Study on Properties of Protein Folding Conformation Network

Wang Nengchao
Parallel Computing Institute
Huazhong University of Science and Technology
Wuhan 430071, P.R.China

Wang Zhongjun
Department of Statistics
Wuhan University of Technology
Wuhan 430063, P.R.China

Abstract - *Protein folding conformations take the form of networks, sets of conformations (vertices) joined together in pairs by links or edges. Protein folding conformation network is a complex network, and has scale-free property that many networks have in common. Understanding the scale-free property in the network of protein folding conformation, one can find the better conformation of protein folding which closes to nature. This paper constructs a complex network of protein folding conformation and analyzes its small world and scale-free properties in the network of protein folding conformation in order to study evolving process of protein structure and predict protein structure.*

Keywords: complex networks; scale-free property; protein folding; conformation network.

1 Introduction

Proteins are of greatly importance in molecular biology. To understand the structure of protein molecules and how the specific functions are formed is one of the hot topics in protein researches.

Forming protein structure is a complex evolutionary process. Understanding the evolution of protein structure is useful for prediction of protein structure. The complex network is an effective tool for simulating a complex biological system[1]. Complex networks' evolutionary technology is conveniently model for the evolving process, and helps for analyzing protein evolution and predicting protein structure [2-3].

Protein folding conformations take the form of networks, sets of conformations or vertices joined together in pairs by links or edges. Protein folding conformation network is a complex network, and has scale-free property [4] that many networks have in common. The degree of a vertex in the network is the number of other vertices to which it is connected, and one finds that there are typically many vertices in the network with low degree and a small number with high degree, the precise distribution often following a power-law form.

If understanding the scale-free property in the network of protein folding conformation, one can find the better conformation of protein folding which closes to nature. This paper constructs a complex network of protein folding conformation and analyzes its properties in the network of protein folding conformation in order to predict protein structure.

2 Building network of protein folding conformations

When protein folding conformation space is very large, if it folds through all conformation, it will need long time, but in fact, protein can finish folding less than 1 second. So ones can guess that protein have an optimal folding path that is shortest.

We wish finding the best conformation in the protein folding conformation space using scale-free property of complex networks. This is because a small number of vertices with high degree may be close to natural conformation.

Folding conformations of a protein take the form of networks. The network is conveniently modeled as a graph $G=\{V,E\}$ which consists of a set V of vertices and a set E of edges which we regard as un-ordered pairs of distinct vertices. Hence, we consider only simple undirected graphs in the network of protein folding conformation. A vertex in G is a conformation, an edge in G is a link that the energy and shape relationship between conformations, where the $e_i=\{x_{i-1},x_i\}$ are the edges connecting vertices, and the x_i are vertices, $i=1,\cdots,n$. The set of neighbors of x is denoted by

$$\partial\{x\}=\{y\in V \mid \{x,y\}\in E\}$$

The degree of a vertex x is the number of edges that contain x; i.e. the number of neighbors of x:

$$\deg(x)=|\{e\in E \mid x\in e\}|=|\{y\in V \mid \{x,y\}\in E\}|=|\partial\{x\}|$$

Equivalently, we may define $\deg(x)$ as the number of edges incident with x.

The potential function which evaluates the potential energy of protein folding conformation is as follows [5]:

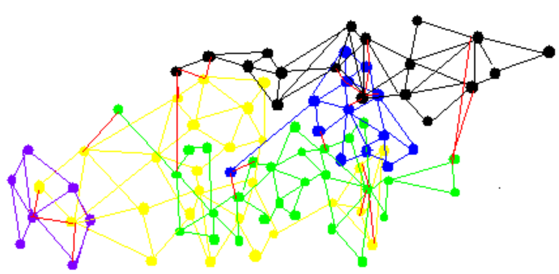
$$E = \sum k_r (r - r_{eq})^2 + \sum k_\theta (\theta - \theta_{eq}) + \sum \frac{V_n}{2}[1 + \cos(n\varphi - r)] + \sum_{i<j}[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}] + \sum[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}]$$

where K_r, K_θ, K_{ij}, B_{ij}, C_{ij}, D_{ij} are force field parameters.

The shape of protein folding conformation which builds the complex network include five shapes, such as **Line type**, **T type**, **Rectangle type**, **Square type** and **Triangle type**. Therefore, we consider two parts of potential energy and conformation shape to make sure whether or not link between two vertices.

In the Figure 1, the case of network is built by 92 vertices, which are conformations of amino acid sequence HPPHHPPPHHPPHP.

The different color vertex stands for different type, namely purple is Line type, yellow T type, green Rectangle type, blue Square type and black Triangle type.



3 General properties on the conforma -tion network

In order to understand protein folding conformation networks, we first analyze network topological properties. Network topological properties include degree distribution, clustering, shortest path, betweenness, and spectrum.

3.1 Network topological properties

Degree is the number of edges that a vertex has, and corresponds to the local centrality in social network analysis, and measures how important are a vertex with respect to its nearest neighbours.

In the network of protein folding conformations, every vertex's degree K_i is different. The mean degree of the protein folding conformation network is as follow:

$$< K > = \frac{\sum_{i=1}^{N} K_i}{N}$$

where degree of vertex i is K_i. In the case, $< K > = 3.86$.

The spread in the vertex degree is characterized by a distribution function $P(k)$, which gives the probability that a randomly selected vertex has exactly k edges. In a random graph of the type studied by Erdos and Renyi, each edge is present or absent with equal probability, and hence the degree distribution is Binomial or Poisson in the limit of large graph size. Real-world networks are mostly found to be very unlike the random graph in their degree distributions. The degrees distribution has a power-law tail:

$$P(k) \sim k^{-r}$$

In the case, $P(k) \sim k^{-2.1}$.

Clustering is a common property of social networks, represents that cliques form. The inherent tendency to cluster is quantified by the clustering coefficient. Clustering coefficient of a vertex is as follow:

$$C_i = \frac{E_i}{k_i(k_i - 1)/2}$$

Clustering coefficient C of the whole network is the average of all individual C_i:

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i$$

In a random graph, since the edges are distributed randomly, the clustering coefficient is $C = p$. However, in most, if not all, real networks the clustering coefficient is typically much larger than it is in a comparable random network. We calculate the clustering coefficient as fellow:

$$C = \frac{3 \times number\ of\ triangles\ on\ the\ graph}{number\ of\ connected\ triples\ of\ vertices}$$

In the case, the clustering coefficient $C = 0.683$.

The shortest (i.e. geodesic) path length is the number of edges that make up the path between two points. It can measure global centrality which points that are "close" to many other points in the network. Global centrality defined as the sum of minimum distances to any other point in the networks.

In the case, the average shortest distance of pair of vertices in the network is:

$$l = \frac{1}{\frac{1}{2}N(N+1)}\sum_{i \geq j} d_{ij} = 5.222$$

where d_{ij} is the shortest distance from vertex i to vertex j.

Between measures the "intermediary" role in the network. It is a set of matrices, B_{ij}^{k} of each vertex is a ratio of shortest paths between i and j that go through k.

There can be more than one geodesic between i and j. The vertex with the maximum betweenness plays a central role.

Spectrum of the adjacency matrix is a set of eigenvalues of the adjacency matrix. Spectral density (density of eigenvalues) is

$\rho(\lambda) = \frac{1}{N}\sum_{j=1}^{N}\delta(\lambda - \lambda_j)$ A symmetric and real network corresponds eigenvalues are real and the largest is not degenerate. The largest eigenvalue shows the density of links. The second largest is related to the conductance of the graph as a set of resistances.

3.2 Three main models

According to their topological properties, complex networks can be divided three main classes of modeling. First, random graphs, which are variants of theErdos-Renyi model, are still widely used in many fields and serve as a benchmark for many modeling and empirical studies. Second, motivated by clustering, a class of models, collectively called small-world models, has been proposed. These models interpolate between the highly clustered regular lattices and random graphs. Finally, the discovery of the power-law degree distribution has led to the construction of various scale-free models that, by focusing on the network dynamics, aim to offer a universal theory of network evolution. The three main models correspond three different networks, i.e. random network, small-world network, and scale-free network.

In the protein folding conformation network of our case, there is smaller average shortest distance $l = 5.222$, and larger clustering coefficient $C = 0.683$ in the network. Therefore, the protein folding conformation network has small world property. The small world property means that the relation between vertices is closer and vertices related will be closer and form communities. And the degree distribution follows power law appreciatively: $P(K) \approx K^{-2.1}$. So the protein folding conformation network has free-scale property.

Consequently the protein folding conformation network we have constructed is a complex network, and has small world and free-scale properties. These properties are useful for advance studying protein structure.

During folding a protein takes up consecutive conformations. Representing with a vertex each distinct state, two conformations are linked if they can be obtained from each other by an elementary move. Scala, Amaral, and Barthelemy (2001) studied the network formed by the conformations of a two-dimensional lattice polymer, finding that it has small-world properties. In particular, evolving networks can be modeled through growing graphs to simulate protein evolution, i.e., graphs to which continuously new vertices and new links (edges) are added.[7]

4 Conclusions

A number of biological systems can be usefully represented as complex networks. Many biological networks, such as metabolic interaction and protein-protein interaction networks have been shown to be scale free, with different vertices having widely different connectivity. Here, the complex networks' evolutionary technology applies to the protein evolving process, and helps for analyzing protein evolution and predicting and restructuring protein structure.

5 Acknowledgements

The authors thank Prof. Yuan Bo, Prof. Di Zengru and Dr. Fan Ying for useful conversations. This work was funded in part by the National Natrual Science Foundation of China under grant number 70371063.

6 References

[1] Reka Albert and Albert-Laszlo Barabasi, Statistics Mechanics of Complex Networks, Reviews of modern Physics, volume 74, January 2002, P47-97.

[2] Wang Zhongjun, Nengchao Wang, and Xingqin Cao, Research on Complex Biological Systems with Automata Network Based on the Evolution Technology of Bisection, DCABES2002, P303-305.

[3] Wang Zhongjun, Nengchao Wang, Protein Evolution Based on Complex Networks, Proceedings of DCABES 2004, Conference, 2004，P311.

[4] Michelle Girvan, M. E. J. Neewman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA99, 7821-7826(2002).

[5] Eric Alm and Adam P Arkin, Biological Networks, current Opinion in Structural Biology 2003,13,P193-202.

[6] Söding J, Lupas AN. (2003). More than the sum of their parts: On the evolution of proteins from peptides. Bioessays 25:837-846.

[7] J.Jost and M.P.Joy, Evolving networks with distance preferences, PHYSICAL REVIEW, E66, 036126 (2002)

[8] Yan Li, Scale-free Networks in Biological Complexity, Phycs498BIO,Assignment 4

[9] Eli Eisenberg and Erez Y. Levanon, Preferential Attachment in the Protein Network Evolution, Physical Review Letters, volume 91,number 13, P138701-4