Monte Carlo Simulation of Protein Folding

Yangrui Hu

2018-May-03

PHYS 2600 Computational Physics Final presentation

Introduction

- Protein folding is one of the most important topics in biology and biophysics.
- When a protein works in a biological activity, its structure is folded, aka **folded state**
- Heating or some chemical environments can cause protein denaturation and unfolding. *[1]*
- Monte Carlo method can give us a rough result quickly and help us understand this system.

- This project: simulate the process of protein folding and study the effect of temperature i.e. the denaturation process

Model

- Assume the protein is an amino acid chain consisting of **N** amino acids and located on a 2D grid. Each amino acid is on a grid point, forming a **self-avoiding walk** pattern.
- Neighboring amino acids are linked by covalent bonds.
- The total interaction energy of the protein:*[2]*

$$E = \sum_{<m,n>} J_{A(m),A(n)}$$

<m,n> means a pair of **non-covalent neighboring** amino acids. $A(m)$ represents the type of m-th amino acid. J is the interaction coefficient between two amino acids.

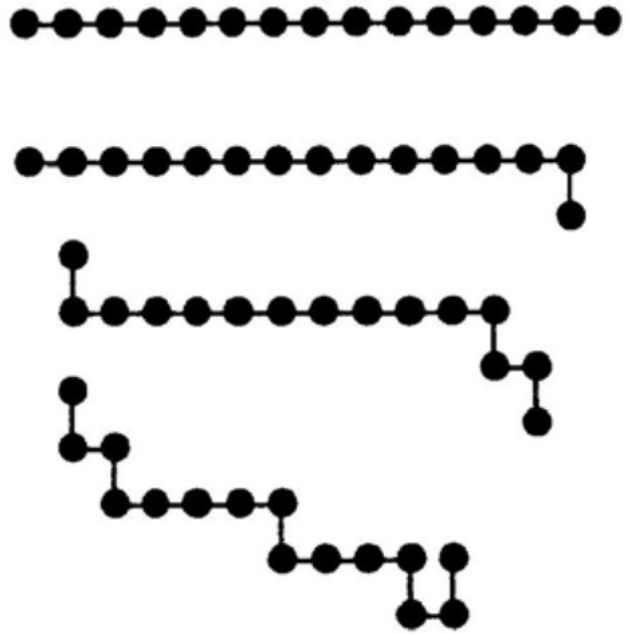


Figure 1: Four examples of possible states for an amino acid chain[2].

Algorithm

- Metropolis*[3][5]* and simulated annealing algorithm*[4]* is applied.
- In high T:
  1. Choose a node randomly: (x_{old}, y_{old})
  2. Choose one of its **next nearest neighbors** randomly: (x_{new}, y_{new})
  3. Test SAW.
  4. Calculate the change in energy ΔE after the movement.
  5. The acceptance probability is $e^{-\beta \Delta E}$, $\beta = \frac{1}{k_B T}$
- In low T: Set high T initially, then decrease T gradually

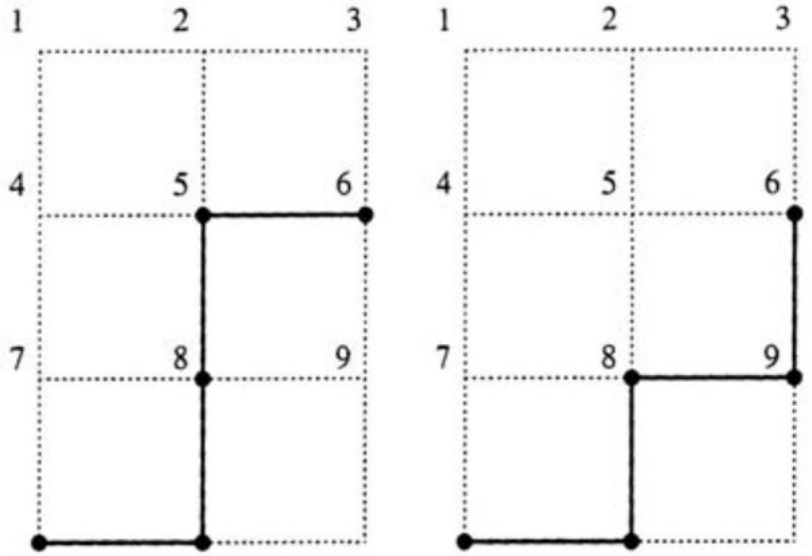


Figure 2: An example for an iteration[2].

Simulation Results: Equilibrium

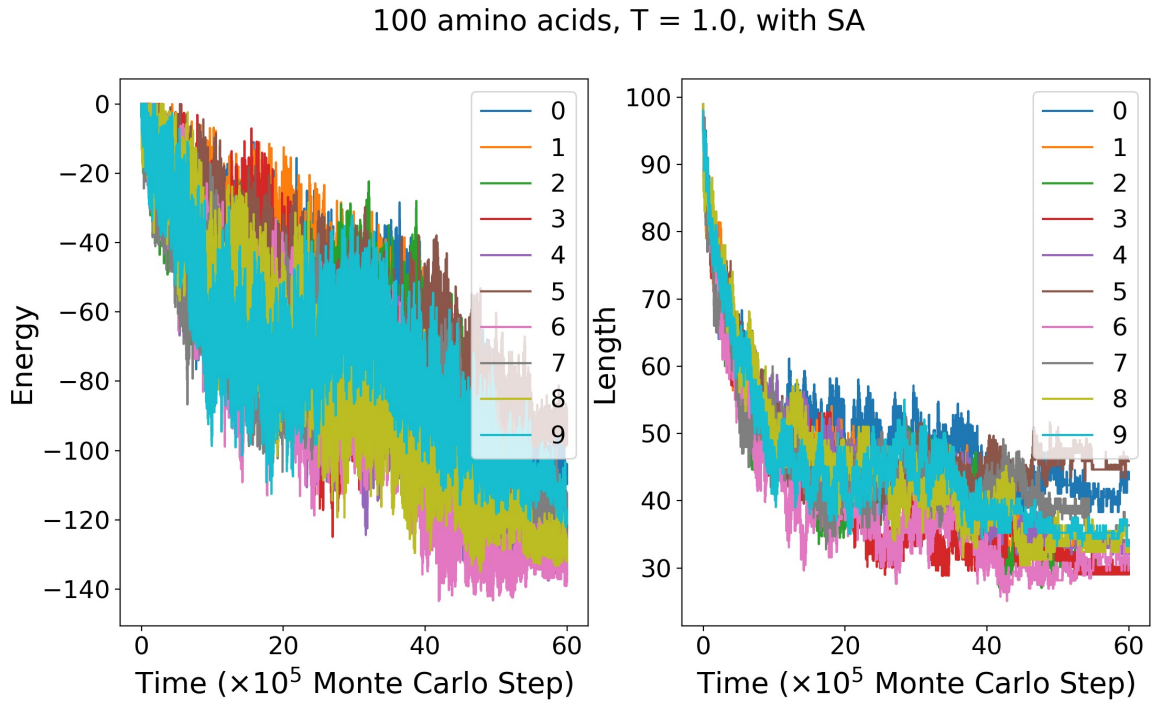


*Figure 3: Evolution of the energy and **end-to-end distance** (length) of the protein consisting of 100 amino acids when T = 1. Simulated annealing method is applied in this simulation.*

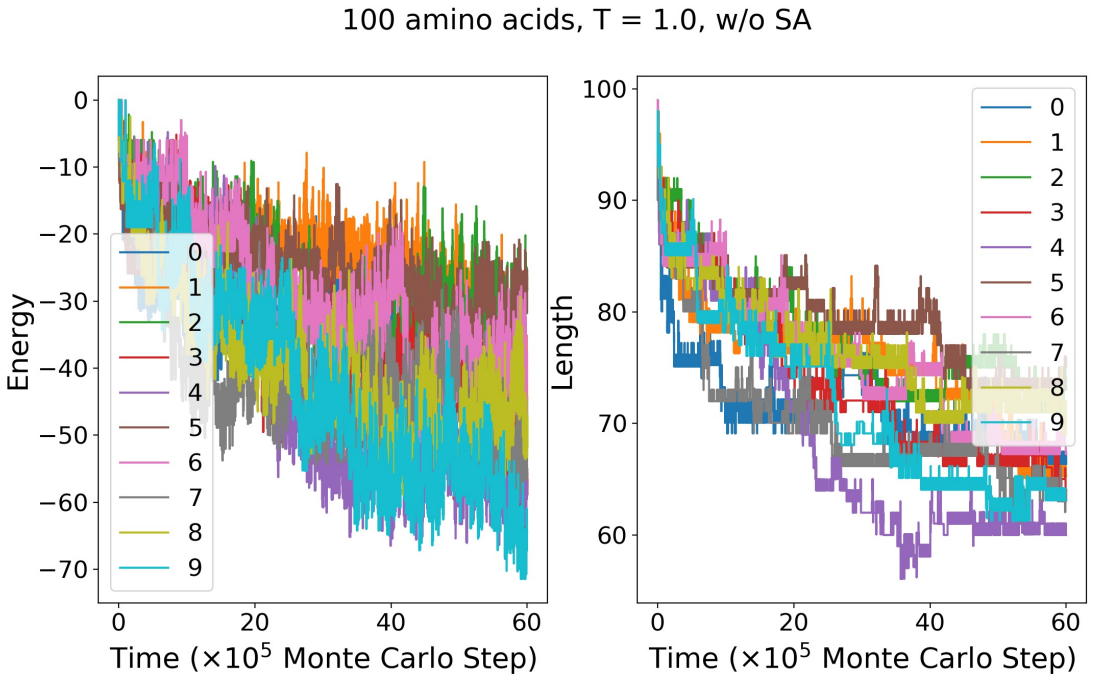


Figure 4: Evolution of the energy and end-to-end distance (length) of the protein consisting of 100 amino acids when T = 1.

- Simulated annealing method does work in low temperature.
- Simulating the system for many times is needed to improve statistical accuracy

Simulation Results: Effect of T & J

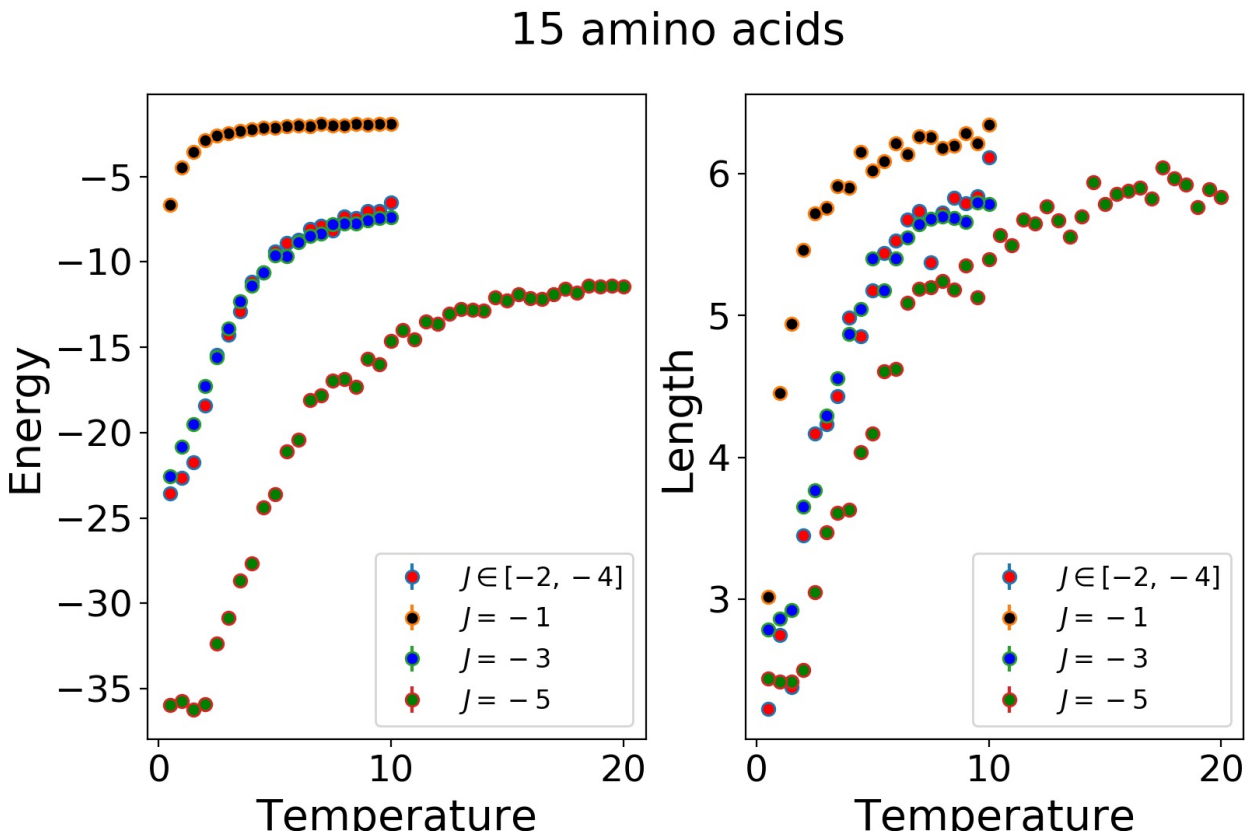


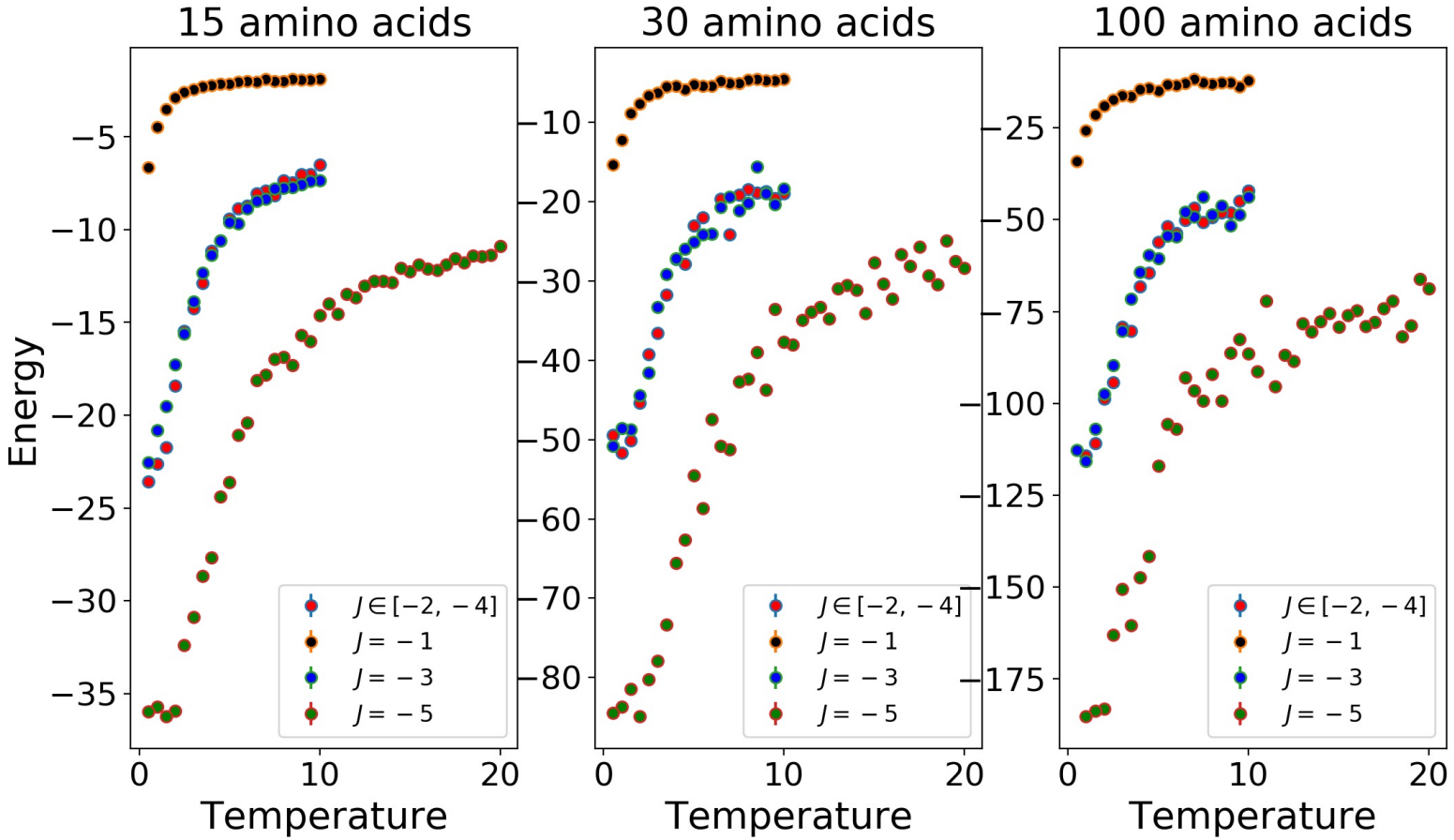
Figure 5: the relationship between temperature and the energy (left) and end-to-end distance (right) of the protein consisting of 15 amino acids with different interaction coefficient J.

- The role of J is to rescale the temperature: $e^{-\Delta E/k_B T} = e^{-J\Delta n/T}$

Temperature:

- In low T, the extent of folding tends to be high.
- There is a **phase transition** as raising temperature.
- When $T \uparrow$, the energy and length stop to increase and approach to **the certain value**.

Simulation Results: Effect of N

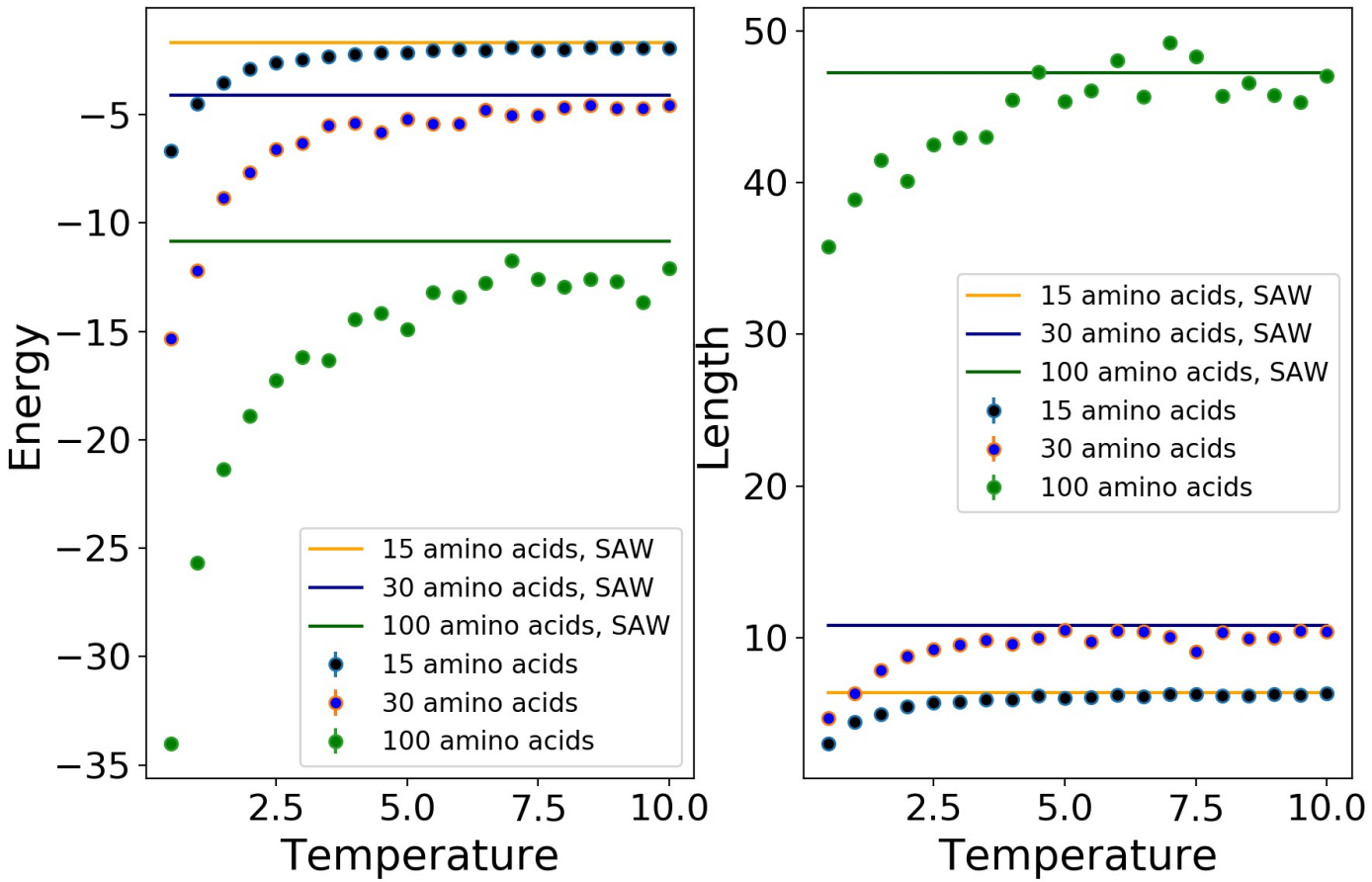


- Rules discussed before are **general**
- N **doesn't change** the transition temperature.

Figure 6: The relationship between temperature and the energy of the protein consisting of 15, 30, 100 amino acids with different interaction coefficient J.

The Certain Value: Self-Avoiding Random Walk

$J = -1$



- This rule is general and applies to all chains with different number of amino acids.
- The effect of thermal fluctuation dominates: $e^{-\Delta E/k_B T} \rightarrow 1$ (large T)

Figure 7: The comparison of SAW simulation and protein folding simulation for chains with 15, 30, 100 amino acids. The interaction coefficient J is -1.

Phase Transition

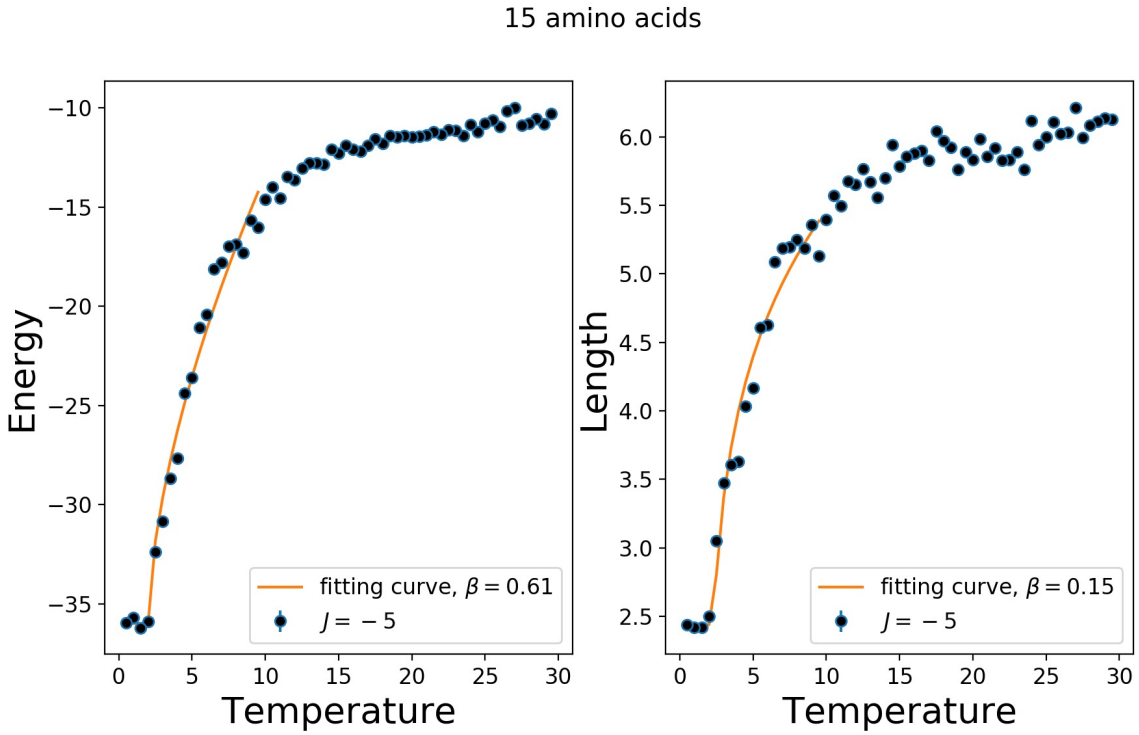


Figure 8: The fitting results of critical exponents for chains with 15 amino acids. The interaction coefficient J is -5.

Fitting function: $y = a(\frac{T}{T_C} - 1)^{\beta}$

- **Critical exponent**: $k = \lim_{\tau \to 0} \frac{\log(|f(\tau|))}{\log(|\tau|)}$ *[6]*
- Reduced temperature: $\tau = \frac{T - T_C}{T_C}$
- This results in the power law: $f(\tau) \propto \tau^k$

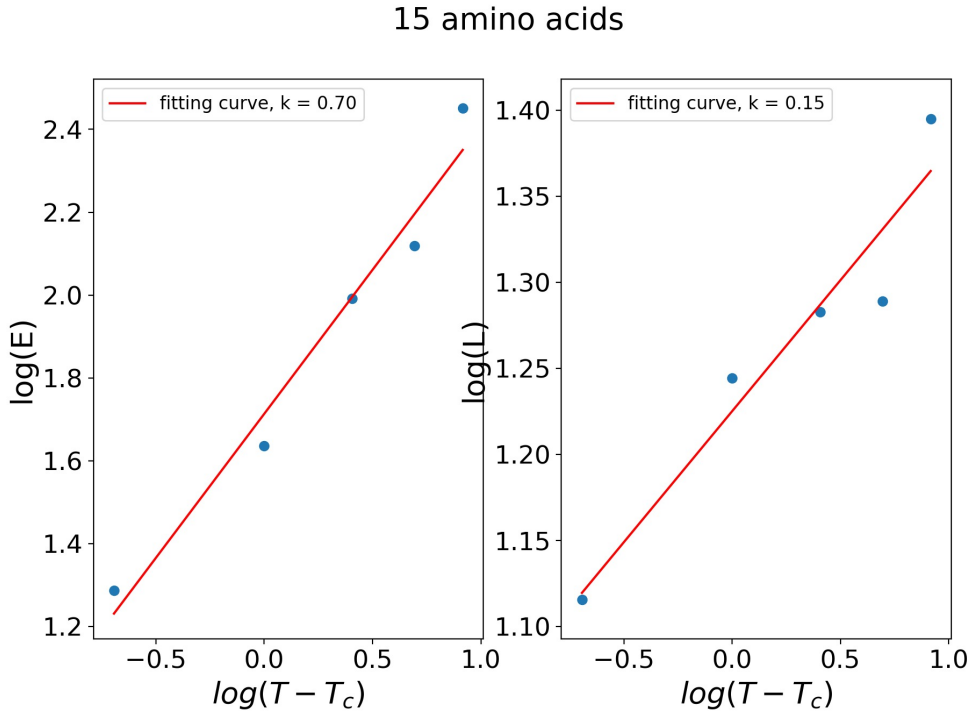


Figure 9: The linear fitting results of ln(E), ln(L) vs ln(T-Tc). (χ^2 plot shows that Tc=2.)

Conclusion & Discussion

1) The extent of folding tends to be high in low temperature.
2) There is a phase transition as gradually changing temperature in the protein system. The transition temperature is independent of N.
3) The energy and the end-to-end distance approaches to the results of self-avoiding random walk when temperature is high enough, which indicates that the effect of thermal fluctuation dominates in high temperature.

Improvements:

a) Increase the Monte Carlo step for N=100 case to get better curve.
b) Further study is needed to explain the phase transition phenomena.

References:

[1] Christian B. Anfinsen. The formation and stabilization of protein structure. *Biochemical Journal.,* 128(4):737–749, 1972.
[2] Zejun Ding. Lecture Notes for Computational Physics. *manuscript.*
[3] WK Hastings. Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–&, 1970.
[4] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
[5] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
[6] https://en.wikipedia.org/wiki/Critical_exponent

Thank You!