Lattice Peptide Simulations using the Wang-Landau Monte Carlo Method

M P Allen

Department of Physics
University of Warwick

MIRAW, Warwick, 7 March 2011

THE UNIVERSITY OF
WARWICK
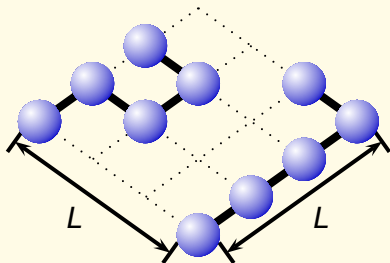
Centre for Scientific Computing
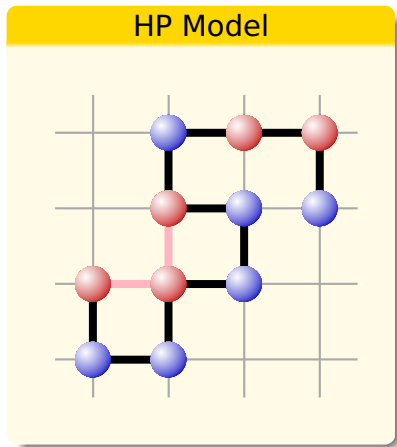
Outline

Simple Polymer Models

Lattice Polymer

- Defined on a lattice
- Sites may be occupied or unoccupied
- Nearest-neighbour interactions
- Connectivity defines polymer chains
- Chains may not cross

- Off-lattice versions are also commonly used.

The HP Model

HP Model

- Self-avoiding chain of hydrophobic $\mathbf{H}$ and polar $\mathbf{P}$ residues,
- living on a 2D (square) or 3D (cubic) lattice.
- Each contact pair of non-bonded H residues contributes one unit $-\epsilon$ of favourable energy.
- Encapsulates the basic problems of folding.

K Lau, KA Dill, *Macromolecules,* **22,** 3986 (1989).

Folding Speed

- Proteins fold fast - sometimes in microseconds
  - Not by random searching of conformations!
  - How do they avoid the wrong routes?
  - "Levinthal paradox"
- Computer simulations need to be faster!

Dill has proposed a zipping-and-assembly model.

- local structures form in independent fragments
- then grow (zip) or coalesce with other structures (assemble) so as to optimise local contacts.
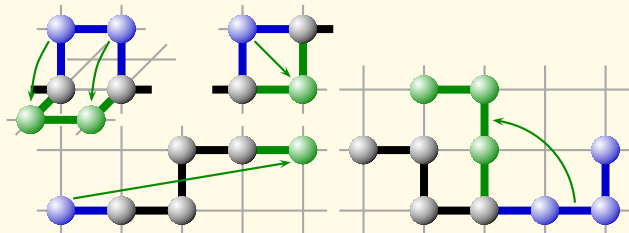
📄 VA Voelz, KA Dill, *Proteins: Structure, Function, and Bioinformatics*, **66,** 877 (2007).

MC Moves for Lattice Polymers
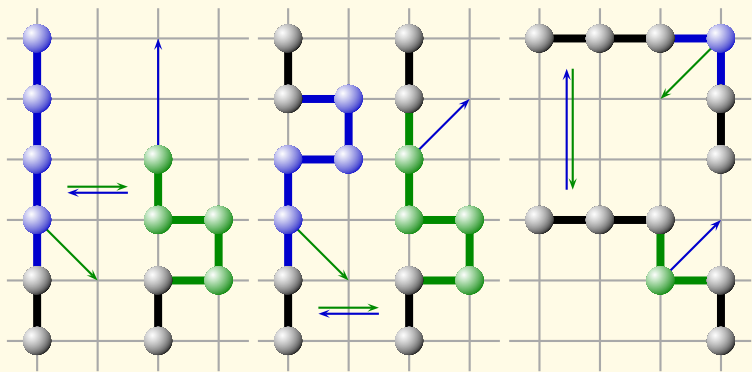

Lattice Polymer Moves

- Conventional lattice polymer Monte Carlo moves: crankshaft, corner flip, slithering snake, pivot.
- However, in dense phases, dramatic moves like pivot are likely to result in overlaps with surrounding atoms.

Novel Pull Moves

Defined by Single-Atom Move

Novel Pull Moves

- New pull moves allow local contacts to form.
- They are ergodic, and improve efficiency.
- Especially important when chain is closely packed.
- Include some conventional moves, e.g. corner-flip.
- This is a *biased* sampling method, and the way the moves are selected must be included in the acceptance/rejection criterion.
- Counting the available pull moves (forward and reverse) is a critical part of the method.

📄 M Lesh, M Mitzenmacher, S Whitesides, *Proc. 7th Ann. Int. Conf. on Research in Computational Molecular Biology,* p188 (2003).

Novel Pull Moves

Pull Move Metropolis Equation

$$\mathcal{P}_{\mathsf{acc}}(\boldsymbol{r'} \leftarrow \boldsymbol{r}) = \min\left(1, \frac{\alpha_{\mathsf{pull}}(\boldsymbol{r} \leftarrow \boldsymbol{r'})}{\alpha_{\mathsf{pull}}(\boldsymbol{r'} \leftarrow \boldsymbol{r})}\mathrm{e}^{-\Delta E/k_{\mathsf{B}}T}\right)$$

- ΔE is change in energy associated with $\boldsymbol{r'} \leftarrow \boldsymbol{r}$
- $\alpha_{\mathsf{pull}}(\boldsymbol{r'} \leftarrow \boldsymbol{r}) = \mathcal{N}_{\mathsf{pull}}(\boldsymbol{r'} \leftarrow \boldsymbol{r})/\mathcal{N}_{\mathsf{pull}}(\boldsymbol{r})$
  - $\mathcal{N}_{\mathsf{pull}}(\boldsymbol{r'} \leftarrow \boldsymbol{r}) =$ number of pull moves to $\boldsymbol{r'}$ from $\boldsymbol{r}$
  - $\mathcal{N}_{\mathsf{pull}}(\boldsymbol{r}) =$ total number of pull moves from $\boldsymbol{r}$
- $\alpha_{\mathsf{pull}}(\boldsymbol{r} \leftarrow \boldsymbol{r'}) = \mathcal{N}_{\mathsf{pull}}(\boldsymbol{r} \leftarrow \boldsymbol{r'})/\mathcal{N}_{\mathsf{pull}}(\boldsymbol{r'})$ similarly.
- Need to consider both forward and reverse moves.
- Counting these available pull moves is quite time consuming.
- We improve the method by simplifying the counting (and rejecting some moves).

Outline

Density of States Sampling

Efficient MC methods have been developed to map out behaviour of systems over a wide temperature range, including identification of low-E structures. With $\beta = 1/k_\mathrm{B}T$, $\mathbb{W}(E) =$ density of states:

Thermodynamics from Density of States

$$\text{partition function:} \quad \mathbb{Q} = \sum_{\boldsymbol{r}} \mathrm{e}^{-\beta E(\boldsymbol{r})} = \sum_{E} \mathbb{W}(E)\,\mathrm{e}^{-\beta E}$$

$$\text{average energy:} \quad \langle E \rangle = \mathbb{Q}^{-1}\sum_{E} \mathbb{W}(E)\,E\,\mathrm{e}^{-\beta E}$$

$$\text{heat capacity:} \quad C_V = \left(\langle E^2 \rangle - \langle E \rangle^2\right)/k_\mathrm{B}T^2$$

Density of States Sampling

- $\mathbb{W}(E)$ increases very rapidly with E.
- A density-of-states simulation samples states with a probability or weight $P(E) \propto 1/\mathbb{W}(E)$.
- This covers the entire energy scale uniformly.
- Allows low-energy configurations to exchange with higher-energy ones and vice versa.

📄 FG Wang, DP Landau, *Phys. Rev. E*, **64**, 056101 (2001).

Acceptance-Rejection Criterion

$$P_{\text{acc}}(\boldsymbol{r}' \leftarrow \boldsymbol{r}) = \min\left(1, \frac{P(E(\boldsymbol{r}'))}{P(E(\boldsymbol{r}))}\right) = \min\left(1, \frac{\mathbb{W}(E(\boldsymbol{r}))}{\mathbb{W}(E(\boldsymbol{r}'))}\right)$$

Wang-Landau Sampling

$\mathbb{W}(E)$ is built up progressively during the simulation.

Wang-Landau Algorithm

1. Set all $\mathbb{W}(E) = 1$, set factor $f = e = 2.71828$.
2. Set all $h(E) = 0$.
3. Conduct MC moves using $\mathbb{W}(E)$, accepted states t
   - After each move $\mathbb{W}(E_t) \rightarrow f \times \mathbb{W}(E_t)$, $h(E_t) \rightarrow h(E_t) + 1$
4. Repeat from step 3 until $h(E)$ is "sufficiently flat"
5. $f \rightarrow \sqrt{f}$ (f decreases exponentially with time)
6. Repeat from step 2 until f is "sufficiently small".

Hope $\mathbb{W}(E)$ converges to correct density of states.

Density of States

- M Bachmann, W Janke, *J. Chem. Phys.,* **120,** 6779 (2004).

- T Wüst, DP Landau, *Comp. Phys. Commun.,* **179,** 124 (2008).

- Wang-Landau covers the entire energy scale uniformly.
- Can map out behaviour of lattice polymers and peptides over a wide temperature range.
- Allows identification of low-E structures.

Pull Move Acceptance-Rejection Criterion

$$\mathcal{P}_{\text{acc}}(\boldsymbol{r}' \leftarrow \boldsymbol{r}) = \min\left(1, \frac{\mathbb{W}\big(E(\boldsymbol{r})\big)}{\mathbb{W}\big(E(\boldsymbol{r}')\big)} \frac{\alpha_{\text{pull}}(\boldsymbol{r} \leftarrow \boldsymbol{r}')}{\alpha_{\text{pull}}(\boldsymbol{r}' \leftarrow \boldsymbol{r})}\right)$$

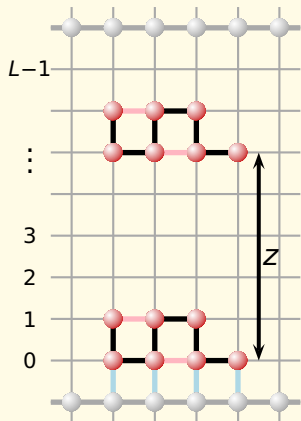# Example: 103-bead Lattice Protein

Successive f Iterations
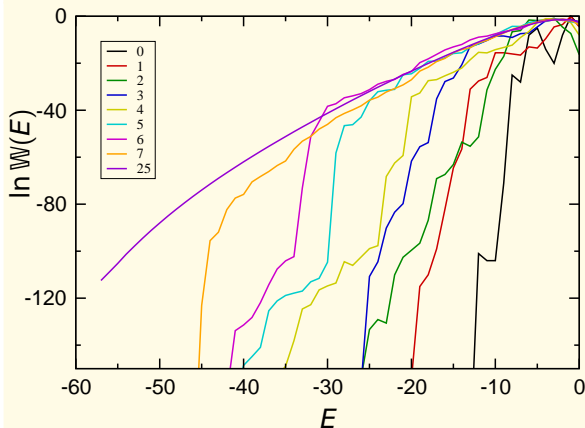
"cytochrome C"
PPHHPPPPPH...
HPPHHPHPPH...
PPPPPPPHPP...
PHHPHHPPPP...
PPHPPHPHPP...
HPPPPPHHHP...
PPPHHPHHPP...
PPPHHPPPPH...
HHHPHPPPPP...
PPPHHHHHPP...
HPP

Polymers on Surfaces

Slit Geometry

- Confining wall at L
- Energy $E = -n\epsilon - s\sigma$
- $n = n(\boldsymbol{r})$ = number of H-H contacts for internal configuration $\boldsymbol{r}$;
- $\boxed{-\epsilon}$ = contact energy.
- $s = s(\boldsymbol{r}, z)$ = number of surface contacts for given $\boldsymbol{r}$, and position z;
- $\boxed{-\sigma}$ = surface energy.
- Density of states $\mathbb{W}_L(n, s)$
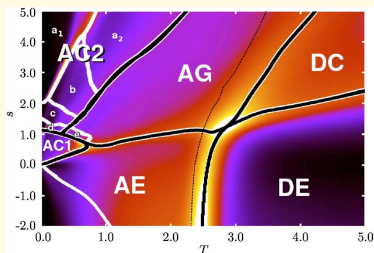- $\mathbb{Q} = \sum_n \sum_s \mathbb{W}_L(n, s) e^{+n\beta\epsilon + s\beta\sigma}$

Polymers on Surfaces

📄 M Bachmann, W Janke, *Phys. Rev. E,* **73,** 020901 (2006); *ibid,* **73,** 041802 (2006).

- Studied H_{179} homopolymer in cavity with $L = 200$.
- Used chain-growth (PERM) rather than MC moves.
- Pseudo-phase diagram in T and $\chi = \epsilon/\sigma$.
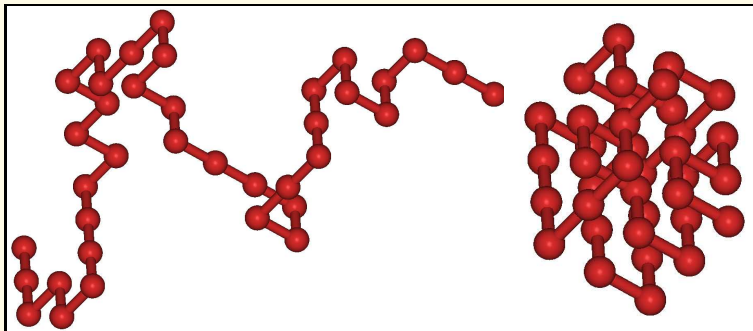- Transitions marked by peaks in "heat capacities".

Phase Diagram

States Observed

AE: Adsorbed-Expanded

DE: Desorbed-Expanded

AC: Adsorbed Compact

DC: Desorbed Compact
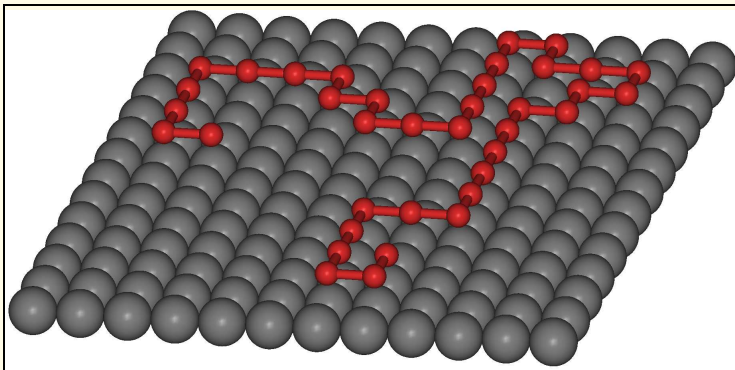
AG: Attached Globule

H_{37} Homopolymer

Desorbed Configurations

- Desorbed Expanded
- High T, low $\chi = \epsilon/\sigma$.

- Desorbed Compact
- High T, high $\chi = \epsilon/\sigma$.

H$_{37}$ Homopolymer

Adsorbed Configurations

- Adsorbed Expanded
- Low T, low $\chi = \epsilon/\sigma$.

H$_{37}$ Homopolymer

Adsorbed Configurations
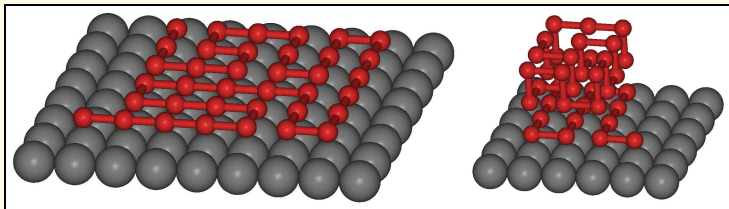
- Adsorbed Compact
- Low T, high $\chi = \epsilon/\sigma$.

- Adsorbed Globule
- Medium T, high $\chi = \epsilon/\sigma$.

Polymers on Surfaces

Really interested in polymer adsorption on a single wall. But cannot allow $L \to \infty$.

- Significant improvement of the method.
- There is no need for the second confining wall.

Redefine $s = s(\boldsymbol{r})$, as the number of beads on the lower bounding surface z of a polymer configuration $\boldsymbol{r}$.

New Surface Monte Carlo Algorithm

1. Standard pull move on isolated polymer
   - Count the monomer-monomer interactions n
2. Translate the surface to the plane of contact
   - Count the monomer-surface interactions s
3. Look up density of states $\mathbb{W}(n, s)$, accept or reject
4. NB: $\sum_s \mathbb{W}(n, s) = \mathbb{W}(n)$ for free polymer, as before.

Polymers on Surfaces

Relate To Slit Geometry

$$\mathbb{W}_L(n, s) = \begin{cases} A\mathbb{W}(n, s) & s > 0 \\ A(L - D(n))\,\mathbb{W}(n) & s = 0 \end{cases}$$

$$\mathbb{Q}_L = A \sum_n \sum_{s>0} \mathbb{W}(n, s) e^{+n\beta\epsilon} e^{+s\beta\sigma} + AL\mathbb{Q}_{\text{int}}$$

$$- A \sum_n \mathbb{W}(n)D(n)e^{+n\beta\epsilon}$$

- $A =$ cross-sectional area, $L =$ wall separation.
- $\mathbb{Q}_{\text{int}} =$ internal partition function of free polymer.
- $D(n)$ is the average z-extent of polymers with n internal contacts

Outline

Wang-Landau Peptides   MIRAW, Warwick, 7 March 2011
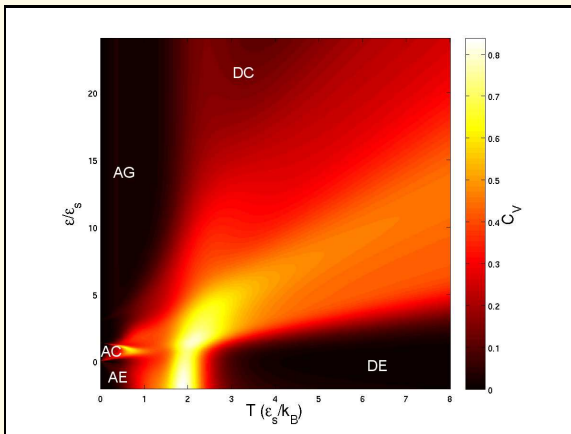
H$_{37}$ Homopolymer

Density of States
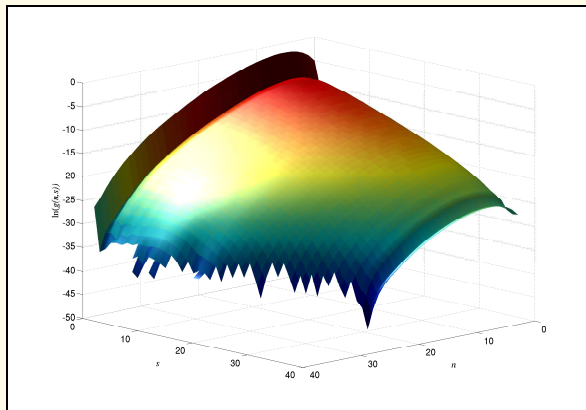
AD Swetnam,
MP Allen,
*Phys Chem
Chem Phys*,
11, 2046
(2009).

H$_{37}$ Homopolymer
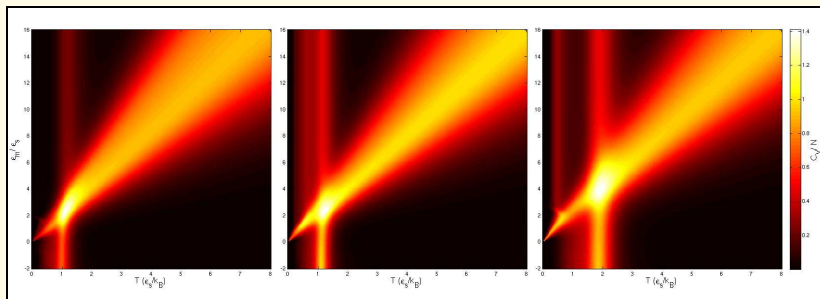
Heat Capacities

36-bead HP Peptide

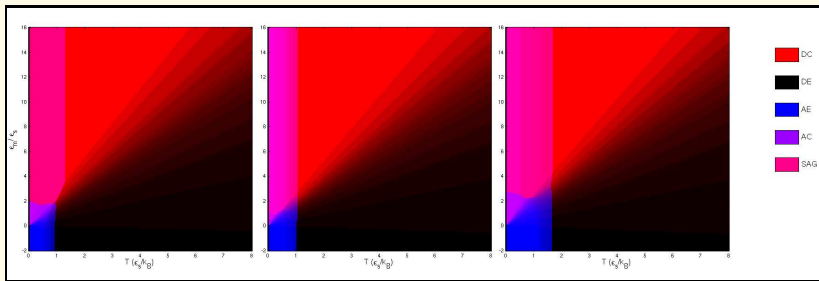PPPHHPPHHPPPPPHHHHHHHPPHHPPPPPHHPPHPP

Heat Capacities

Left to right: surface attracts H only, P only, both H&P.
Pseudo-transitions indicated by regions of high C_V.

36-bead HP Peptide

PPPHHPPHHPPPPPHHHHHHHPPHHPPPPHHPPHPP

Phase Diagrams

Left to right: surface attracts H only, P only, both H&P.
Phases are identified by monomer and surface contacts
of free-energy-minimizing conformations.

Outline

Wang-Landau Peptides   MIRAW, Warwick, 7 March 2011

Optimizing Wang-Landau

Is it best to make f change exponentially?

- At each stage of Wang-Landau $f \rightarrow \sqrt{f}$.
- It may be better to make $\ln f \propto 1/t$

📄 RE Berardinelli, VD Pereyra, *J. Chem. Phys.,* **127,** 184105 (2007).

📄 C Zhou, J Su, *Phys. Rev. E,* **78,** 046705 (2008).

- To test properly requires exactly soluble model.
- Exact enumeration, isolated 18 bead homopolymer.
- Fractional error $= \epsilon_{\mathrm{rms}} = \sqrt{\overline{\epsilon(E)^2}}$ where

$$\epsilon(E) = \frac{\ln \mathbb{W}(E) - \ln \mathbb{W}_{\mathrm{exact}}(E)}{\ln \mathbb{W}_{\mathrm{exact}}(E)}$$

Optimizing Wang-Landau

Conventional Wang-Landau

- Specify a flatness criterion: 80%. . . 99%
- When $\min(H) > \text{flatness} \times \overline{H}$, $f \to \sqrt{f}$.

Modified Wang-Landau

- Initially change modification factor in the normal way (in this case with flatness=80%)
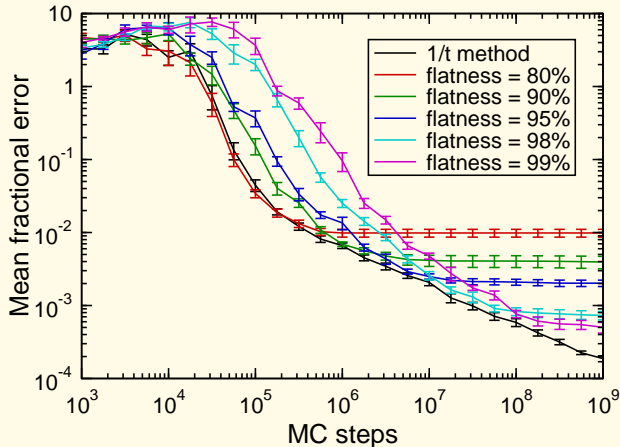- After the point where $\ln f < 1/t$, set $\ln f = 1/t$.

Here t is defined as the number of MC steps M divided by the number of energy levels N_E.

Optimizing Wang-Landau

10 runs were performed to get each set of data.

Comparison of Wang-Landau Methods

HP peptides are challenging!

- For HP sequences, some runs fail to converge.
- Adjust modification factor according to progress.
- Hard-to-access low-lying states
- Even the $1/t$ method has trouble

Proposed Improvement

$$\ln f = p \sum_E \left(\frac{H(E)}{\sum_E H(E)} - \frac{1}{N_E} \right)^2 = p \left(\frac{\sum_E H(E)^2}{M^2} - \frac{1}{N_E} \right)$$
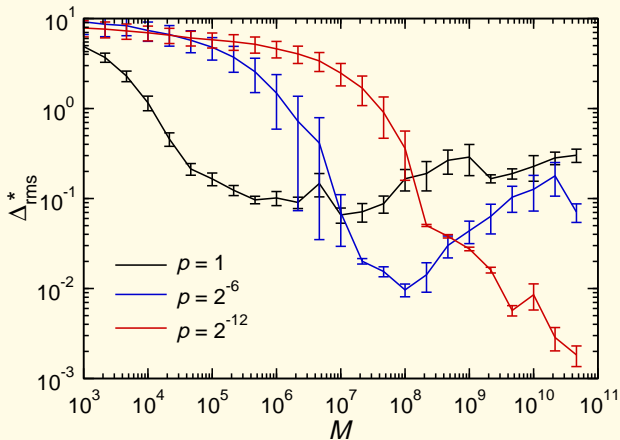
$p =$ adjustable prefactor.

C Zhou, J Su, *Phys. Rev. E,* **78,** 046705 (2008).

AD Swetnam, MP Allen, *J. Comput. Chem.,* **32,** 816 (2011).

Example: 103-bead Lattice Protein

Various prefactors p

Outline

Wang-Landau Peptides   MIRAW, Warwick, 7 March 2011
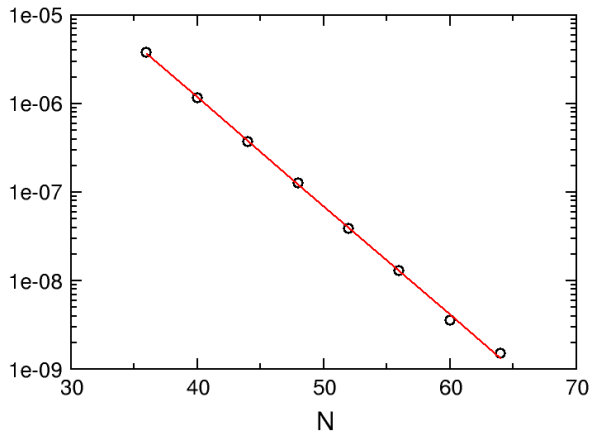
Ring Polymers

- Pull moves preserve knot topology.
- Wang-Landau sampling of $\mathbb{W}(E, L)$.
- $L =$ z-height (or wall separation).
- Plot C_V vs (T, F), F is tensile/compressive force.
- Combinatorial effects diminish number of attempts $L = 2 \rightarrow L = 1$ for unknotted polymer and large N.
- Effectively a physical bottleneck in E, L-space.
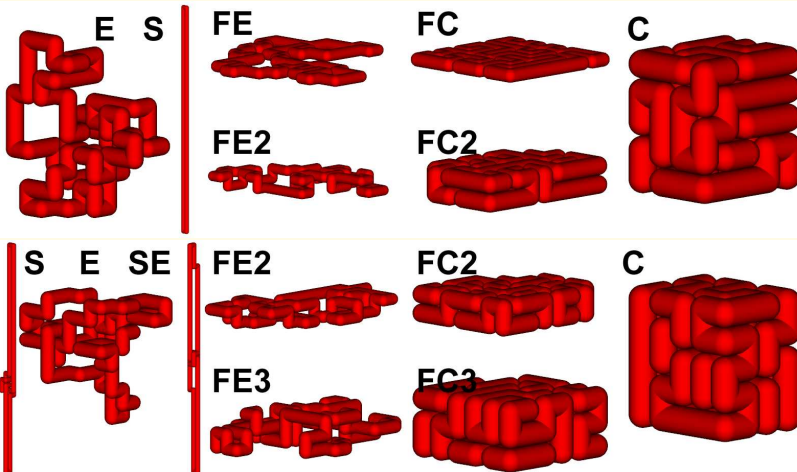- Sample the number of beads in each layer.

Ring Polymers

Attempt fraction for $L = 2 \rightarrow L = 1$ vs N

100-bead Ring Polymer

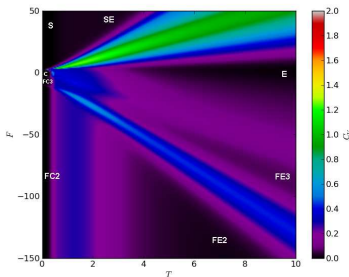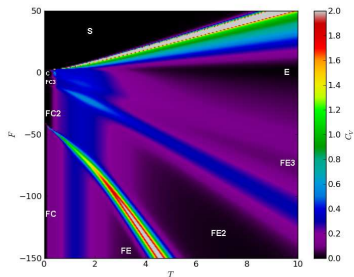Un-knotted (above) and Trefoil-knotted (below)

100-bead Ring Polymer

- Heat capacity maps (pseudophase diagrams).
- Strong transition to flat state in unknotted case.

Unknotted (left) and Trefoil-knotted (right)

Acknowledgements