Leipzig Spring School 2008

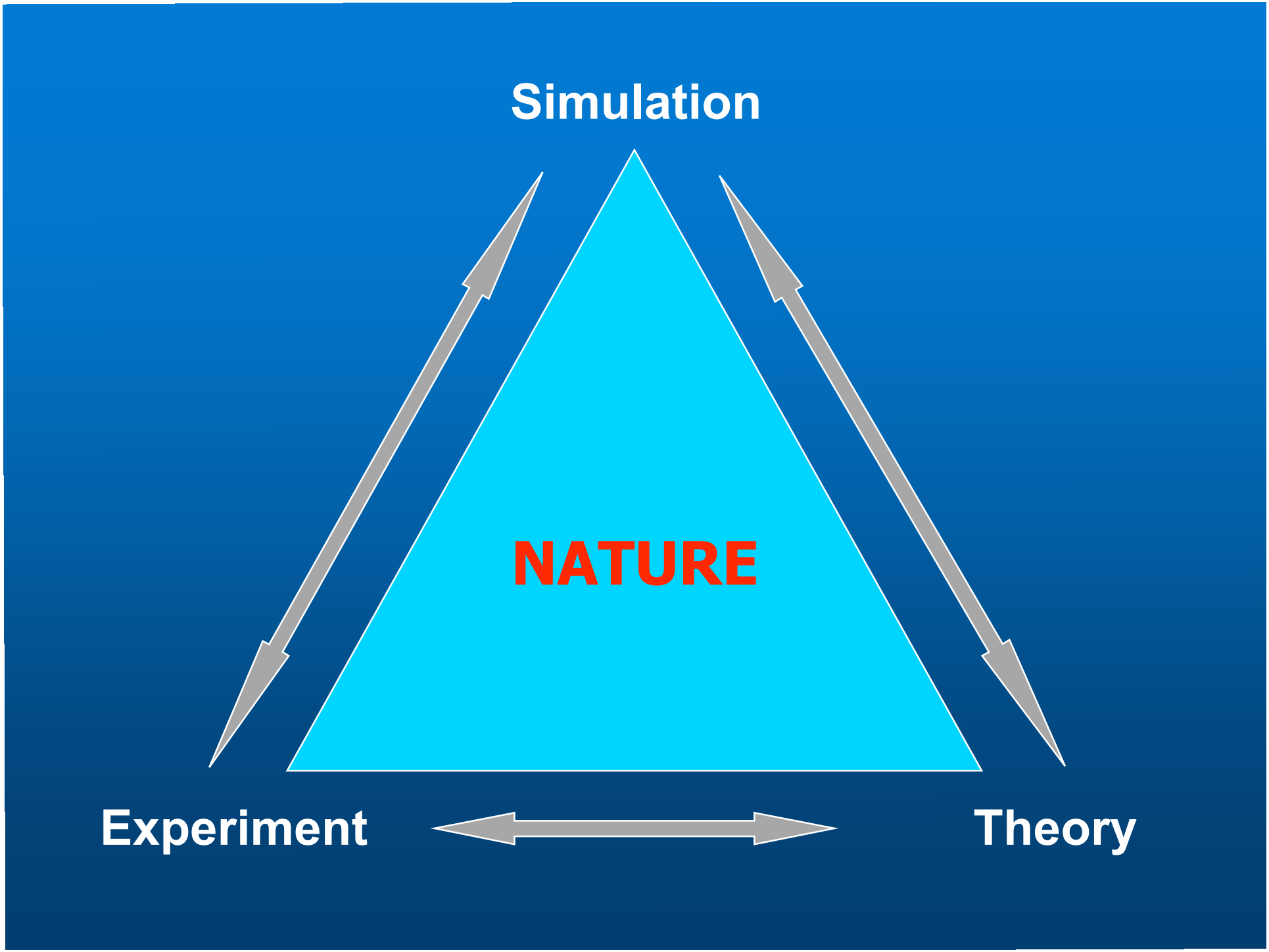
Applications of Monte Carlo Simulations to Systems with Complex Energy Landscapes

D. P. Landau

- **Introduction**
- **Review of Algorithms**
- **Applications**
  - *Spin glasses*
  - *"Lattice proteins"*
  - *"Real" proteins*
- **Conclusions**

Review: Algorithms for “complex” systems

- **Multicanonical sampling**
- **Parallel tempering**
- **Wang-Landau sampling**


Simulation
NATURE
Experiment
Theory

Reminder

The *Partition function* contains all thermodynamic information:

$$Z = \sum_{all\ states} e^{-\mathcal{H}/k_B T}$$

Simple Monte Carlo approach: sample via a random walk in probability space

Single spin-flip sampling for the Ising model

Produce the n^{th} state from the m^{th} state … relative probability is $P_n / P_m \rightarrow$ need only the *energy difference, i.e.* $\Delta E = (E_n - E_m)$ between the states

Any transition rate that satisfies ***detailed balance*** is acceptable, usually the Metropolis form *(Metropolis et al, 1953).*

$$W(m \rightarrow n) = \tau_o^{-1} \exp(-\Delta E / k_B T), \quad \Delta E > 0$$
$$= \tau_o^{-1}, \quad \Delta E < 0$$

where τ_o is the time required to attempt a spin-flip.

MC Problems and Challenges

***Statics*: Monte Carlo methods are valuable, but near T_c**

$\Rightarrow$ ***critical slowing down*** **for** *2^{nd} order* **transitions**

$\Rightarrow$ ***metastability*** **for** *1^{st} order* **transitions and for systems with complex energy landscapes**

$\therefore$ ***Try to reduce characteristic time scales or circumvent them***

***"Dynamics"*: stochastic vs deterministic**

Review: Wang-Landau sampling

Random Walk in Energy Space with a Flat Histogram

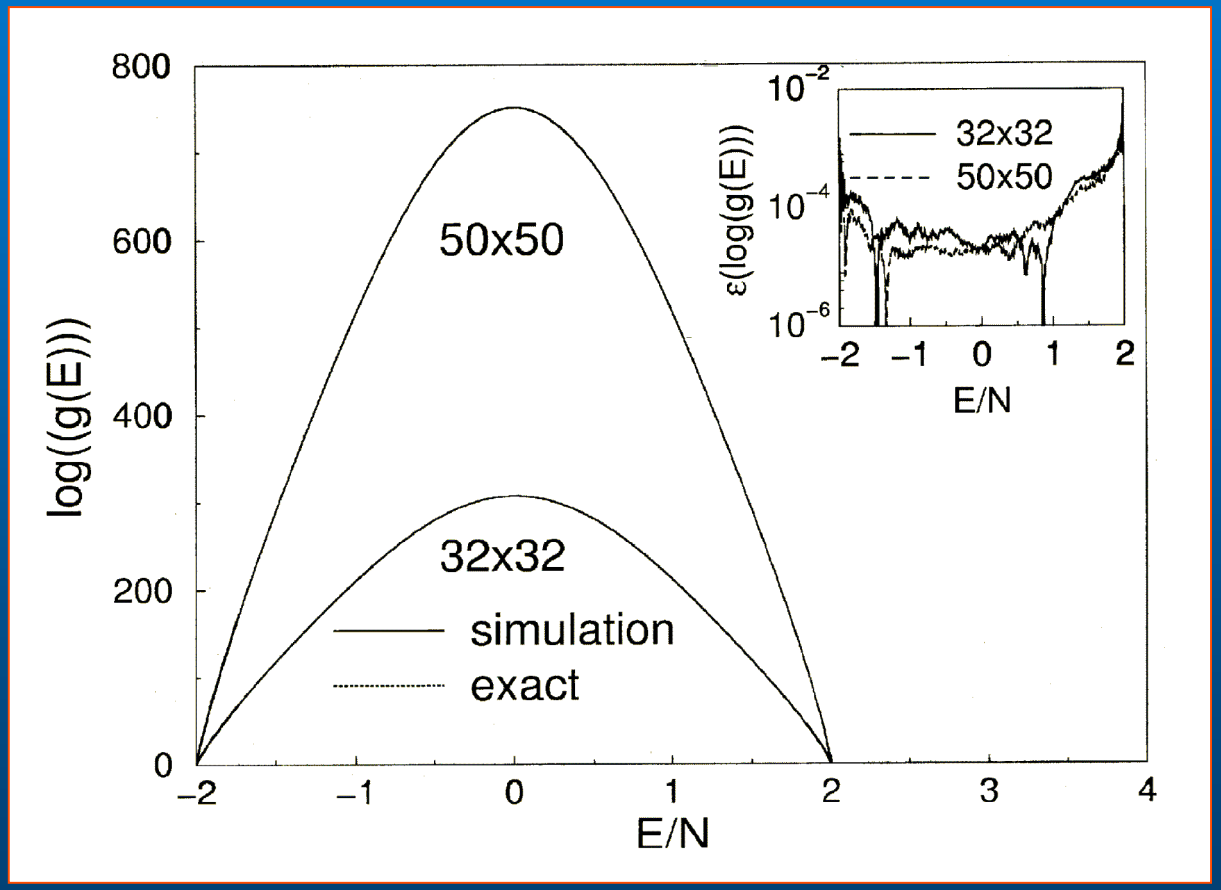
$$Z = \sum_{\substack{all \\ states}} e^{-\mathcal{H}/k_B T} \equiv \sum_{\substack{all \\ energies}} g(E) e^{-\mathcal{H}/k_B T}$$

Estimate the *density of states* $g(E)$ directly by performing a random walk in energy space:

1. Set $g(E)=1$; choose a modification factor (e.g. $f_0 = e^1$)
2. Randomly flip a spin with probability: $p(E_1 \rightarrow E_2) = \min\left(\frac{g(E_1)}{g(E_2)}, 1\right)$
3. Set $g(E_i) \rightarrow g(E_i) * f$
4. Continue until the histogram is "flat"; decrease f, e.g. $f_{i+1} = f^{1/2}$
5. Repeat steps 2 - 4 until $f = f_{min} \sim \exp(10^{-8})$
6. Calculate properties using final density of states $g(E)$

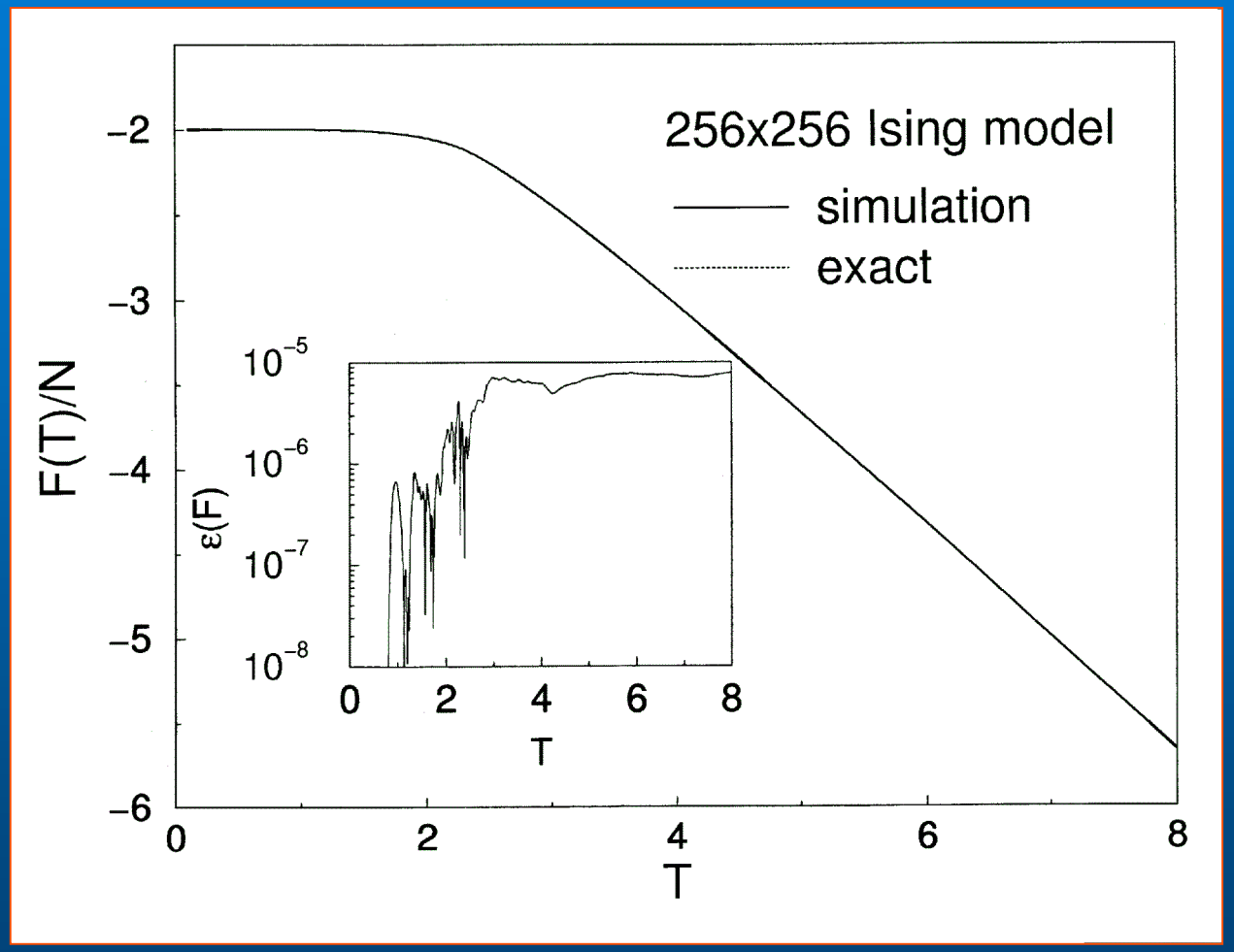
Density of States for the 2-dim Ising model

Compare exact results with data from random walks in energy space: *L×L lattices with periodic boundaries*



ε = relative error (*exact solution is known for L≤ 64*)

Free Energy of the 2-dim Ising Model

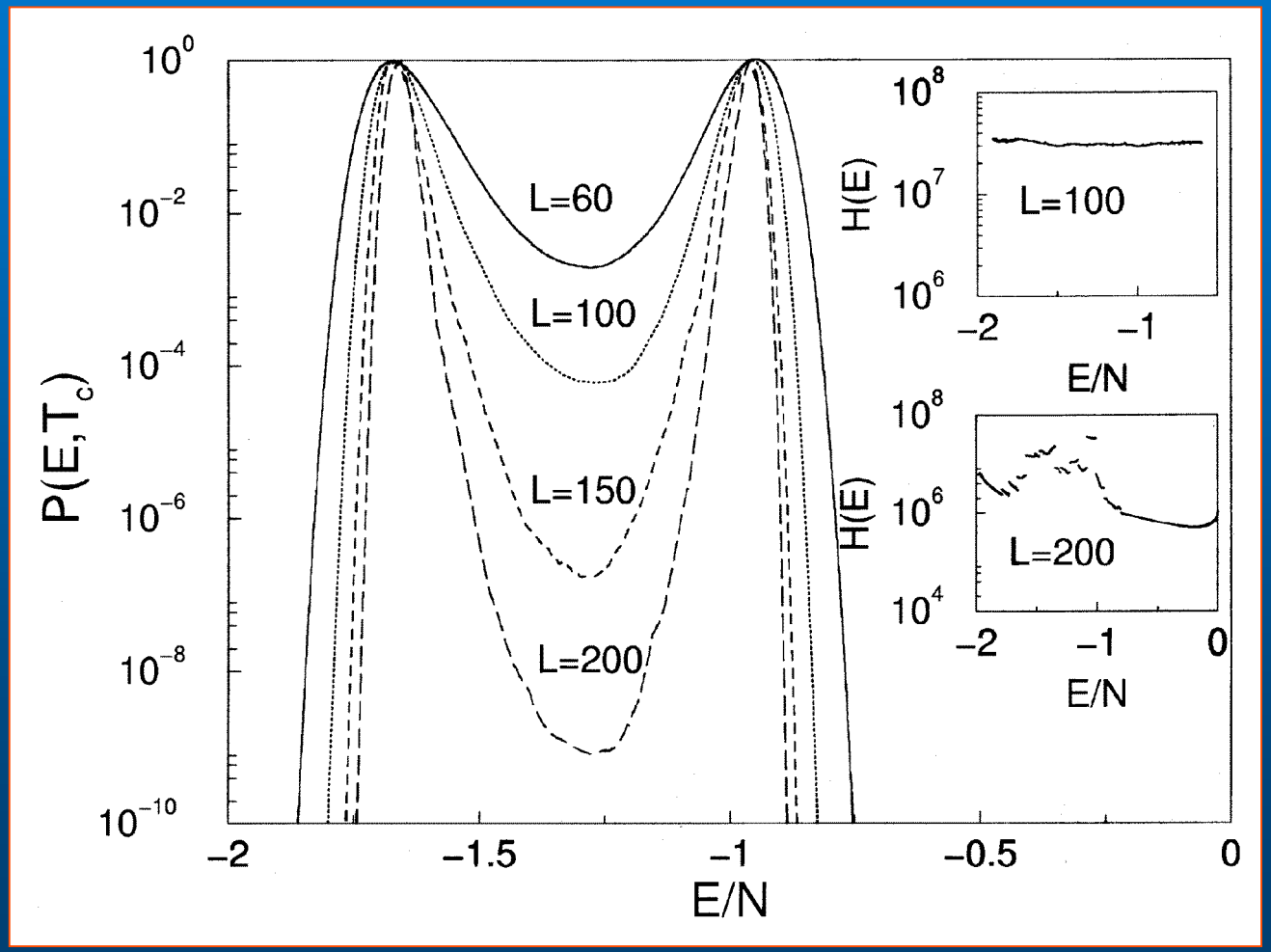


ε = relative error

Wang-Landau sampling at a 1st Order Transition

The q=10 Potts model in 2-dim

At T_c *coexisting states* are separated by an energy barrier



Applications to "Complex" Systems

Applications to "Complex" Systems

- *Spin glasses*
- *"Lattice proteins"*
- *"Real" proteins*

A Magnetic System with Complex "Order"

The EA (Edwards-Anderson) spin glass model in 3 dim:

$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} \sigma_i \sigma_j$

At T_c (if it exists) a spin glass state forms ⇒
get a "rough" energy landscape where multiple minima are separated by high energy barriers

Define an Order Parameter

First choose a finite lattice groundstate , then

$$q = \frac{1}{n} \sum_{\alpha} < \frac{1}{N} \sum_{i} \sigma_i^{\alpha} \sigma_i > \qquad \text{"EA order parameter"}$$

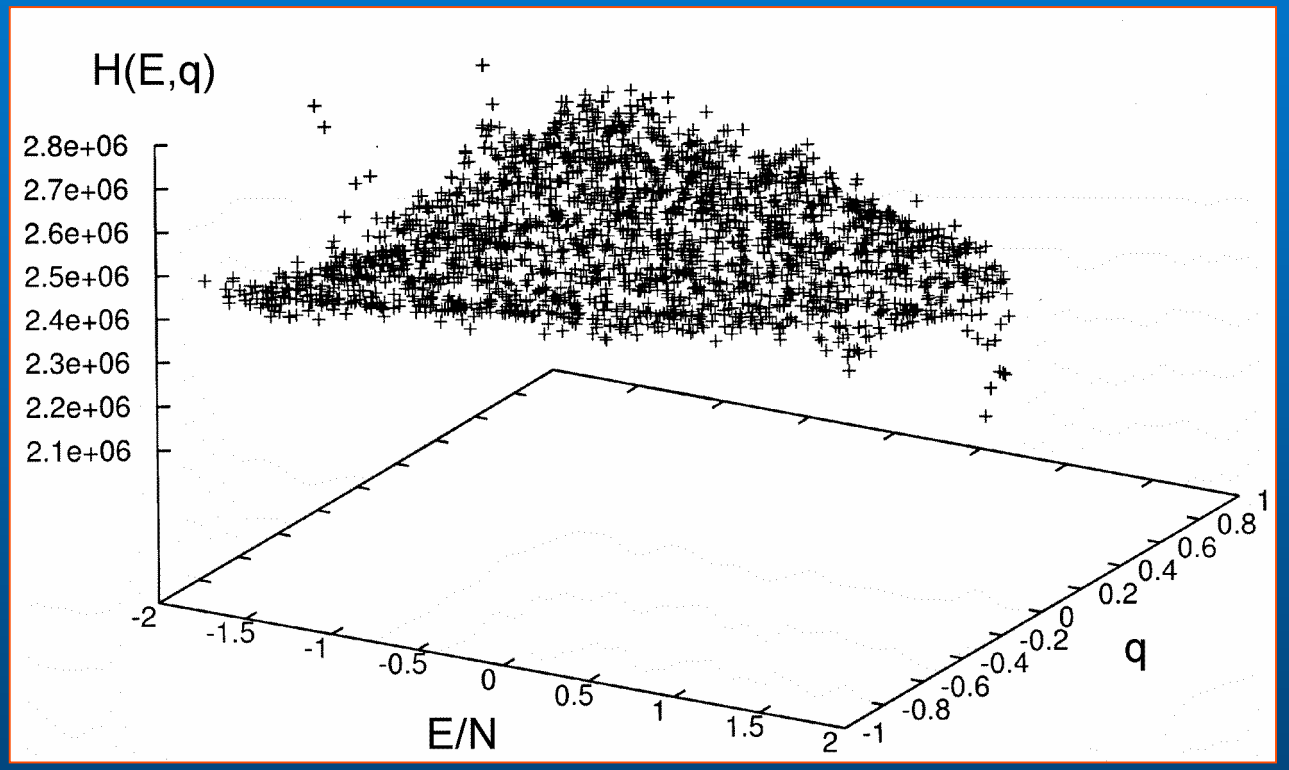
↑

number of bond configurations

***Extend* random walk ⇒ multi-dimensional parameter space**

The 3-dim EA Spin Glass model: A Two-dimensional random walk

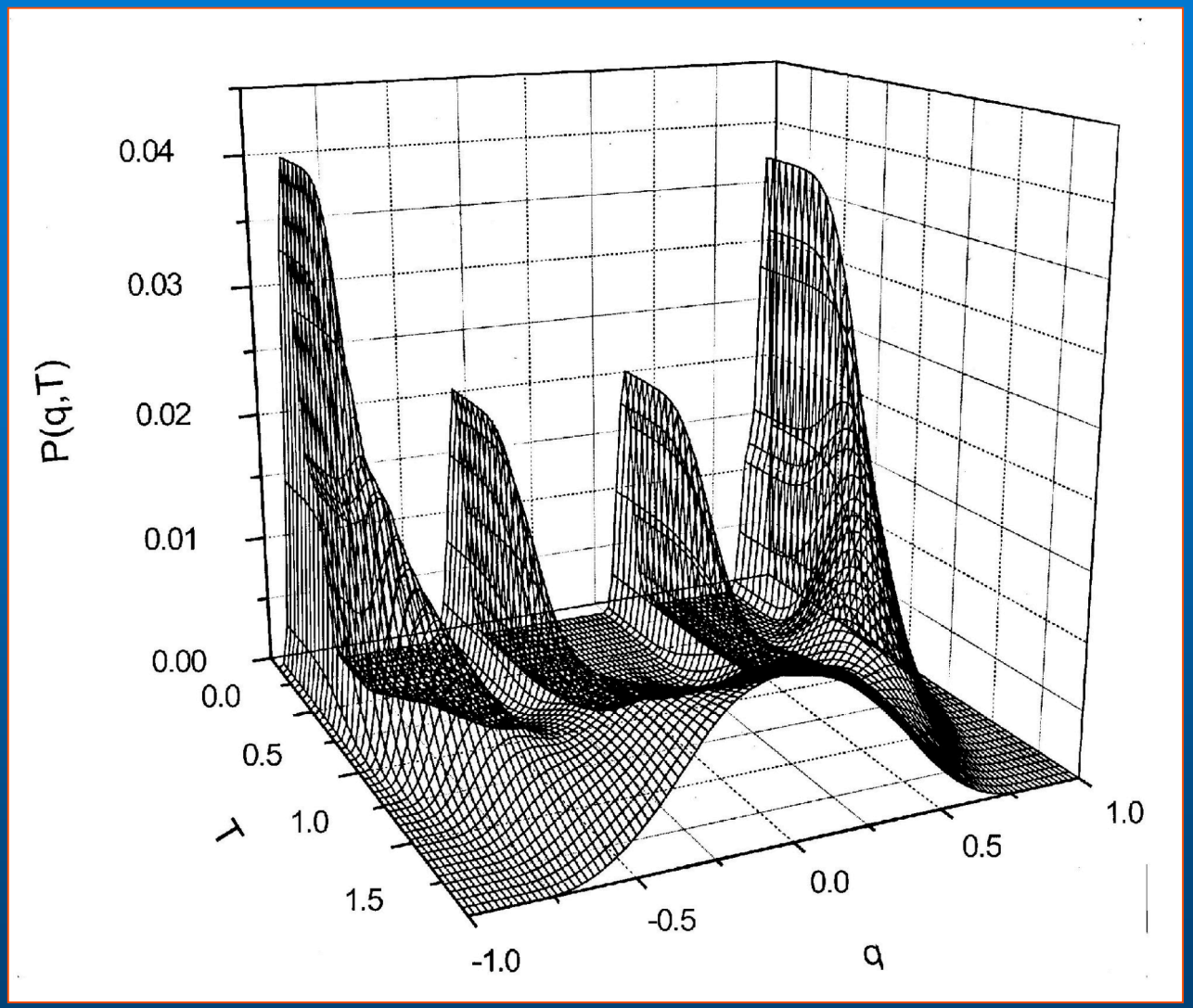
- Energy - order parameter histogram *(L=6)*



Perform a different random walk sequence for each bond distribution

Distribution of States: LxLxL EA Spin Glass

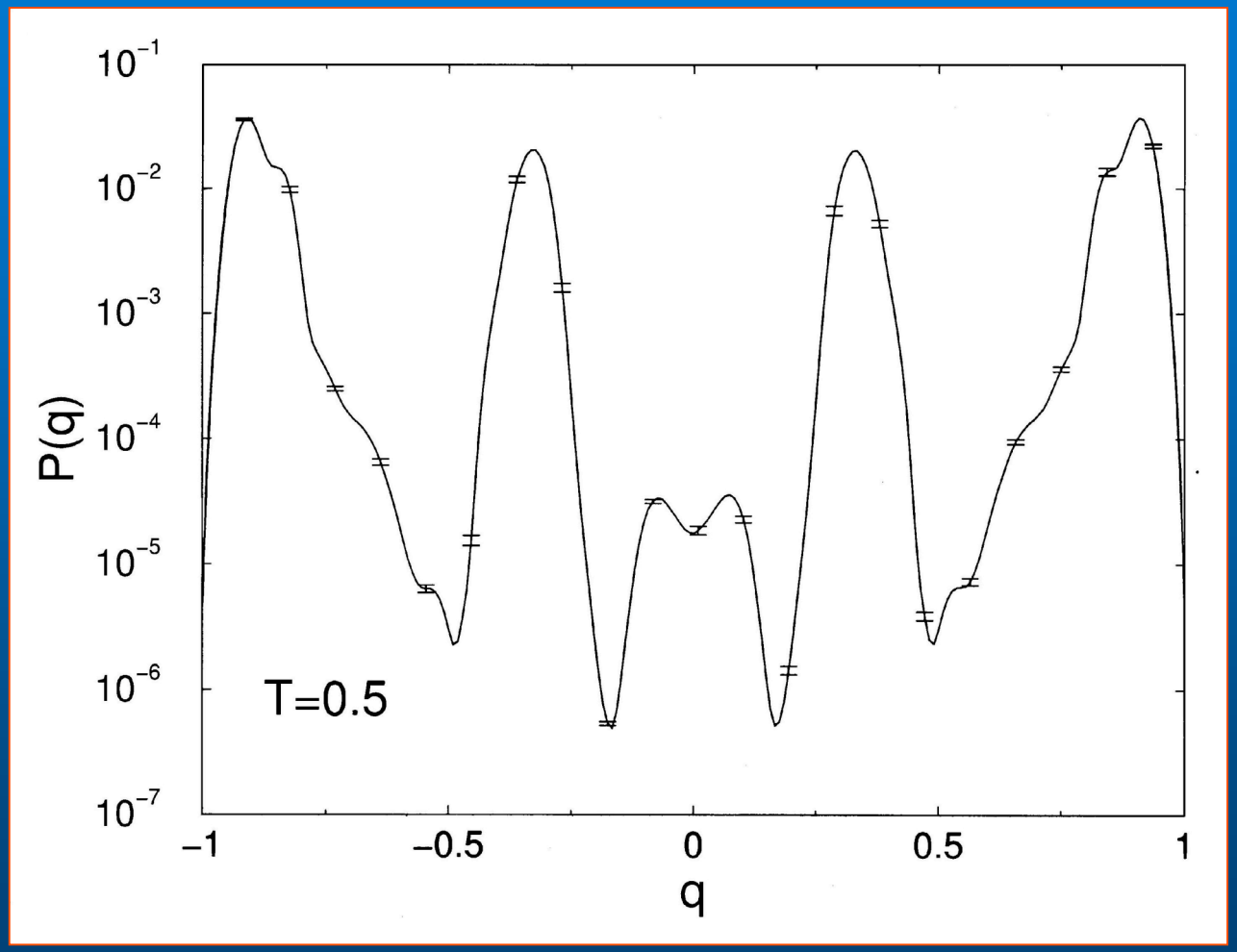
L=6



. . . For larger L, P(q,T) becomes even more complex!

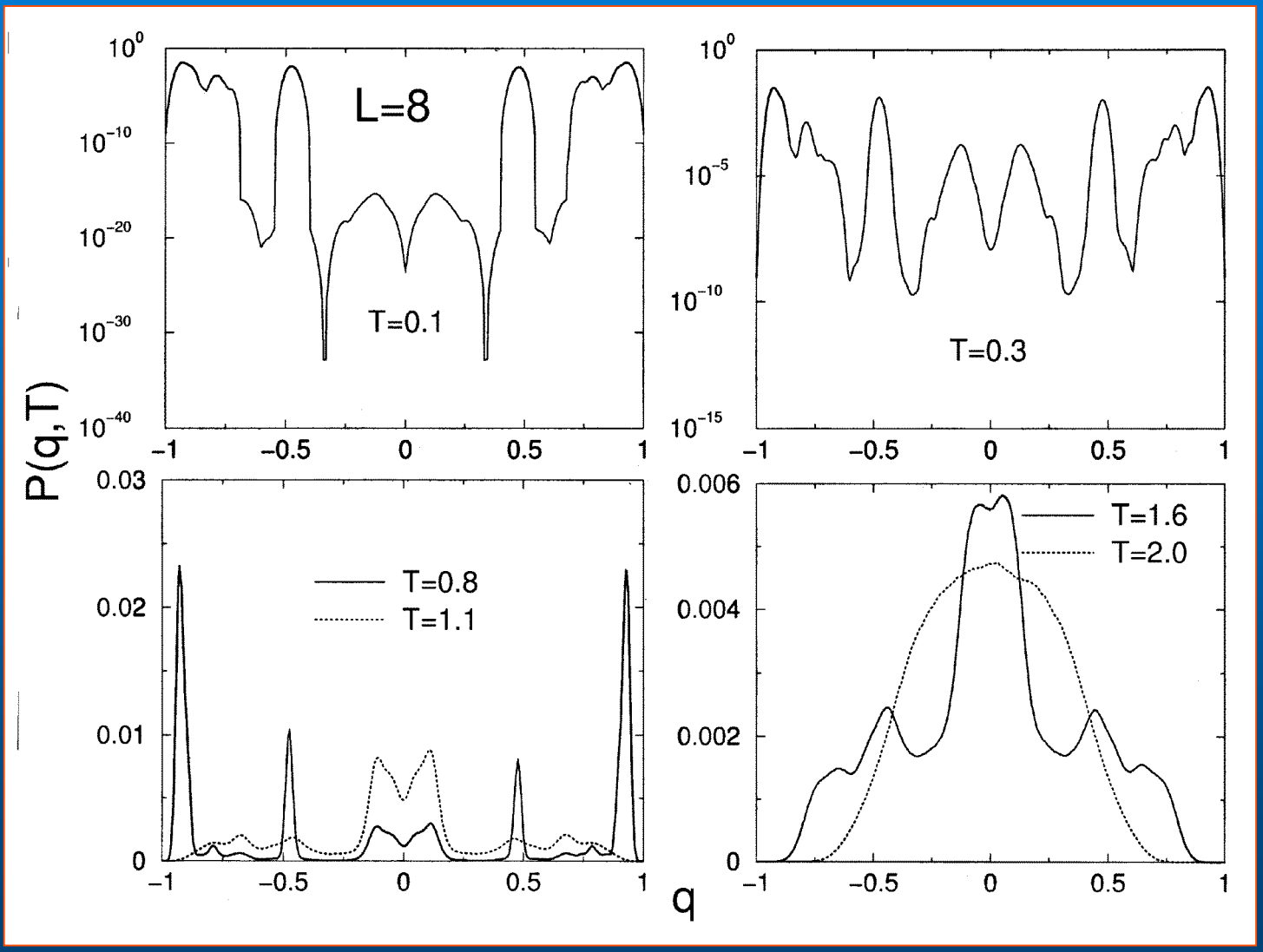
Distribution of States: LxLxL EA Spin Glass

L=6 at low temperature



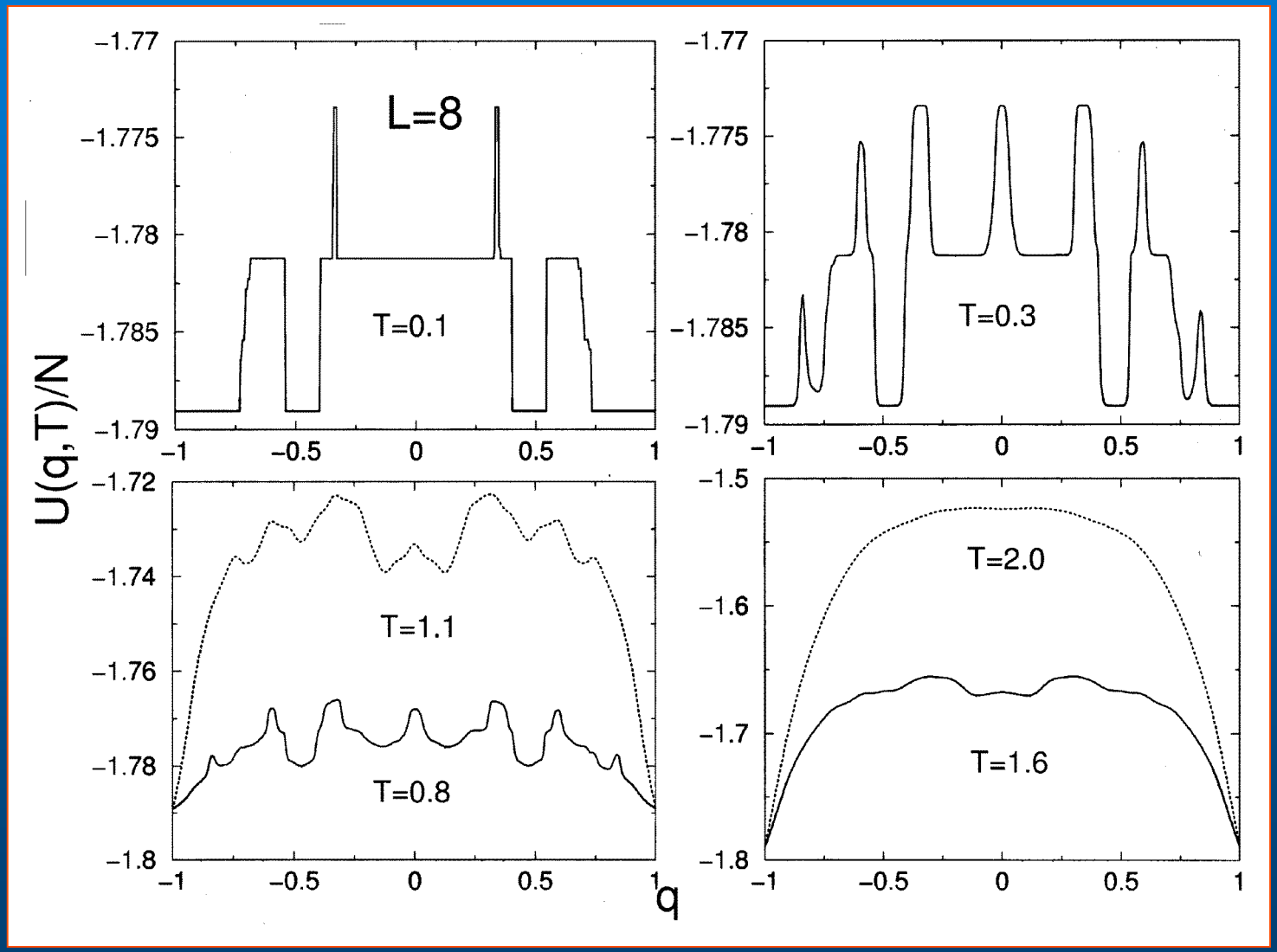
. . . For larger L, P(q,T) becomes even more complex!

Distribution of States: LxLxL EA Spin Glass



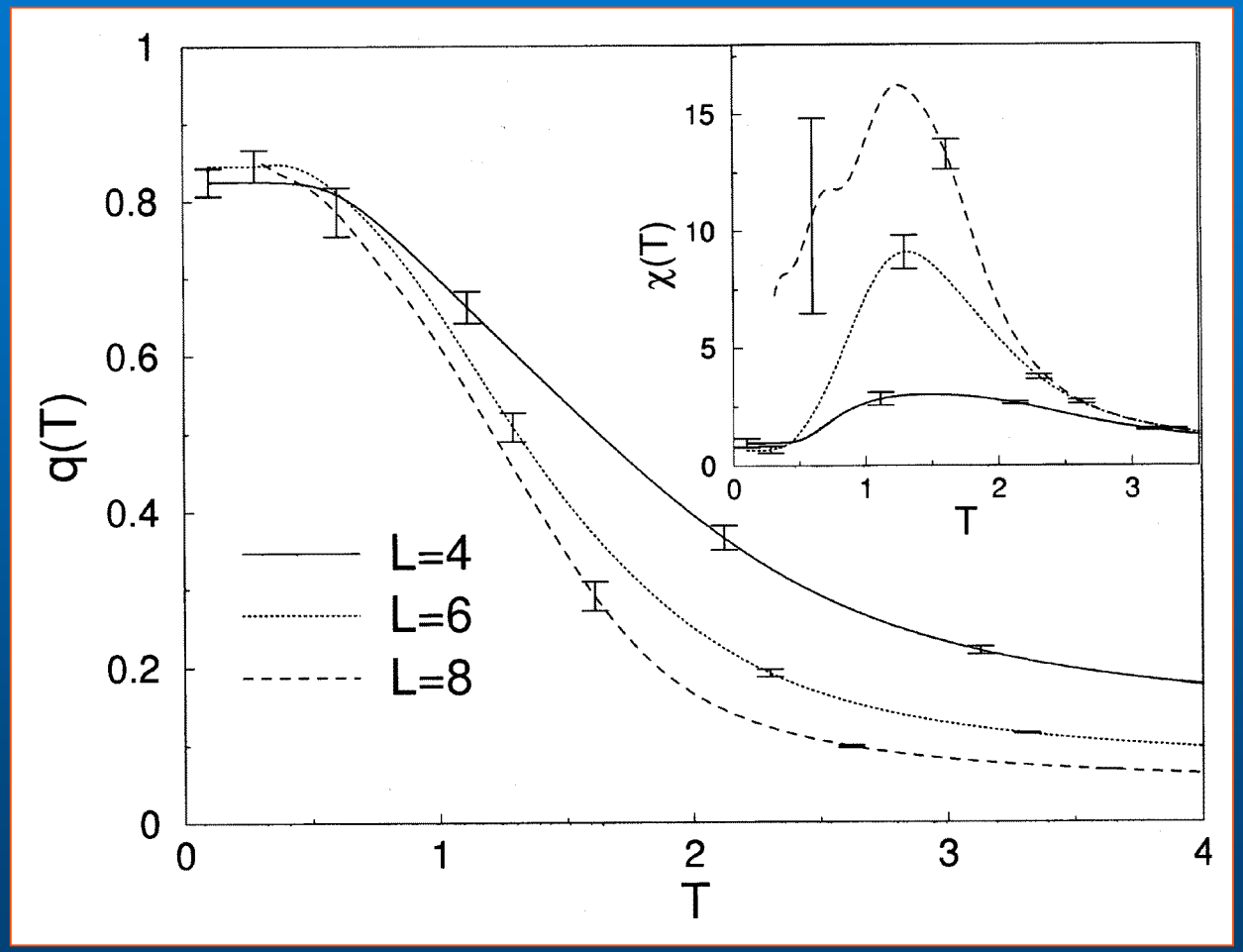
. . . For larger L, P(q,T) becomes even more complex!

Energy landscape – 3d EA Spin Glass Model



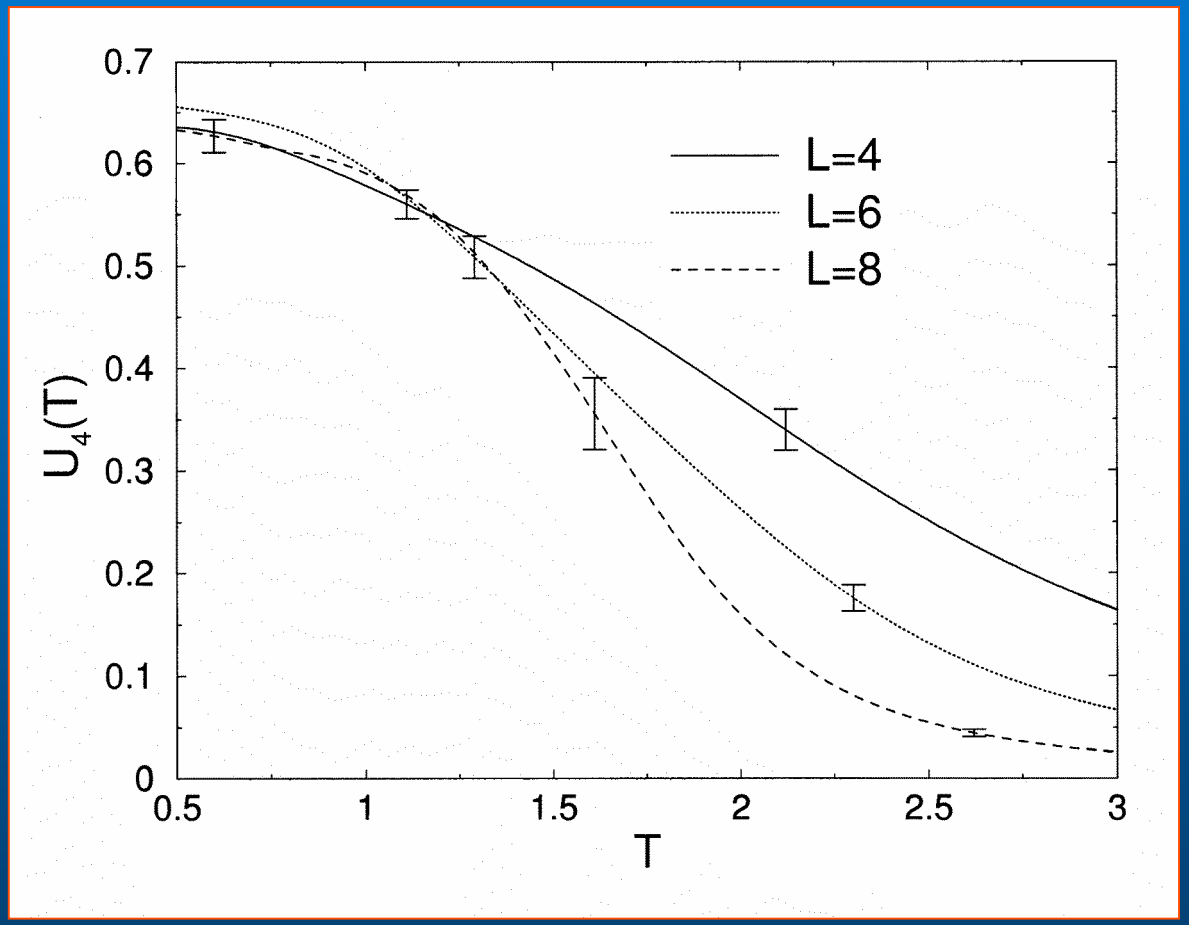
3d EA Spin Glass Model

Look for a phase transition *(Preliminary results)*



3d EA Spin Glass Model

4th order Cumulant crossing *(Preliminary results)*



Groundstate Properties of the 3-d EA Spin Glass

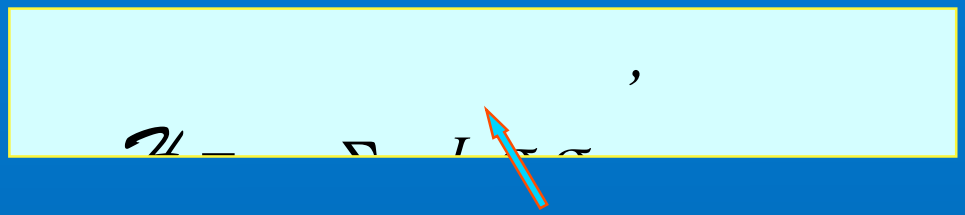
Entropy and energy for the $L \times L \times L$ simple cubic lattice

	Wang-Landau sampling		Multicanonical sampling*	
L	S_0	E_0	S_0	E_0
4	0.075±0.027	-1.734±0.006	0.0724±0.0047	-1.7403±0.0114
6	0.061±0.025	-1.767±0.024	0.0489±0.0049	-1.7741±0.0074
8	0.049±0.007	-1.779±0.016	0.0459±0.0030	-1.7822±0.0081
12	0.053±0.001	-1.780±0.012	0.0491±0.0023	-1.7843±0.0030
16	0.058±0.004	-1.776±0.004		
20	0.056±0.003	-1.774±0.004		

** Berg, Celik, and Hansmann (1993)*

Variation on a Theme

The EA (Edwards-Anderson) spin glass model in 3 dim:

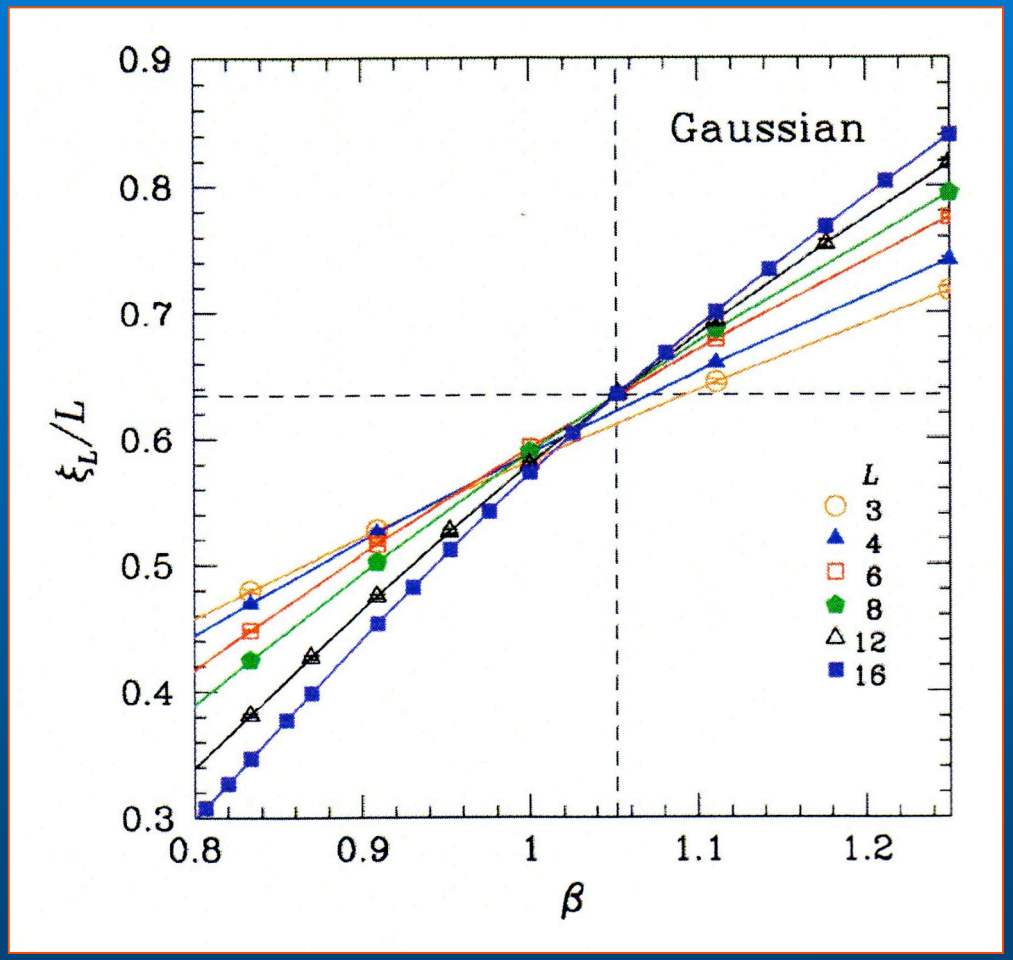


Gaussian distribution

- Use parallel tempering
- Determine the correlation length
- Apply finite size scaling

Katzgraber, Körner, and Young (2006)

The 3d Gaussian EA model: Scaling of the correlation length



$\beta = 1/T$

Katzgraber, Körner, and Young (2006)

Applications to "Complex" Systems

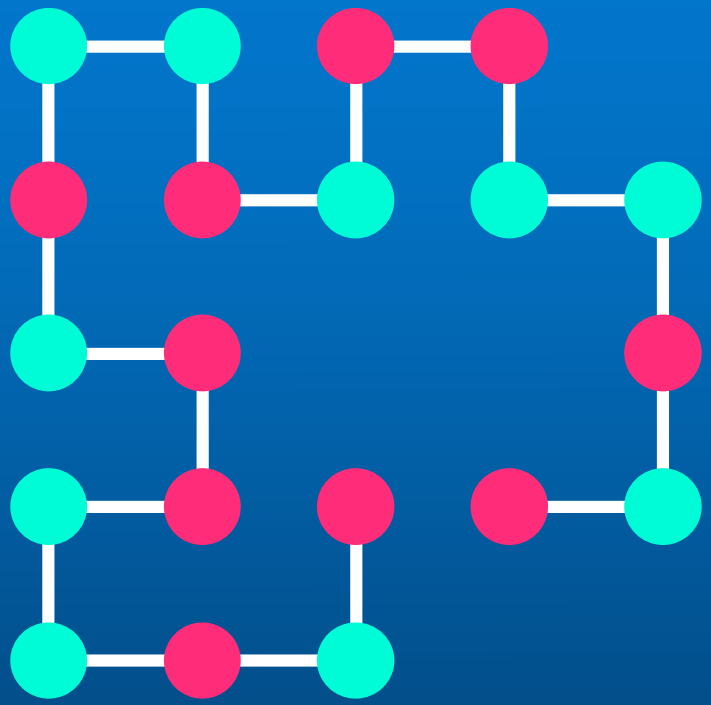
- *Spin glasses*
- *"Lattice proteins"*
- *"Real" proteins*

A Biological "Grand Challenge: Protein Folding

Real proteins are long polymers with side chains of different types and complicated interactions $\Rightarrow$ simplify . . . but how much?

A "Biologically inspired" problem

The HP model of protein folding



Amino acid = "bead"

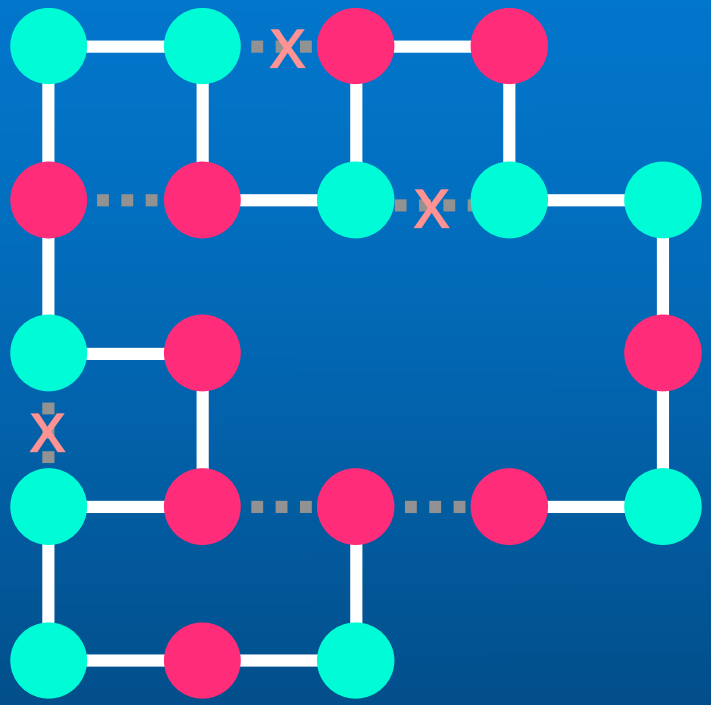
- Hydrophobic (H)
- Polar (P)

Protein sequence = "HPHPPHHPHPP…"

Protein conformation = "self-avoiding walk" on a lattice, e.g. square (2D), cubic (3D)

A "Biologically inspired" problem

The HP model of protein folding



Amino acid = "bead"

- Hydrophobic (H)
- Polar (P)

Protein sequence = "HPHPPHHPHPP…"

Protein conformation = "self-avoiding walk" on a lattice, e.g. square (2D), cubic (3D)

Nearest-neighbor interactions
between non-covalently bound neighbors

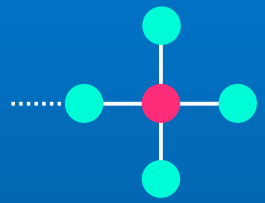
$E_{HH} = -1$, $E_{HP} = 0$, $E_{PP} = 0$ $\Rightarrow$ Compact hydrophobic core / polar (hydrophilic) shell

(Dill, Biochemistry 1985; Lau, Dill, Macromolecules 1989)

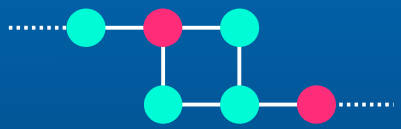
The importance of move sets

Local moves:

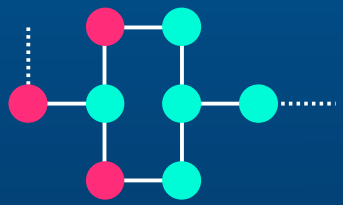
End flip (1 bond)



Kink flip (2 bonds)



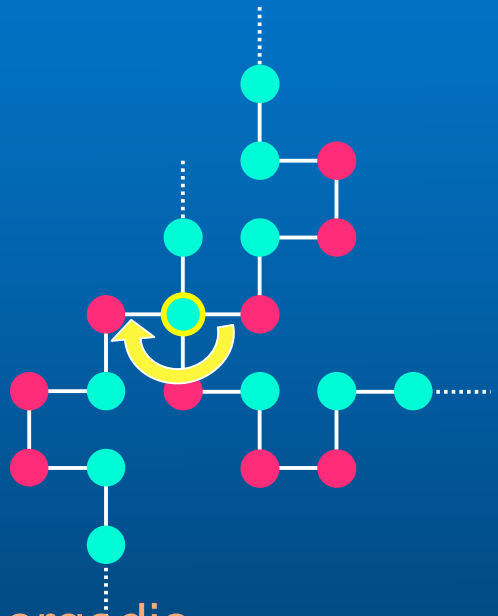
Crankshaft (3 bonds)



⇒ non-ergodic

Non-local moves:

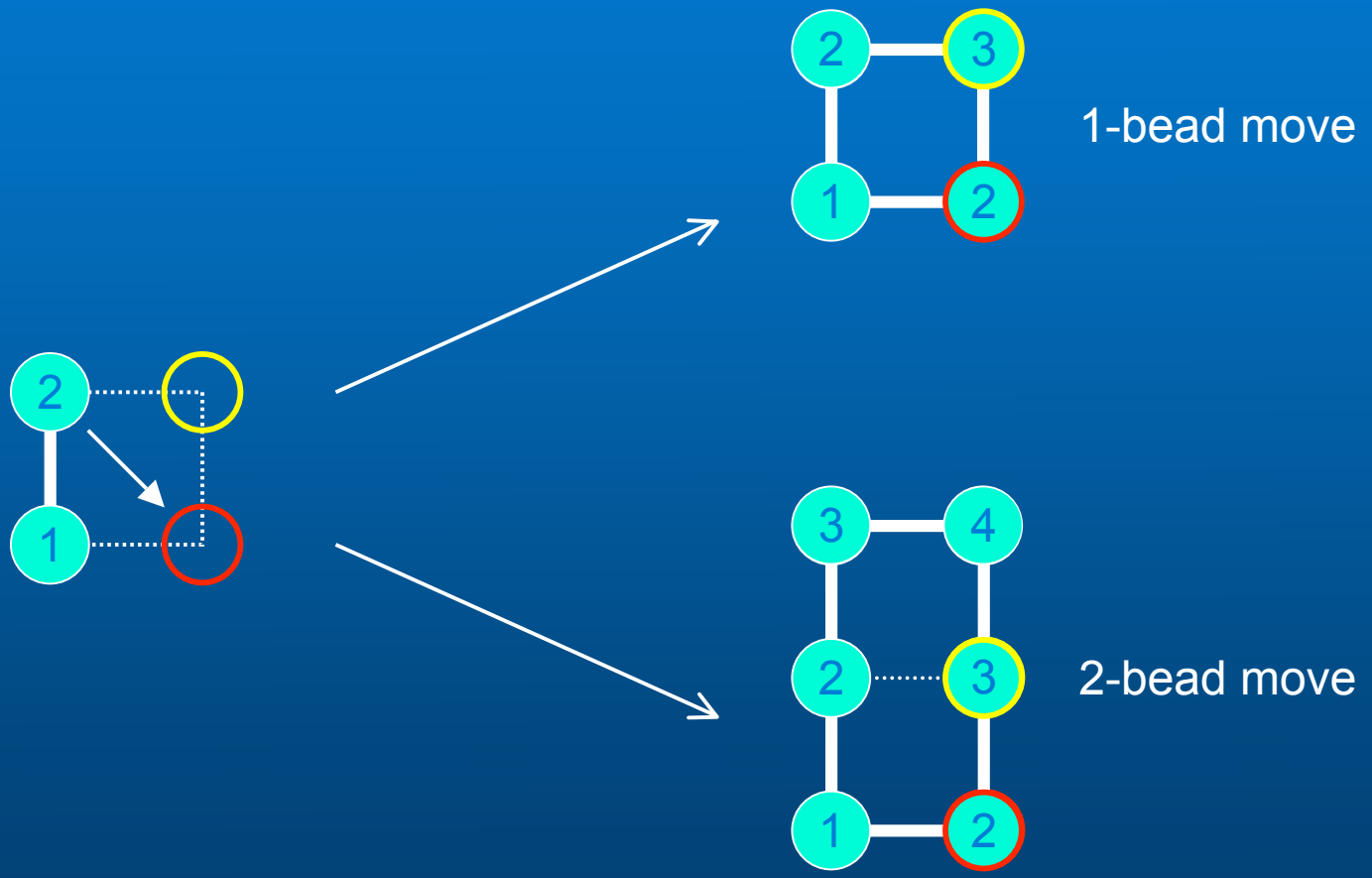
Pivot move



⇒ ergodic

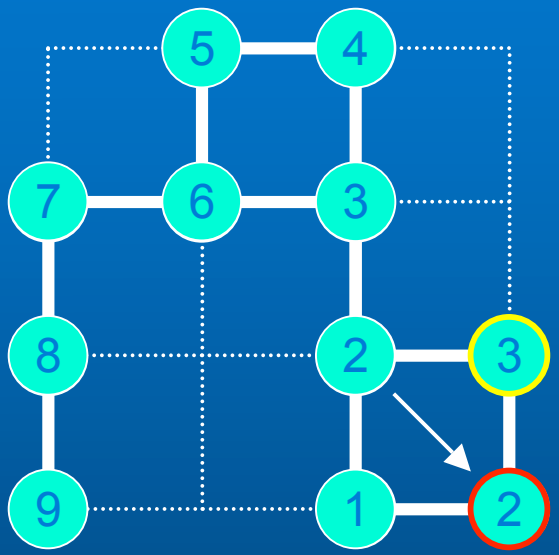
… but inefficient for dense conformations ⇒ high rejection probability

Pull moves

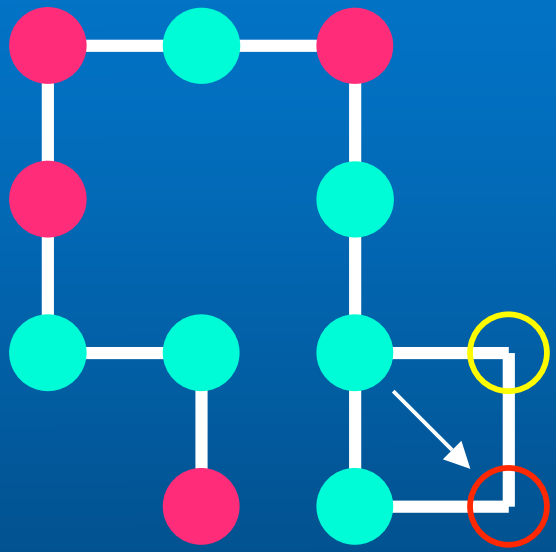


(Lesh, Mitzenmacher, Whitesides, 2003)

Pull moves

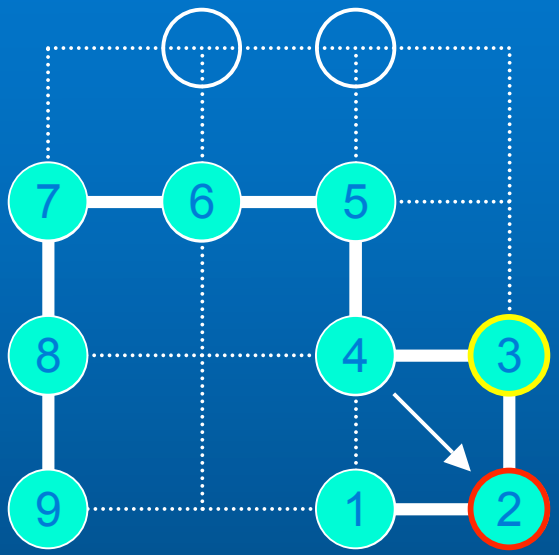


multi-bead move
(Completes internally)



multi-bead move
(Pulls until the end of the sequence)

Pull moves



Extensible to n dimensions

Ergodic (complete)

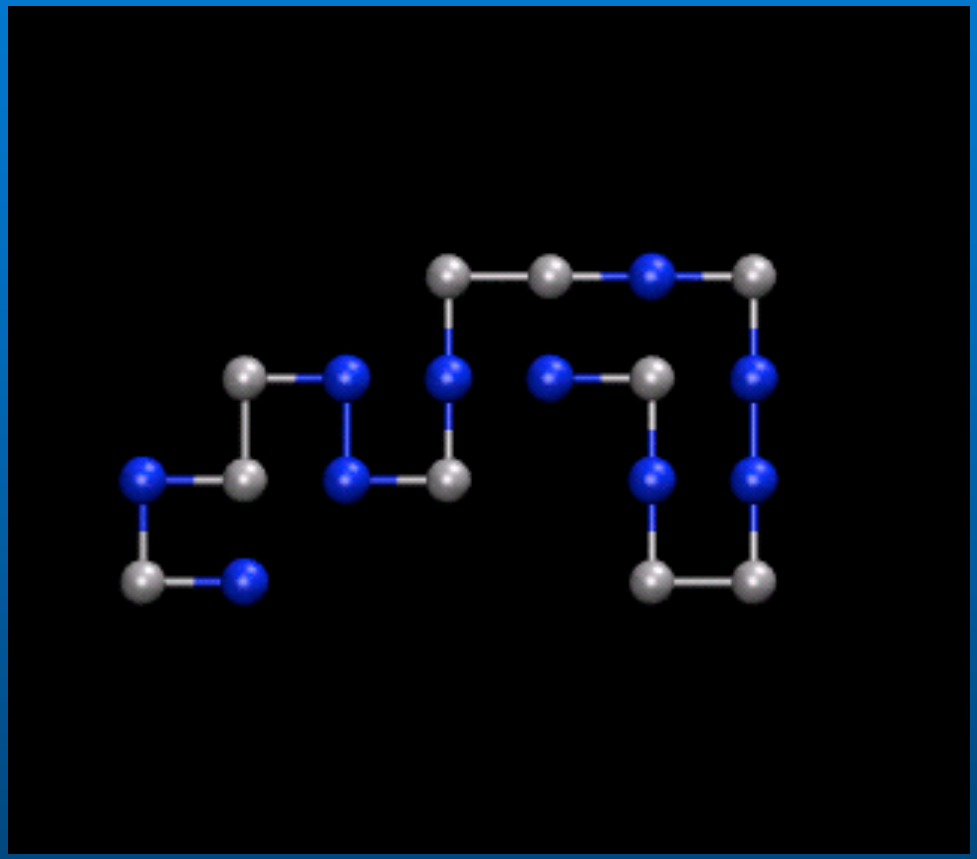
Reversible
$\Rightarrow n(A \rightarrow B) = n(B \rightarrow A)$ (detailed balance!)

No time-consuming self-avoidance test required

Good balance: local $\leftrightarrow$ non-local
"Close-fitting"
$\Rightarrow$ High acceptance ratio

$\Rightarrow$ Ideal for Wang-Landau sampling

Wang-Landau sampling of the HP model



Wang-Landau sampling of the HP Model*

64mer in 2 dimensions (square lattice)

Seq2D64

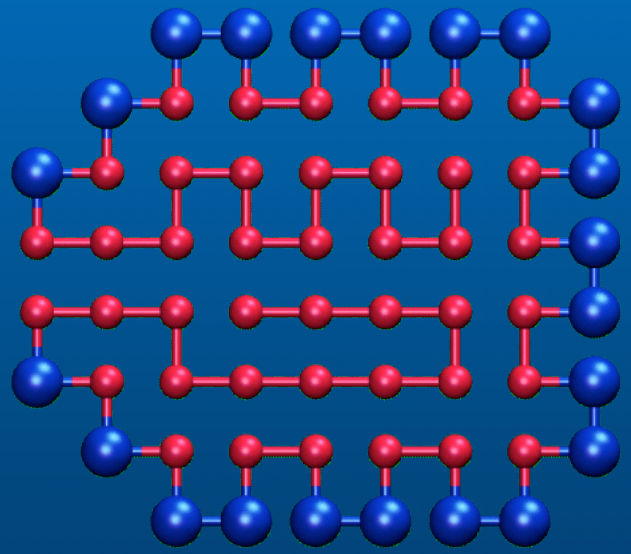
Ground state search

- Core directed chain-growth
  (Beutler, Dill 1996)
- PERM
  (Bastolla et al. 1998)

Density of states

- Multi-self-overlap ensemble (MSOE)
  (Chikenji et al. 1999)
- Equi-energy sampling (EES)
  (Kou et al. 2006)

Ground state (E = -42)

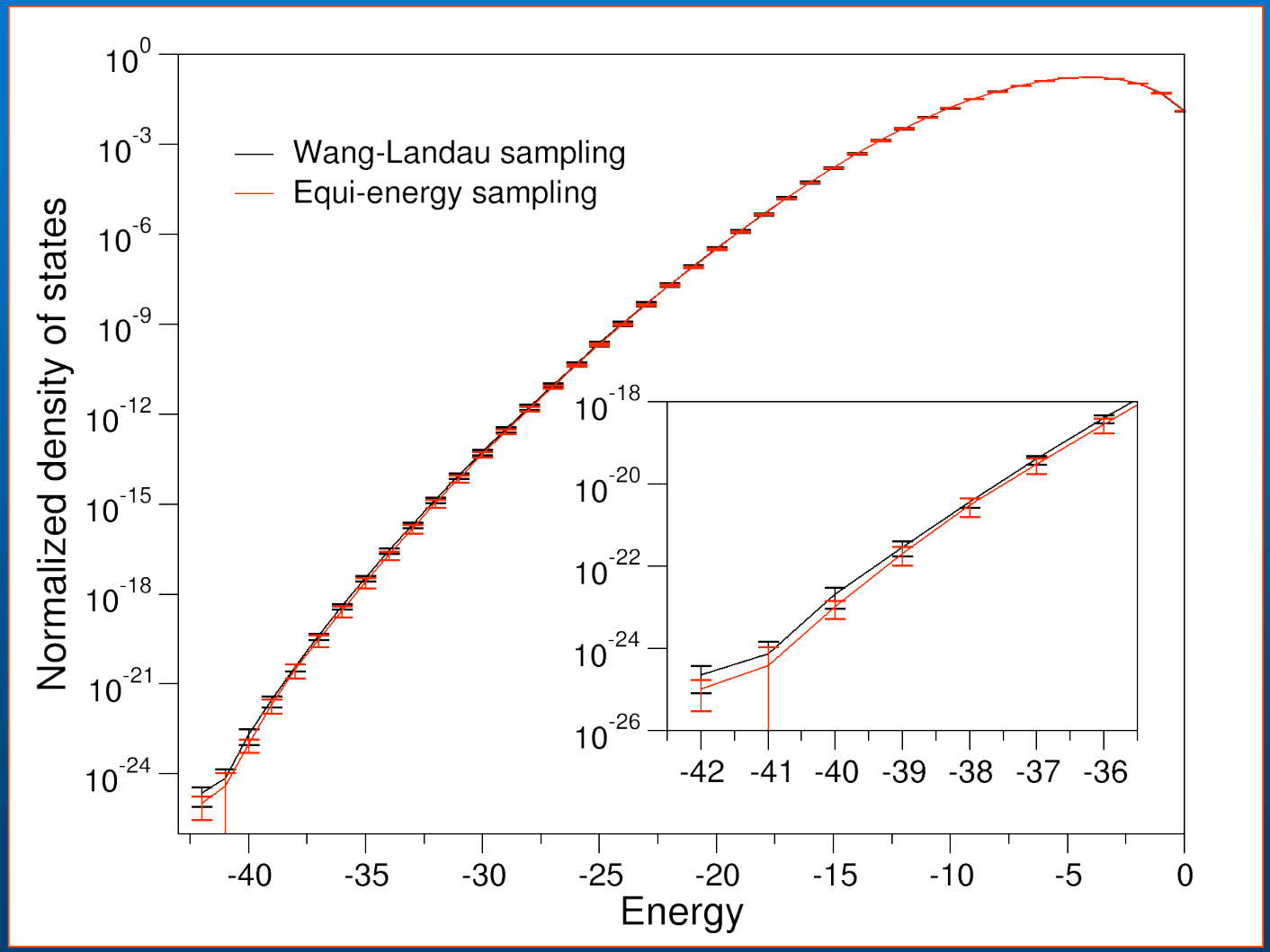


** with T. Wüst*

A "Biologically inspired" problem

The HP Model of Protein Folding

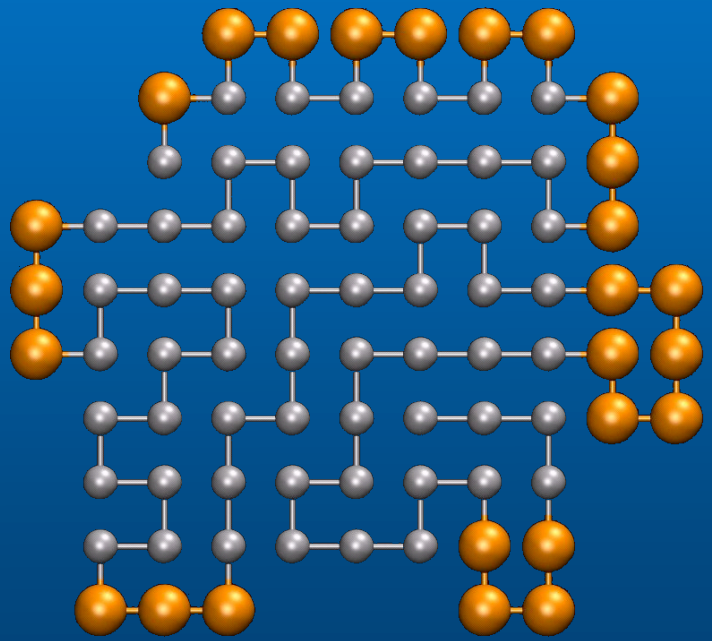
Seq2D64



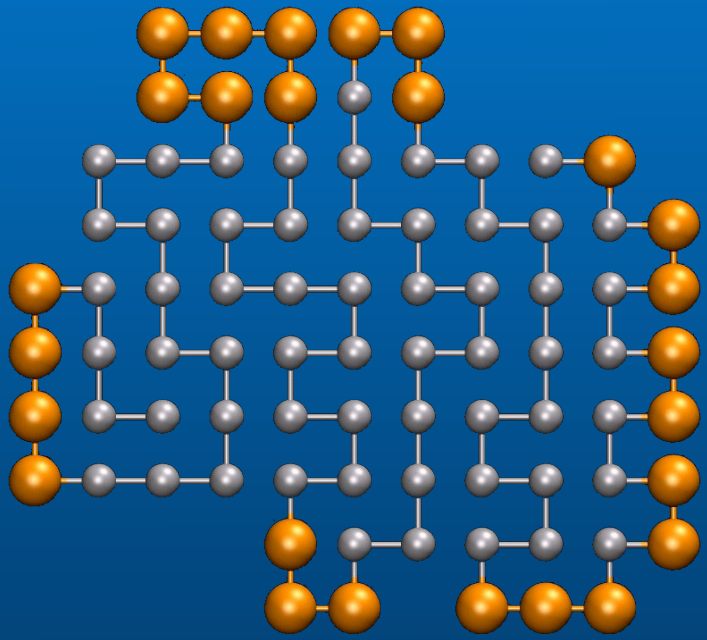
85mer in 2 dimensions (square lattice)

Seq2D85

Ground state (E = -53)



1st excited state (E = -52)



Wang-Landau Sampling of the HP model

103mer in 3 dimensions (simple cubic lattice)

Seq3D103

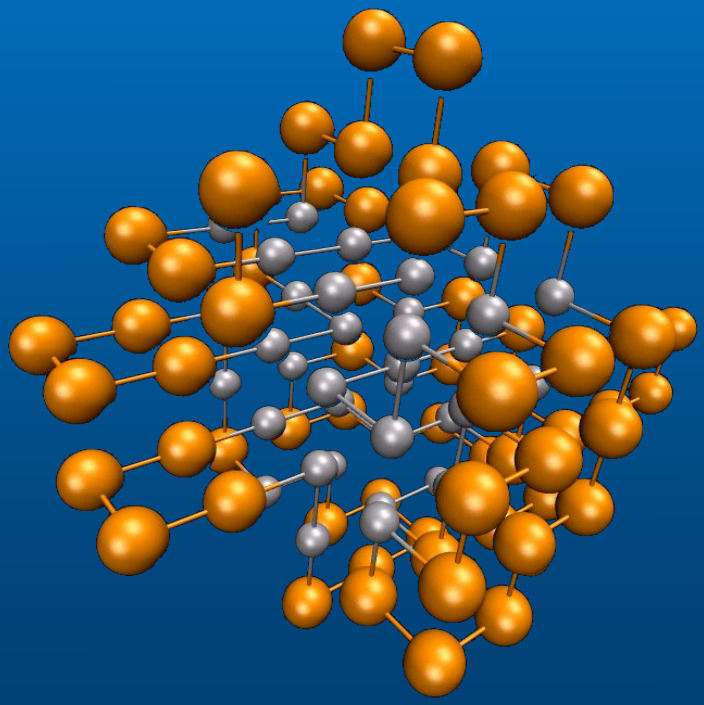
Ground state search

- Fragment regrowth MC
  (Zhang, Kou et al. 2007)

Density of states

- Multicanonical chain-growth (MCCG)
  (Bachmann, Janke 2003 / 2004)

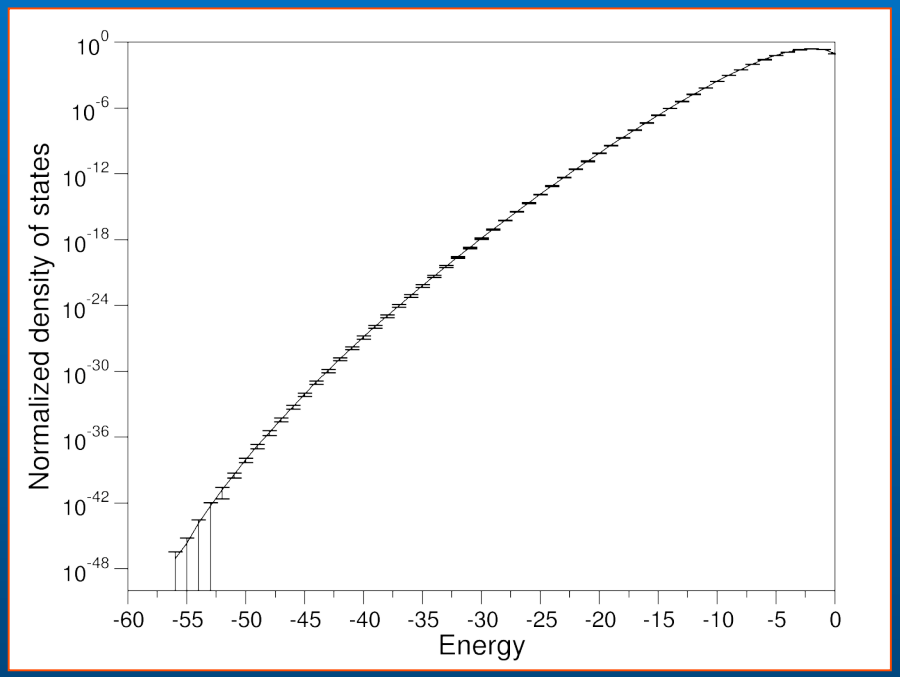
Ground state (E = -57)



Seq3D103: Comparison WLS ⇔ MCCG

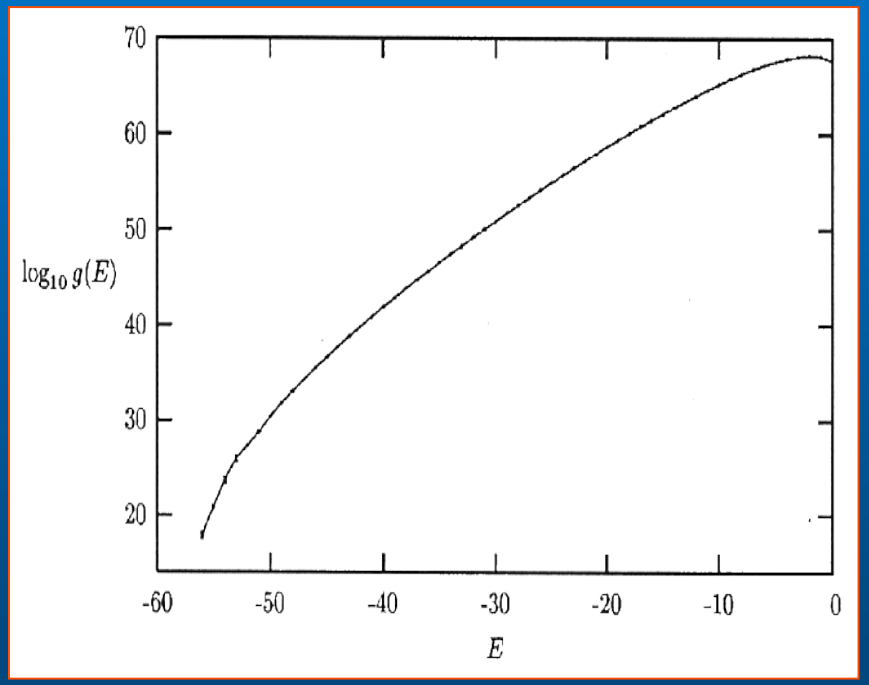
Density of states: -56 to 0 (without E = -57)

Wang-Landau sampling



5 runs
≈ 80h_{CPU} / run

Multicanonical chain-growth

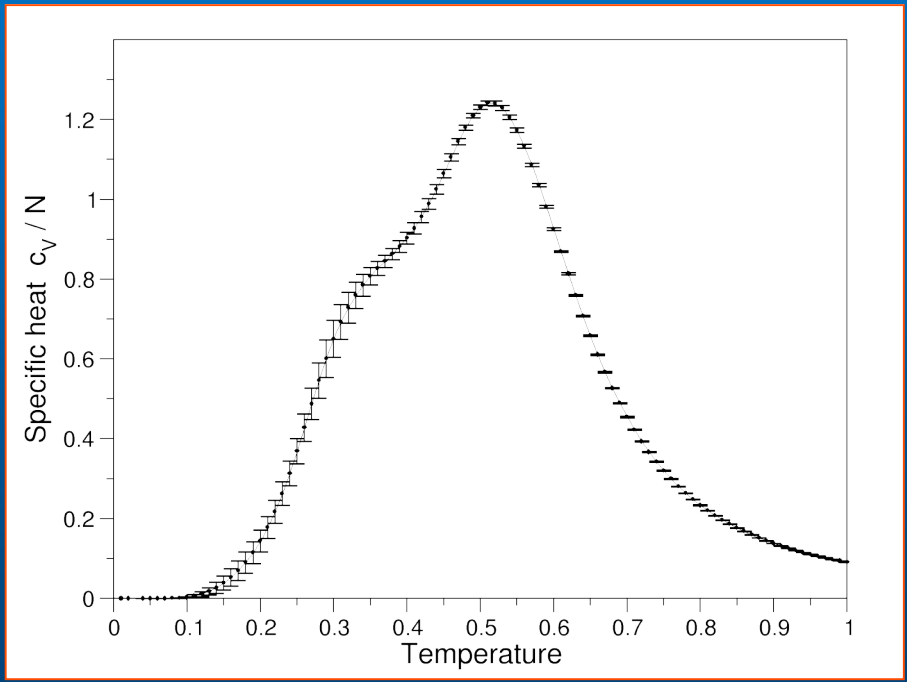


(Bachmann, Janke, J. Chem. Phys. 2004)

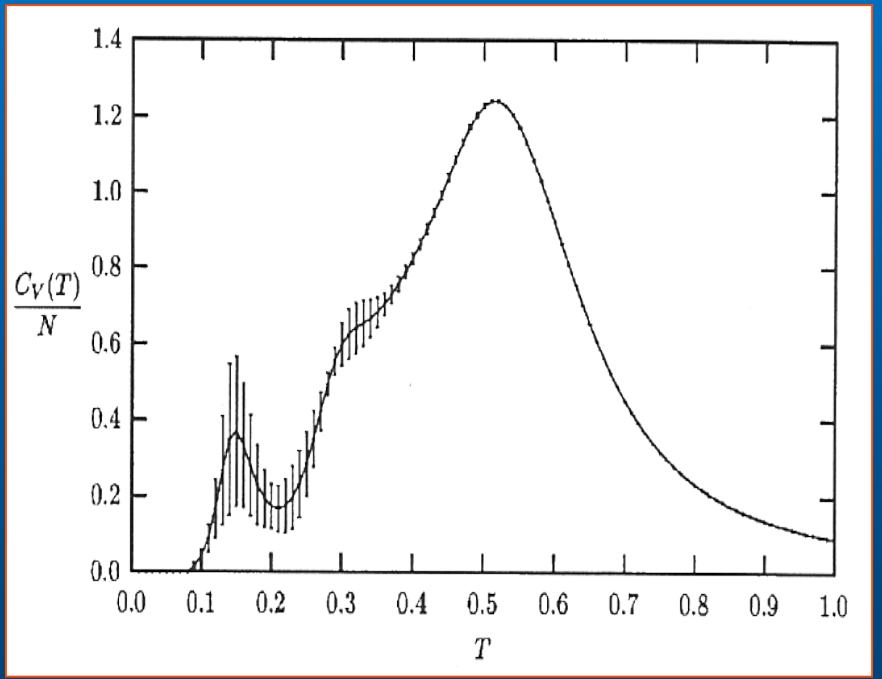
Seq3D103: Comparison WLS ⇔ MCCG

Specific heat c_V (T) / N

Wang-Landau sampling



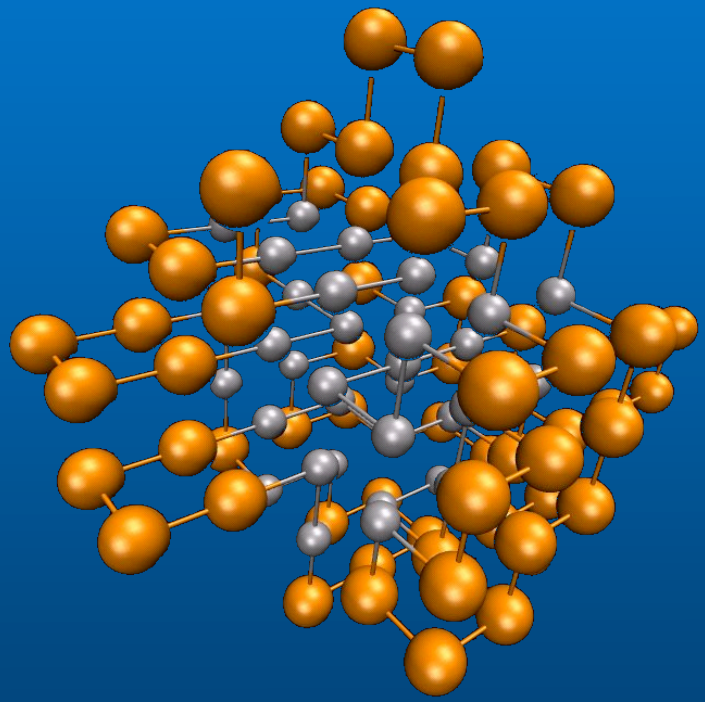
Multicanonical chain-growth



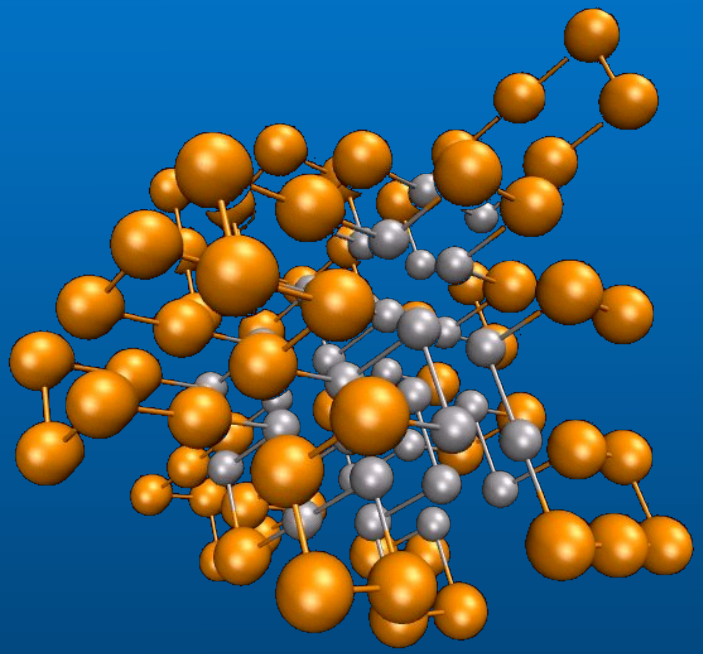
(Bachmann, Janke, J. Chem. Phys., 2004)

103mer in 3 dimensions (cubic lattice)

Ground state (E = -57)



1st excited state (E = -56)



Applications to "Complex" Systems

- *Spin glasses*
- *"Lattice proteins"*
- *"Real" proteins*

A Real Biological Problem: Structure of Membrane Proteins

- A few words of introduction and then results for a few real problems

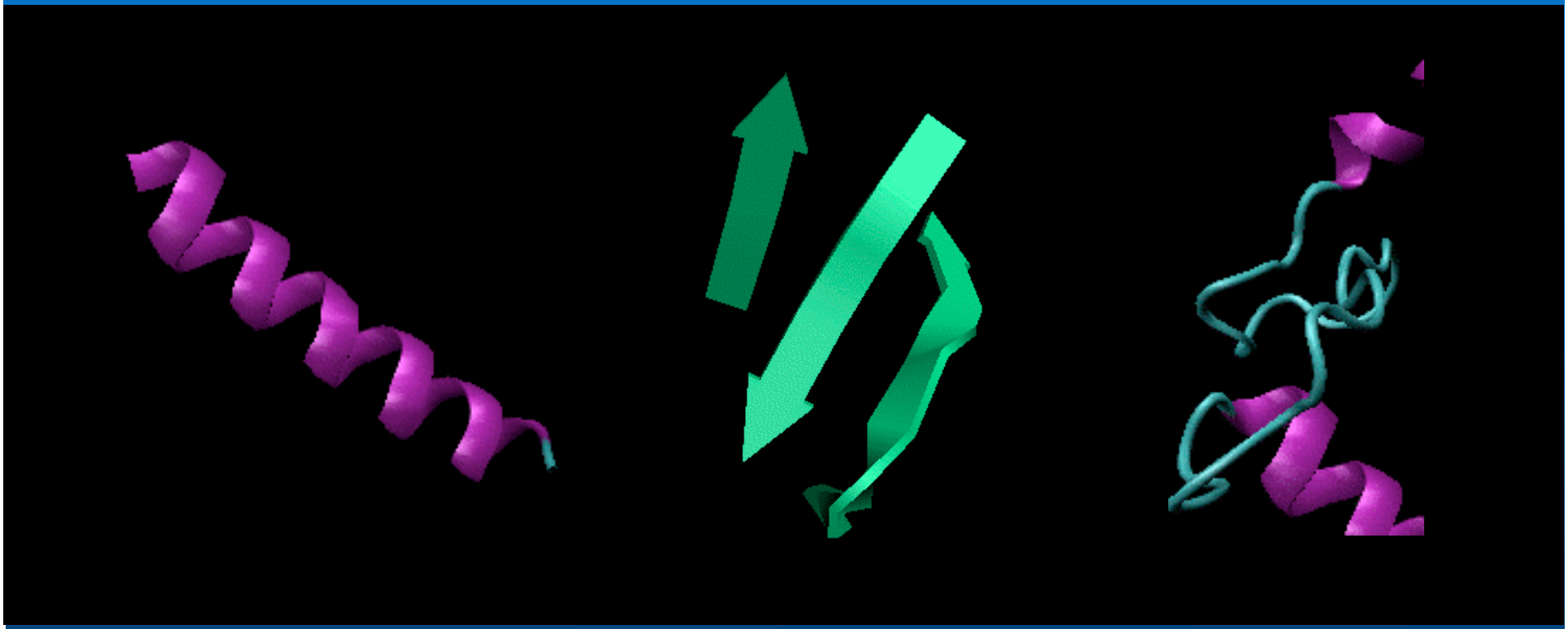
What is a protein?

Primary structure: Sequence of amino acid residues

EITLIIFGVMAGVIGTILLISY

What is a protein?

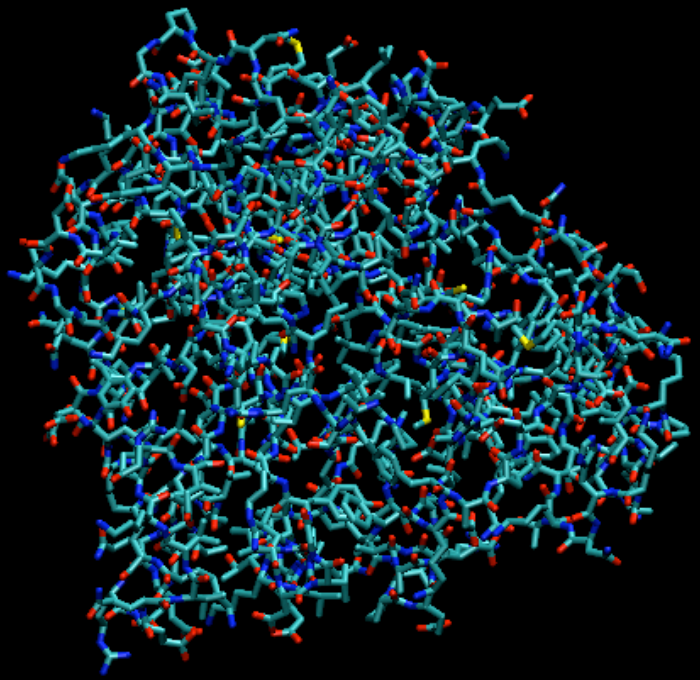
Secondary structure: H-bonds of backbone atoms



Alpha-helix Beta-sheet Loop

What is a protein?

Tertiary structure: 3-dim arrangement of atoms

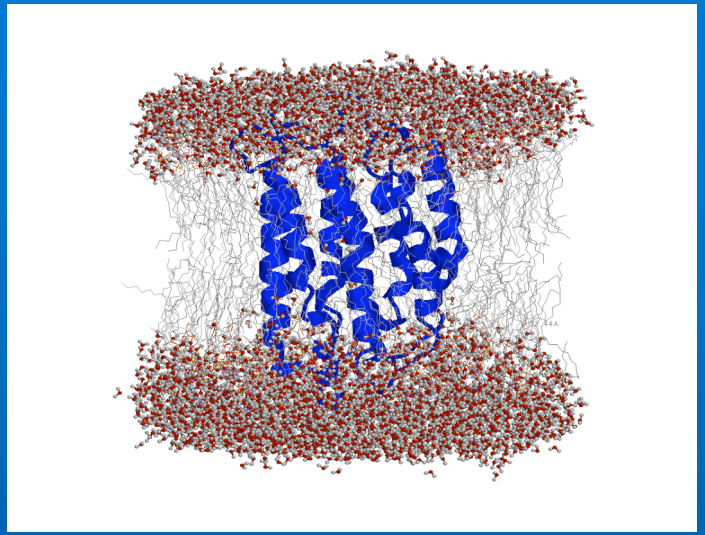


balls & sticks

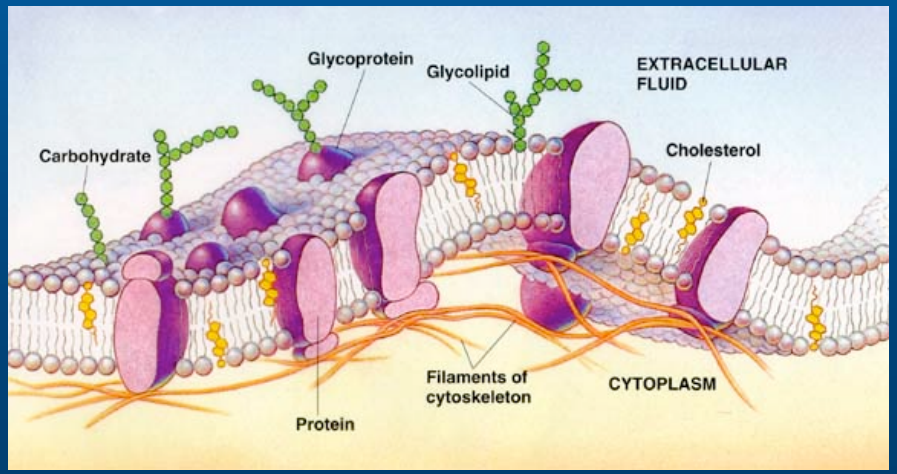
helices & arrows

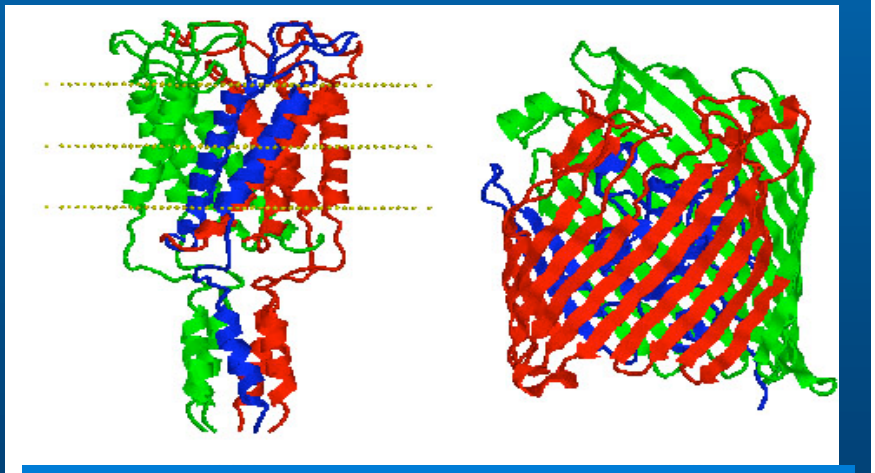
What is a membrane protein?

- **Roles in biological process:**
  - Receptors;
  - Channels, gates and pumps;
  - Electric/chemical potential;
  - Energy transduction



- **> 50% new drug targets are membrane proteins (MP).**

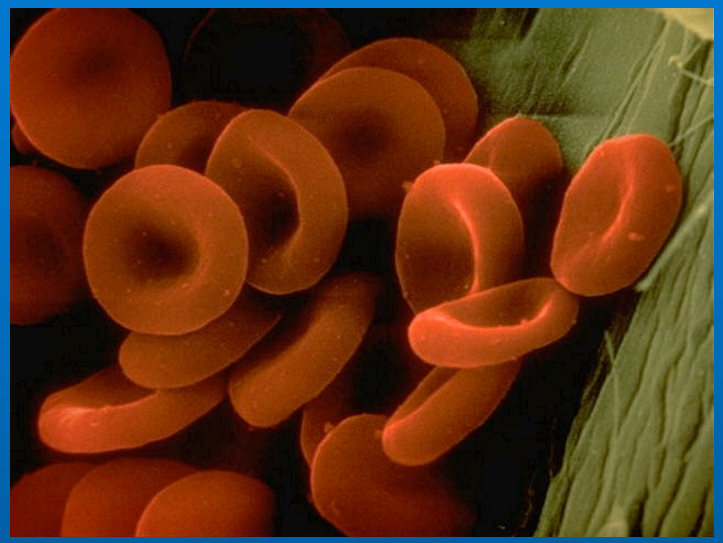


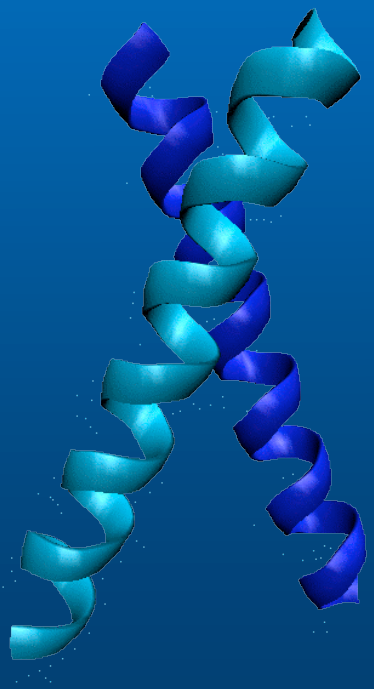


Helical structure Beta structure

A "simple" problem: Glycophorin A (GPA)

- **Function:**
- **Red blood cell membrane**
- **Carries sugar molecules**





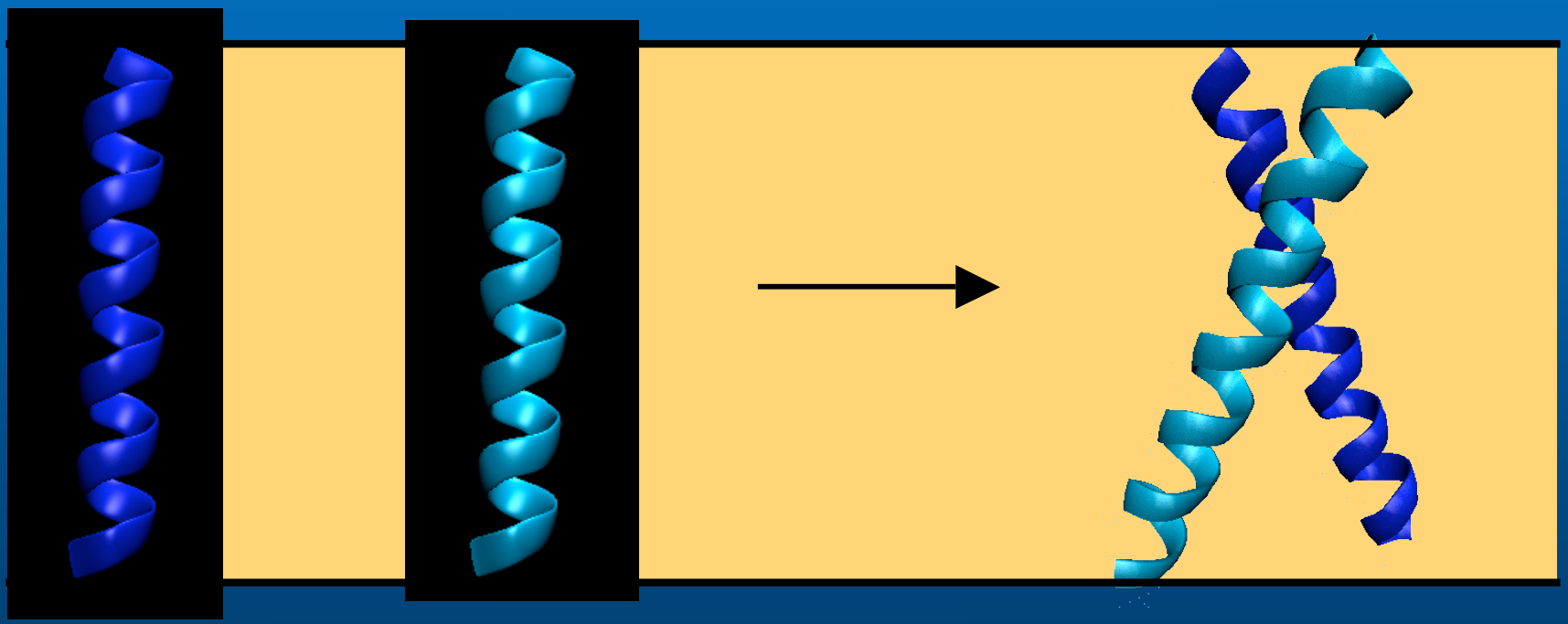
Structure:

- Single alpha-helix
- Symetrical dimer

EITLIIFGVMAGVIGTILLISY

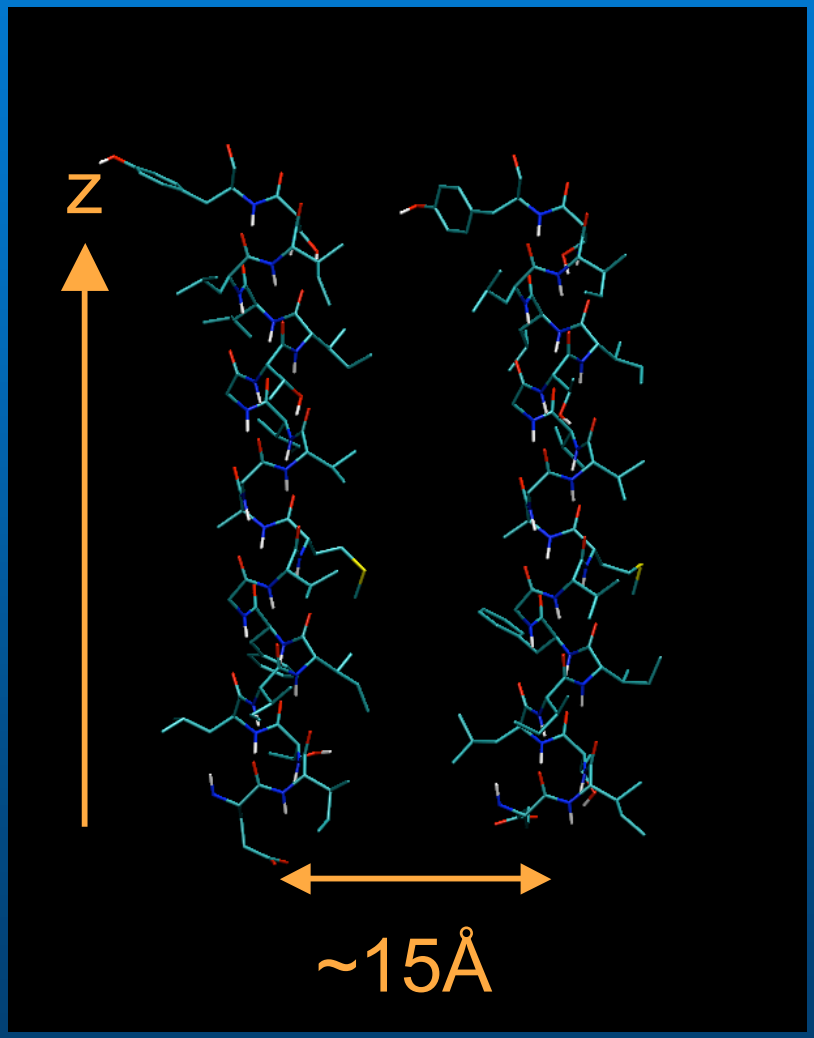
Folding process of GPA

- Single helix stable in the membrane
- Association of helices



(Popot, Engelman, Biochemistry, 1990)

Wang-Landau sampling of a GPA model*

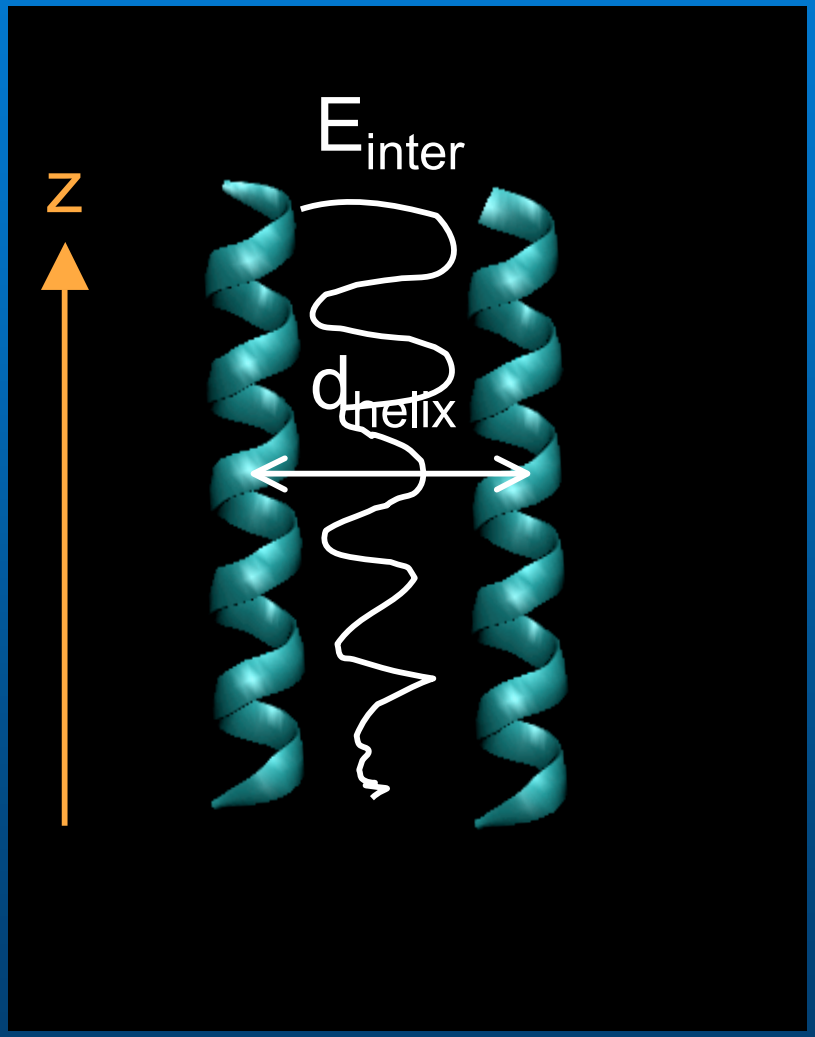


- ➢Unified-atom model
- ➢Total: 368 atoms
- ➢2 helices (22 amino-acids)

- Energy :
  - CHARMM19
  - Lipid potential
- Starting structure: parallel helices
- 7 Monte Carlo Moves: protein, helix, side-chain

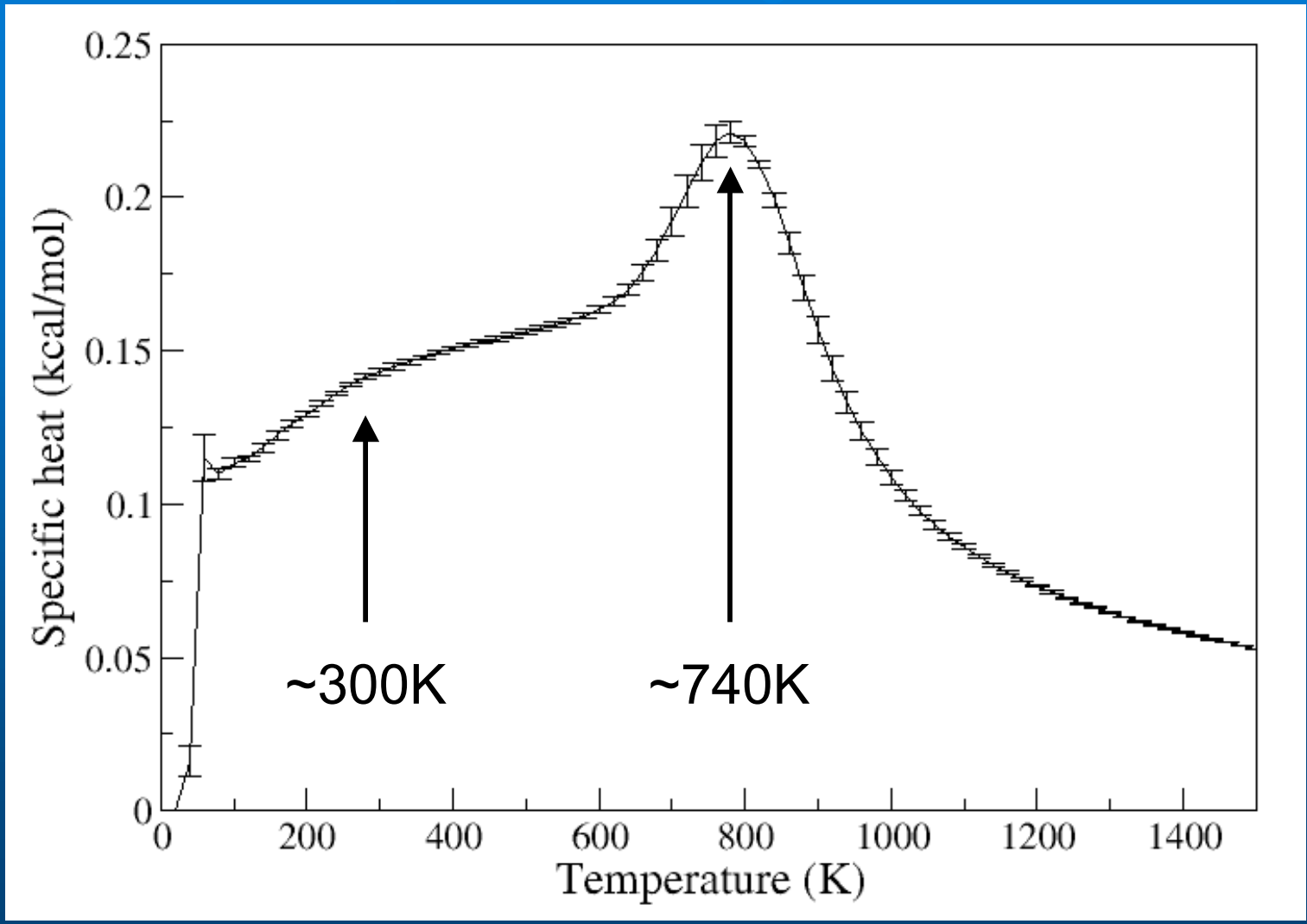
**with Claire Gervais, IOB*

W-L sampling - Observables for GPA

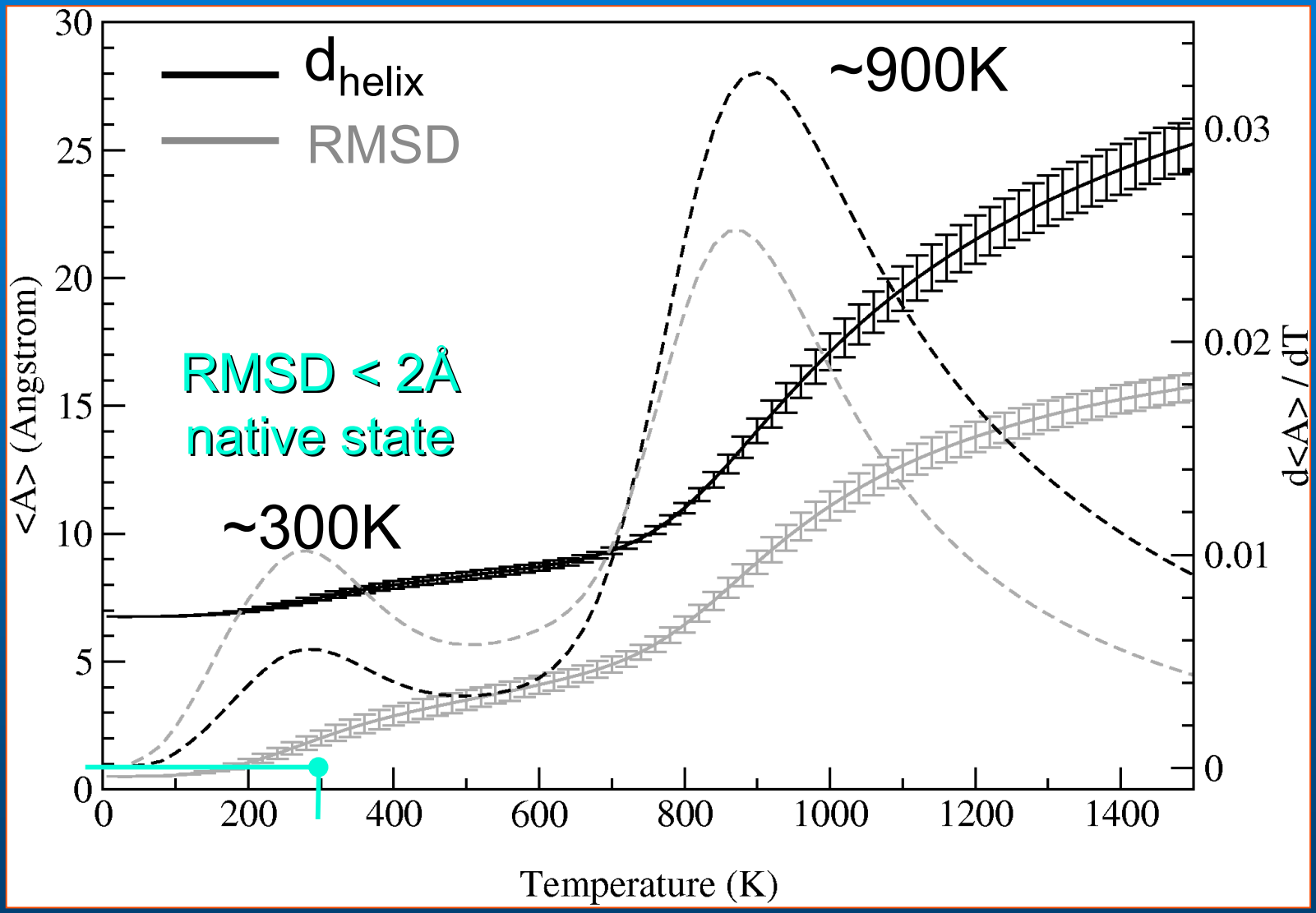


- helix-helix nonbond energy (E_{inter})
- helix-helix distance (d_{helix})
- RMSD of C_α atoms

Specific heat for GPA



GPA: RMSD and d_{helix}



W-L sampling GPA Results

Study residue energies, heat capacities, etc.:

- First native contacts appear at ~740K
- Final native contacts at ~300K

→ **Gradual convergence to the native state**

- Appearance of native contacts: Leucine → Glycine → Threonine

→ **Hierarchical acquisition of the native state**

Another protein folding example:

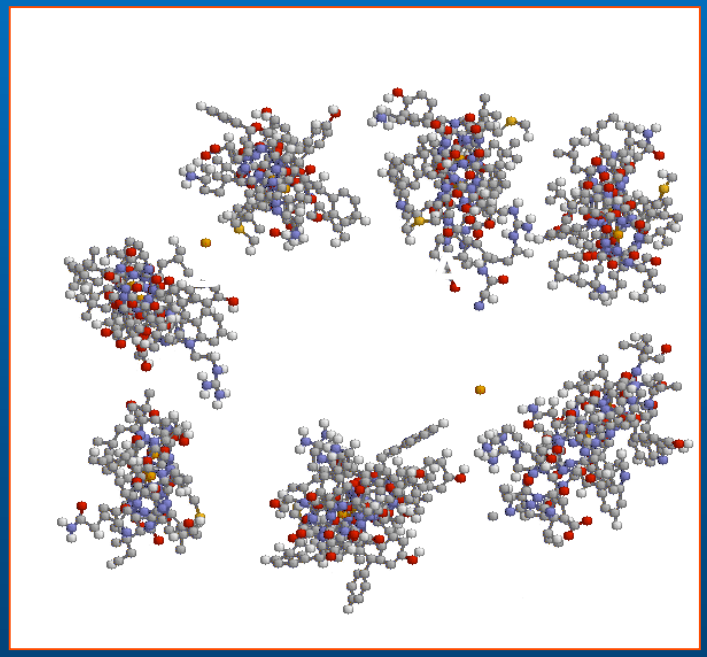
Application of Wang-Landau sampling to Docking of a seven-helix bundle: Bacteriorhodopsin (1QHJ)

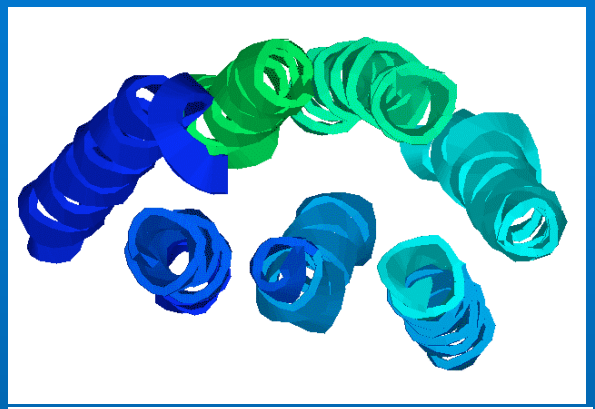
Docking of Bacteriorhodopsin

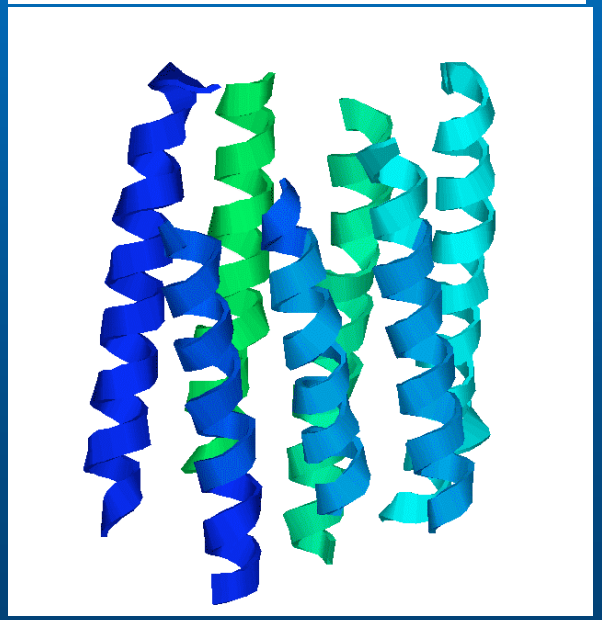
(Zhong Chen, IOB, UGA)

7 helices, 174 residues, 1619 atoms

- Rigid side-chains
- VDW + lipid-helix potential
- One month CPU time at f=2.781







A GEM structure with rmsd=3.0 Å was obtained in the self-assembly simulation of a 7-helix bundle

Overview and Conclusion

There are models with complex energy landscapes that are now amenable to study. Examples include:

- Spin glasses
- “Lattice proteins”
- Real proteins

Appendix

For more examples of the application of MC in Statistical Physics

(More coming in the 3rd edition)