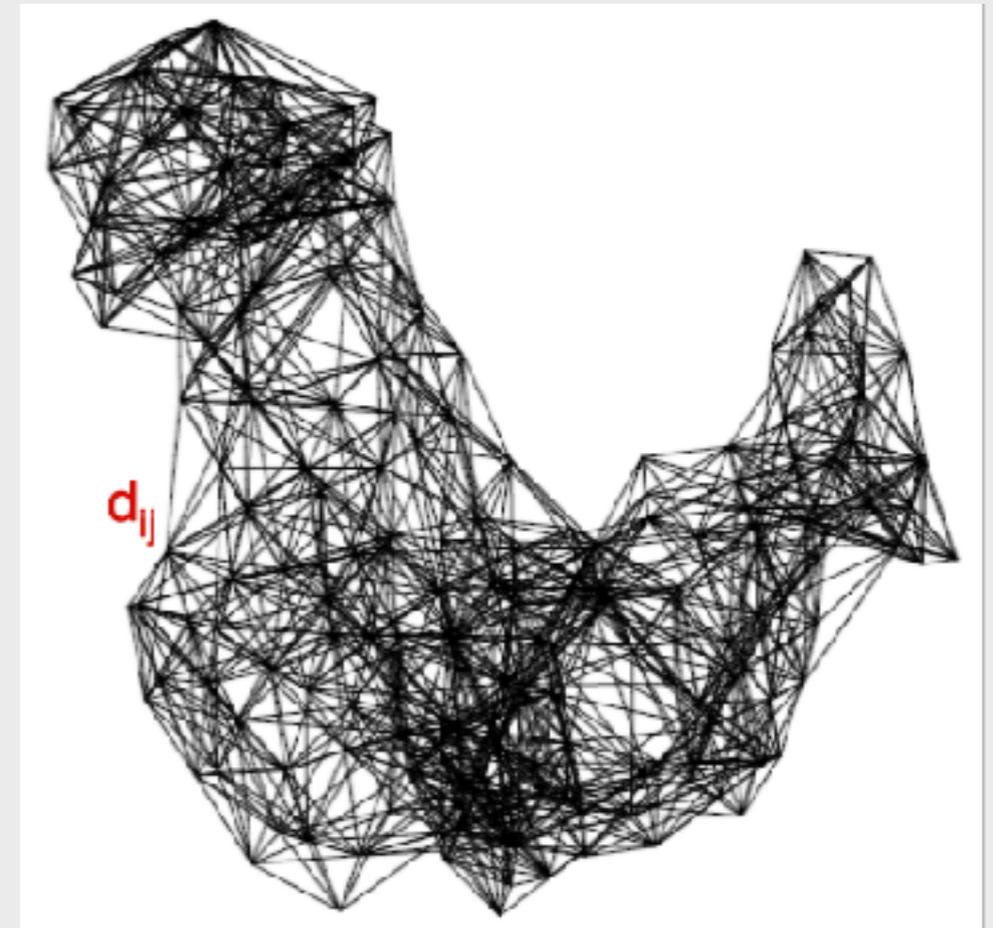- Heuristic distance-dependent dielectric
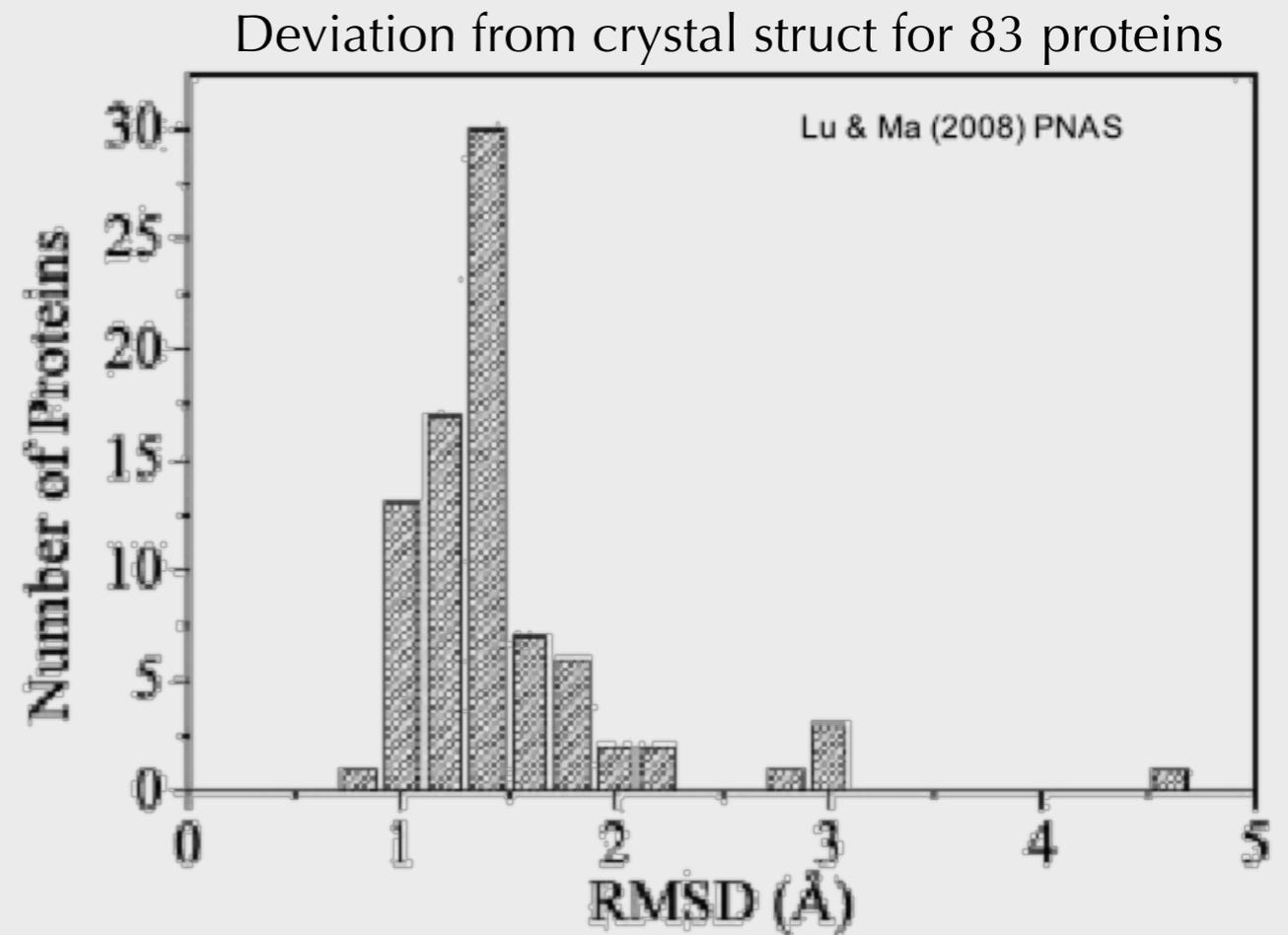
# Elastic Network Models (ENM)

$$V_{\mathrm{ENM}} = \sum_{i,j} C_{ij} \left( d_{ij} - d_{ij}^o \right)^2 \quad \forall \; d_{ij} < d_{\mathrm{cutoff}}$$

Simplified harmonic potential

- Based on Tirion (1996) PRL
- $d_{ij}^o$ are from initial structure
  - **no energy minimization**
  - **no structural deviation**
  - **fast**

# Global exploration of the EL

Ensemble-less methods

Ensemble approaches

# Normal modes information can extend far beyond $x_o$

LAO



In proteins, lowest-frequency NM directions predict directions of conformational change

$$\Delta x = x_{\text{closed}} - x_{\text{open}}$$

mode $v_l$, $\theta$, $\Delta x$
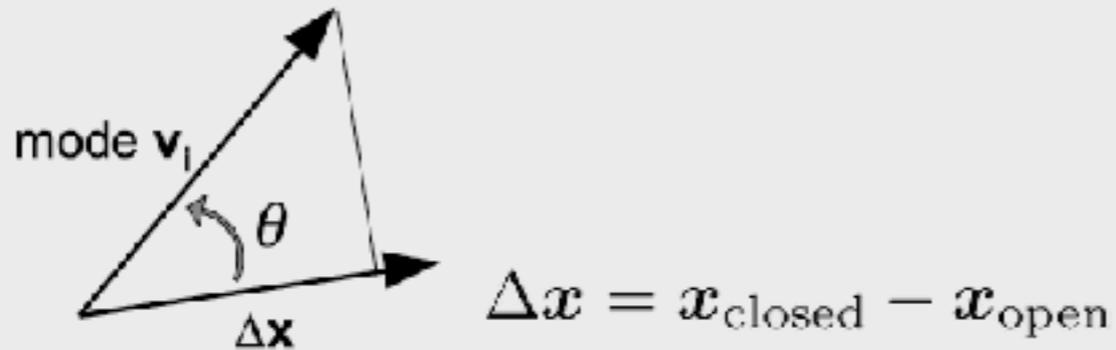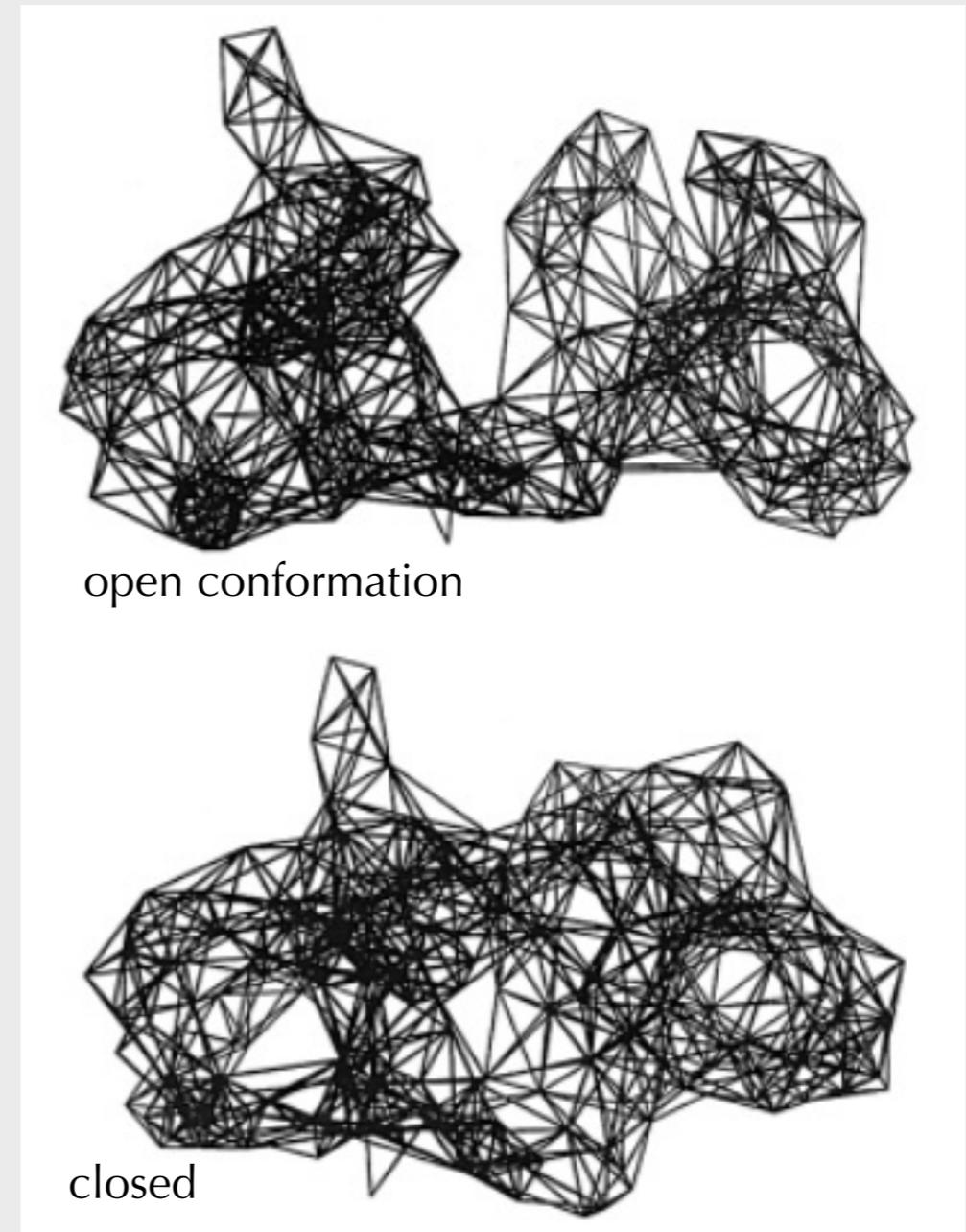


open conformation

---

**Table III.** Overlap of the mode the most involved in the conformational change of proteins of various sizes and motion types, when the 'open' or 'closed' conformations are considered

| Protein | Overlap | |
|---|---|---|
| | Open | Closed |
| Citrate synthase | 0.83 | 0.57 |
| Calmodulin | 0.50 | 0.37 |
| Che Y protein | 0.32 | 0.34 |
| Dihydrofolate reductase | 0.72 | 0.64 |
| Diphtheria toxin | 0.58 | 0.37 |
| Enolase | 0.33 | 0.30 |
| LAO binding protein | 0.84 | 0.40 |
| Triglyceride lipase | 0.30 | 0.17 |
| Maltodextrin binding protein | 0.86 | 0.77 |
| Thymidylate synthase | 0.56 | 0.40 |

Tama and Sanejouand (2001) Prot Eng 14, 1

closed

# Conformational exploration with NM

Displace along eigenvectors

- Random linear combinations
- Displace along a chosen mode(s)
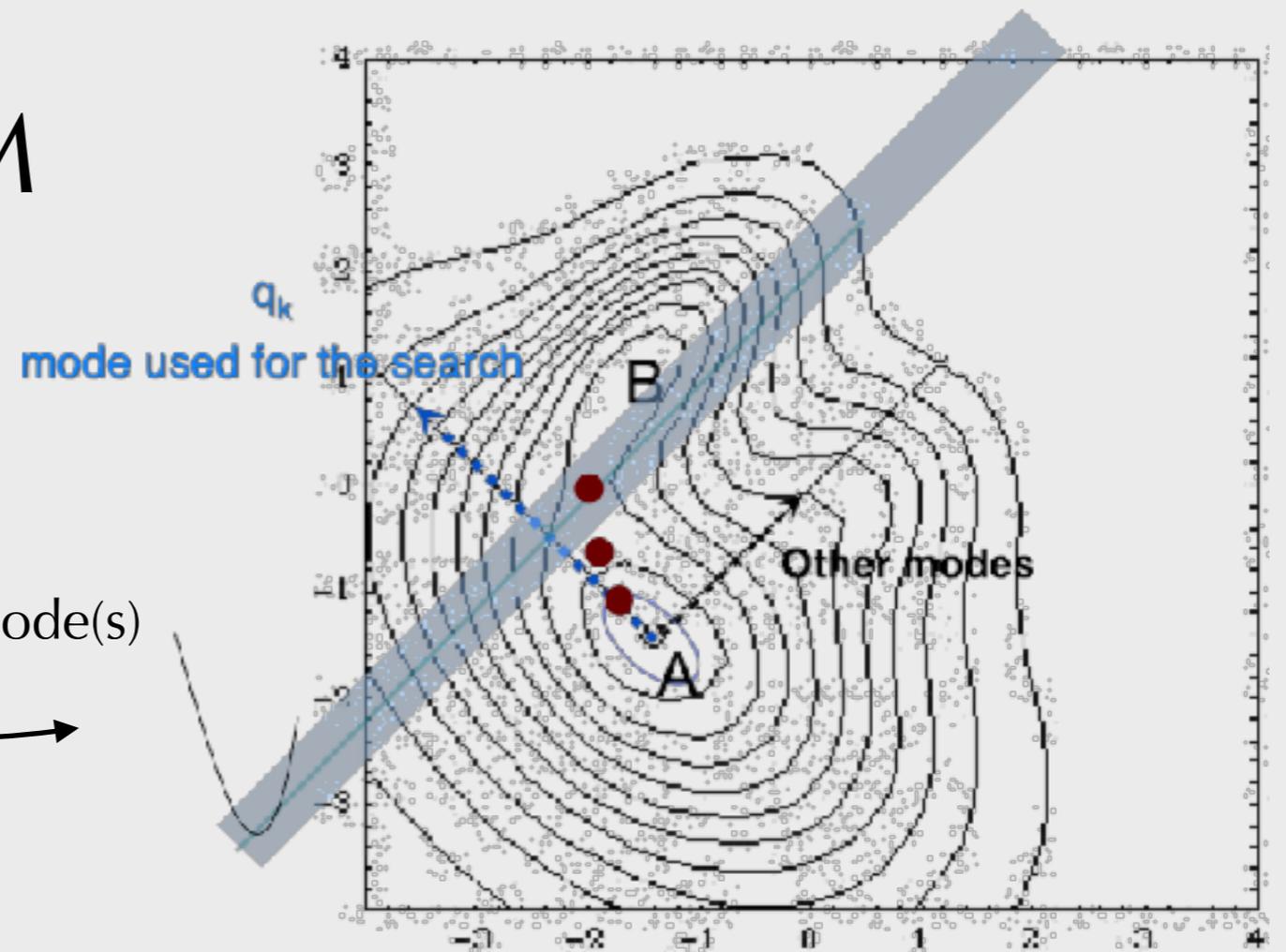- Restrain projection along chosen mode(s)

Typical applications

Studying mechanisms

Flexible docking

Crystal structure refinement

Cryo EM

…



see Perahia, Hinson, Sanejouand, Delarue, Zacharias, Chacon, Grudinin …
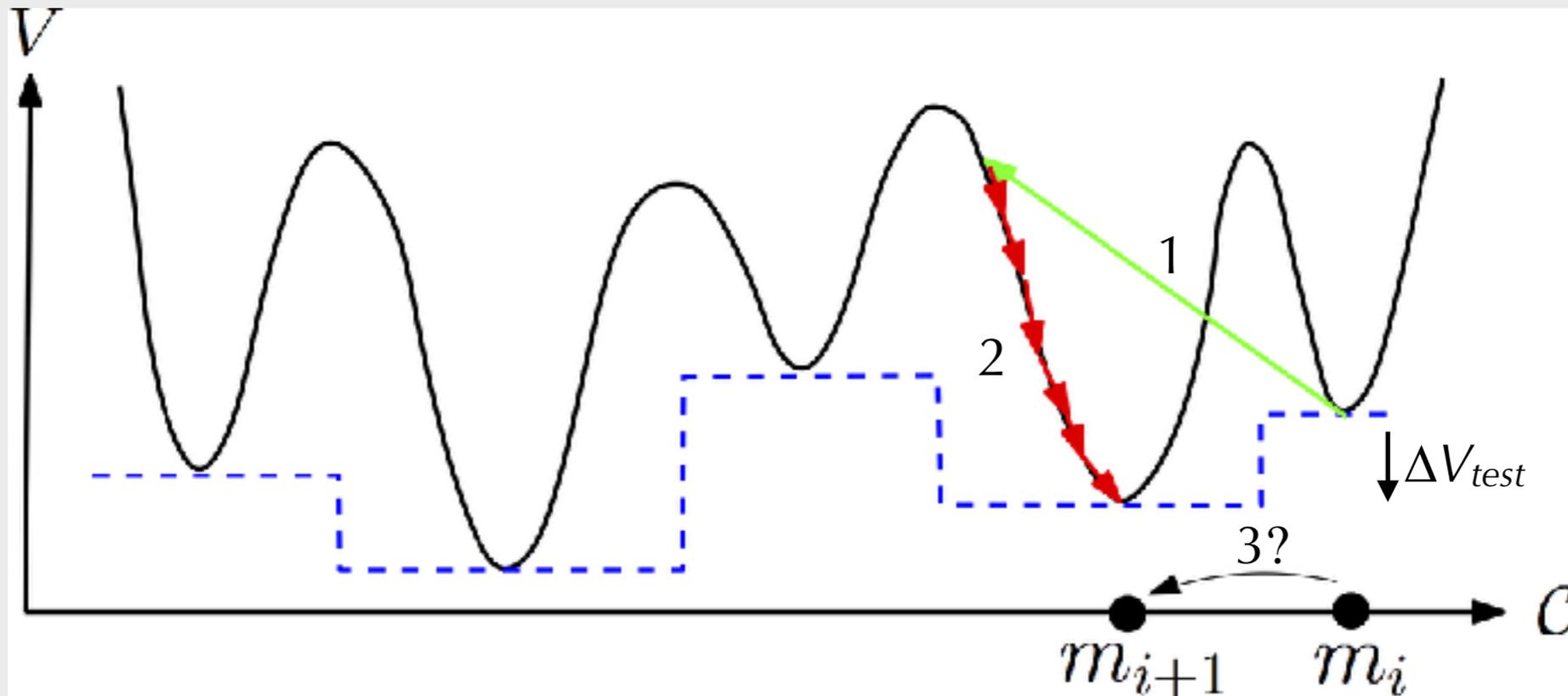
# Global optimization

Explicit goal is to find low-energy structures

 Structure prediction


Implicit goals:

- Exploration
  - explore far reaches of conformational space
- Exploitation
  - find low-energy structures

# Global minimization/exploration with Basin Hopping



aka "Monte Carlo minimization", Li and Scheraga PNAS 1987

Monte Carlo move followed by minimization before Metropolis test

BH transforms EL to remove all barriers

- If unchecked even samples chirality changes
- → Ensemble sampling of transformed surface

Efficient for finding global minimum in somewhat large systems (100 - 1000 atoms)

# Frustrated folding of a model protein

$$V_{BLN} = \frac{1}{2}K_r \sum_{i=1}^{N-1}(R_{i,i+1} - R_c)^2 + \frac{1}{2}K_\theta \sum_{i=1}^{N-2}(\theta_i - \theta_c)^2$$

$$+ \varepsilon \sum_{i=1}^{N-3}[A_i(1 + \cos\phi_i) + B_i(1 + 3\cos\phi_i)]$$

$$+ 4\varepsilon \sum_{i=1}^{N-2}\sum_{j=i+2}^{N} C_{ij}\left[\left(\frac{\sigma}{R_{ij}}\right)^{12} - D_{ij}\left(\frac{\sigma}{R_{ij}}\right)^6\right]$$
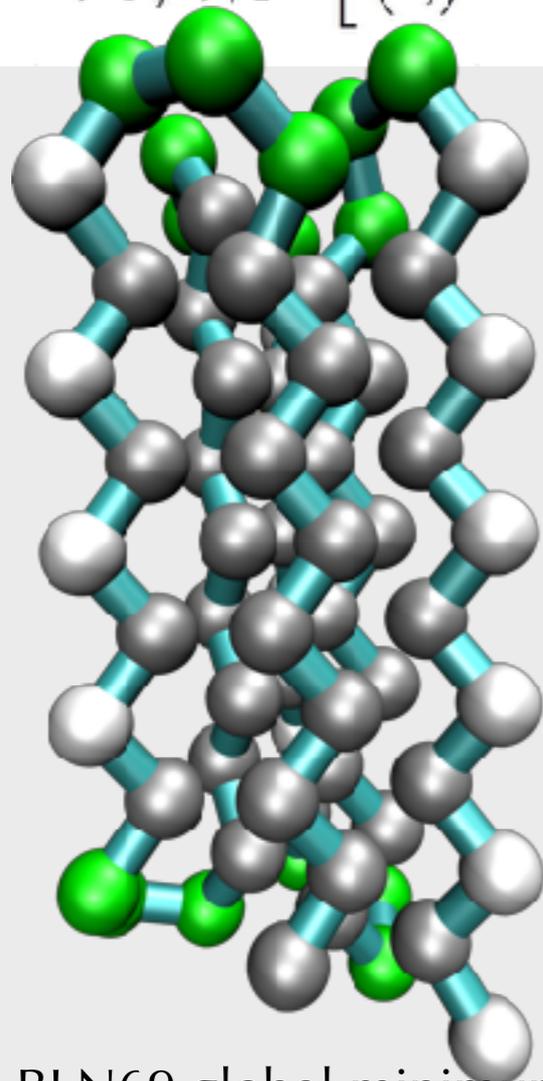


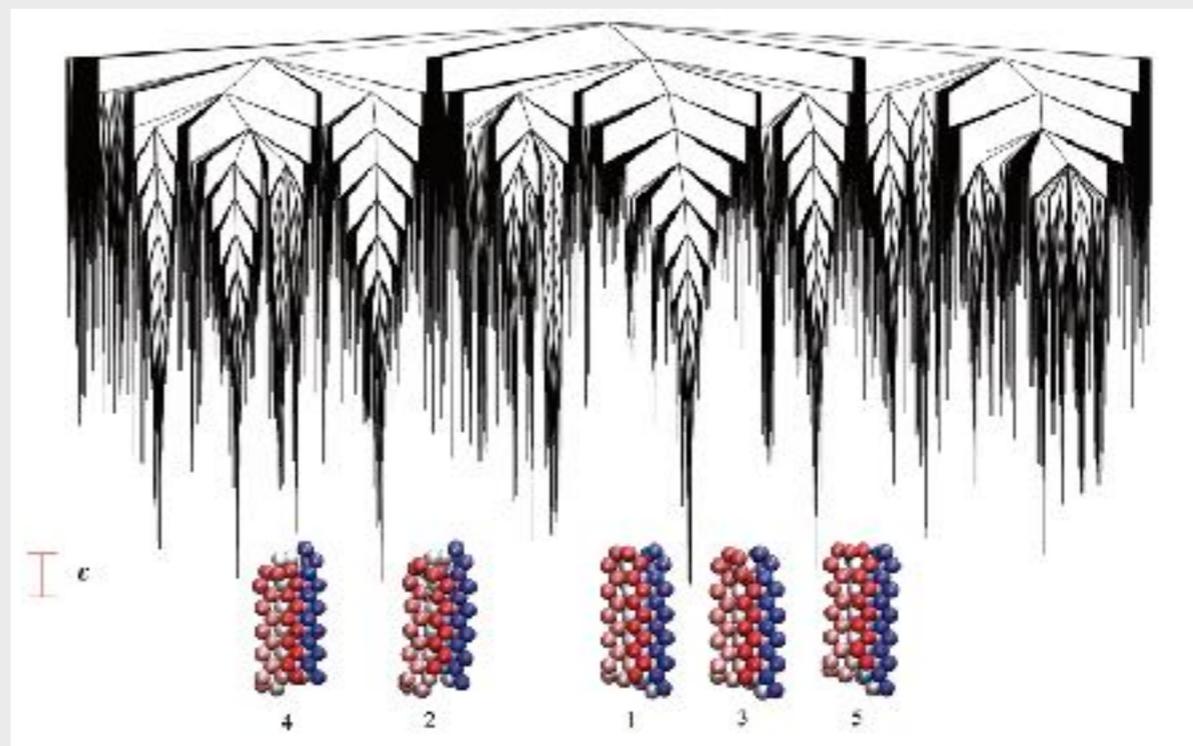BLN69 global minimum:
a ß-barrel

**BLN** coarse-grained model

- HydrophoBic, hydrophiLic, Neutral
- Honeycutt & Thirumalai PNAS (1990)
- Head-Gordon PNAS (2003)

Highly frustrated EL

- Numerous low-energy structures in deep funnels
- Extensive BH studies (~$10^6$ minima)



Oakley, Wales & Johnston J Phys Chem (2011)

# BH + Voronoi bias: *hybrid explorer*



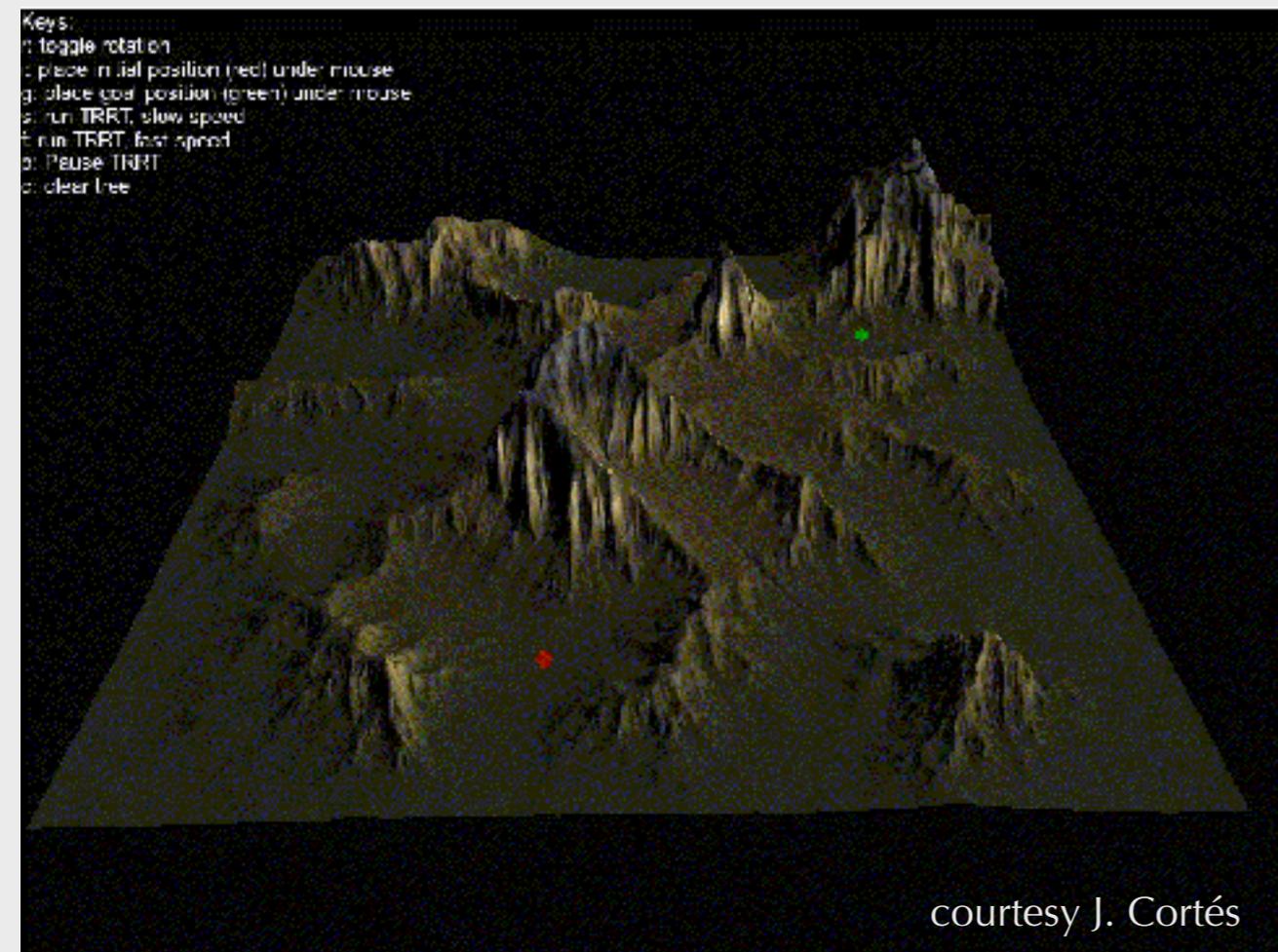courtesy J. Cortés

Combine two strategies

1) **T-RRT**: Voronoï bias *(taboo-ish)*

- Temperature controlled Rapidly-expanding Random Tree
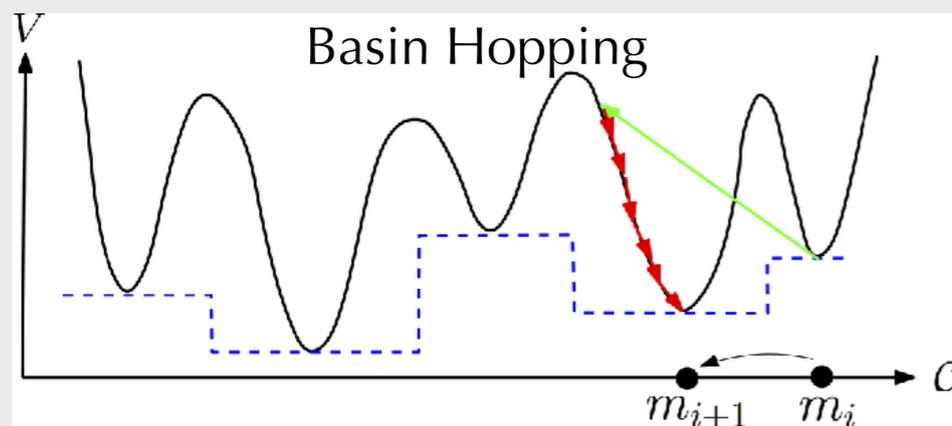- Jaillet, et al. (2011) J Comp Chem

2) Monte Carlo minimization (**Basin Hopping**)

- Li & Scheraga (1987) PNAS

→ Basin hopping threads biased towards unexplored regions

**+**



Basin Hopping

$m_{i+1}$ $m_i$



Cazals et al. (2015) J Comp Chem
Roth et al. (2016) J Comp Chem

# Hybrid explorer efficiently finds low-energy structures

## Model system

BLN coarse grained protein model (69 amino acids)
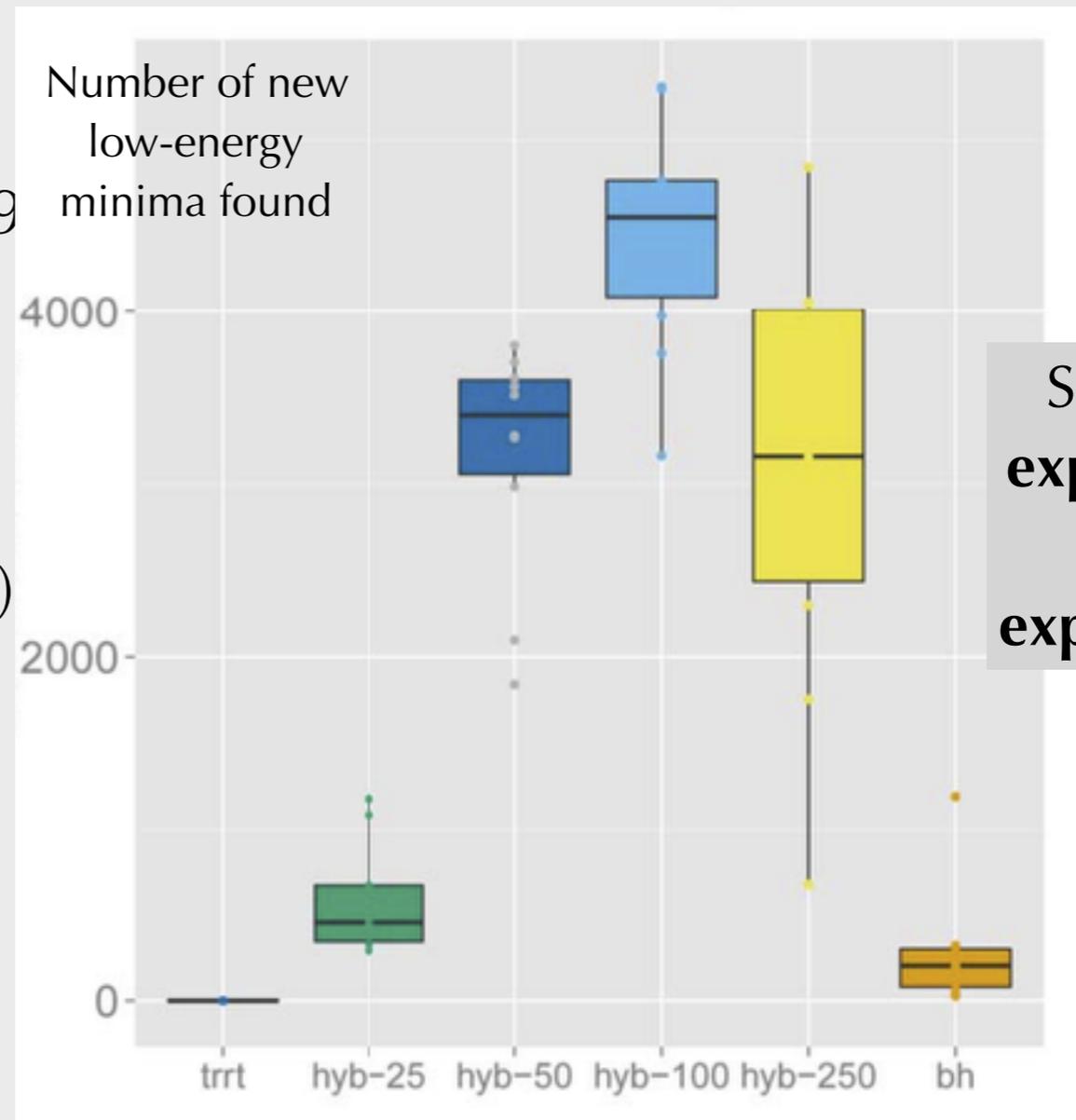
Global minimum known

Gold standard database of minimum-energy structures (Oakley, et al. 2011)

## Hybrid explorer

More low-energy minima

Wider exploration

Competitive run time



Number of new low-energy minima found

Synergy:
**exploration**
*and*
**exploitation**

More T-RRT-like ← → More basin-hopping-like

## …in the SBL

Cazals et al (2015) J Comp Chem; Roth et al (2016) J Comp Chem

# Global minimization algorithms: pluses and minuses

Advantages:

- target low-energy structures (structure prediction)

- broad exploration

Disadvantages:

- no thermodynamic ensemble: averages not related to observables

- In Basin Hopping, MC move sets limiting for compact states

# Thermodynamic sampling of the EL

MC or MD allows sampling in a thermodynamic ensemble

Averaging samples allows calculating observable properties

*n(E)*: density of states (energy states)

- A property of the entire energy landscape
- Grows rapidly with E

Boltzmann factor:

- Diminishes rapidly with E

$$W_{\mathrm{B}}(E) = exp^{-\beta E(X)}$$

Probability distribution:

$$P_{\mathrm{can}}(E;T) \propto n(E) exp^{-\beta E}$$

- Peaks at average energy at T

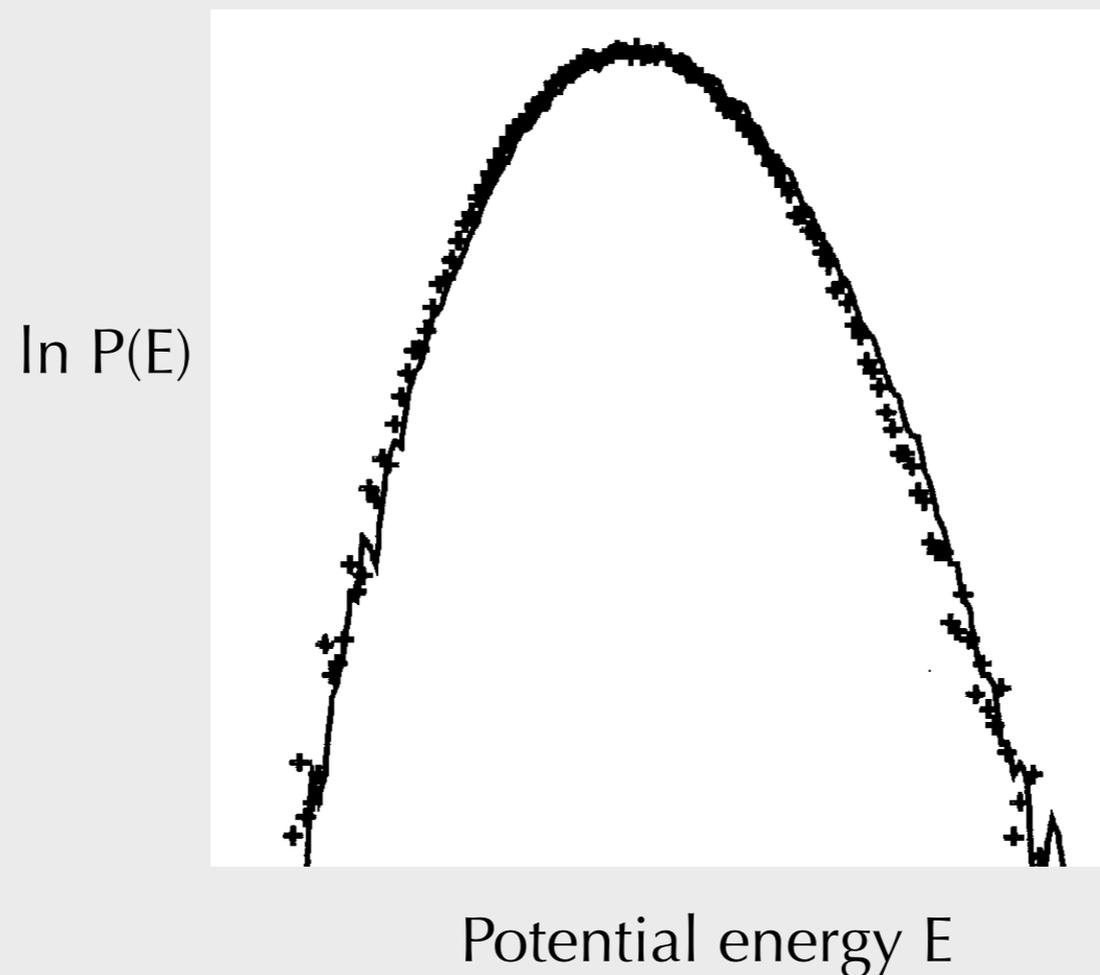Berg and Neuhaus (1992) Phys Rev Lett
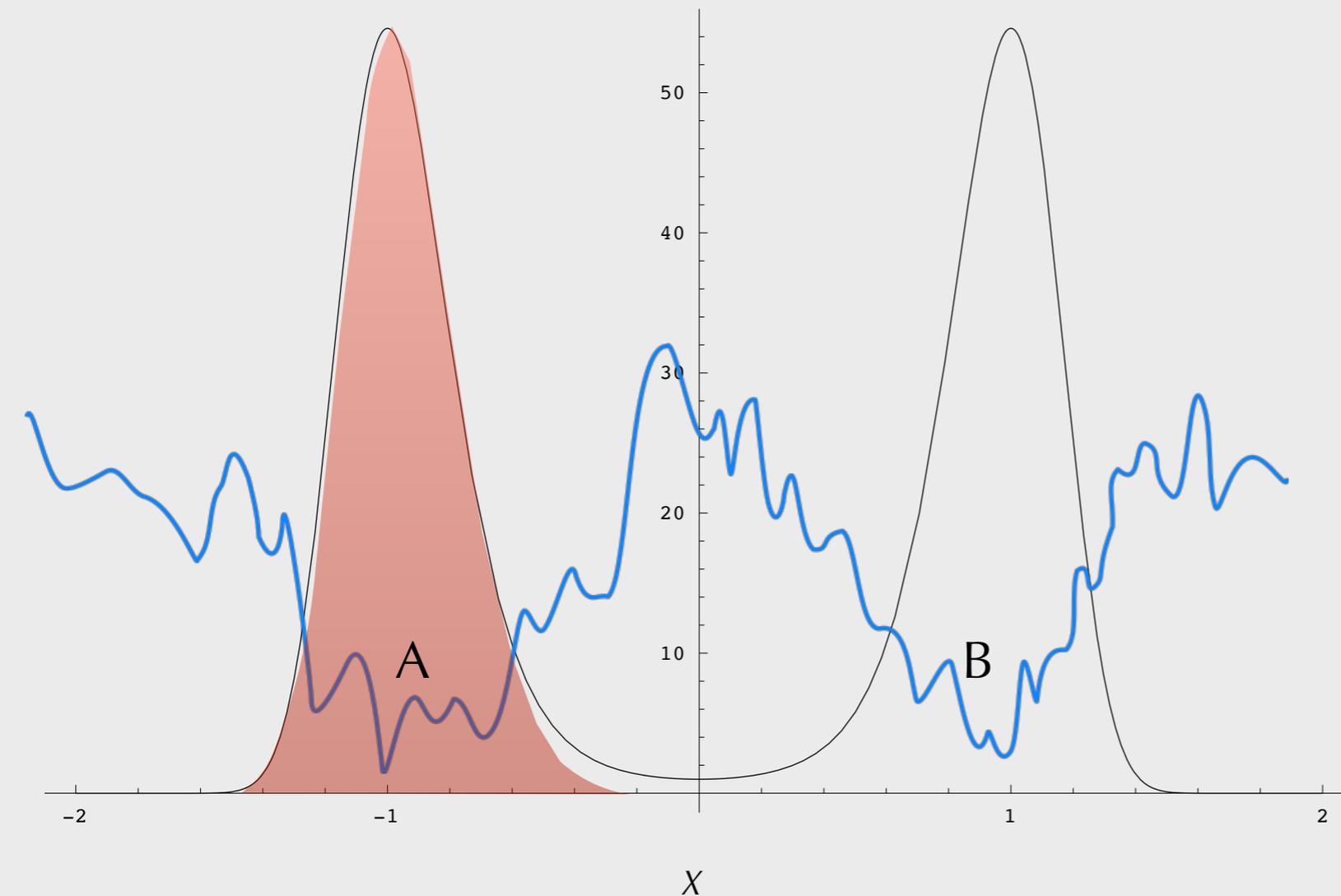Mitsutake, Sugita & Okamoto (2001) Biopolymers

# Canonical sampling

Probability distribution of E(T) is approximately Gaussian around avg E

$$\ln P_{\text{can}} = a - b(E - \bar{E}(T))^2$$

But crossing barriers may require sampling high E transition regions…



ln P(E)

Potential energy E

# Trapping in canonical simulations



- Transition regions are exponentially suppressed
- Simulation remains trapped in local (super-) basin
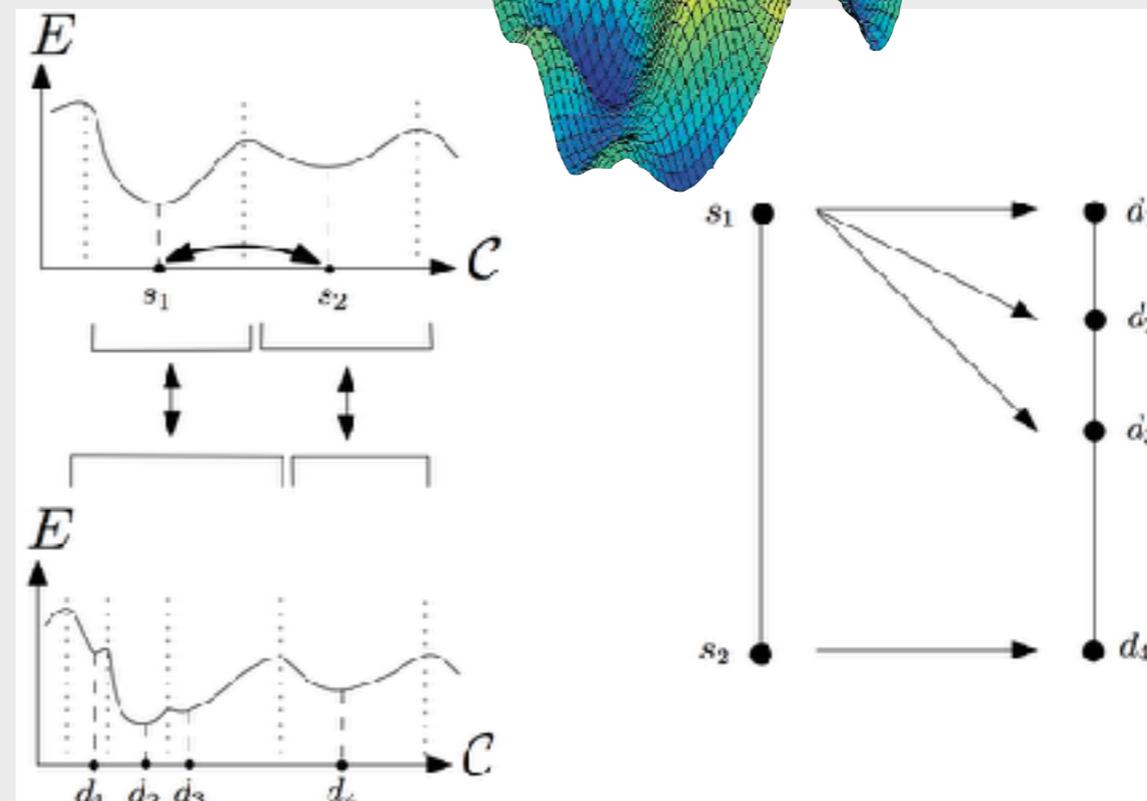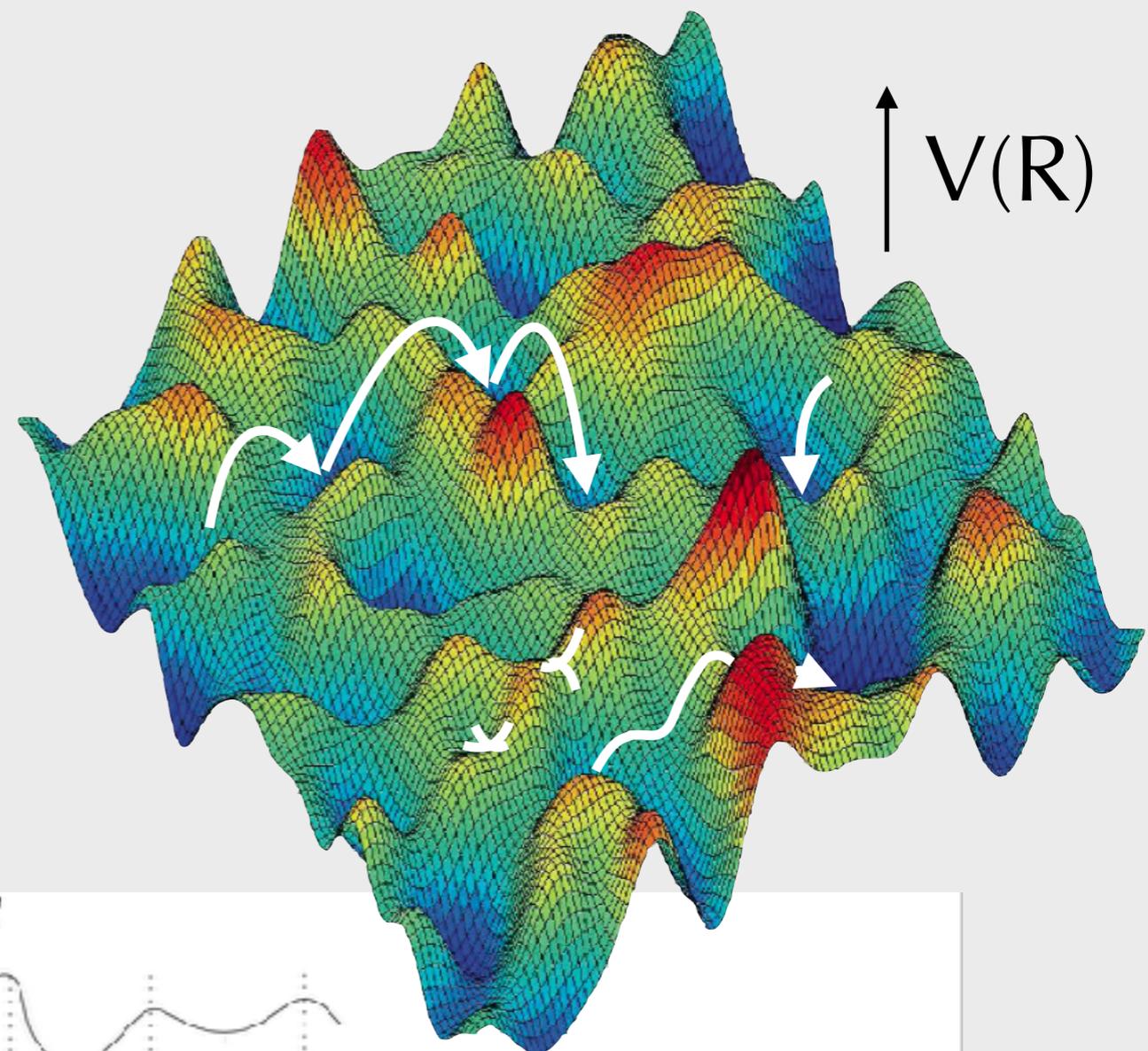
Here the true $P_{can}(T)$ is bimodal

Sampled $P_{sim}(T)$ is not: poor sampling

Large swaths of the energy landscape may be ignored

→ Other approaches (e.g., REMD) sample better [...]

# Summary

1. New and old methods* for exploring the EL of macromolecular systems both locally and globally

2. Ways to accommodate and compare** diverse data in simplified but still rich representations

---

* including robotics [Juan Cortés]

** …also in the SBL [Frédéric Cazals]



$V(R)$

Comparing 2 EL using the earthmover's distance
(Cazals et al. 2015)

# Thanks



Laboratoire de Biochimie Théorique

Institut de Biologie Physico Chimique, Paris

Centre National de la Recherche Scientifique (CNRS)

Joël Janin (Gif-sur-Yvette, Orsay, Paris)

Pinak Chakrabarti (Kolkata)

David Perahia (Orsay)

Frédéric Cazals (Inria Sophia-Antipolis)

Juan Cortés (LAAS-CNRS Toulouse)

and many others…

...you're done, stop talking