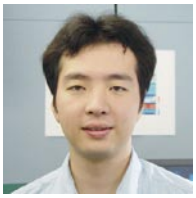
Study of Protein Segment Structure and Implications for Diversity of Protein Folds



Kazuyoshi Ikeda
AIST Research Staff
Molecular Modeling & Design Team

The variety of protein structural patterns (folds) is very large, said to include approximately 1,000 types. The component substructures of the fold (segment conformations), however, are known to share similar structures among the entirely different folds (Fig. 1). That is, protein fold, despite its rich diversity, is made up of combinations of a limited number of segment conformations. A recent study[*1] revealed that the newly found folds (novel folds) mostly consists of the known segment conformations with 10 residues or less. These findings indicate that the segment structure information is important for predicting protein structure.

In this context, our research has aimed at elucidating the structural properties of segments embedded in natural proteins. In general, the tertiary structure of polypeptide is diverse, even at the short segment level (less than 10 residues), requiring a multi-dimensional space to represent it. We use the PCA method, a multivariate analysis technique, to project segments cut out from a diverse set of folds onto a lower-dimensional space, enabling detailed analysis of those distributions (Fig. 2). We have so far constructed conformational spaces of segments with 20 residues or less, and identified the essential structural variables for short segments by interpreting the meaning of conformational axes. We have also identified a possible novel motif suggesting being participated in protein-protein interactions, by applying this method to classification of structural motifs.[*2]

We currently attempt to apply the method to the conformational spaces of longer segments (up to 50 residues). The results of this analysis indicate that the statistical characteristics of conformational distribution vary significantly depending on the length of the segment. We could also identify the structural variables of the segments conserved among various proteins, by comparing conformational spaces of the protein structure classes (All-α, All-β, α/β, α+β) which are defined by amounts and combinations of secondary structures. Furthermore, the analysis has shown that there are several conformational axes particular to each structure class, suggesting these axes are helpful for understanding diversity of protein folds.

The results of our study hold promise for such applications in structural prediction as classifying various outputs generated by powerful sampling method, such as FA (fragment assembly) method, and filtering native-like models. We expect this ongoing study to reveal important clues for better understanding the diversity of protein structures.

References

*1 Du, P., Andrec, M. and Levy, RM.: "Have we seen all structures corresponding to short protein fragments in the Protein Data Bank? An update", *Protein Eng.*, **16**, pp.407-414 (2003).

*2 Ikeda, K., Tomii, K., Yokomizo, T., Mitomo, D., Maruyama, K., Suzuki, S. and Higo, J.: "Visualization of conformational distribution of short to medium size segments in globular proteins and identification of local structural motifs", *Protein Sci.*, **14**, pp.1253-1265 (2005).

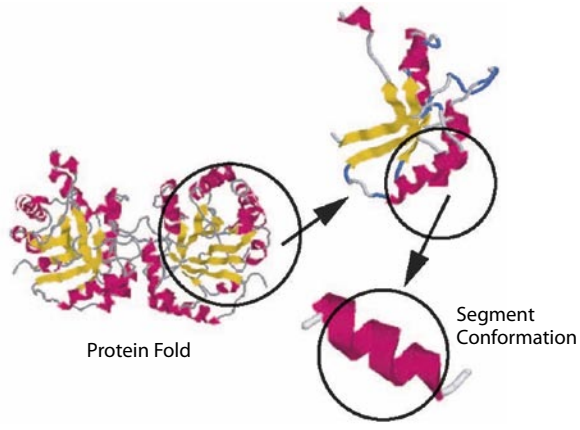


Fig.1 Protein Fold and Segment

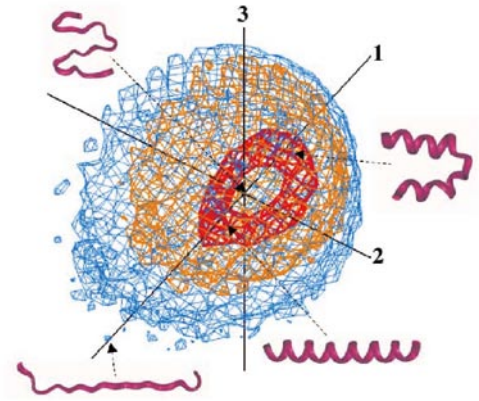


Fig.2 Segment Conformational Space generated by using PCA method