FSU National Magnet Lab Winter School Jan. 12, 2012

Exploring Complex Free Energy Landscapes with Wang-Landau Sampling

D. P. Landau
Center for Simulational Physics
U. of Georgia, U.S.A.

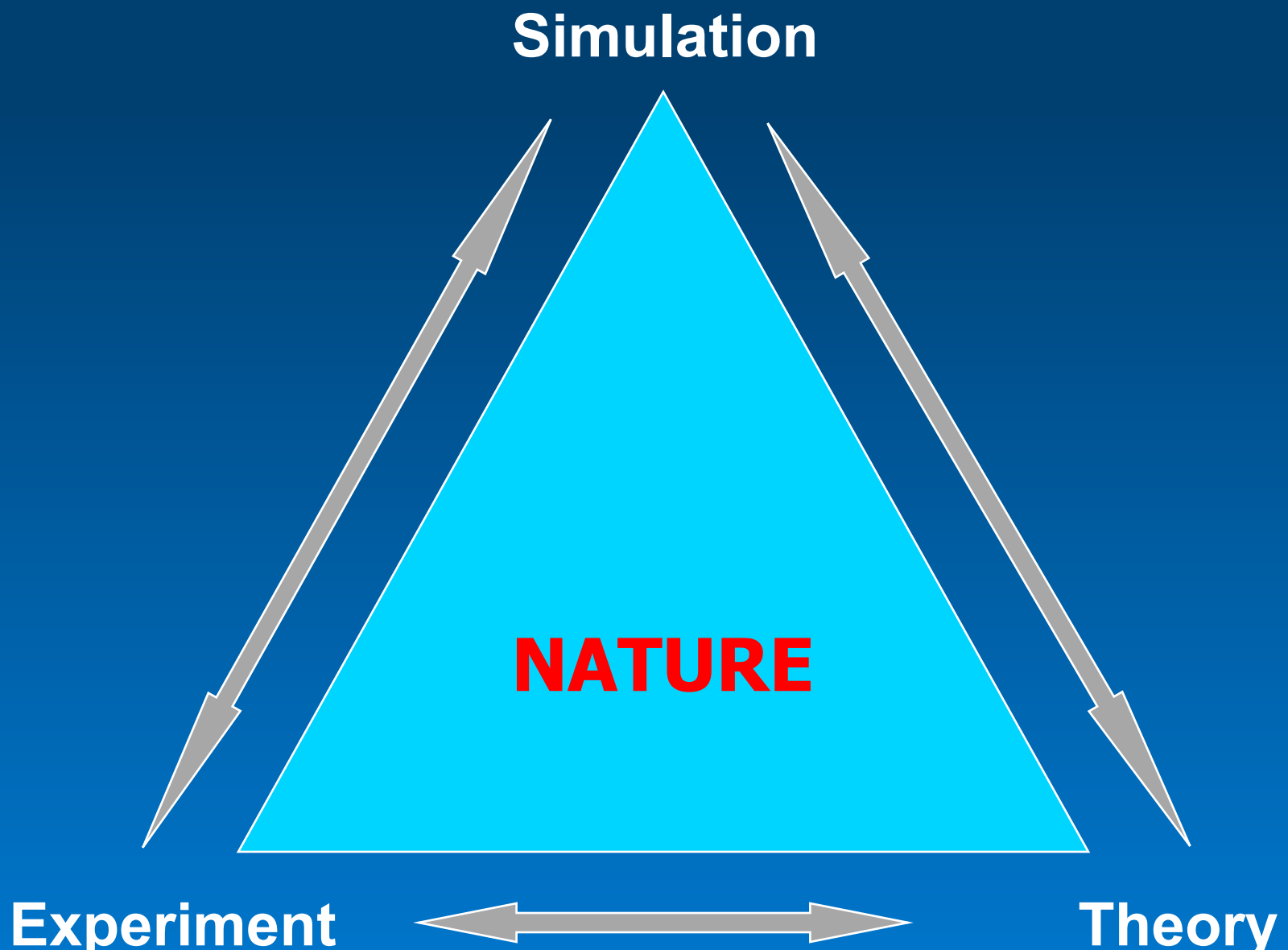
- **Introduction**
- **Brief Review: Wang-Landau Algorithm**
- **Some characteristic problems**
  - *Spin glasses*
  - *"Lattice proteins"*
  - *"Real" proteins*
- **Conclusions/Outlook**

Background and motivation

Many systems in nature have complex or “rough” free energy landscapes in which there are many maxima and minima that may have widely spaced values of relevant thermodynamic parameters.

⇨ At “low” temperature, transitions between minima become **very** infrequent

Almost all models for such systems are impossible to treat analytically!


Simulation
NATURE
Experiment
Theory

Reminder from Statistical Mechanics

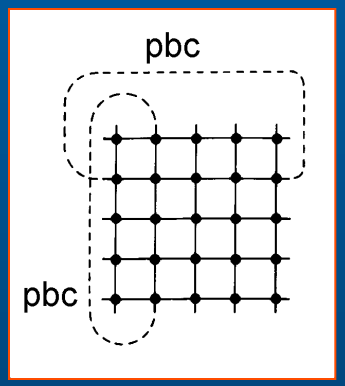
The ***Partition function*** contains all thermodynamic information:

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

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

The "fruit fly" of statistical physics: The Ising model

$$\mathcal{H} = -J \sum_{<i,j>} \sigma_i \sigma_j \,, \quad \sigma_i = \pm 1$$



For a system of N spins have 2^N states!

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*

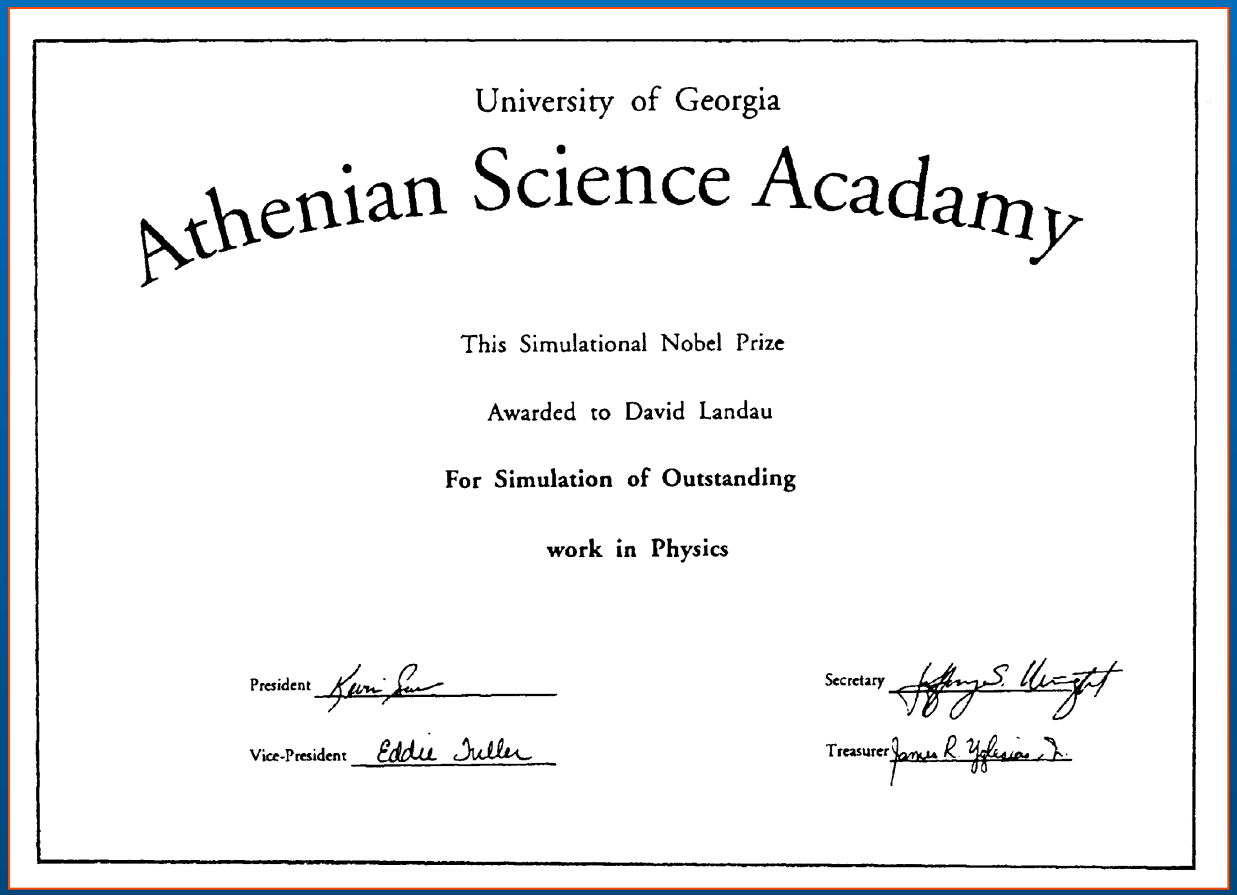
Center for Stimulational Physics

Center for Stimulational Physics

Center for Simulated Physics

Center for Stimulational Physics

Center for Simulated Physics



University of Georgia

Athenian Science Acadamy

This Simulational Nobel Prize

Awarded to David Landau

For Simulation of Outstanding

work in Physics

President Kurt Binder

Vice-President Eddie Muller

Secretary Jeffrey S. Wright

Treasurer James R. Glosser, Jr.

The "Random Walk in Energy Space with a Flat Histogram" method

or

"Wang-Landau sampling"

Wang and Landau, PRL (2001)

Wang-Landau sampling

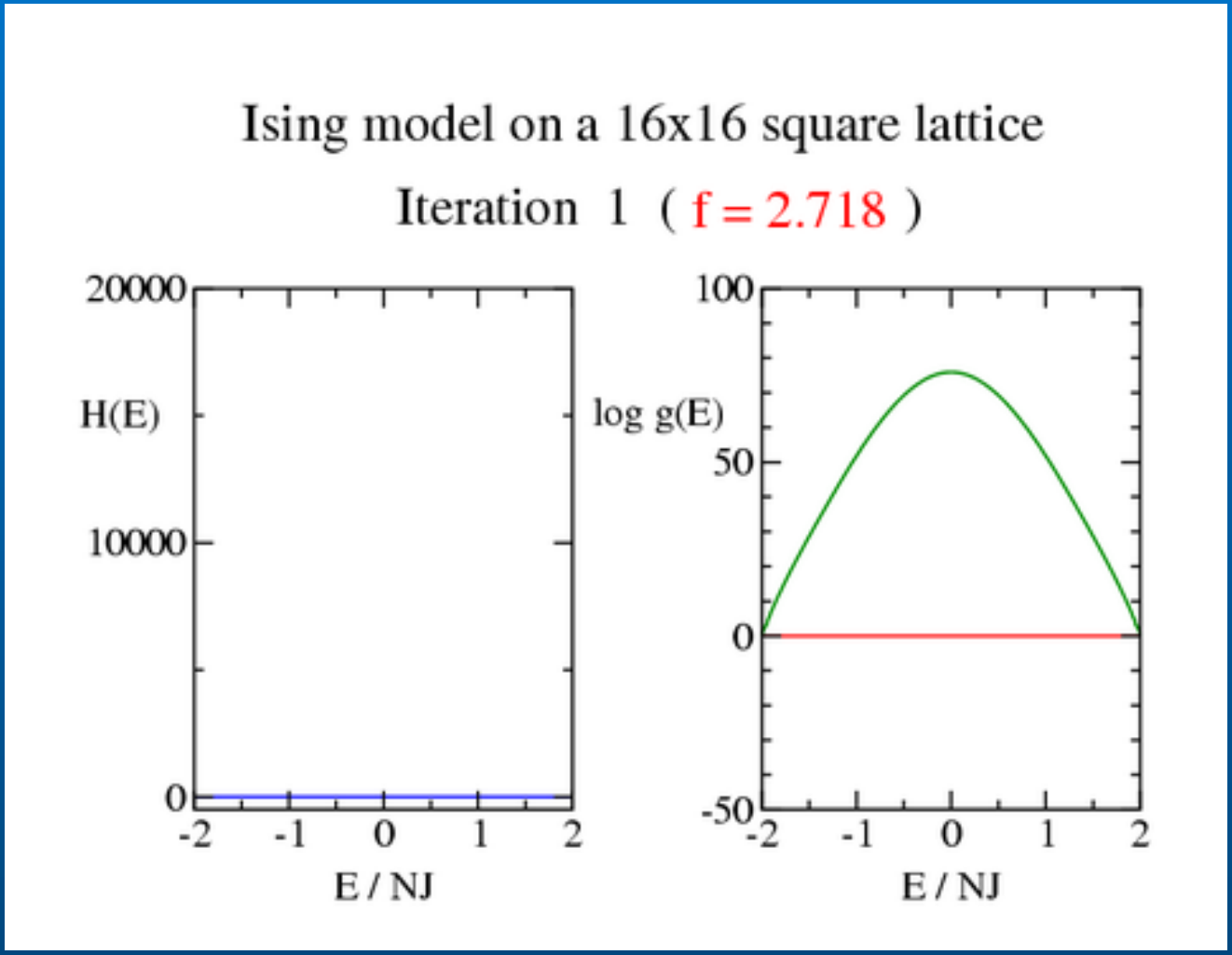
Random Walk in Energy Space with a Flat Histogram

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

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

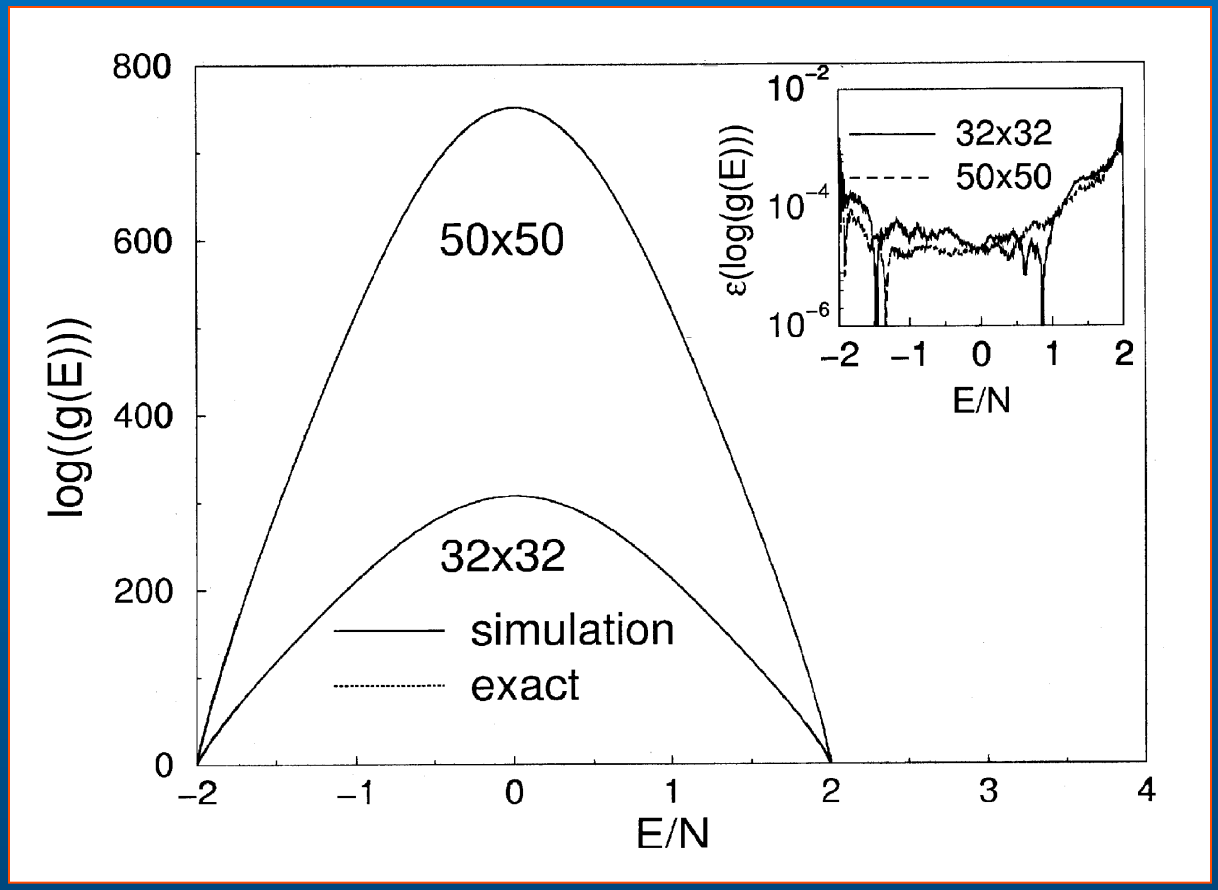
1. Set $g(E)=1$; choose a modification factor (e.g. $f_0 = e^1$)
2. Randomly flip a spin with probability: $p(E_1 \rightarrow E_2) = \min\left(\frac{g(E_1)}{g(E_2)}, 1\right)$
3. Set $g(E_i) \rightarrow g(E_i) * f$, $H(E) \rightarrow H(E)+1$
4. Continue until the histogram is "flat" decrease f, e.g. $f_{i+1} = f^{1/2}$
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



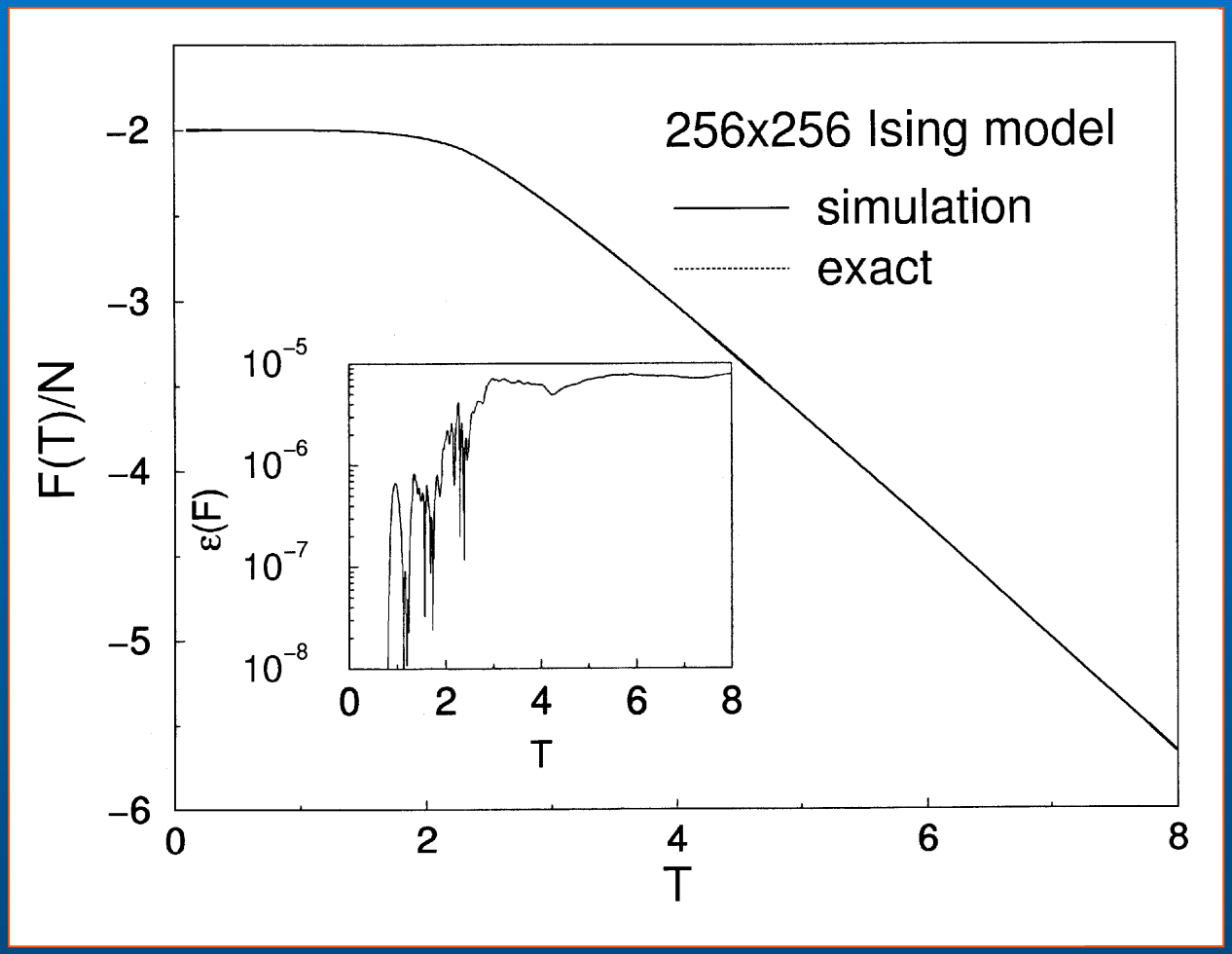
Density of States for the 2-dim Ising model

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

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:

$$\mathcal{H} = -\sum_{<i,j>} J_{ij}\sigma_i\sigma_j,\ \sigma_i = \pm 1 \quad and\ J_{ij} = \pm J.$$

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

A Magnetic System with Complex "Order"

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

$$\mathcal{H} = -\sum_{<i,j>} J_{ij}\sigma_i\sigma_j,\ \sigma_i = \pm 1 \quad and\ J_{ij} = \pm J.$$

At T_c (if it exists) a spin glass state forms ⇒ get a "rough" energy landscape with many local minima are separated by barriers

This is just an Ising model with random interactions

A Magnetic System with Complex "Order"

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

$$\mathcal{H} = -\sum_{<i,j>} J_{ij}\sigma_i\sigma_j, \; \sigma_i = \pm 1 \quad and \; J_{ij} = \pm J.$$

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

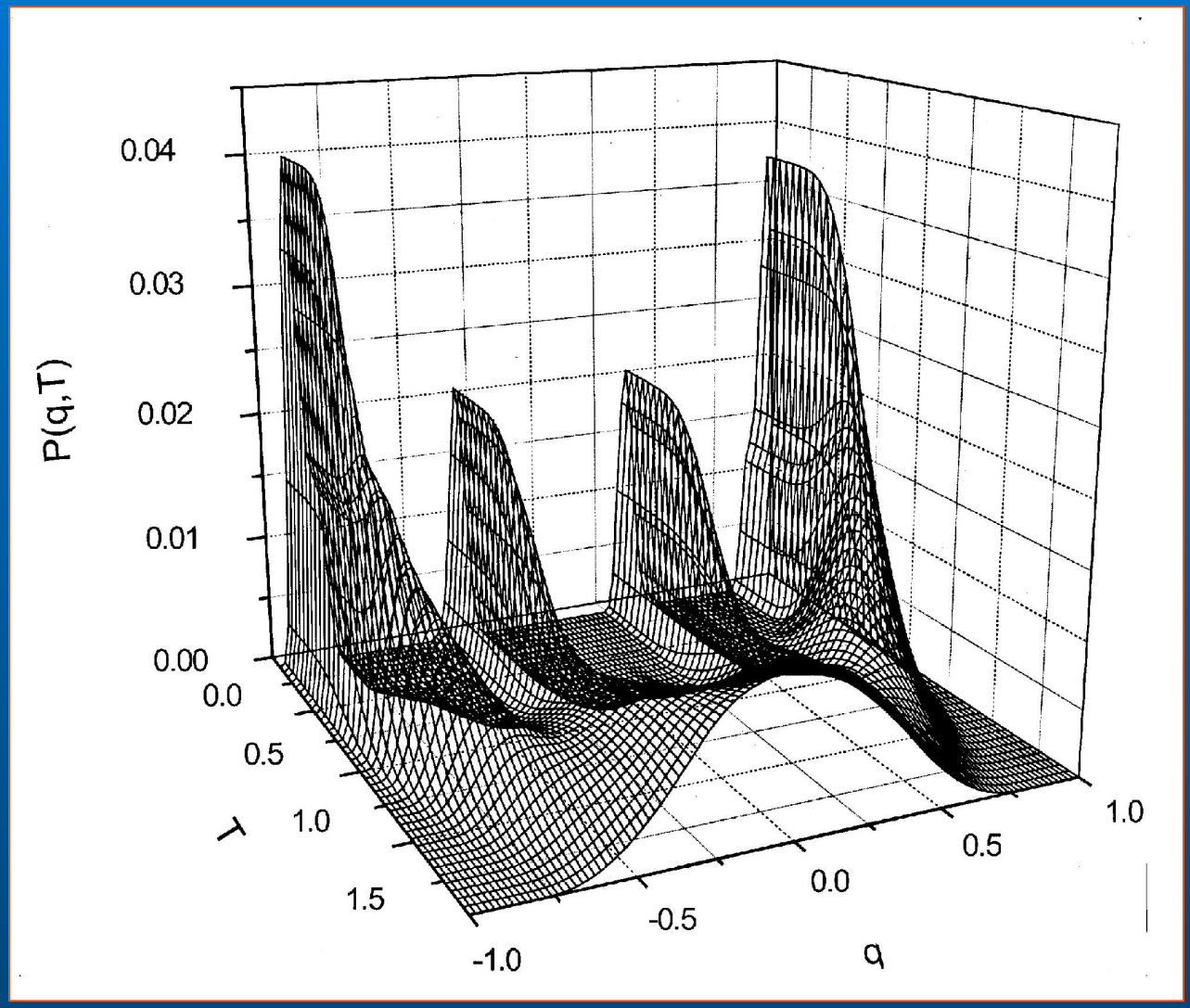
Define an Order Parameter 'q'

- Overlap of a configuration with a groundstate . . . but must also take the configurational average over different bond distributions

Extend random walk ⇒ two-dimensional parameter space

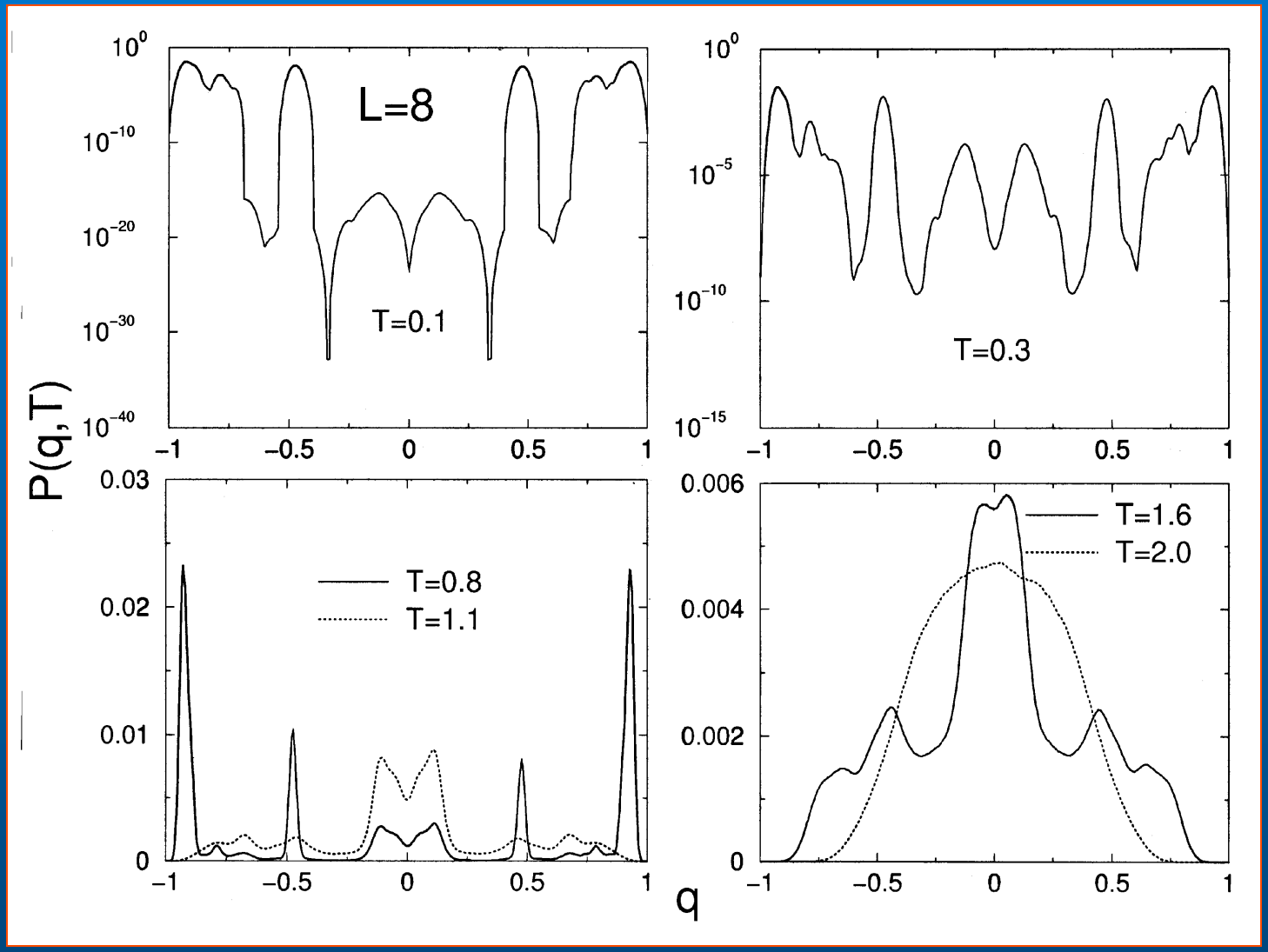
Distribution of States: LxLxL EA Spin Glass

L=6



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

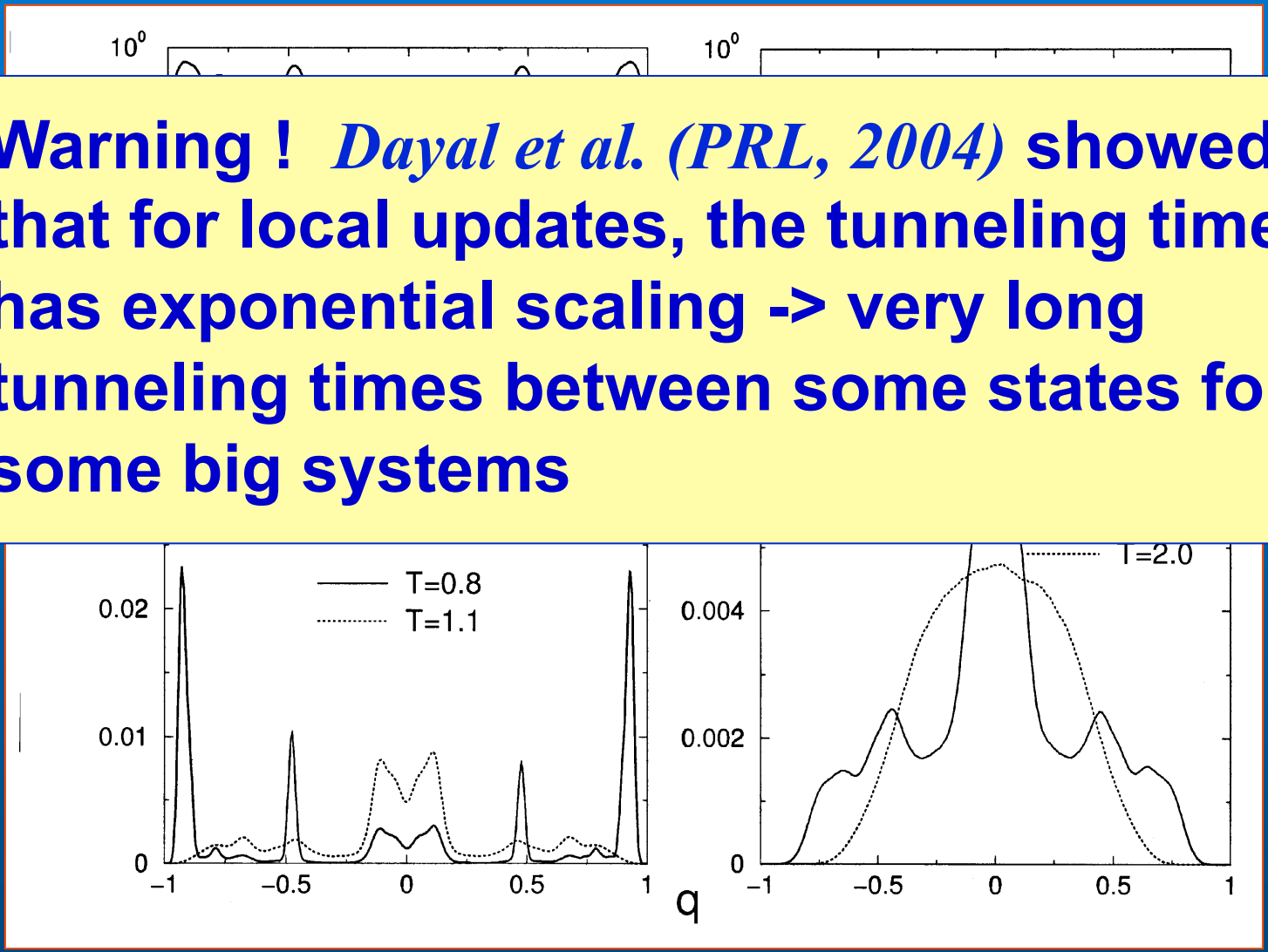
Distribution of States: LxLxL EA Spin Glass



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

Distribution of States: LxLxL EA Spin Glass

Warning ! ***Dayal et al. (PRL, 2004)*** **showed that for local updates, the tunneling time has exponential scaling -> very long tunneling times between some states for some big systems**



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

Groundstate Properties of the 3-d EA Spin Glass

Entopy 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)*

Applications to “Complex” Systems

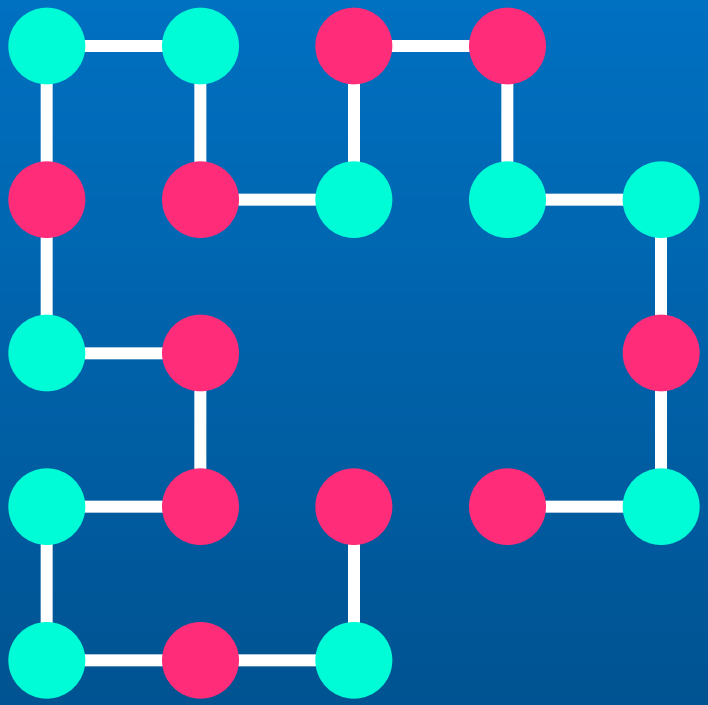
- *Spin glasses*
- *“Lattice proteins”*
- *“Real” proteins*

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"

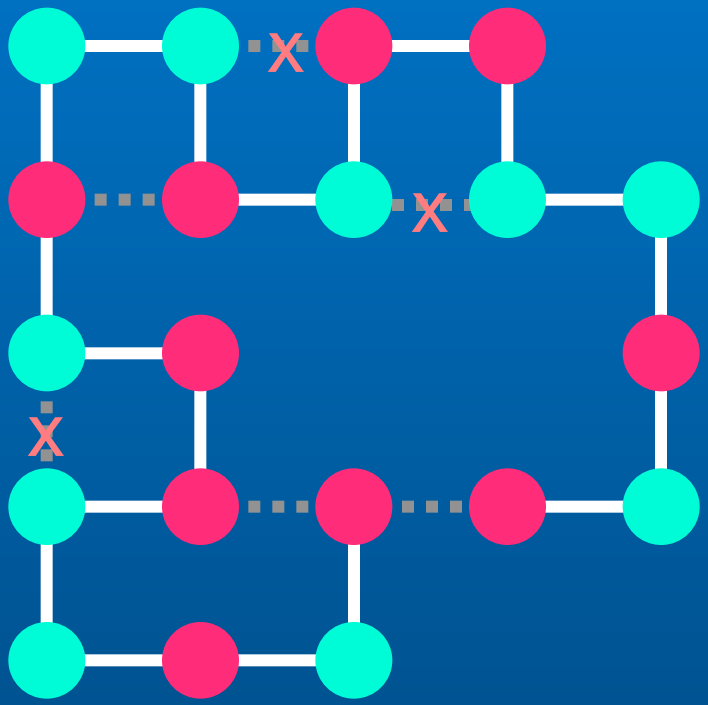
- Hydrophobic (H)
- Polar (P)

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)

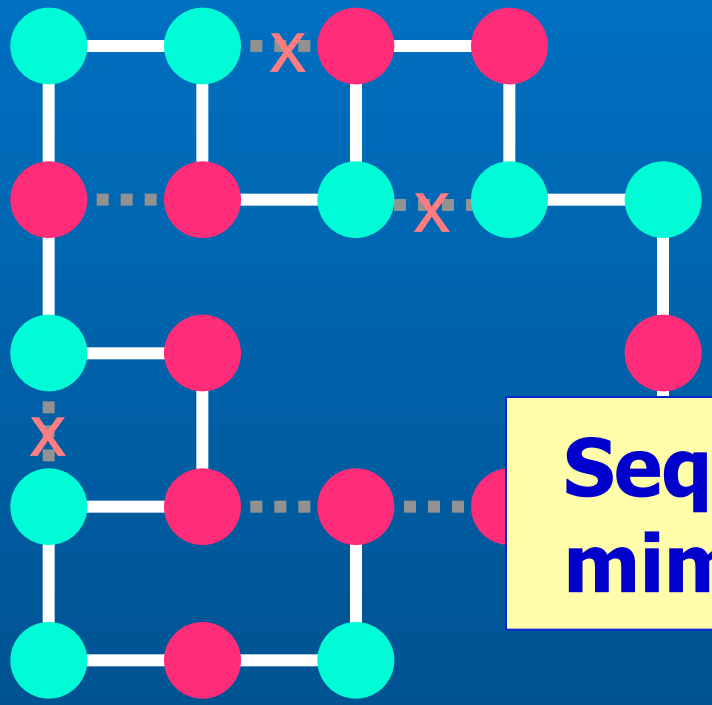
Nearest-neighbor interactions between non-covalently bound neighbors

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

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

A "Biologically inspired" problem

The HP model of protein folding



Amino acid = "bead"

- Hydrophobic (H)
- Polar (P)

Protein sequence = "HPHPPHHPHPP…"

Sequences are chosen to mimic real proteins

Nearest-neighbor interactions
between non-covalently bound neighbors

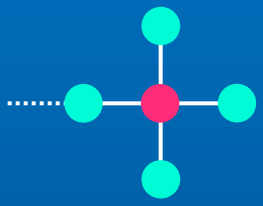
$E_{HH} = -1, E_{HP} = 0, E_{PP} = 0 \Rightarrow$ Compact hydrophobic core / polar (hydrophilic) shell

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

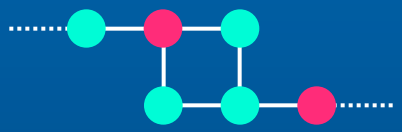
The importance of move sets

Local moves:

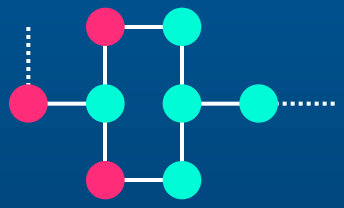
End flip (1 bond)



Kink flip (2 bonds)



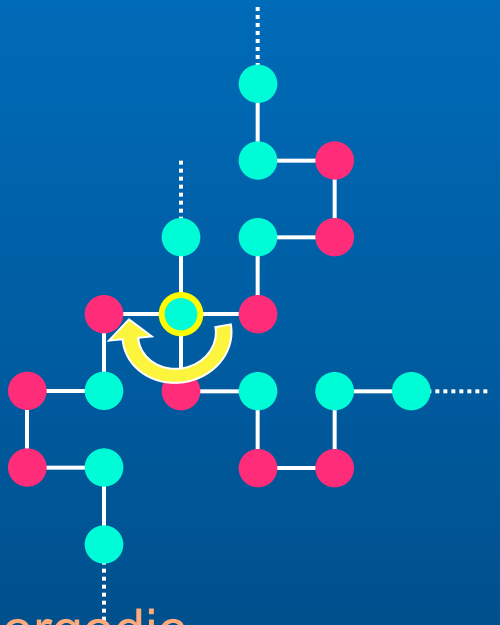
Crankshaft (3 bonds)



⇒ non-ergodic

Non-local moves:

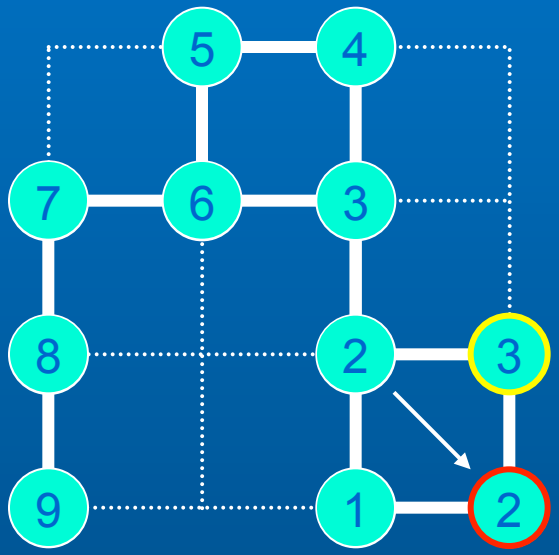
Pivot move



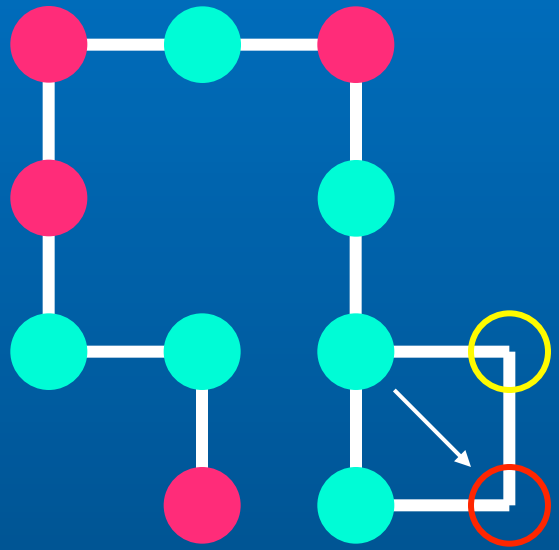
⇒ ergodic

… but inefficient for dense conformations ⇒ high rejection probability

Pull moves



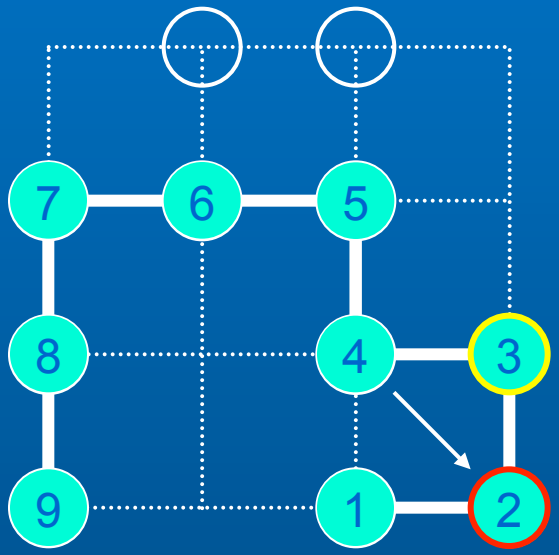
multi-bead move
(Completes internally)



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

(Lesh et al., 2003)

Pull moves



Extensible to any dimension

Ergodic (complete)

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

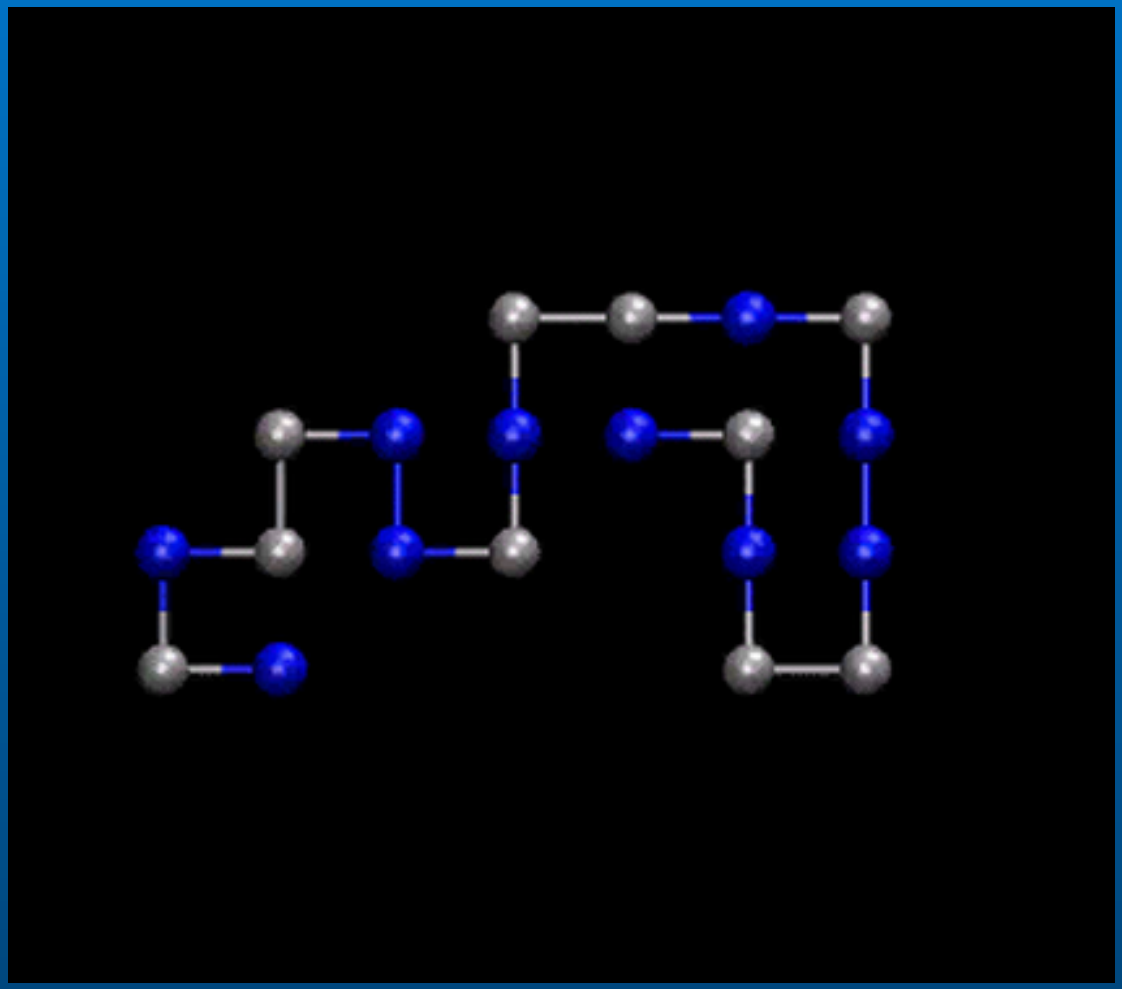
No time-consuming self-avoidance test required

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

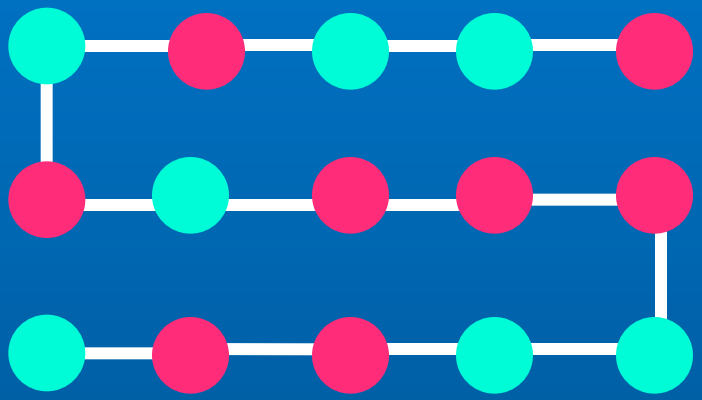
$\Rightarrow$ Ideal for Wang-Landau sampling

Wang-Landau sampling of the HP model

- With pull moves



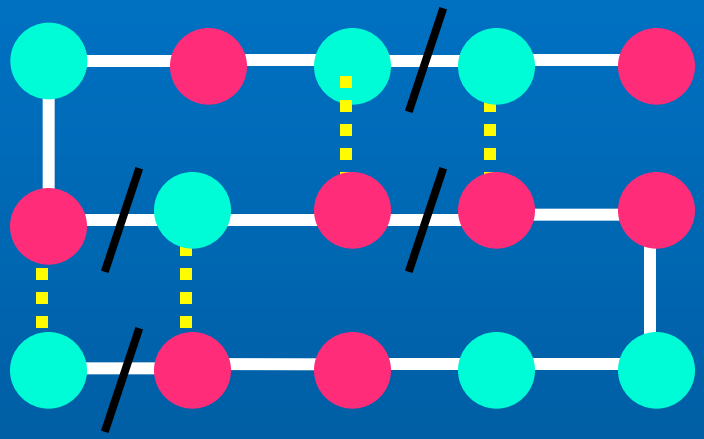
“Cut and join” moves



Initial HP configuration

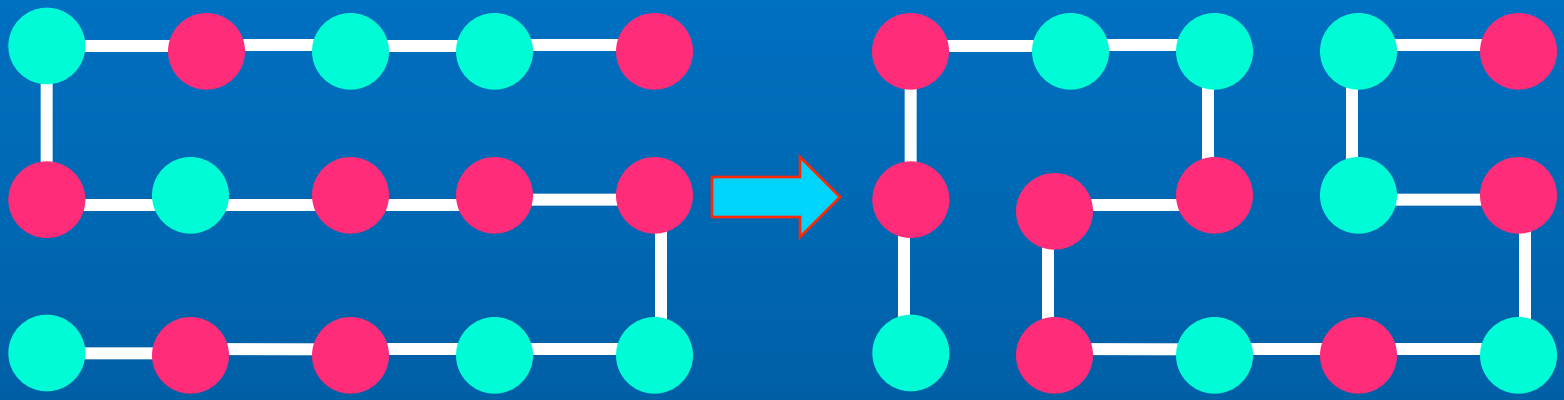
(Deutsch, J. Chem. Phys. 1997)

“Cut and join” moves



(Deutsch, J. Chem. Phys. 1997)

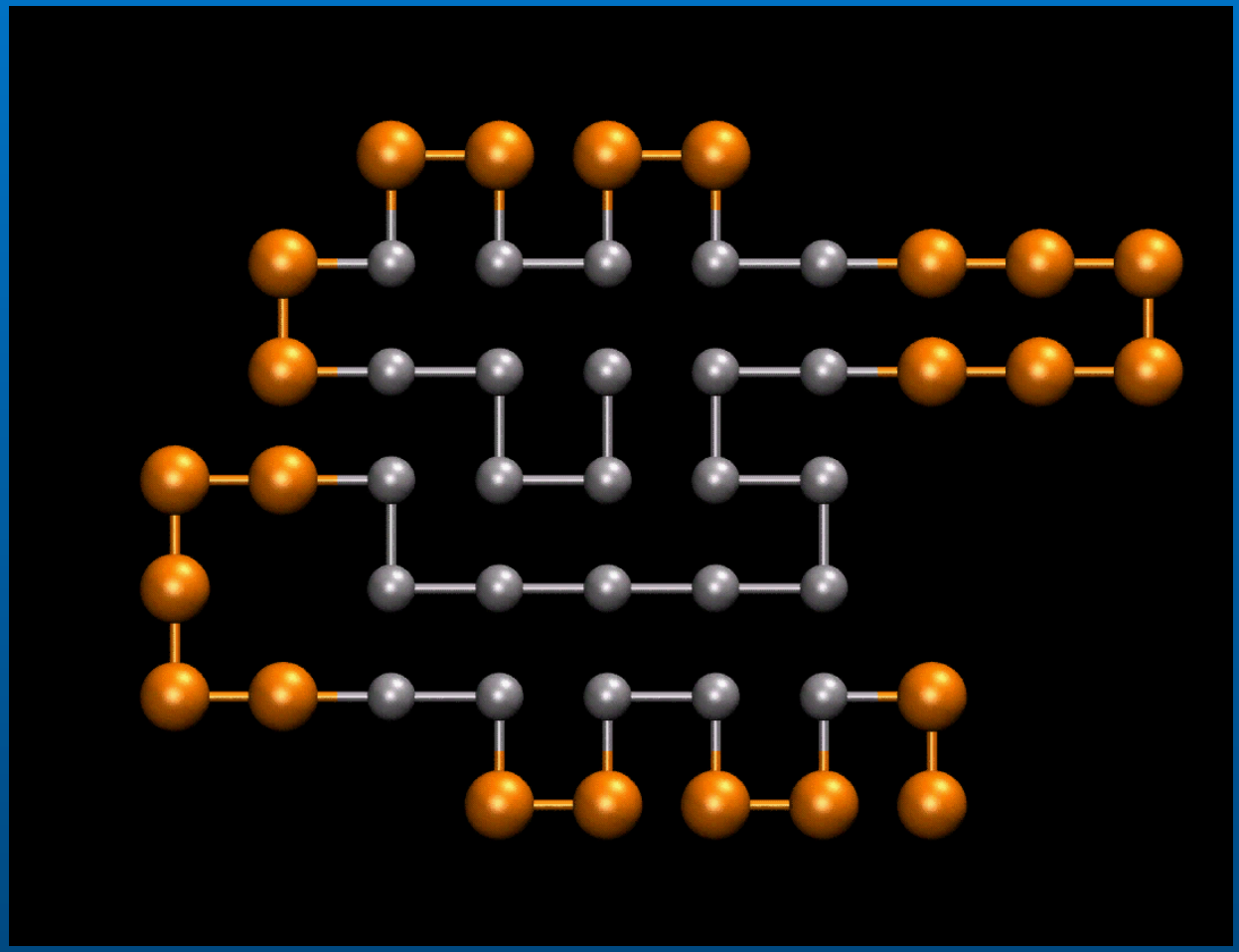
“Cut and join” moves



Note: For the HP model, the sequence of H’s and P’s **must** be maintained

(Deutsch, J. Chem. Phys. 1997)

“Cut and join”



Wang-Landau sampling of the HP Model

64mer in 2 dimensions (square lattice)

Seq2D64

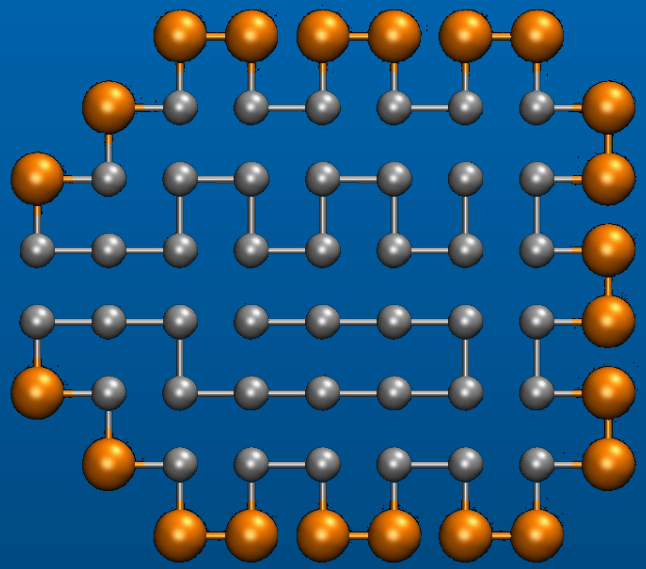
Ground state search

- 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)



Wang-Landau sampling of the HP Model

64mer in 2 dimensions (square lattice)

Seq2D64

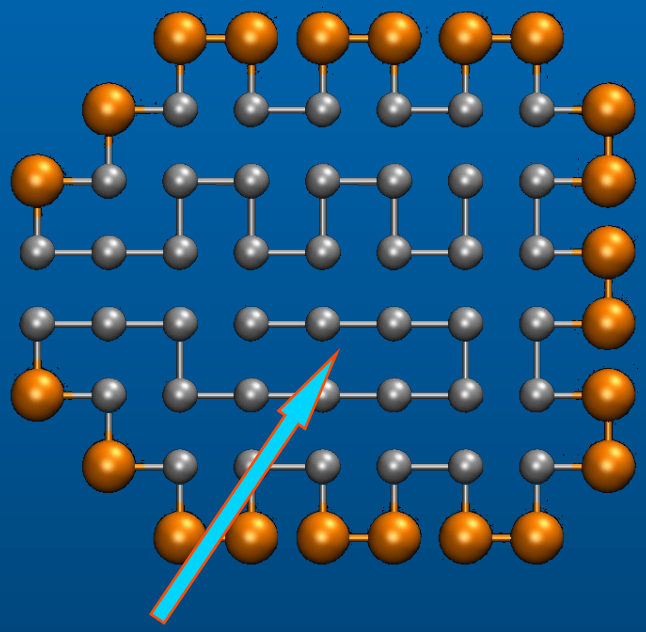
Ground state search

- 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)



Hydrophobic core

Wang-Landau sampling of the HP Model

64mer in 2 dimensions (square lattice)

Seq2D64

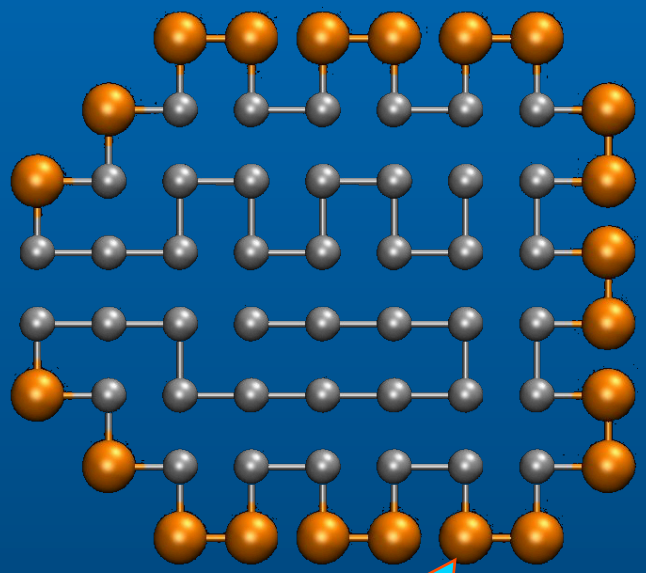
Ground state search

- 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)

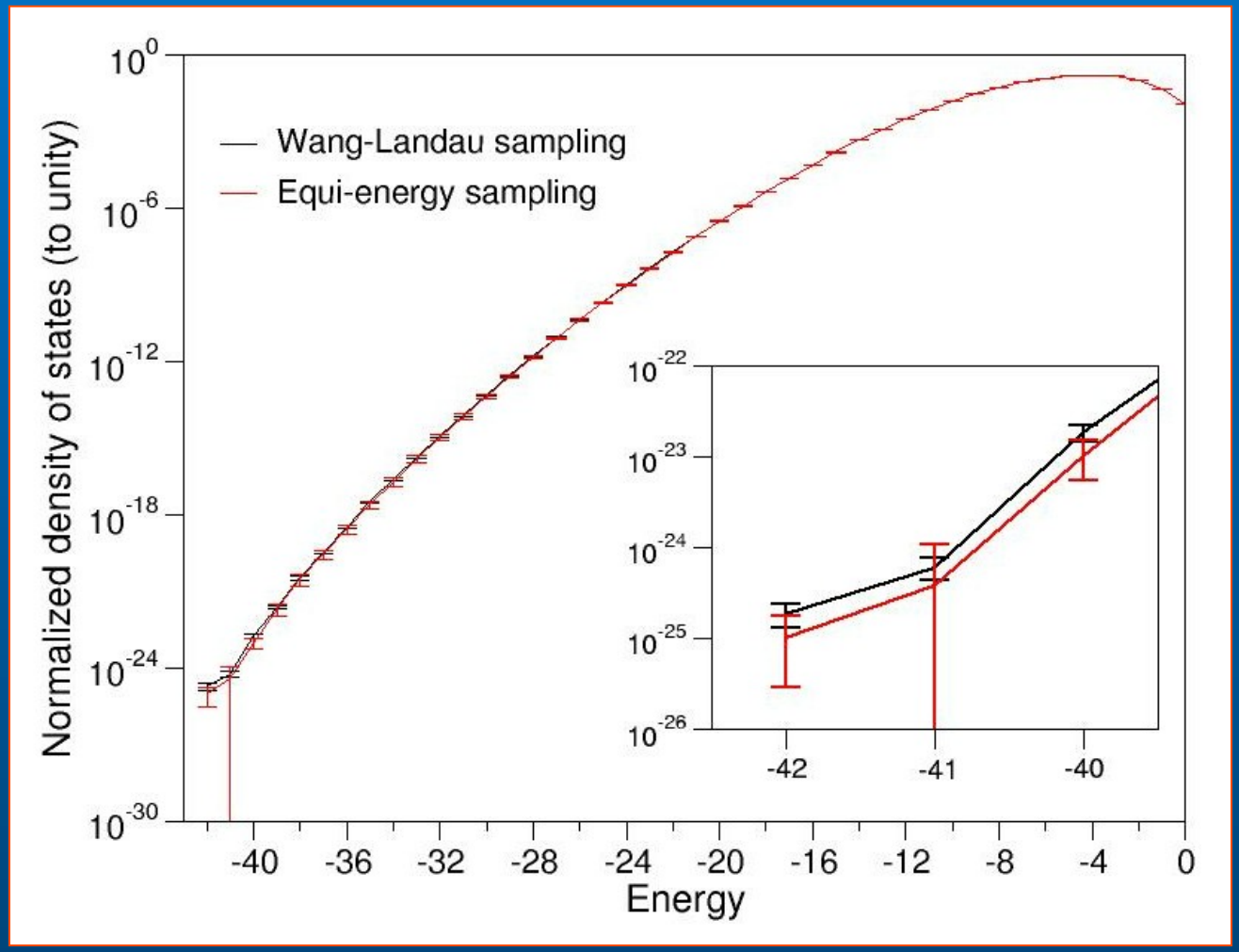


Hydrophilic surface

A "Biologically inspired" problem

The HP Model of Protein Folding

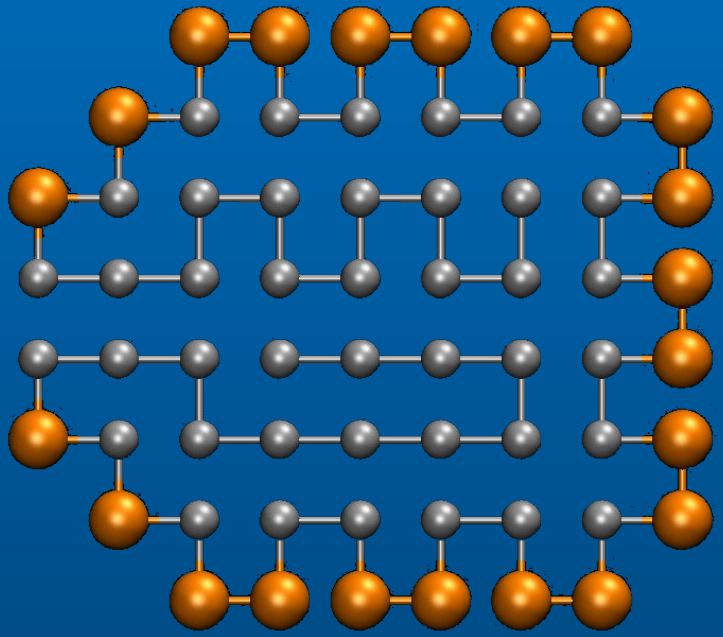
Seq2D64



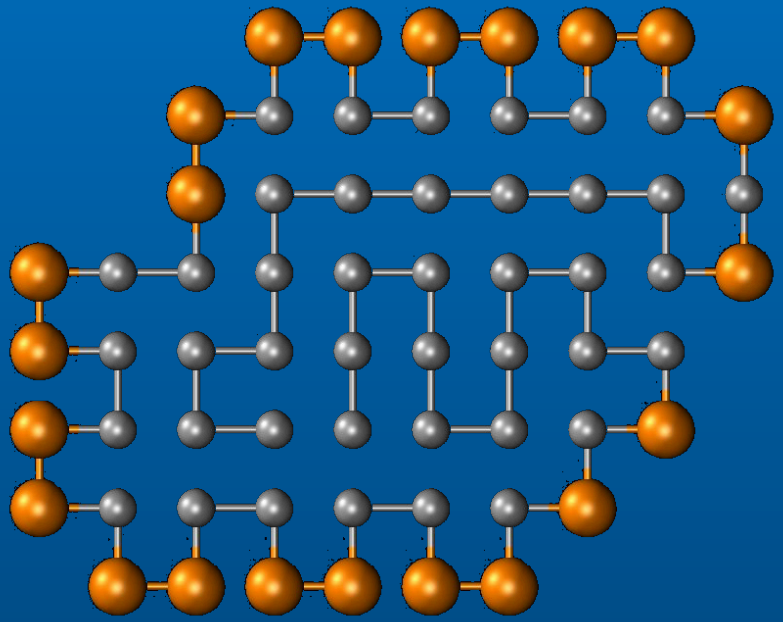
64er in 2 dimensions (square lattice)

Seq2D64

Ground state (E = -42)



1st excited state (E = -41)**



** *highly degenerate*

Wang-Landau Sampling of the HP model

103mer in 3 dimensions (simple cubic lattice)

Seq3D103

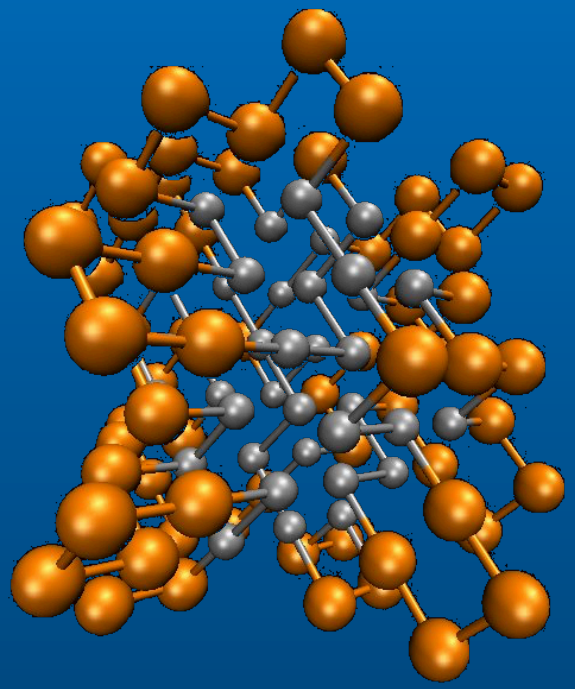
Ground state search

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

Density of states

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

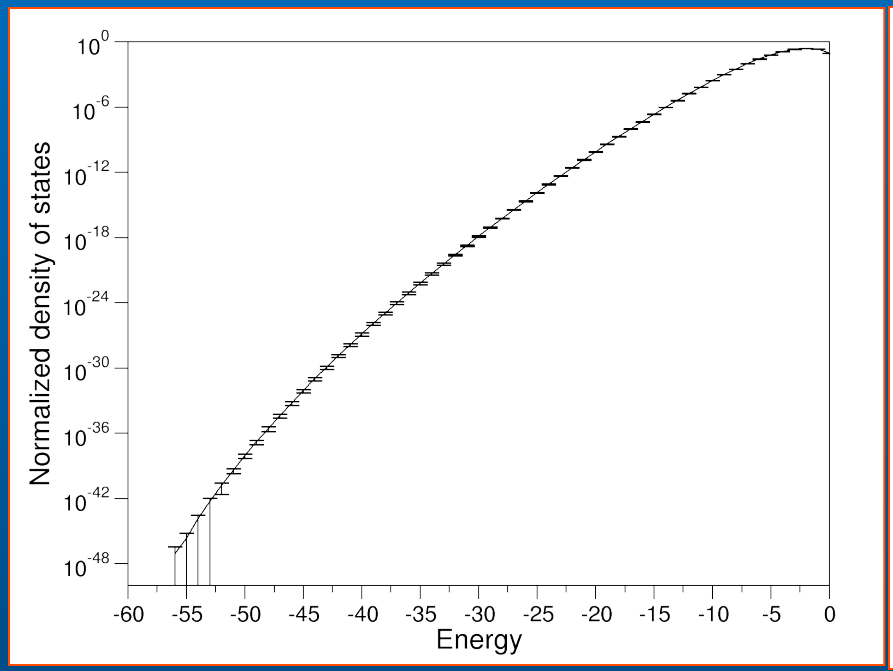
Ground state (E = -58)



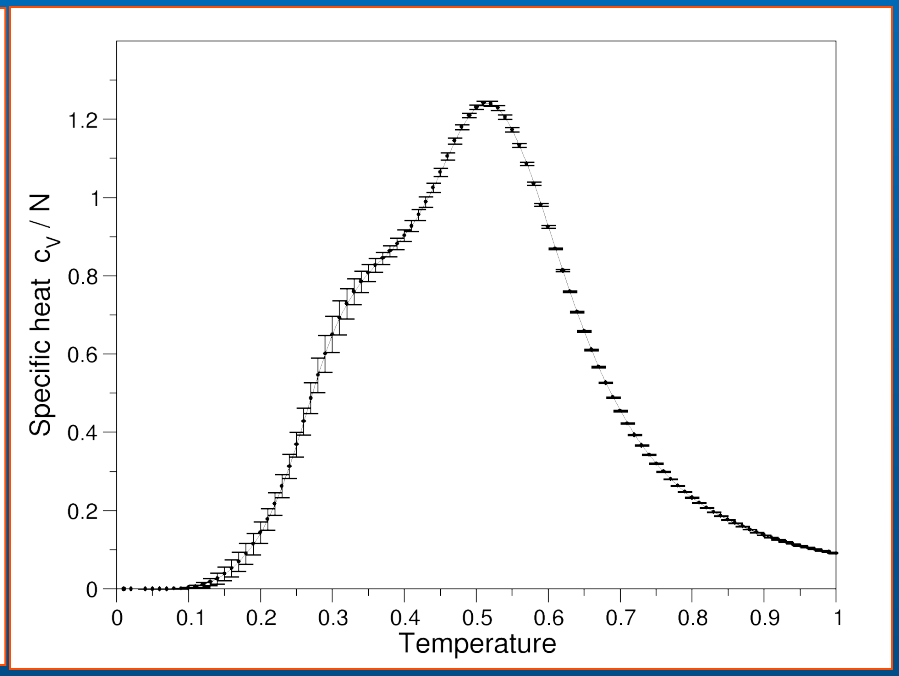
Seq3D103: Thermodynamic properties

Two-step folding process

Density of states



Specific heat

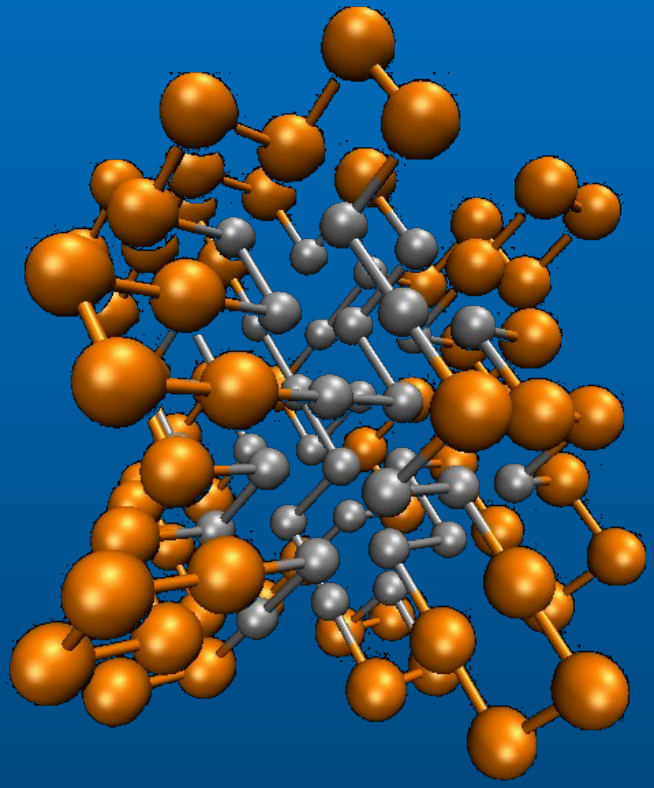


5 runs

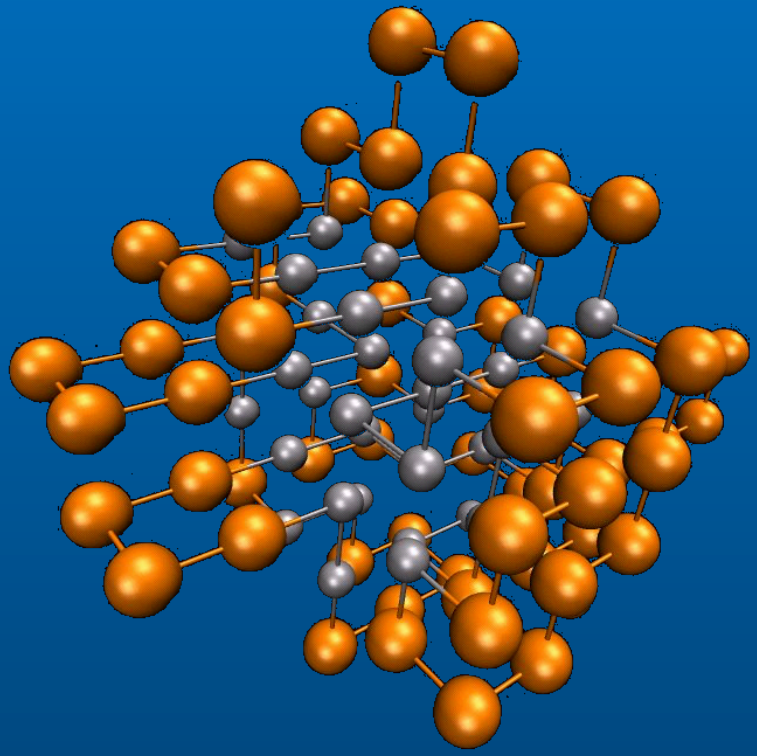
≈ 80h_{CPU} / run

103mer in 3 dimensions (cubic lattice)

Ground state (E = -58)



1st excited state (E = -57)



Now introduce an attractive surface:

The HP model of protein folding

E_{HH} = H – H bond interaction energy

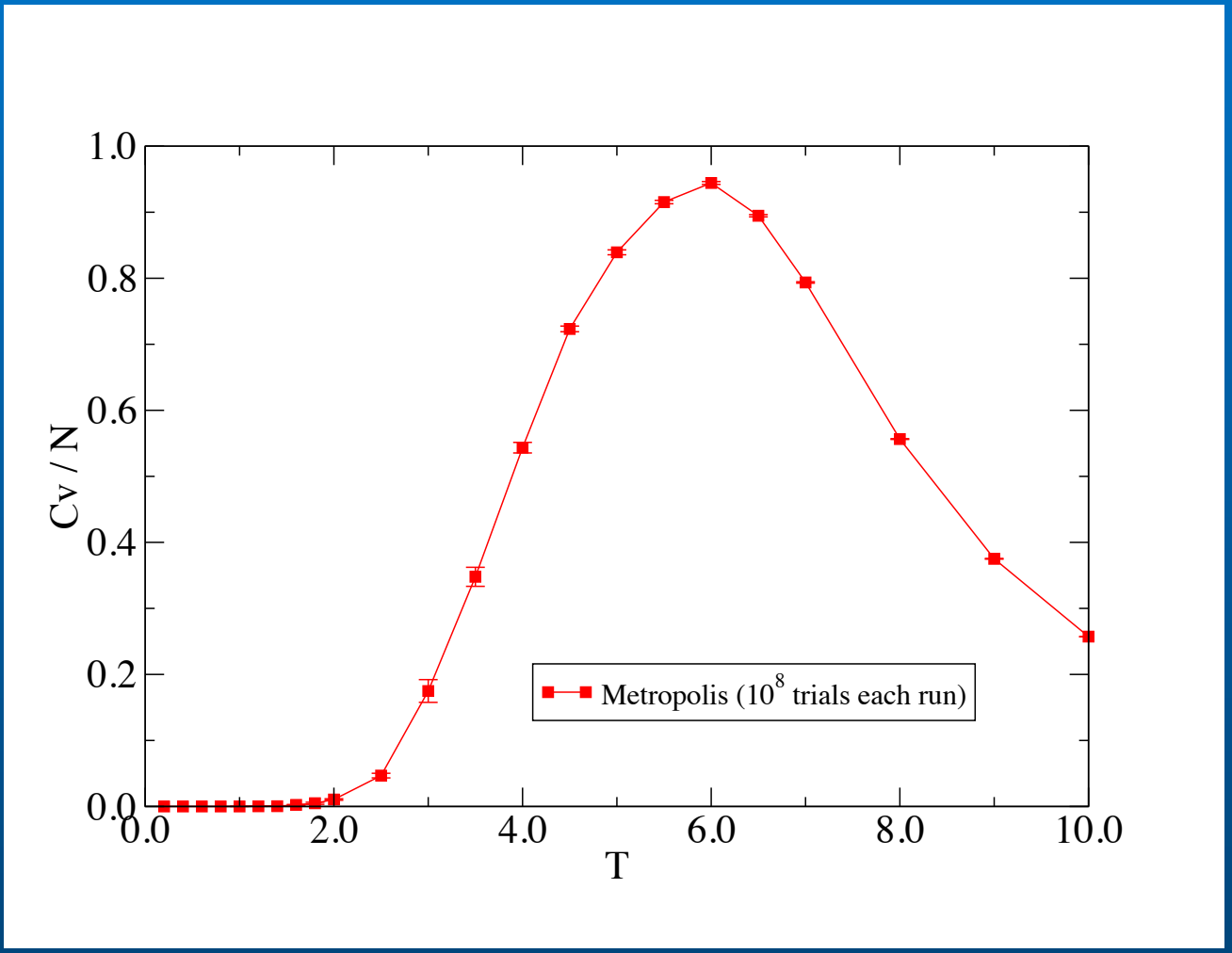
E_{SH} = H – surface interaction energy

E_{SP} = P – surface interaction energy

We have studied $E_{SH} = E_{SP}$, but other surfaces can be easily simulated!

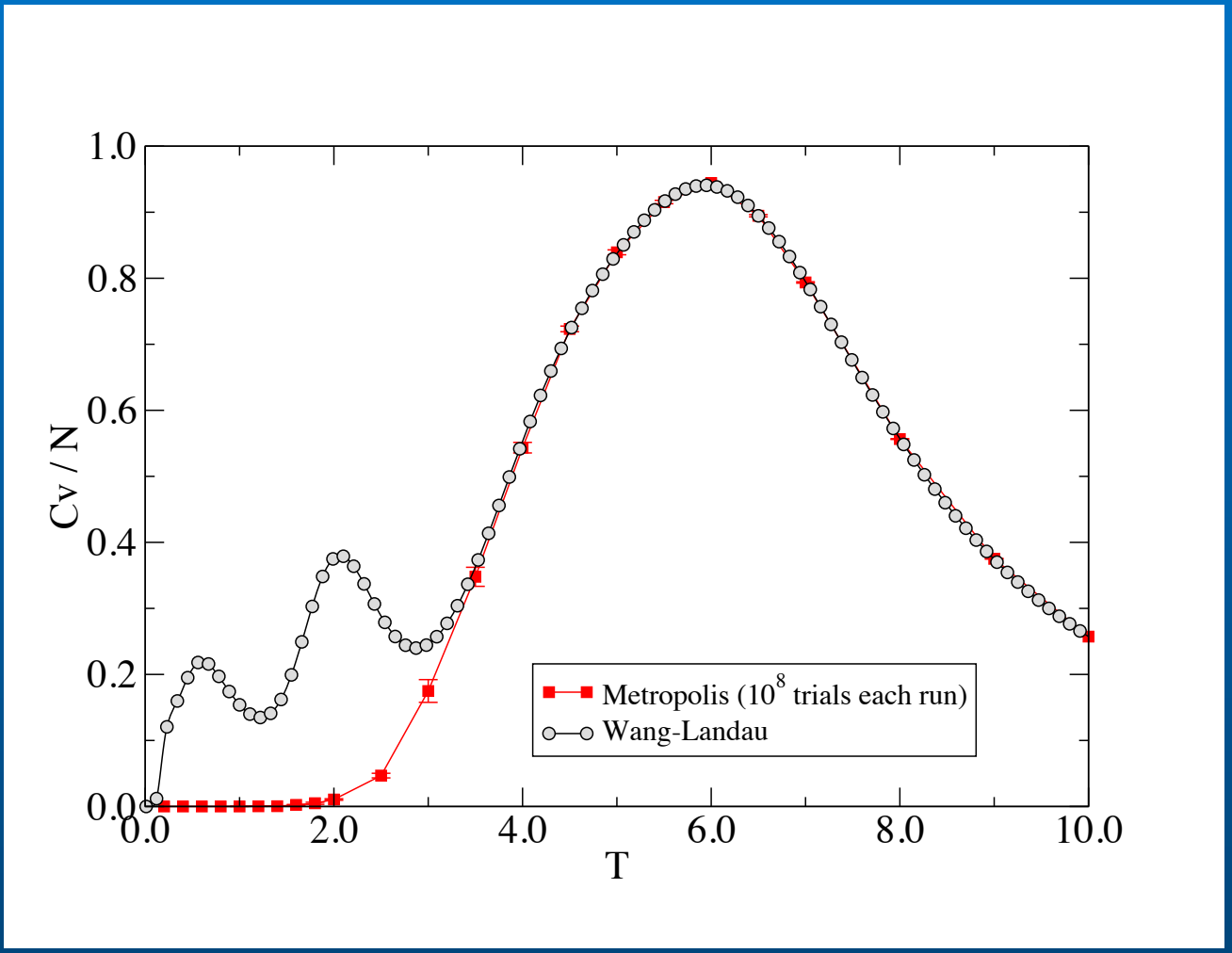
36mer in 3 dimensions (cubic lattice) with a surface

$E_{SH} = E_{SP} = E_{HH}/12$



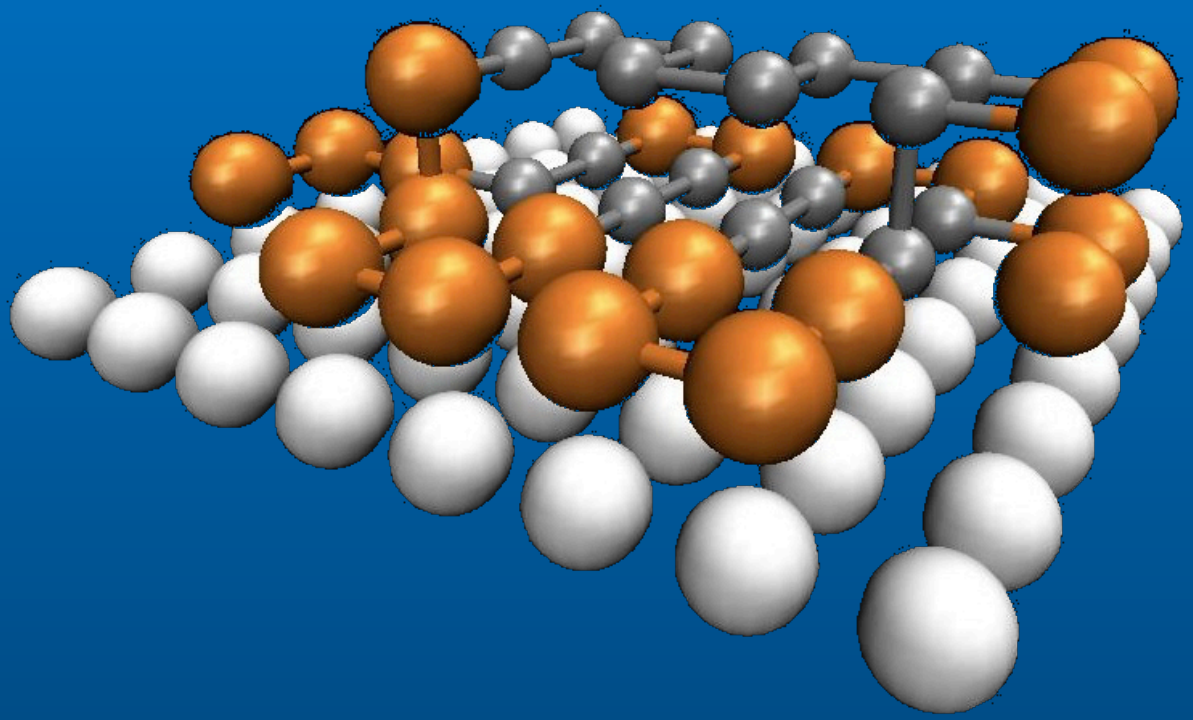
36mer in 3 dimensions (cubic lattice) with a surface

$E_{SH} = E_{SP} = E_{HH}/12$



36mer in 3 dimensions (cubic lattice) with a surface

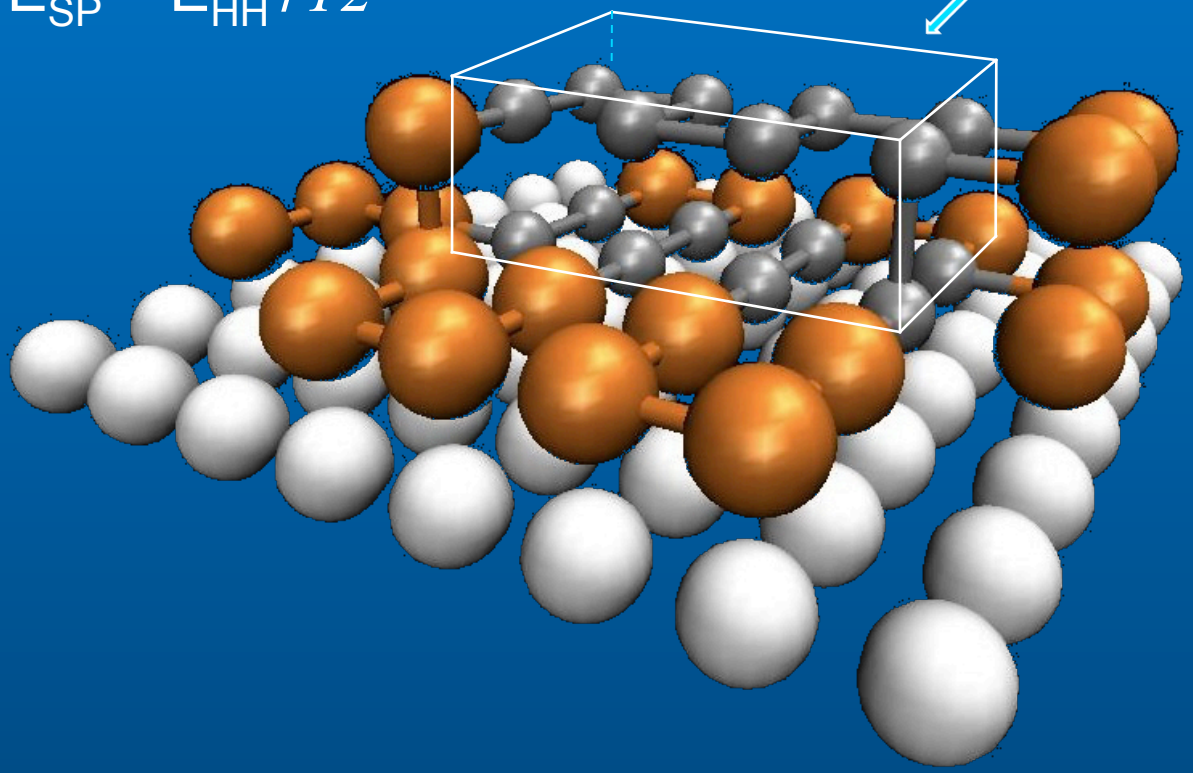
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

$E_{SH} = E_{SP} = E_{HH}/12$

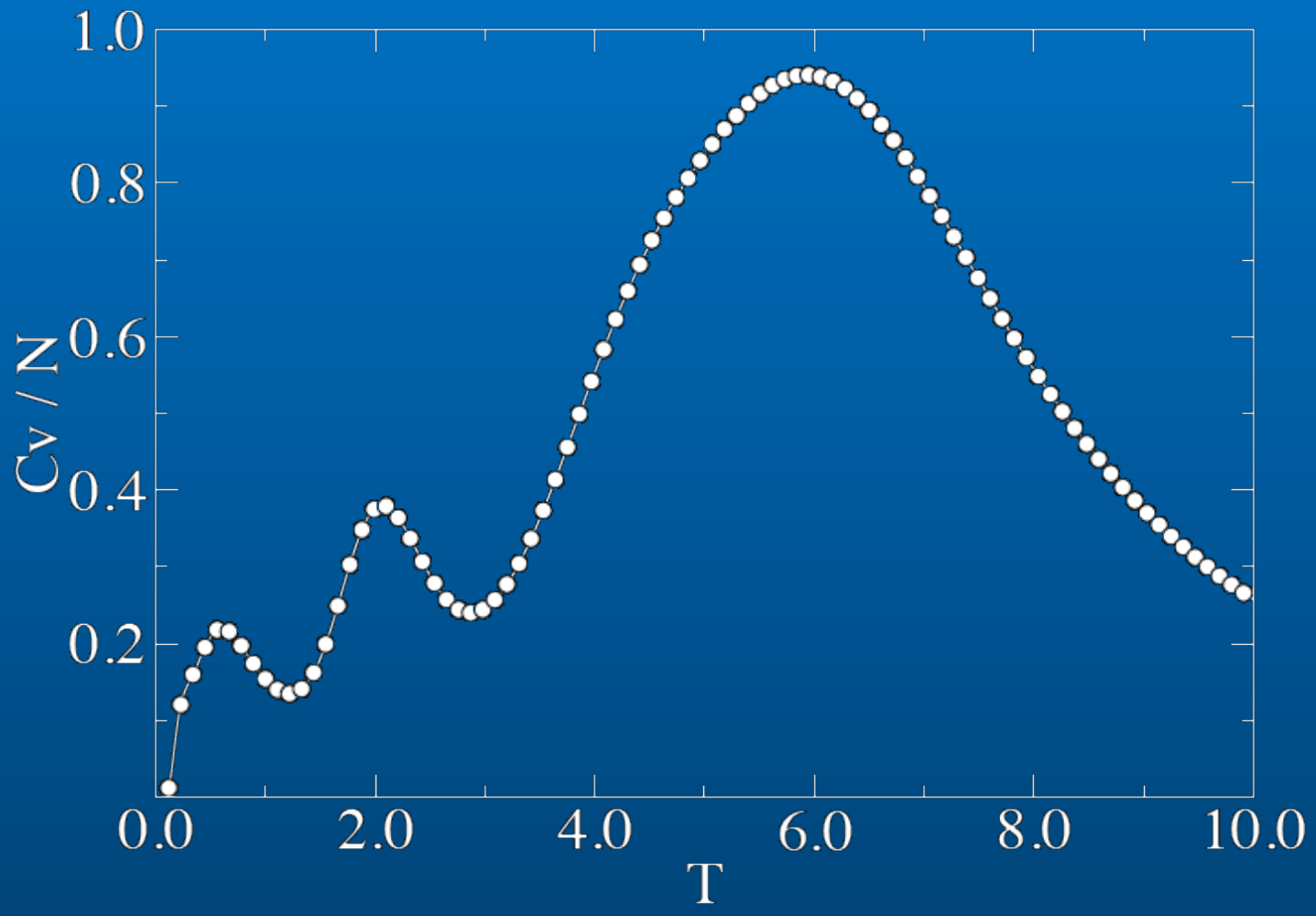
$2 \times 4 \times 2$ hydrophobic core



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

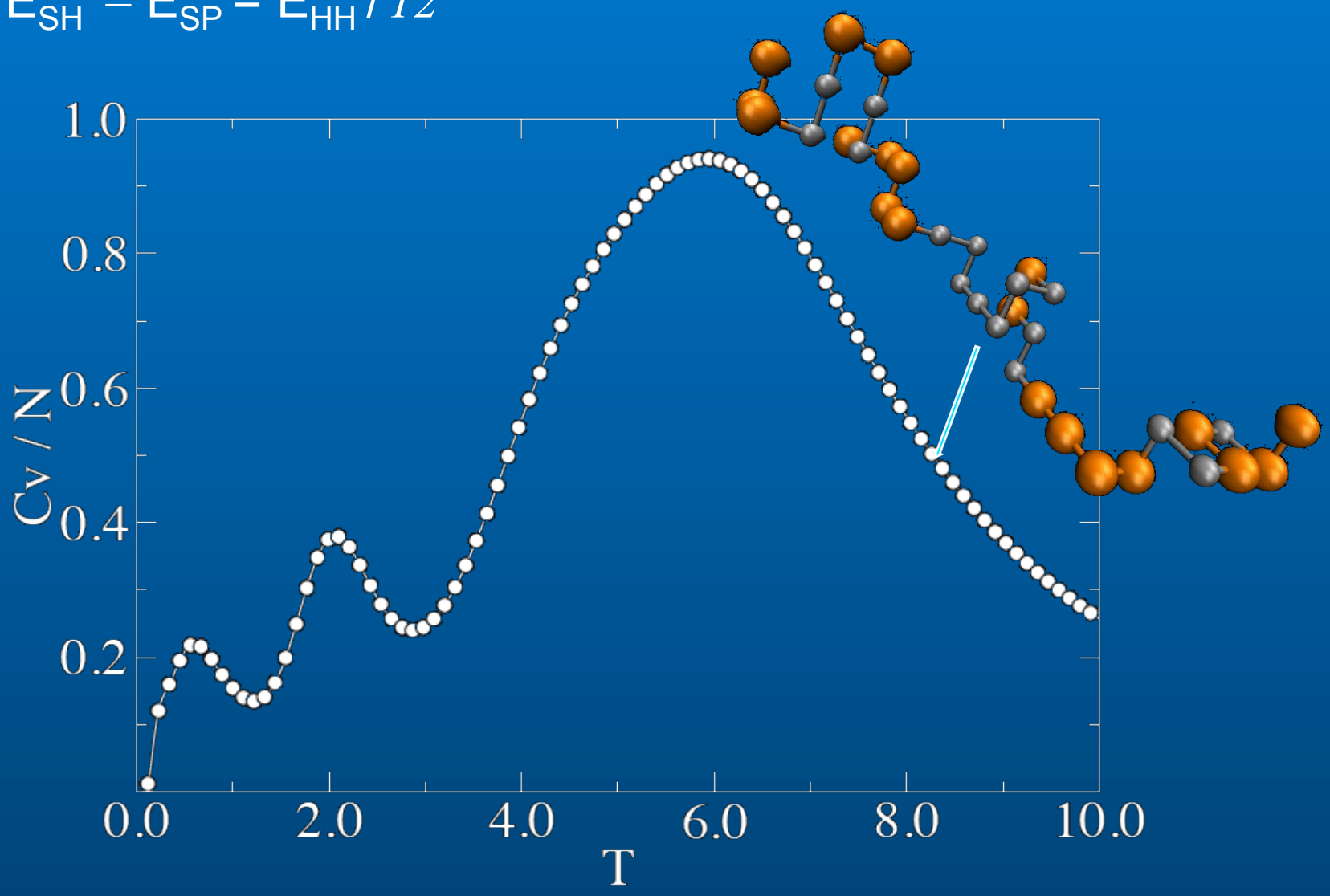
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

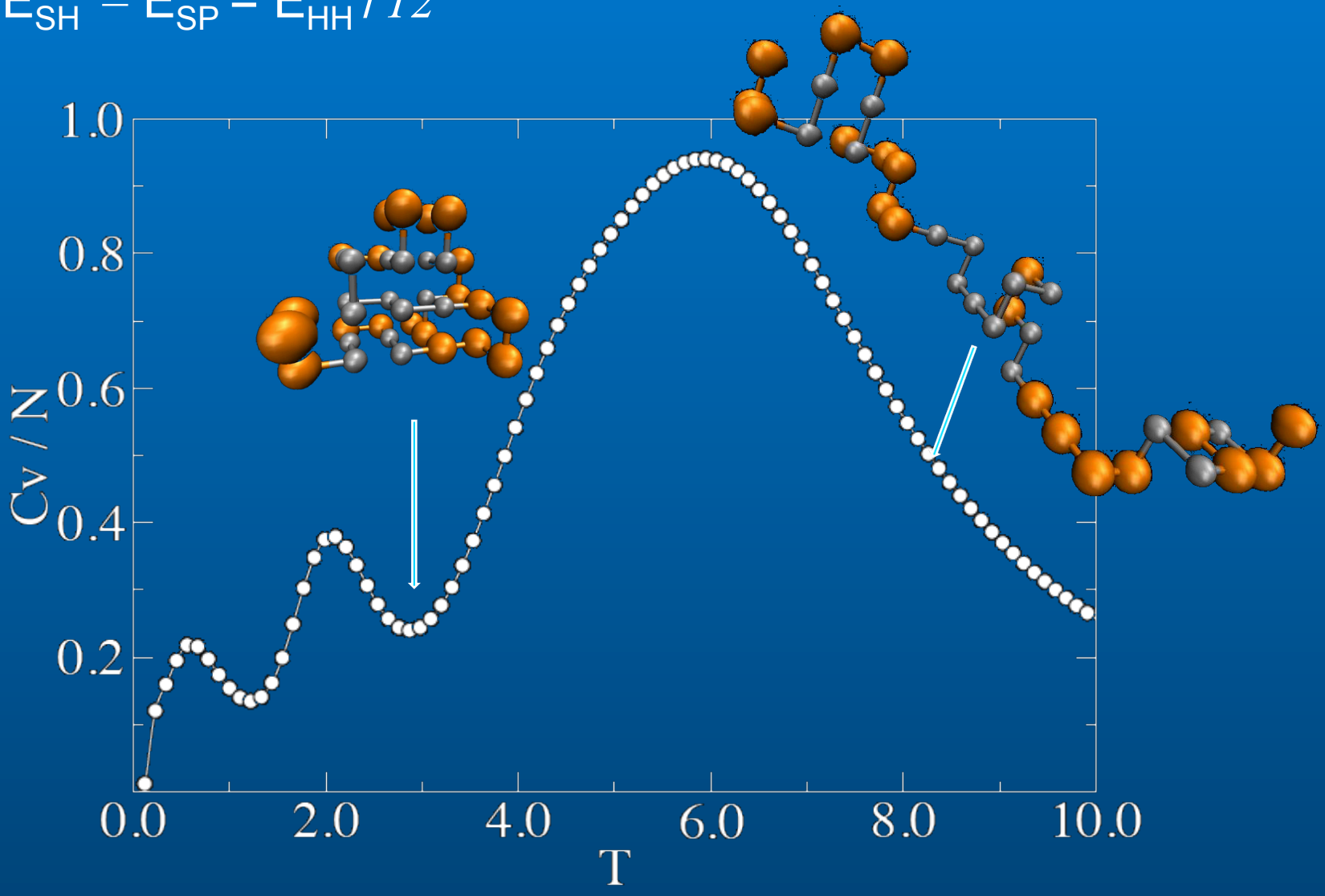
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

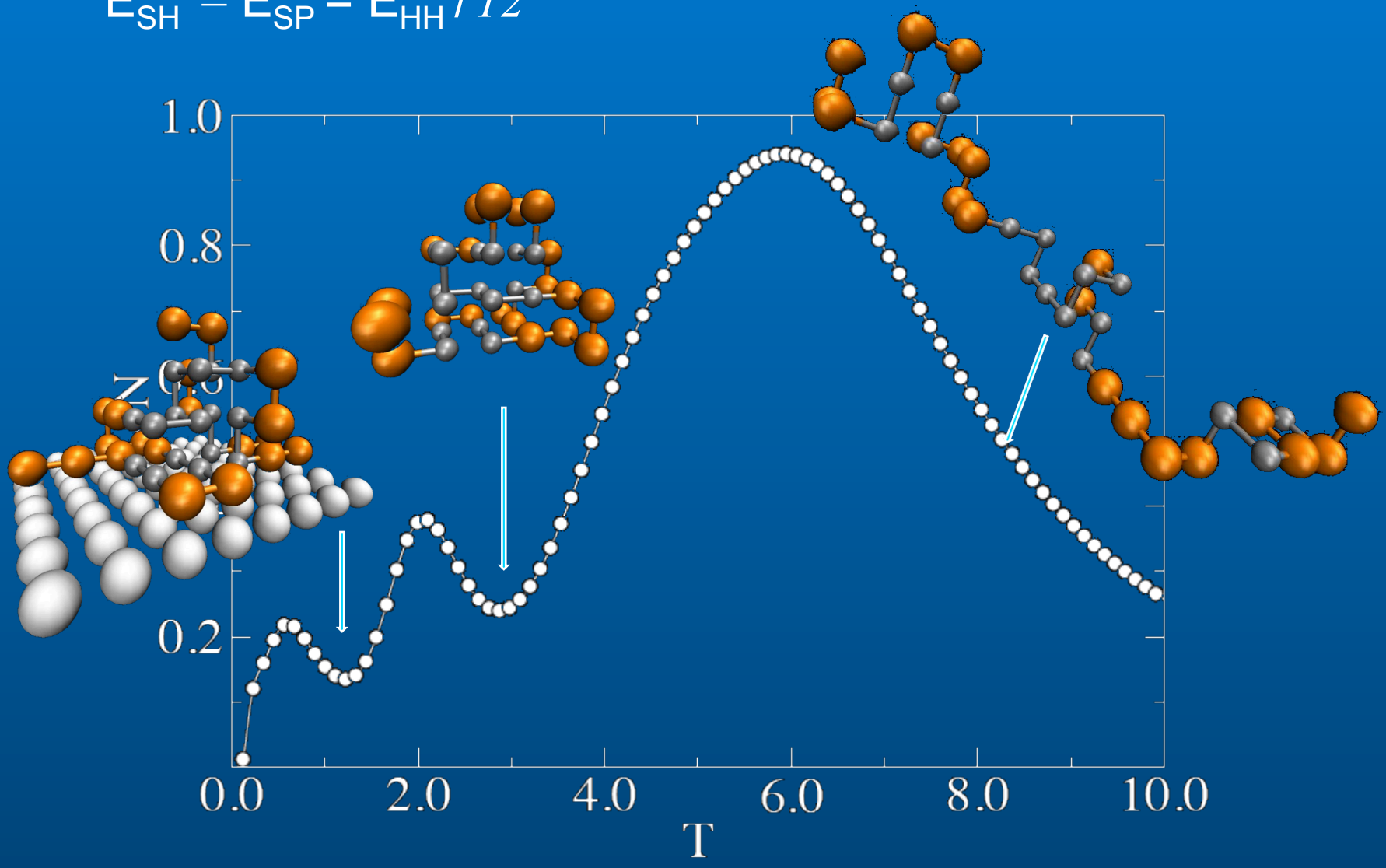
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

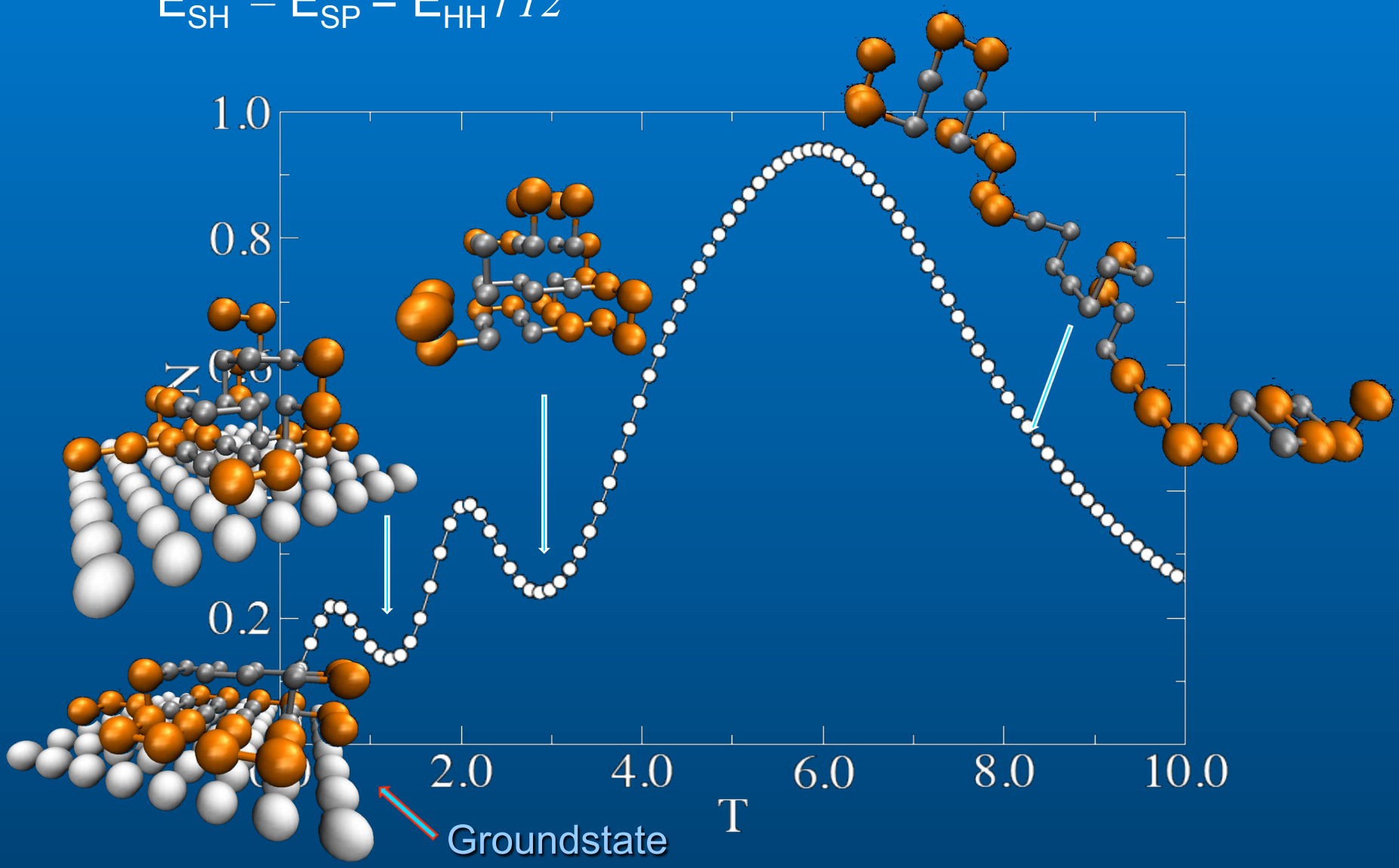
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

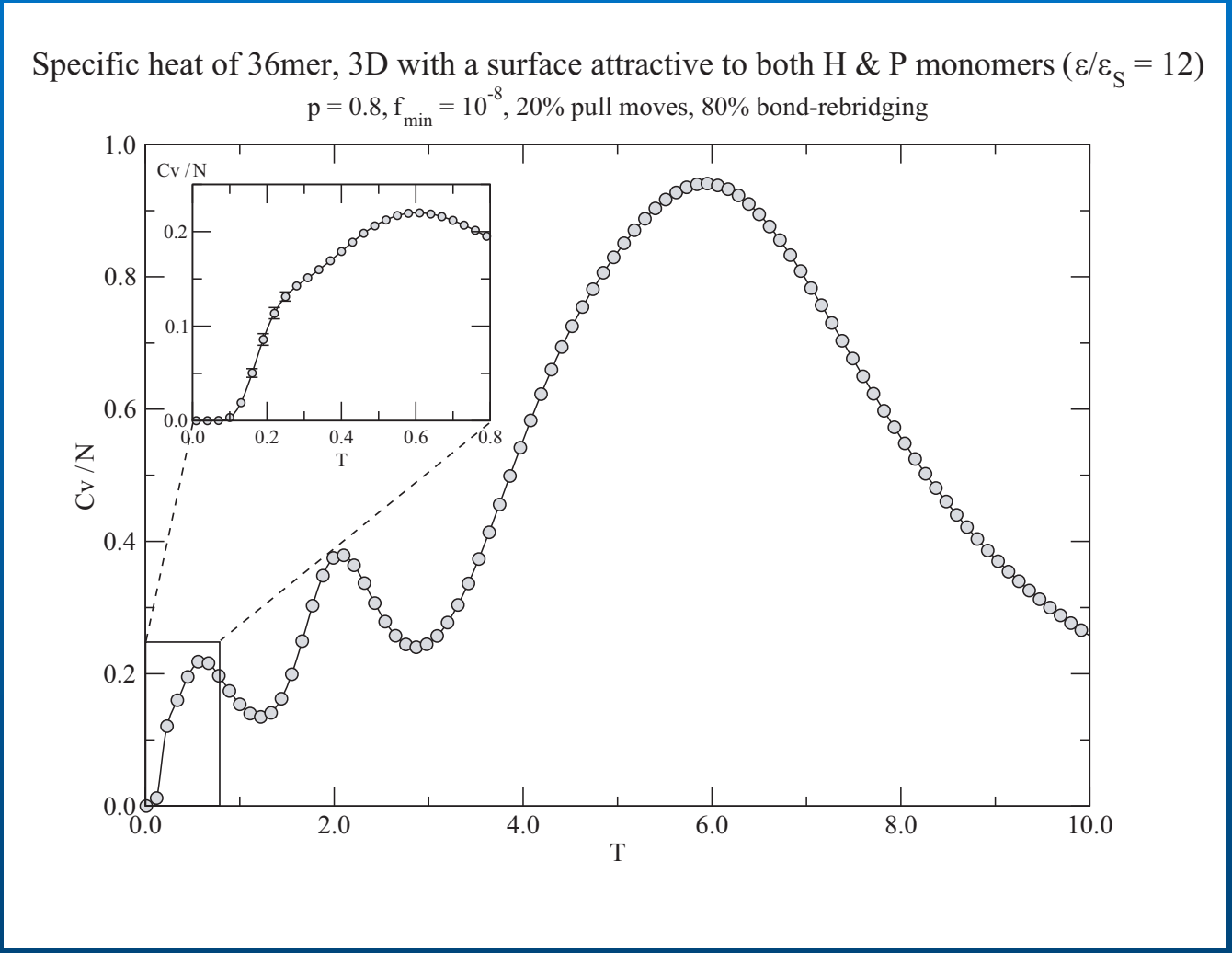
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

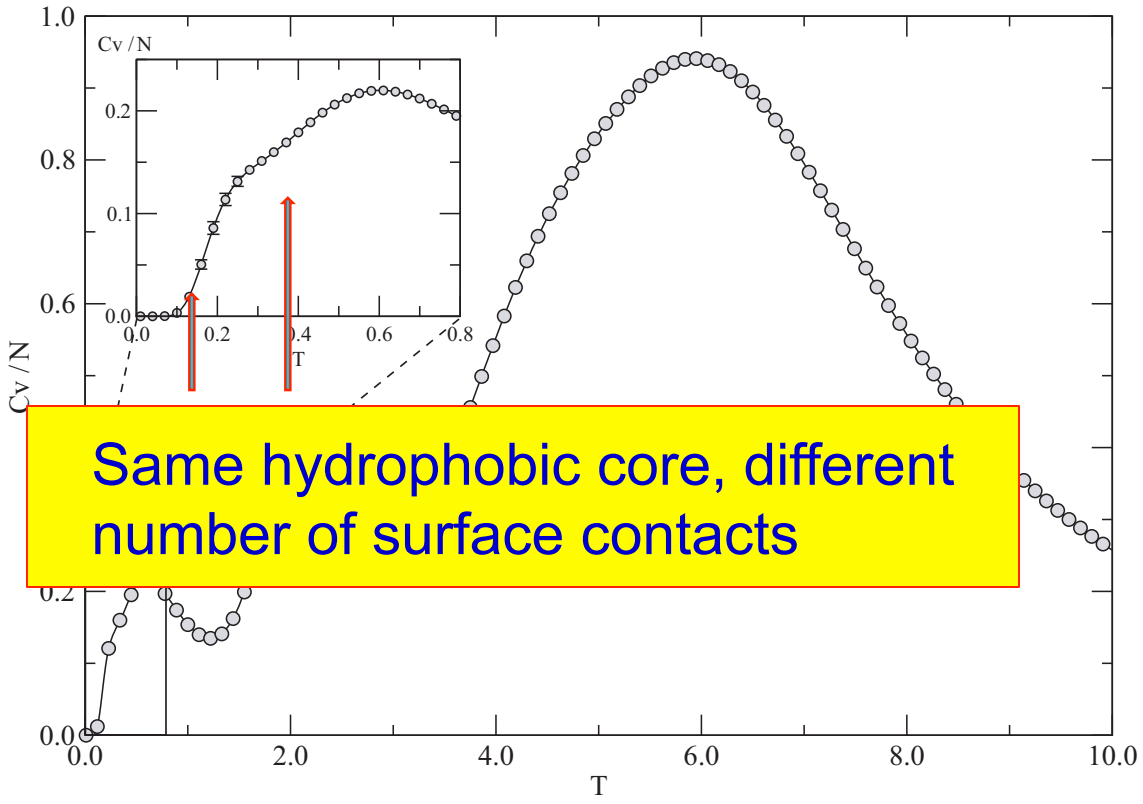
$E_{SH} = E_{SP} = E_{HH} / 12$



36mer in 3 dimensions (cubic lattice) with a surface

The HP model of protein folding

$E_{SH} = E_{SP} = E_{HH} / 12$



Applications to “Complex” Systems

- *Spin glasses*
- *“Lattice proteins”*
- *“Real” proteins*

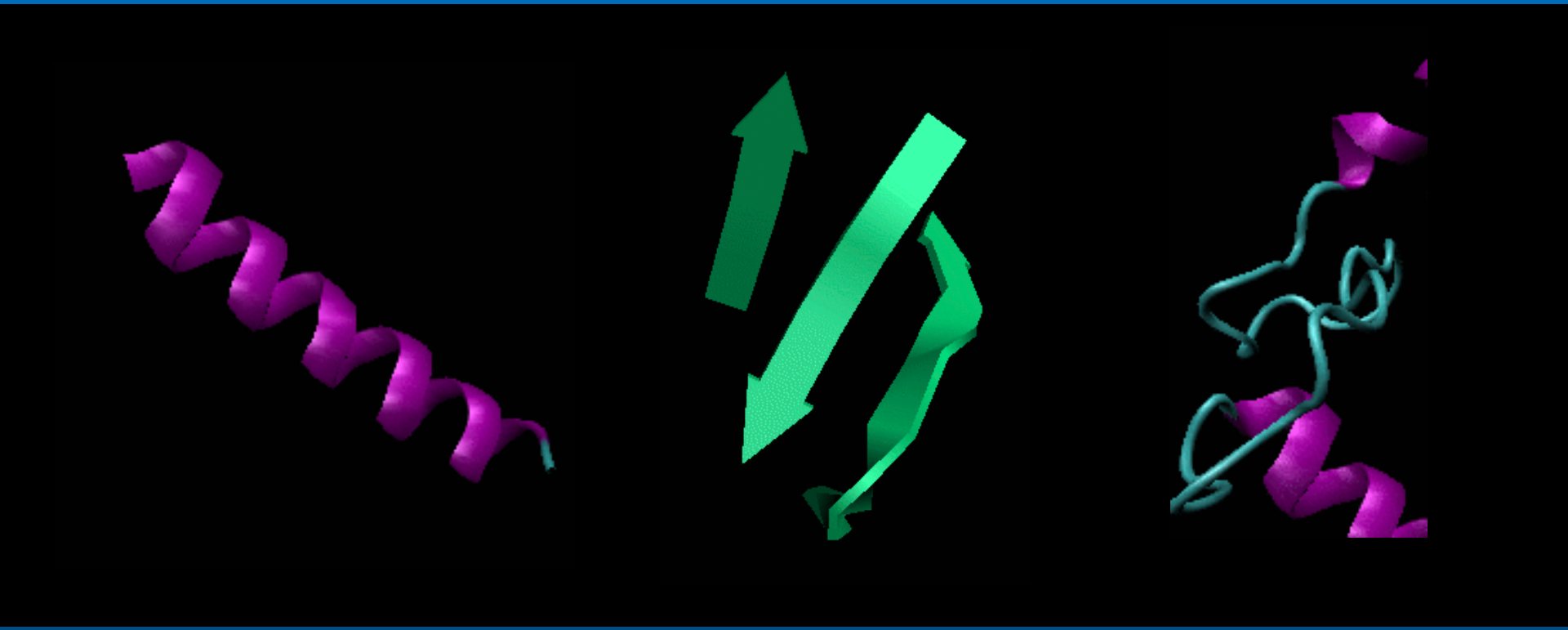
What is a protein?

Primary structure: Sequence of amino acid residues

EITLIIFGVMAGVIGTILLISY

What is a protein?

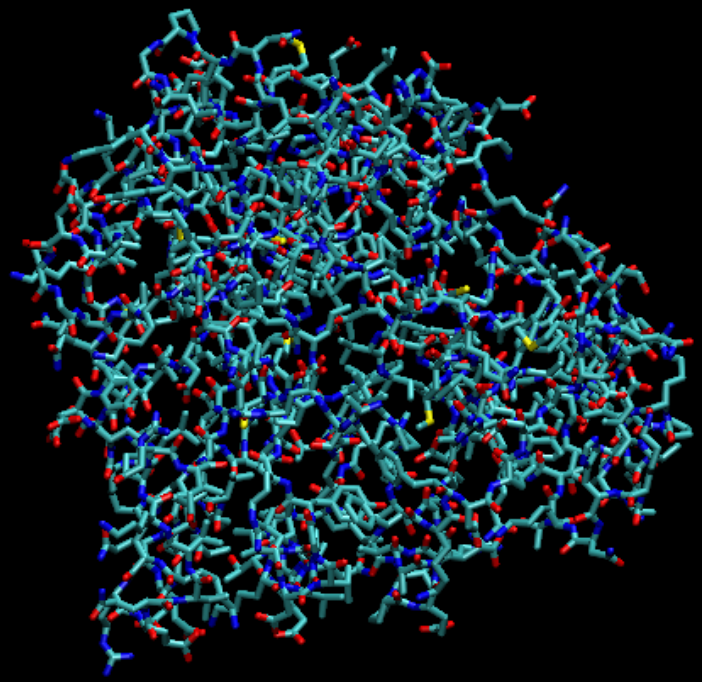
Secondary structure: H-bonds of backbone atoms



Alpha-helix Beta-sheet Loop

What is a protein?

Tertiary structure: 3-dim arrangement of atoms

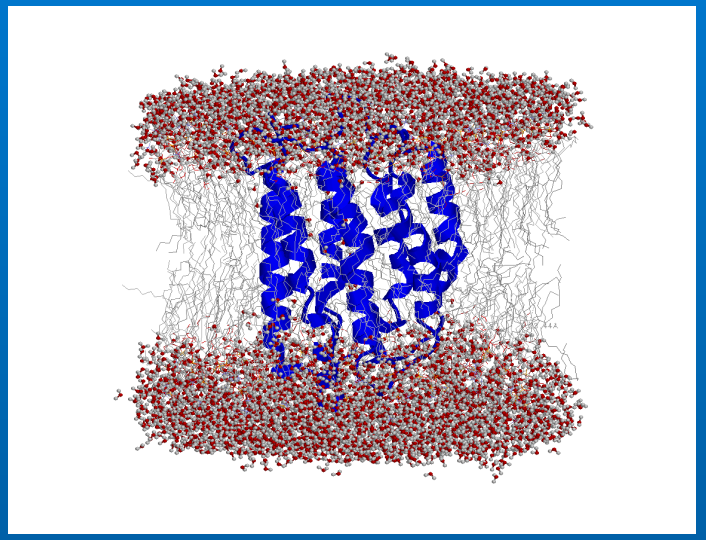


balls & sticks

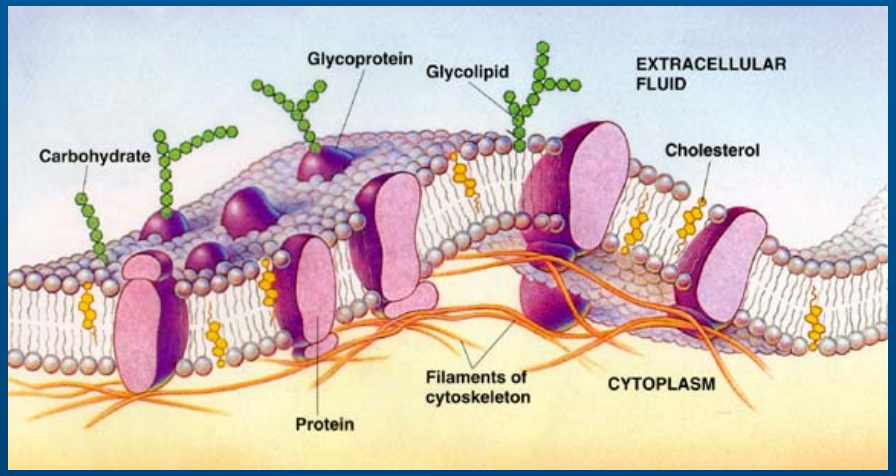
helices & arrows

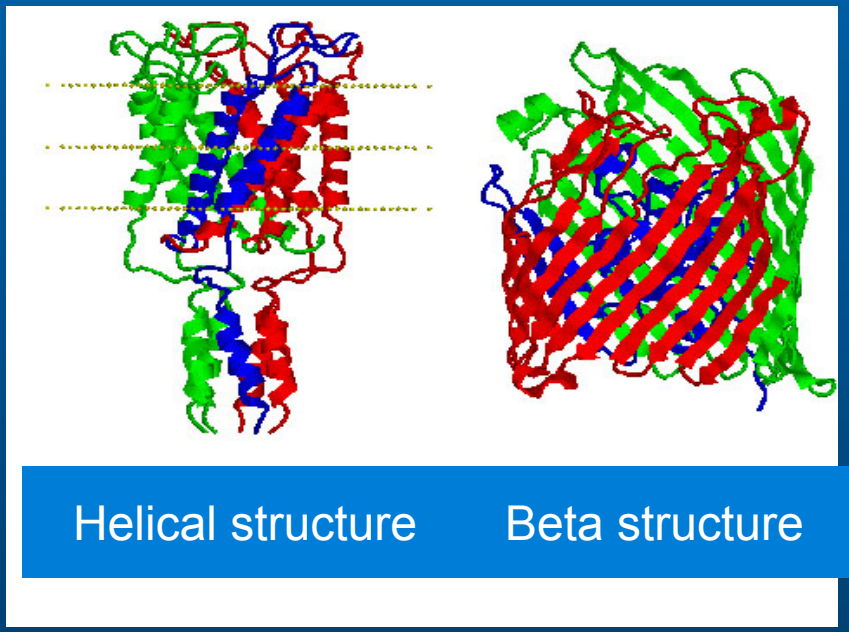
What is a membrane protein?

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



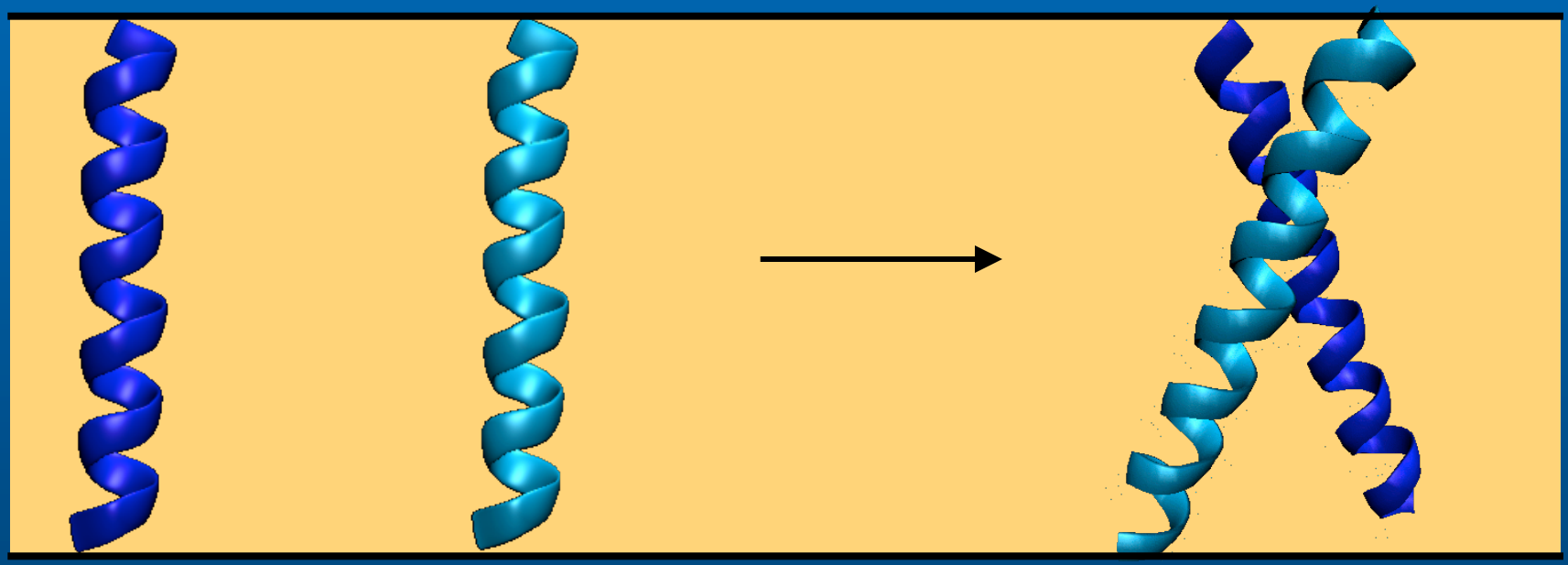
- **> 50% new drug targets are membrane proteins (MP).**





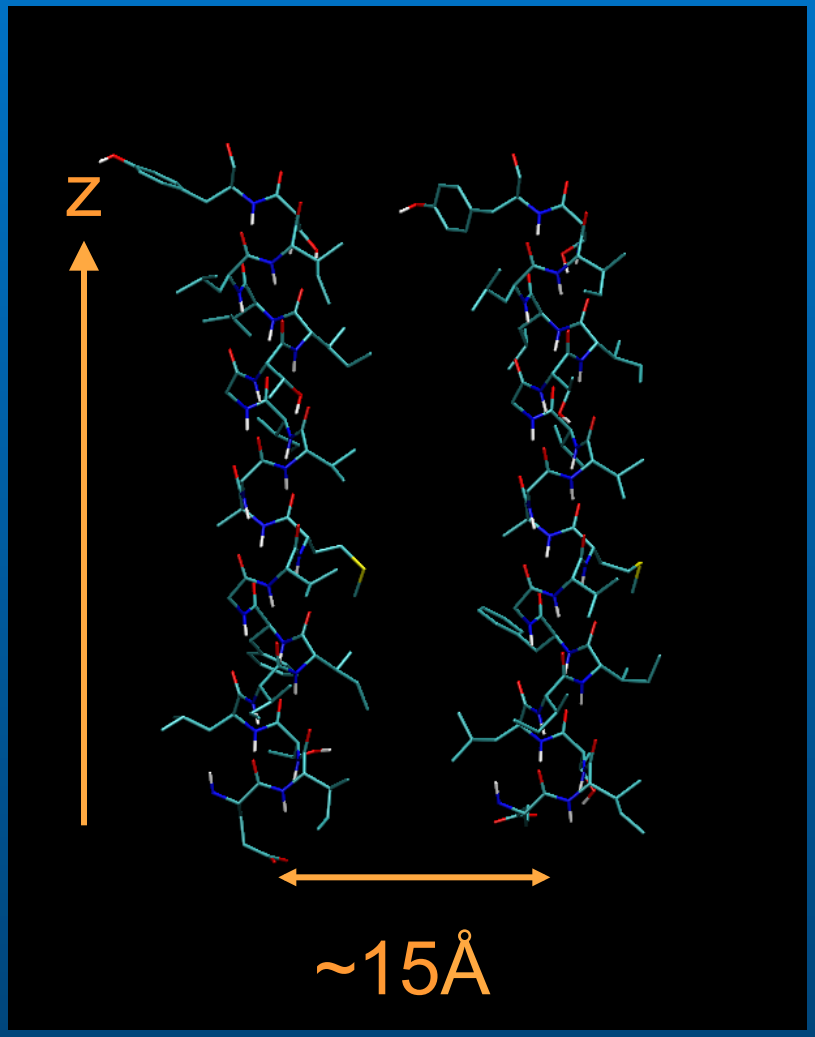
Folding process of GPA

- Single helix stable in the membrane
- Association of helices



(Popot, Engelman, Biochemistry, 1990)

Wang-Landau sampling of a GPA model*

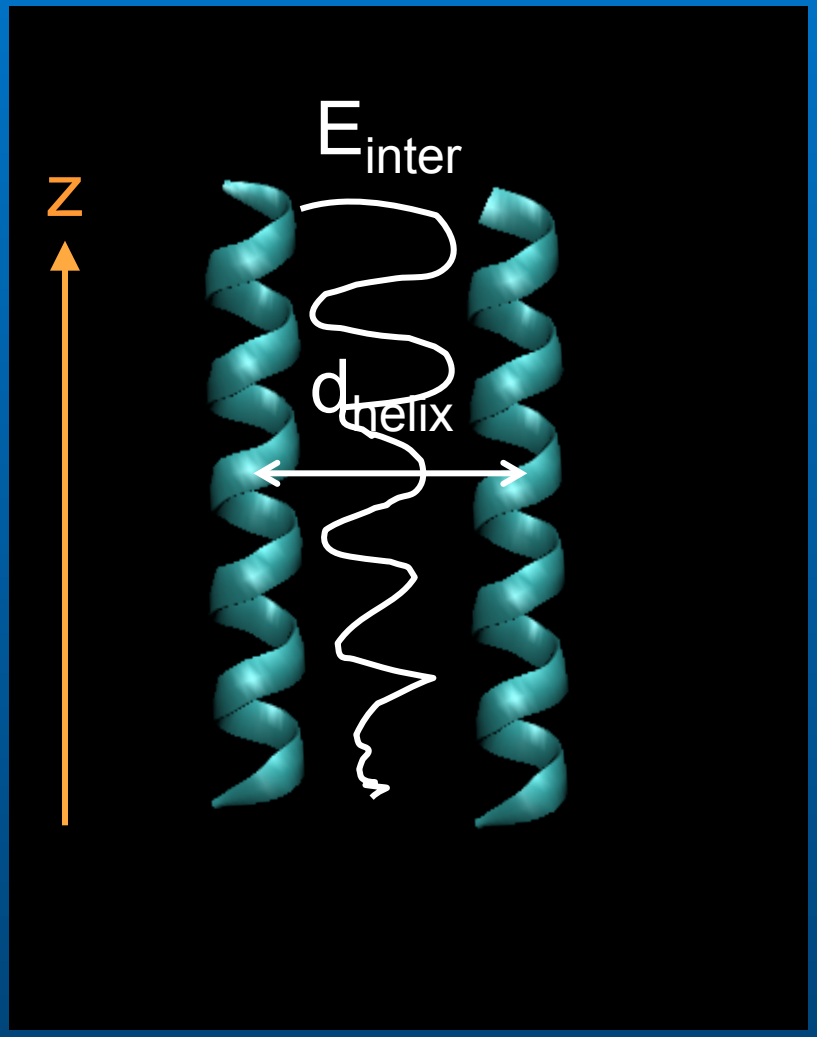


- **Unified-atom model**
- **Total: 368 atoms**
- **2 helices (22 amino-acids)**

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

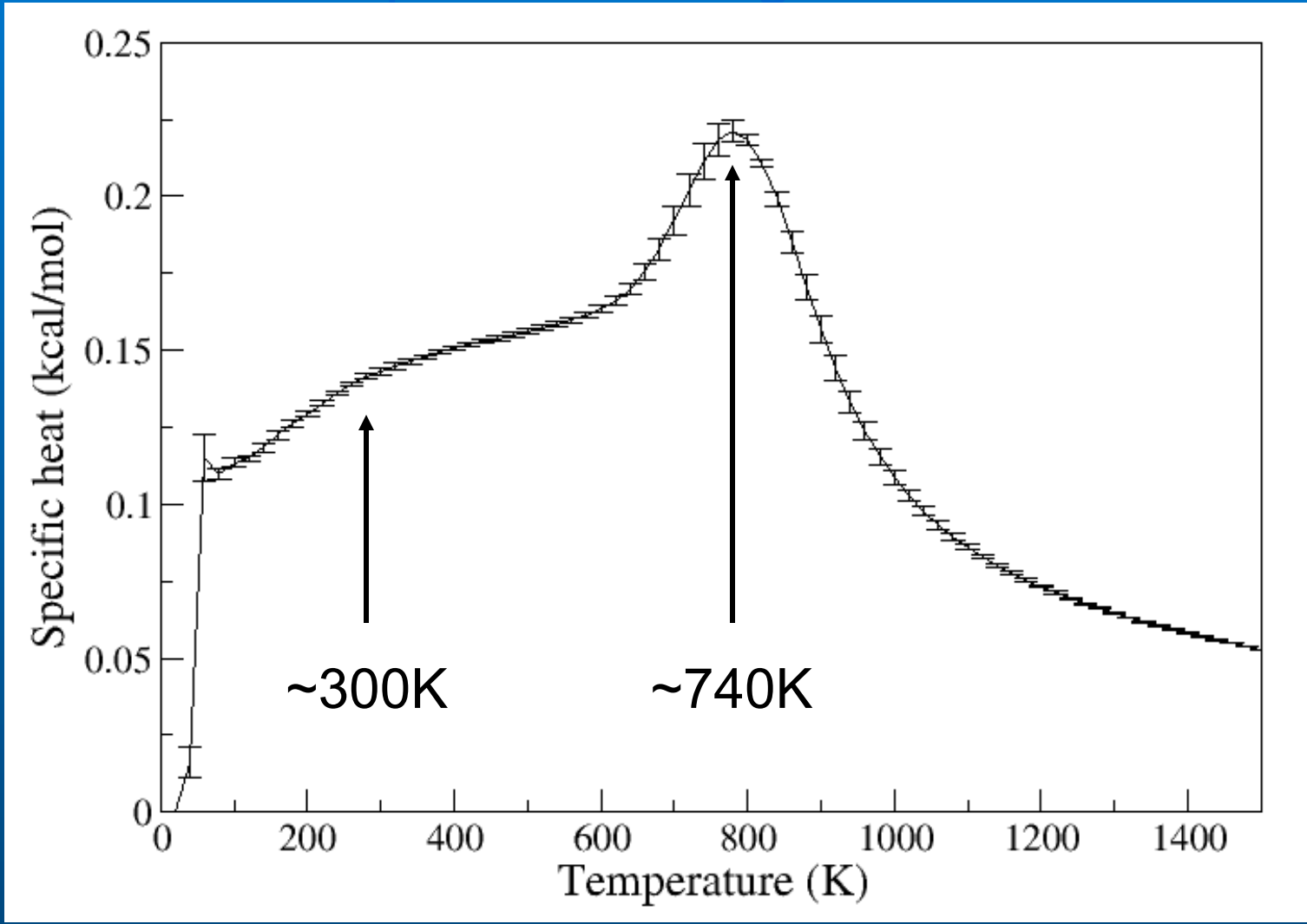
**with Claire Gervais, IOB*

Observables for GPA

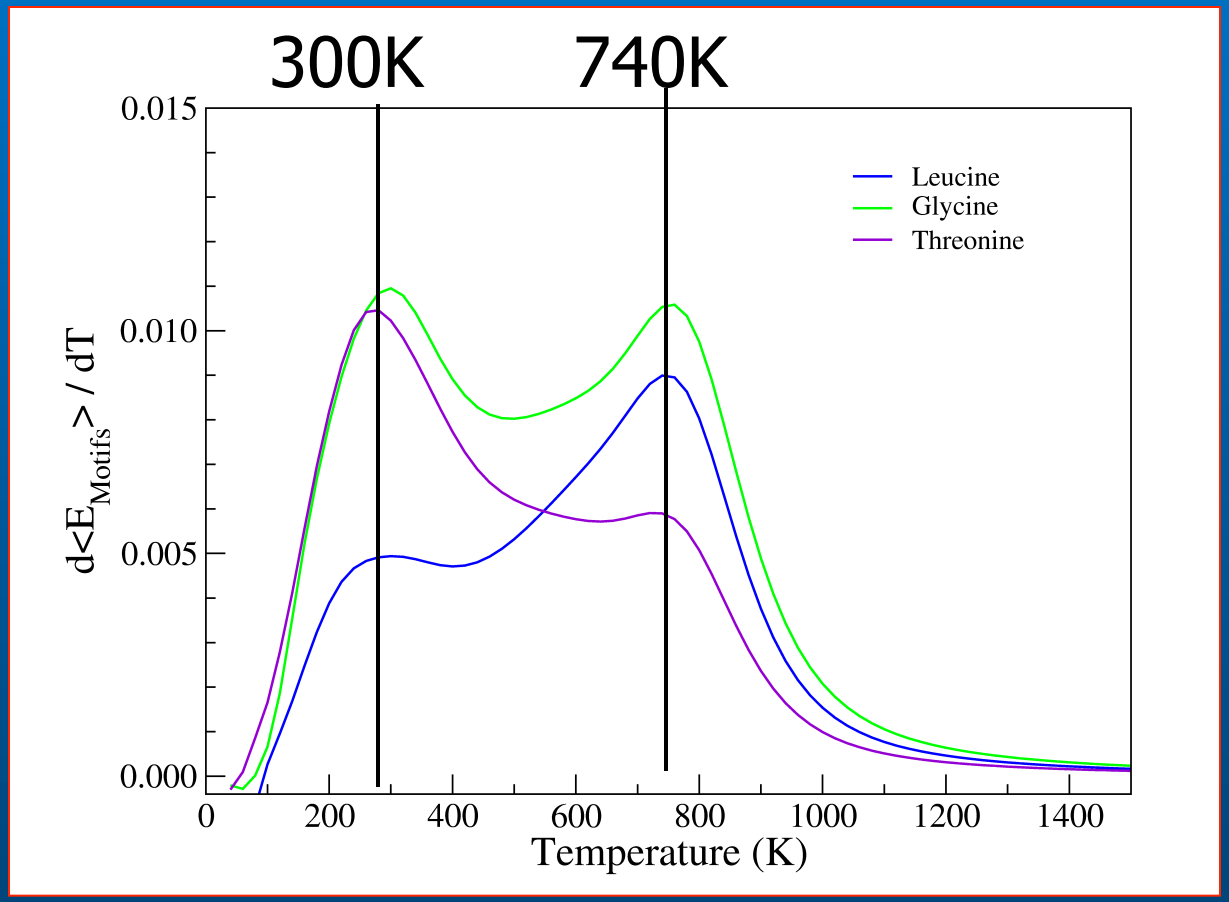


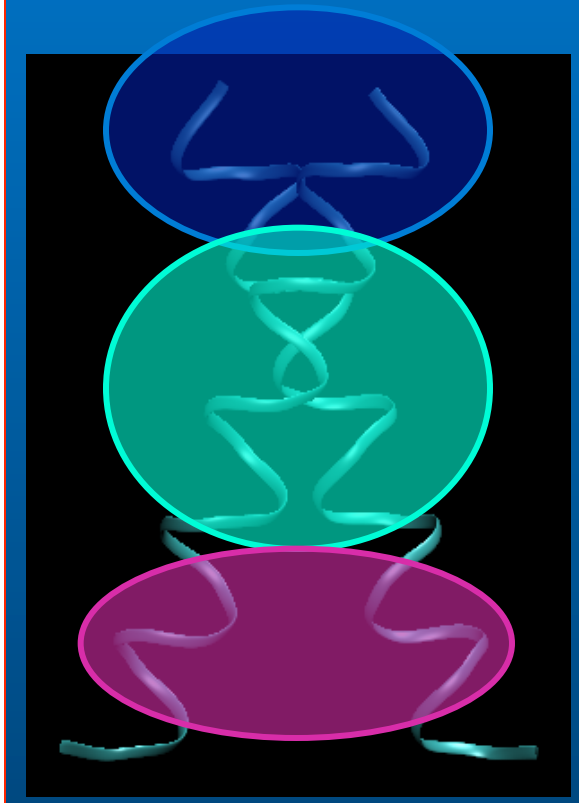
- helix-helix nonbond energy (E_{inter})
- helix-helix distance (d_{helix})
- RMSD of C_α atoms

Specific heat



Results for GPA





Convergence towards native contacts dependent on position in structure

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

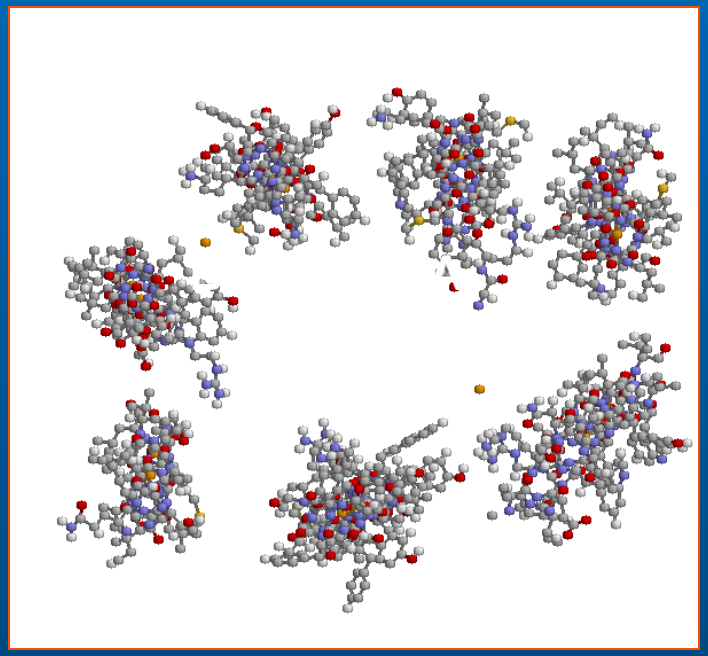
→ Hierarchical acquisition of the native state

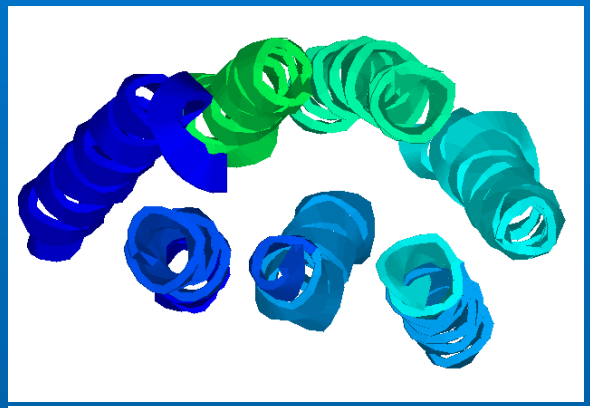
Docking of Bacteriorhodopsin

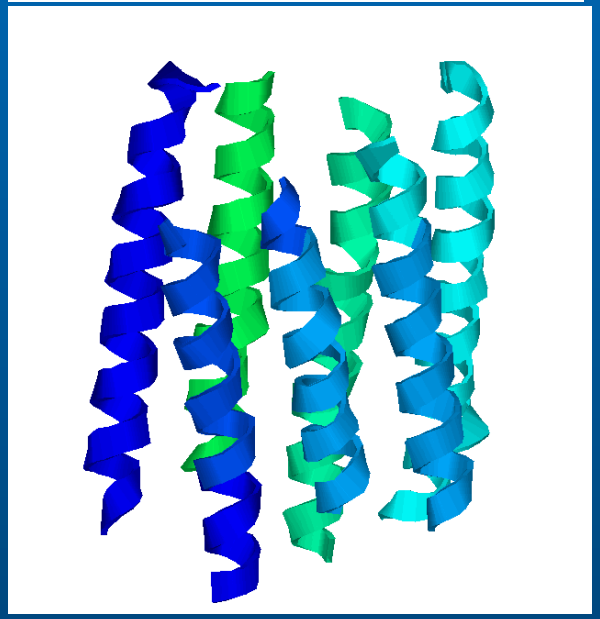
(Zhong Chen, IOB, UGA)

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

Summary

Systems with complex free energy landscapes are particularly challenging for computer simulations. Inventive Monte Carlo algorithms are beginning to make them accessible.

Characteristic examples include:

- Spin glasses (no connectivity constraints)
- "Lattice proteins" (rich, minimalistic model)
- Real proteins (force fields remain a problem)

Summary/Outlook

Systems with complex free energy landscapes are particularly challenging for computer simulations. Inventive Monte Carlo algorithms are beginning to make them accessible.

Parting thought: Research in coming decades will focus on complex systems. Inventive algorithms, together with petaflop/exaflop computers, offer great promise for the future.

Acknowledgments

Collaborators:

Y.-W. Li
F. Wang
T. Wüst
} CSP, U. of Georgia

C. Gervais
Y. Xu
} IOB, U. of Georgia

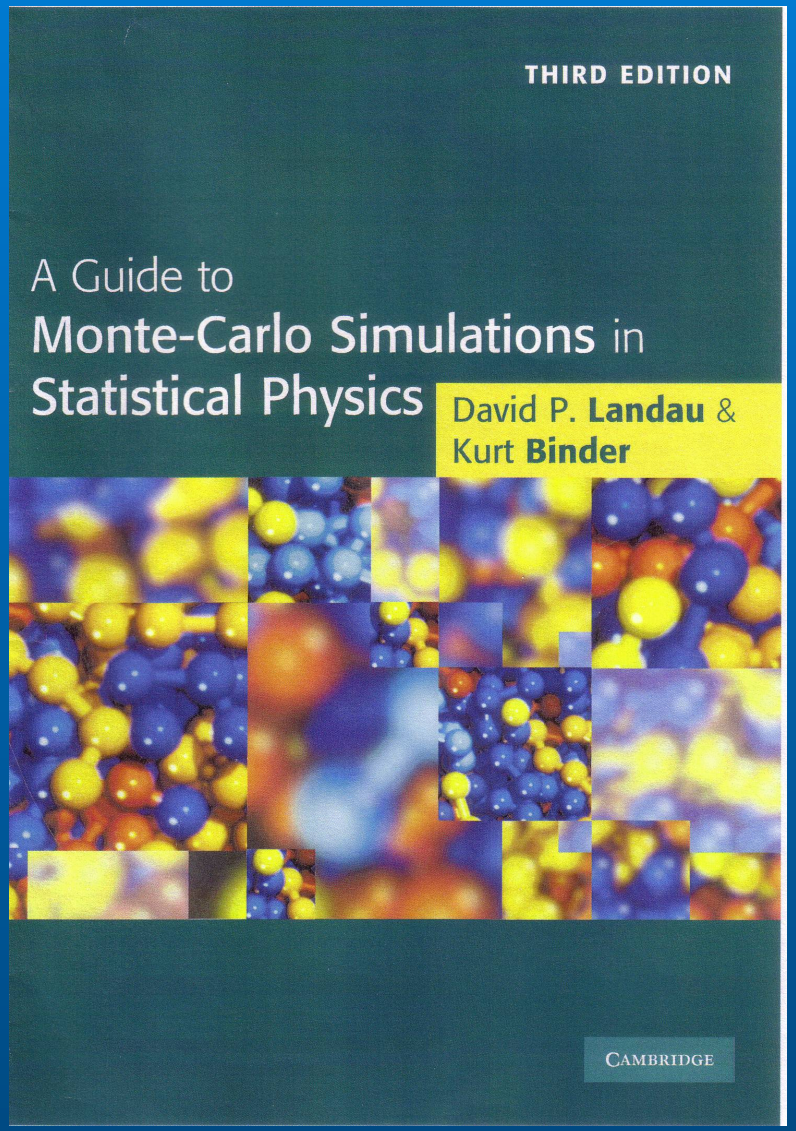
$$$:

NSF (National Science Foundation)
NIH (National Institutes of Health)

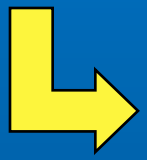
Bits, bytes, and cycles

Research Computer Center at U. of Georgia

Learn more about Monte Carlo,



Please remember



25th Annual Workshop

Recent Developments in Computer Simulation Studies in Condensed Matter Physics

Come help us celebrate the 25th Workshop of the CSP

February 20-24, 2012

Sponsored by the
Center for Simulational Physics
The University of Georgia

Some grad student support is available