MODELING OF PROTEIN STRUCTURE AND PROTEIN FOLDING

ALEXANDER V. EFIMOV

Institute of Protein Research, Russian Academy of Sciences,

142290 Pushchino Moscow Region, Russia

E-mail: efimov@protres.ru

The amino acid sequence is the main determinant of the protein 3D structure. On the other hand, sequence of events in protein folding is very important determinant of unique protein tertiary structure. However, until now there are no experimental methods to observe protein folding in real time. The problem of searching of all possible protein folds is also important for protein 3D structure prediction. Theoretical modeling of protein folds and folding pathways based on construction and analysis of structural trees of proteins is a promising approach to solving the problems.

The approach is based on the hypothesis that at the first step of protein folding a nucleus is formed and then the remaining part of the molecule or domain is folded around it. In modeling, the structural motif having a unique overall fold and handedness is taken as the starting structure of the root structure of the tree. The larger protein structures are obtained by stepwise addition of α-helices and/or β-strands to the growing structure taking into account a restricted set of rules inferred from the known principles of protein structure. Among these rules, attention to handedness and compactness, prohibition of crossing connections, and a requirement to α-helices to be packed in α-helical layers and β-strands in β-layers are the most important. The number of allowed overall folds that can be obtained from one structural motif is limited since the rules drastically reduce the number of allowed pathways of growth of intermediate structures. A general scheme that represents the root structural motif, all the intermediate and completed structures connected by lines showing allowed pathways of structure growth, is referred to as the structural tree.

To date several structural trees for the largest protein superfamilies such as β-proteins containing abcd-units and 3β-corners, (α+β)-proteins containing abCd-units, α/β-proteins containing five-segment and seven-segment α/β-motifs and others have been constructed (see,e.g. [1-3]) . Some of them are available at the web-site (http://strees.protres.ru). In my opinion, the structural trees are a good tool for searching of all possible protein folds, for

modeling of folding pathways of proteins, for protein structure comparison, structural classification of proteins etc.

This work was supported by the Russian Foundation for Basic Research (Projects No. 10-04-00727 and 10-04-92804) and DST (India).

[1] A.V. Efimov. Structural trees for protein superfamilies. *Proteins*, vol. 28, pp. 241-260, 1997.
[2] A.V. Efimov. A structural tree for proteins containing S-like β-sheets. *FEBS Letters,* vol. 437, pp. 246-250, 1998.
[3] A.B Gordeev, A.M. Kargatov, and A.V. Efimov. PCBOST: Protein classification based on structural trees. *BBRC,* vol. 397, pp. 470-471, 2010.