The Six Memorial Workshop for Kazuhide Mori on Computational Sciences

An Application of Graphs to Protein Structures

Yoshinori NAGAI[1], Stephen HYDE[2], and Hiroshi WAKO[3]

We have studied the protein structures in order to search for structural signals beyond the usual measures, such as α-helices and β-sheets, well characterised as secondary structures in proteins. We calculated the number of paths from α-carbon of the i-th amino acid to α-carbon in the j-th amino acid on the amino acid sequence or a protein, as a function of a threshold ''cut distance''. We link i-j α-carbons with an edge if their distance is less than this threshold. This approach affords a graph description of proteins. In graph terminology, each ''node'' is an α-carbon, and each edge is a connection from α-carbon to α-carbon. Our treatment of protein graph can be described a scheme shown in Fig. 1. Note that our graph construction analyses only the α-carbon string.

We recognize that the graph structure is sensitive to this cut-distance, varying from a linear chain at small distance to a complete graph at large distances. The number of edges in a protein is the frequency of paths for the length of paths in a protein. The derivative of this frequency distribution as a function of distance exhibits a single well defined maximum in cases that we call ''structured proteins'', of the form shown below. We take the cut-distance corresponding to this maximum as the protein diameter, and use this distance to construct the structural graph of the protein. This measure of protein size is typically close to its usual diameter. Note that ''structureless'' pro-

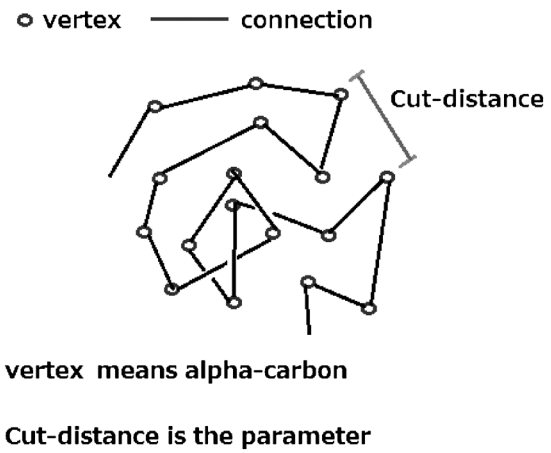

Fig. 1 An image of a protein string

[1] Faculty of Business, Kokushikan University, 3–4–28 Setagaya, Tokyo 154–8515, Japan
[2] School of Applied Mathematics, Research School of Physics and Engineering, Australian National University, Canberra, ACT 0200, Australia.
[3] School of Social Science, Waseda University, Nishi-Waseda 1–6–1, Shinjuku, Tokyo 169–8050, Japan.

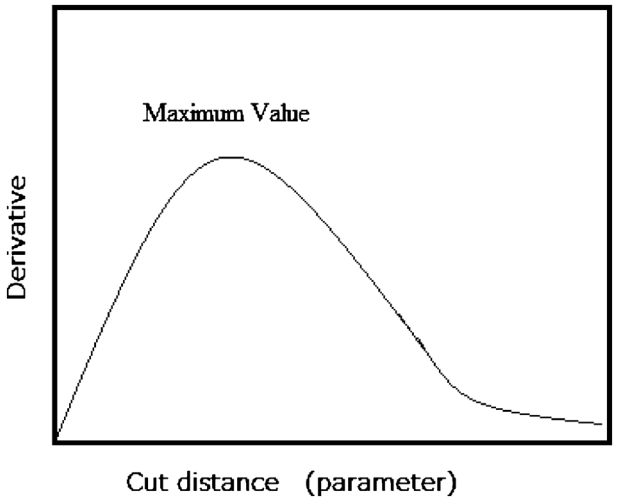

Fig. 2 Schematic shape of distribution

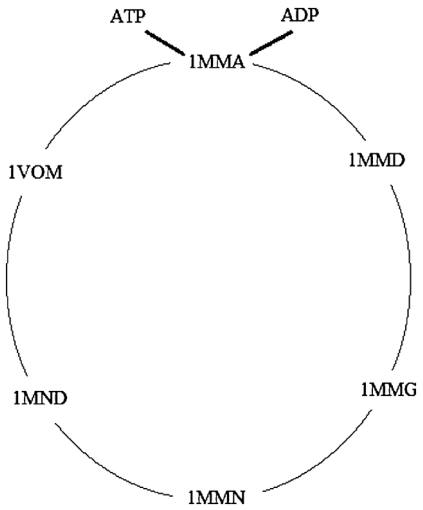

Fig. 3 ATP hydrolysis scheme using crystallized PDB data
1MMA is a rigol state of myosin that crystallized with ADP. Other PDB data (1MMD, 1MMG, 1MMN, 1MND, 1VOM) are crystallized with different ATP analogs.

teins typically exhibit no such single maximum, in common with random Gaussian chains.

In fact, the cut-distance vs. degree is the same as considering a ball of radius given by the cut-distance located sequentially on each α-carbon along the chain. The degree is simply the number of other α-carbons found in that ball. Choosing the ball radius equal to the diameter as defined in the previous paragraph gives the most sensitive measure of protein interior and surface, and local clustering - analogous to more conventional measures of secondary structure.

People gathered at this sixth memorial workshop of Kazuhide Mori have an interest in the heart: an organ consisting of heart muscles like the skeletal muscle. We therefore apply our graph

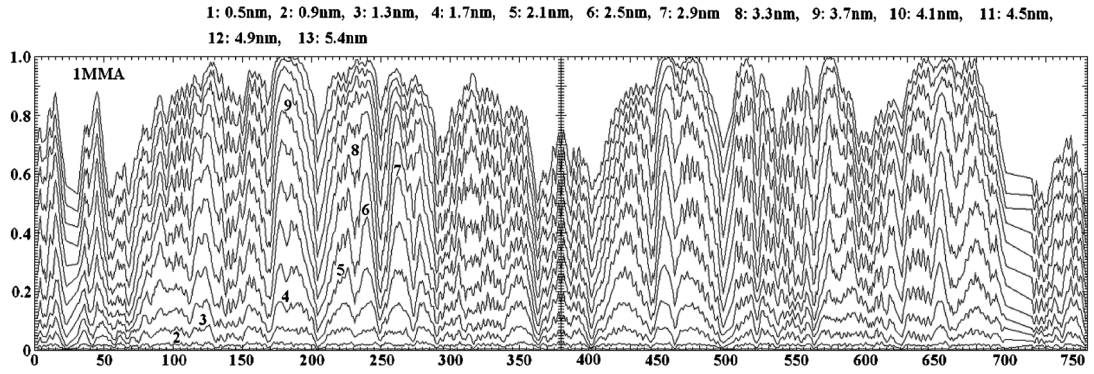

Fig. 4 Normalised graph degree along the protein α-carbon chain in 1MMA. Stacked plots are for increasing cut-distance.

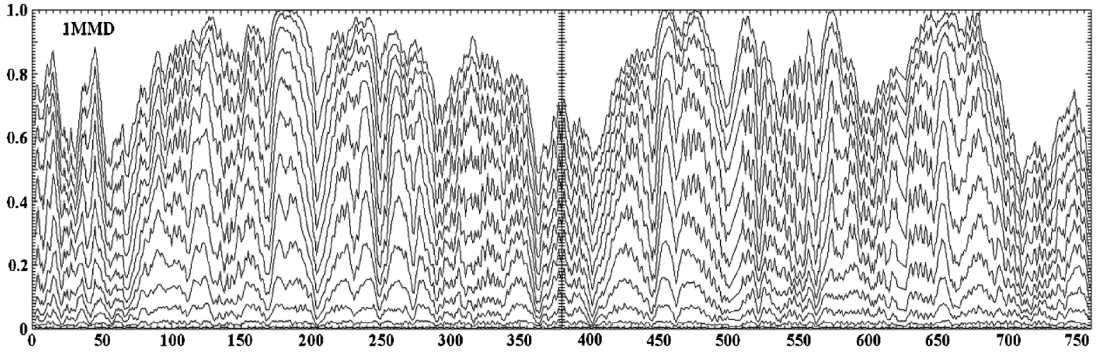

Fig. 5 Normalised graph degree along the protein α-carbon chain in 1MMD. Stacked plots are for increasing cut-distance.

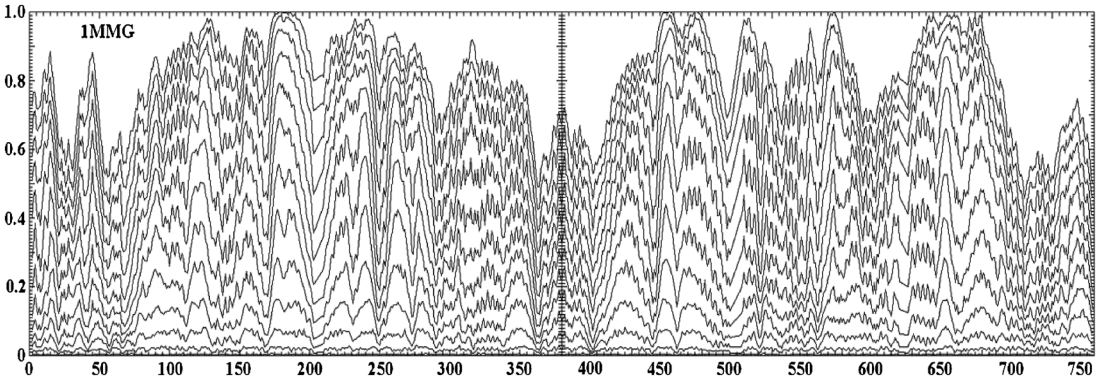

Fig. 6 Normalised graph degree along the protein α-carbon chain in 1MMG. Stacked plots are for increasing cut-distance.

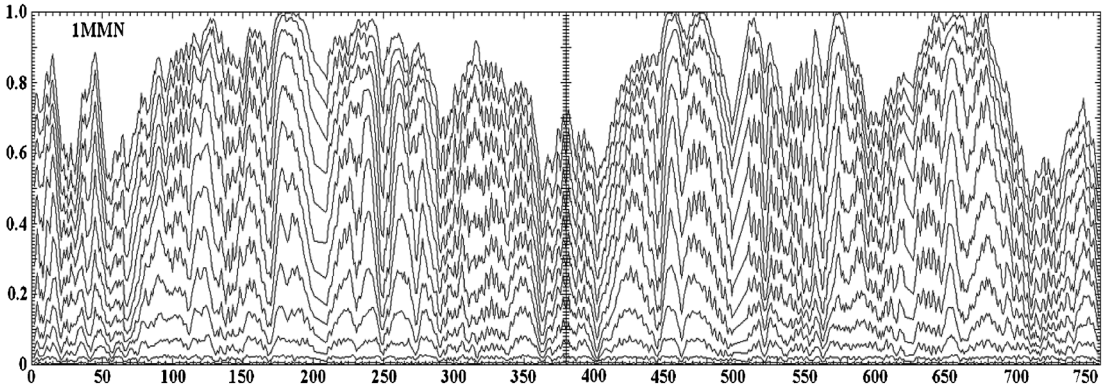

Fig. 7 Normalised graph degree along the protein α-carbon chain in 1MMN. Stacked plots are for increasing cut-distance.

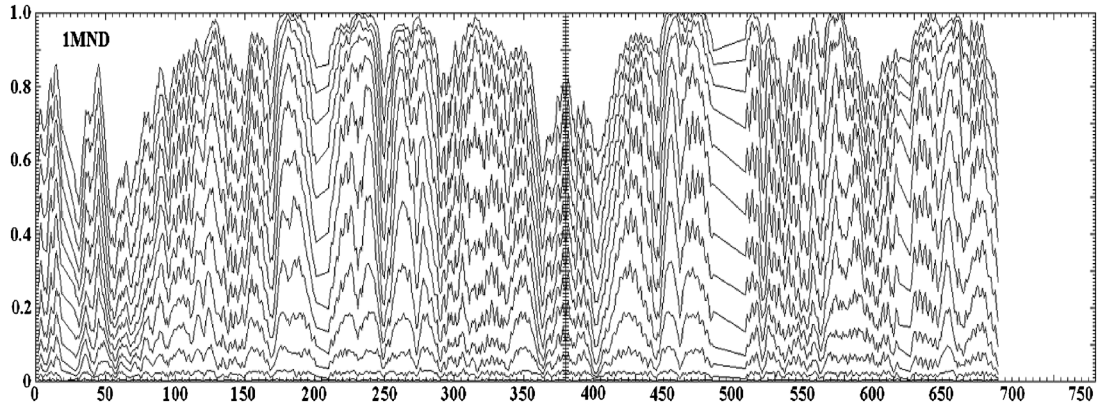

Fig. 8 Normalised graph degree along the protein α-carbon chain in MND. Stacked plots are for increasing cut-distance.

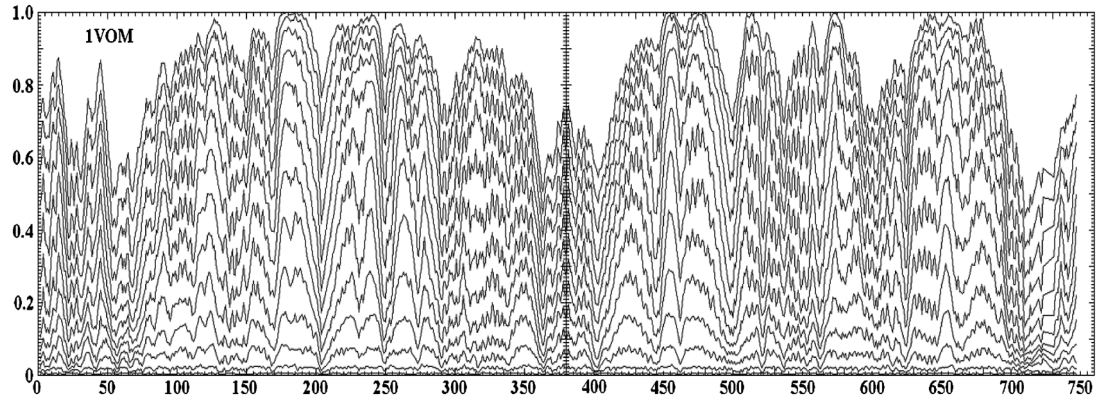

Fig. 9 Normalised graph degree along the protein α-carbon chain in 1VOM. Stacked plots are for increasing cut-distance.

analysis to explore a major component of muscle, myosin. The ATP hydrolysis cycle and its effects on protein structure can be interpreted using the PDB protein structural data according to the schema shown in Fig. 3.

We investigated the PDB data that gives data for different crystallized ATP analogs, arriving at the ATP hydrolysis scheme of Fig. 3. Careful exploration of PDB data led us to conclude that the six data are enough to describe the ATP hydrolysis cycle.

Fig. 4 to Fig. 9 show that the plots of (y-axis) number of graph edges incident to the (x-axis) i-th α-carbon (or graph degree, normalised by the total number of α-carbons) for cut-distances 0.5 nm, 0.9 nm, 1.3 nm, 1.7 nm, 2.1 nm, 2.5 nm, 2.9 nm, 3.3 nm, 3.7 nm, 4.1 nm, 4.5 nm, 4.5 nm, 4.9 nm, and 5.4 nm. Note that as the cut-distance becomes large, the normalized degree becomes unity. The graphs reveal local maxima interrupted by featureless flat regions. The maxima correlate well with regions of defined secondary structure. We interpret flatter featureless regions as portions of the protein chain that are structureless, marking ill-condensed areas of myosin with ATP analogs. For example, in 1MMA, the alpha-carbons of the range near the 700 to 725 are non-crystallized. In the figures the values appeared in horizontal axis are residue numbers.

The PDB data for 1MND do not give positions of alpha-carbons from 490 to 515 and those over the 690. Likewise, 1VOM PDB data lacks positions of alpha-carbons near 750. We ascribe this lack of data in the crystals to the likely rapid motion of these regions, preventing crystallization. We conclude that ATP hydrolysis is occurred using these areas.

More recently, our interest has shifted to the dynein molecule. We have applied our graph concepts to the dynein molecule. PDB data for dynein has two chains, chain A and chain B. Each dynein unit consists of six or seven units, though we have no idea of the meaning of these units. We

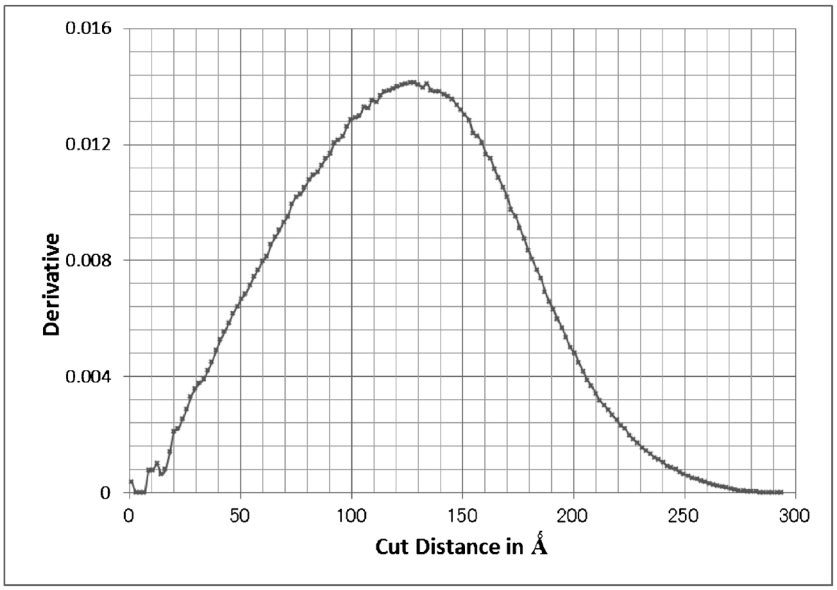

Fig. 10 Derivative of normalised number of alpha-carbons in the ball.
The radius of ball is given by the maximum value of horizontal axis.
The figure is drawn using PDB data 4AKG chain A.

concentrate on chain A of 4AKG (PDB id). The resulting derivative of normalised degree vs. cut-distance are shown in Fig. 10. (Fig. 10 is conceptually equal to Fig. 1.) From Fig. 10, the diameter of chain A of dynein according to our new method is near 130Å. The structure of one-side chain of dynein molecule is like a computer disc, and our diameter measures the diameter of that disc. As anyone knows, a dynein is a big protein.

Reference

S. Hyde, Y. Nagai, ''Folding of Proteins and Neighbourhood Clustering'', in preparation.