

Finding Protein Folding Funnels in Random Networks

Macoto Kikuchi (菊池誠)

Cybermedia Center, Osaka Univ. (大阪大學)

DDAP9 (17 Dec. 2016)

Outline

- 1 Introduction: Rareness, Funnel picture, Network representation
- 2 Model and Method
- 3 Results
 - Probability distribution of the ideal funnels
 - Network size dependence of the probability and S.D.
- 4 Summary and Discussions

Introduction: rareness

Proteins are not at all typical random polypeptide

- If we make a random polypeptide, its low-energy state will become glassy.
 - Many different conformations have low energy close to each other.
- A protein folds into its specific native conformation spontaneously under the physiological conditions.
 - Anfinsen's dogma: The native conformation is determined by the amino acid sequence.

Such a special property of proteins have been developed through Darwinian evolution.

Proteins are good examples of the fact that the evolution can create very rare states of matter.

Introduction: funnel picture

Funnel picture of the energy landscape has been accepted widely after 1990s as a mechanism behind the protein folding

- Minimum frustration principle (Bryngelson and Wolynes)
- Consistency principle (Go)

funnel picture

The number of conformations decreases monotonously as energy lowers and approaches to the native state.

Introduction: motivation of this study

We consider the following question.

Main question

How rare are the funnel-like energy landscapes?

To answer this question, at least partly, we introduce a simple and abstract model based on [the random energy model on random networks](#), which expresses the energy landscape of the proteins.

Introduction: network description

Network model has widely been used so far to understand the protein folding dynamics.

- Markov state model has been used to describe the relationship between many conformations obtained during the MD simulations.
- Hori *et al.* (PNAS 2009) tried to determine interconnection between all the conformations of some proteins including the conformations that do not appear in MD. And compare the obtained network with that of random polypeptide.

Model

conformation network

- 1 Give a random network, which represent the connections between conformations.
 - Node: metastable ensemble of conformations.
 - Edge: possible transitions between the nodes
 - We consider that the network structure is determined by the native conformations
 - 1 to 1 correspondence between the native conformation and the network structure

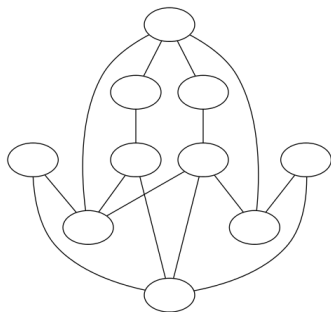
We assume any random network corresponds to some native state.

random order model

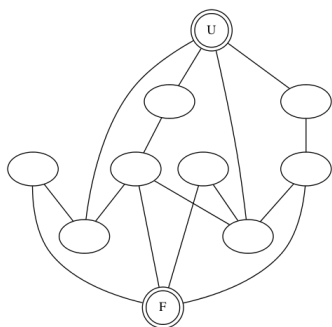
- 1 Each node is assigned an integer randomly.
 - Integer represents the energy of the node (Random Energy Model)
 - We need only the order of the nodes according to the energy.
 - We consider that the arrangement of the numbers is determined by the amino acid sequence

We assume all the arrangements of the numbers are possible.

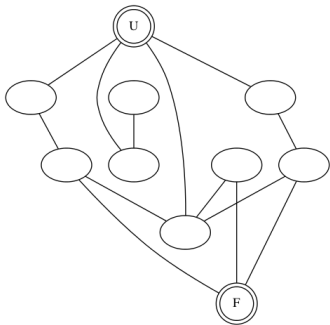
Construction of the model: 1



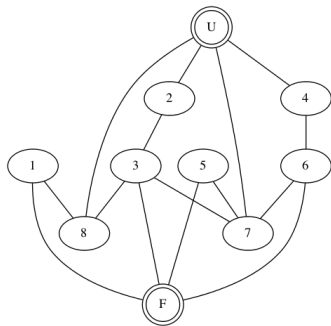
- Make a simple random graph
 - N : number of nodes
 - L : average number of edges connected to each node



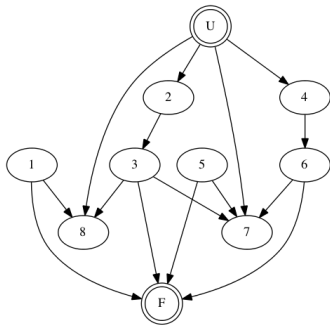
- Select one of the nodes having the largest number of edges as **U** (unfolded state)
- Select one of the farthest nodes from U as **F** (native folded state)



- The network is rejected if it is separated when U is deleted,



- Assign integers from 1 to $N - 2$ randomly to remaining nodes.
 - U and F are fixed to 0 and $N - 1$, respectively



- Draw arrows from the edge of smaller number to that of larger number, if two node are connected.

- The arrows represent the directions of transitions.
- We assume only the energy-lowering transitions

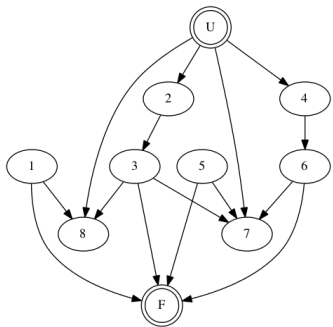
Ideal Funnel

We introduce a concept of **ideal funnel**

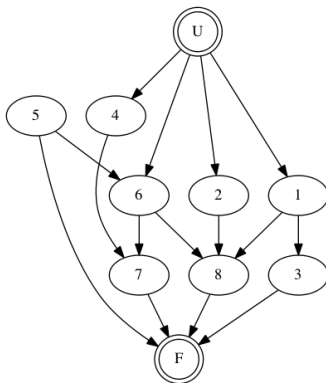
definition

Starting from U node, if all the energy-lowering path lead to F node, we call the network is **ideal funnel**

The energy-lowering transitions between nodes eventually leads to the native state without being trapped by misfolded state.



non-ideal funnel
(misfolded states 7 and
8 exist)



ideal funnel

Question again

In the language of proteins

Given one native state, how rare are the amino acid sequence that the energy landscape becomes the ideal funnel among all the possible sequences.

In terms of the model

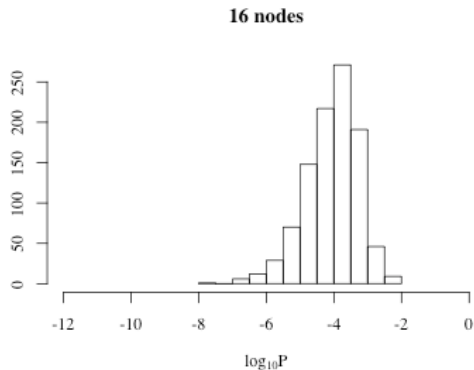
Given a random graph, estimate the appearance probability of the ideal funnel among all the possible arrangement of integers.

Method

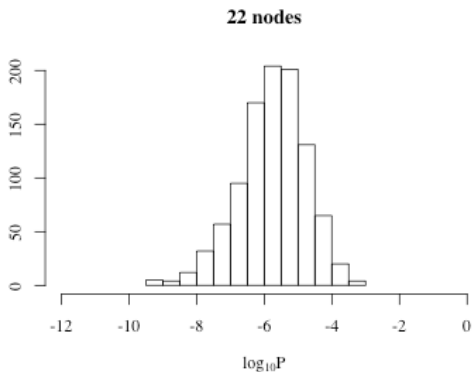
- $N = 8 \sim 27$
- $L = 3, 4, 5$ (We will present the result only for $L = 3$)
- Generate 1000 random graphs for each (N, L)
- Estimate the appearance probability of ideal funnel among all the possible arrangement of numbers for each graph.

- Exact enumeration for $N \leq 14$
- Multicanonical Monte Carlo method for $N > 14$
 - This method with Wang-Landau weight learning is powerful for counting the number of rare states of combinatorial problems (application to counting the number of the magic squares is: A. Kitajima and MK, PLOS One 2015) .

Results

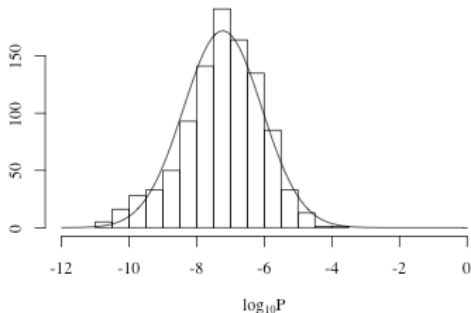


PDF of ideal funnels ($N = 16, L = 3$)



PDF of ideal funnels ($N = 22, L = 3$)

27 nodes

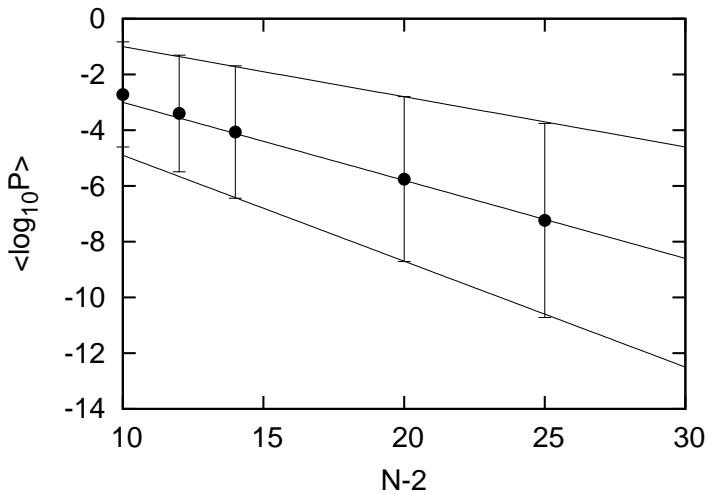


PDF of ideal funnels ($N = 27, L = 3$). The solid line represents the Log-Normal distribution

We expect that the appearance probability of the ideal funnels approaches the Log-Normal distribution for large N

Implication of Log-Normal

- Very small number of networks have a large probability
 - Since the different arrangement of the number corresponds to the different sequence, such networks are robust against mutation.
 - Possible explanation for the fact that there are only a relatively small number of **folds** for known proteins.



Ndependence of $\log P$ (bars: $3 \times \text{S.D.}$)

- P decreases exponentially with N (as expected)
- Number of robust networks decreases more slowly than typical networks
 - Evolutionally favorable

Summary and Discussion

- We estimated the rareness of the folding funnels using the random-energy model on the random network.
- PDF of appearance probability of the ideal funnel is close to Log-Normal type.
 - There are very small number of the native conformations that are robust against mutations.
- Typical networks decrease exponentially with N
 - The robust networks decrease also exponentially but more slowly.

Remark

The model is very simple, abstract, and rather arbitrary. So is far from reality. We, however, still consider that we can get some insight on rareness of the foldable proteins and its implication to evolutions from this study.