

Finding Protein Folding Funnels in Random Networks

Macoto Kikuchi
kikuchi@cmc.osaka-u.ac.jp

Cybermedia Center, Osaka Univ.

CCP2017 (11 Jul. 2017)
paper in preparation

Outline

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

Introduction: rareness

- A protein folds into its specific native conformation spontaneously under the physiological conditions.
 - Anfinsen's dogma: The native conformation is the thermodynamically stable one determined by the amino acid sequence.

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.



ex. The native conformation of lysozyme

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

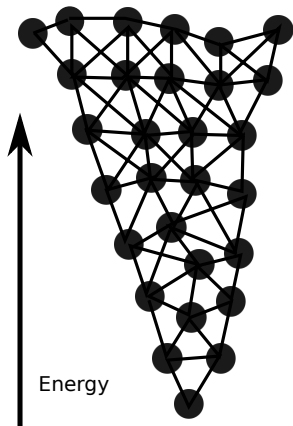
Proteins are good examples of the fact that the evolution can create very rare states of matter.

Introduction: funnel picture

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

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

The energy landscape is determined by the native structure.



funnel picture

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

It is well known that the GO-like model, in which the interaction between the residues are determined by the native structure, can nicely reproduce the experimentally observed folding processes,

- GO-like model is the simplest model that realizes the funnel-like energy landscape.

The reason is not yet so clear. But anyway the funnel picture well describes the protein folding.

Introduction: folds

Number of "folds" is very small compared to the number of proteins.

- Fold: skeleton of the native structures.

ex. classification by SCOP reported the number of folds is 1195 in the year 2009.

- <http://scop.mrc-lmb.cam.ac.uk>

Although the definition of the "fold" is still ambiguous, it is clear that many proteins have the same "fold".

- Why do the folds so scarce?

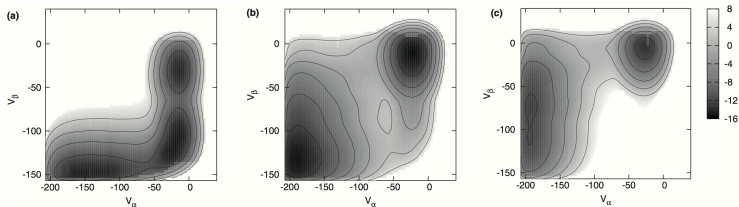
Introduction: variation of the funnel structure

Free-energy landscape can have a variety even within the framework of the GO-like model.

- Folding pathways are restricted by the native structure, but not uniquely determined.

ex. All the experimentally observed variety of the folding processes of lysozyme family can be reproduced by the extended GO-like model, in which the relative interaction strength of two domains are tuned as a parameter.

- Kanzaki and MK, Chem. Phys. Lett. 427



Variety of the free-energy landscape at the folding temperature for lysozyme by the extended GO-like model

- Kanzaki and MK, Chem. Phys. Lett. 427 (2006) 414.

Three points to be considered

Proteins are rare states of polypeptides.

Folds are very scarce.

Free-energy landscape and folding pathway can have a variety within the funnel picture.

Motivation of the 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.

Structural network

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

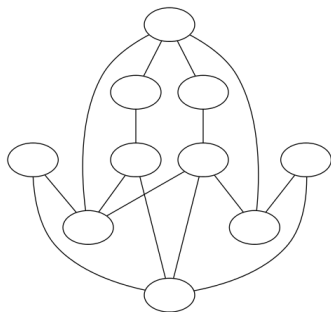
We assume any random network corresponds to some native state.

random order model

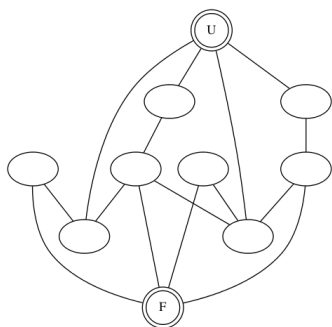
- 1 Each node is assigned an integer randomly.
 - Integer represents the energy of the node (Random Energy Model)
 - We need only the order of the nodes according to the energy.
 - We consider that the arrangement of the numbers is determined by the amino acid sequence

We assume all the arrangements of the numbers are possible.

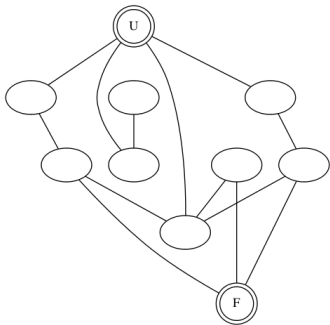
Construction of the model 1



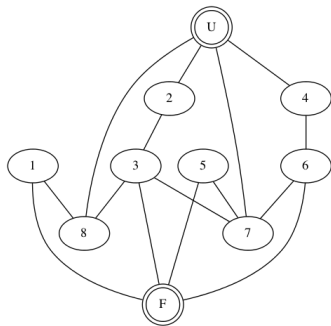
- Make a simple random graph
 - N : number of nodes
 - L : average number of edges connected to each node



- 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)

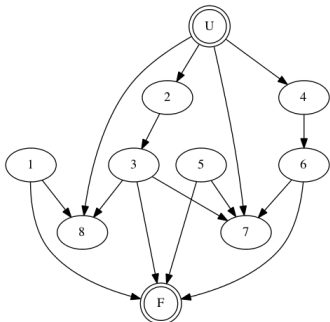


- The network is rejected if it is separated when U is deleted,



- 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 node of smaller number to that of larger number, if two nodes are connected (DAG).



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

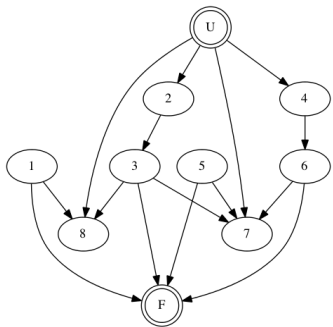
Ideal Funnel

We introduce a concept of **ideal funnel**

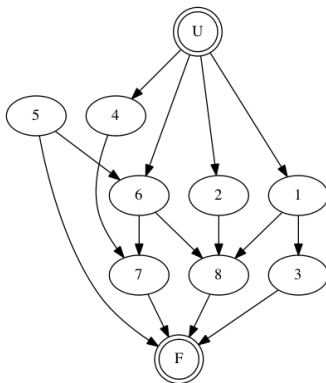
definition

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

Rare event sampling

We estimate the appearance probability of ideal funnel among all the possible arrangement of numbers using the **Multicanonical Ensemble** Monte Carlo method with parameters determined by **Wang-Landau** procedure.

The smallest number of the dead-end node that can be reached from the unfolded state is used as **energy** of the arrangement for Multicanonical method.

Numerous but Rare: An Exploration of Magic Squares

Akimasa Kitajima^{1,2*}, Macoto Kikuchi^{3,2}

How rare are magic squares? So far, the exact number of magic squares of order n is only known for $n \leq 5$. For larger squares, we need statistical approaches for estimating the number. For this purpose, we formulated the problem as a combinatorial optimization problem and applied the Multicanonical Monte Carlo method (MMC), which has been developed in the field of computational statistical physics. Among all the possible arrangements of the numbers $1; 2, \dots, n^2$ in an $n \times n$ square, the probability of finding a magic square decreases faster than the exponential of n . We estimated the number of magic squares for $n \leq 30$. The number of magic squares for $n = 30$ was estimated to be $6.56(29) \times 10^{2056}$ and the corresponding probability is as small as 10^{-212} . Thus the MMC is effective for counting very rare configurations.

We estimated the appearance probability of large magic squares. (PlosONE 10(5) e0125062)

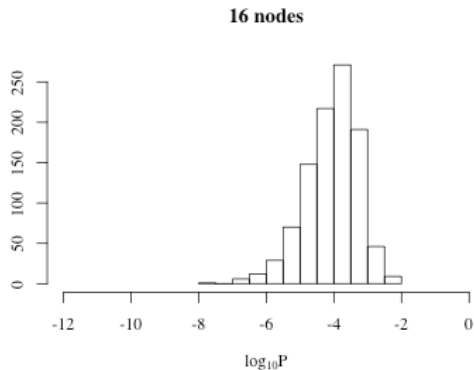
Multicanonical ensemble method with the Wang-Landau learning is highly suitable for estimating the appearance probability of very rare conformations

Since the total number of conformations is known for the present mode, we can estimate "absolute" appearance probability of the ideal funnels.

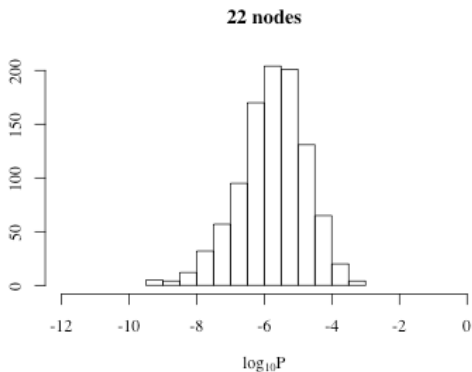
Computation Detail

- $N = 8 \sim 27$
 - Exact enumeration for $N \leq 14$. Multicanonical for $N > 14$.
- $L = 3$
- Generate 1000 random graphs for each N
- Estimate the appearance probability of ideal funnel for each graph.

Results

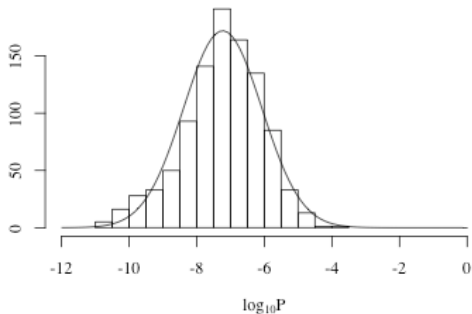


PDF of ideal funnels ($N = 16, L = 3$)



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

27 nodes

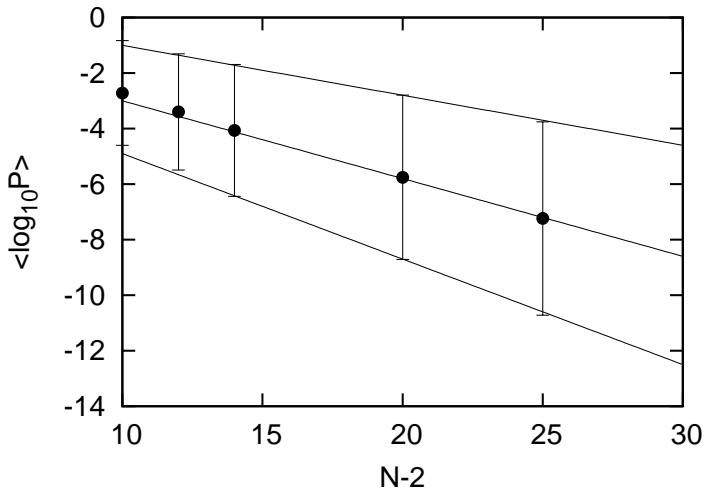


PDF of ideal funnels ($N = 27, L = 3$). The solid line represents the Log-Normal distribution

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.
 - The robust networks decrease also exponentially but more slowly.
- Multicanonical and Wang-Landau method is suitable for estimating the probability of rare conformations.

Remark

- The model is very simple, abstract, and rather arbitrary.
 - We did not consider special features of structural graphs, such as small-world, scale-free, hub structure etc.
 - The present study, however, will serve as a start point to consider rareness of the foldable proteins and its implication to evolutions from this study.
- Study of similar direction is on the way.
 - Gene regulation network, Genetic code etc.
- **Rareness** will be an important keyword in considering **life** related phenomena in the field of Biophysics.

- The paper is in preparation.
- Contact kikuchi@cmc.osaka-u.ac.jp if you're interested in the rareness of life-related phenomena.