Native and non-native oligopeptide fragments biased to alpha-helical formation

Gelena Kilosanidze[1], Alexey Kutsenko[2]

Keywords: protein secondary structure, structure prediction, molecular mechanics, energy contributions

1 Introduction.

Secondary structure prediction from protein sequences is the most studied problem in computational biology. Current matter of thinking implies that a complete solution to all aspects of the protein structure problem require that the information available from the analysis of large structural and sequence databases be combined with the physical-chemical principles of structure formation.
The approach of secondary structure prediction proposed by us [2,3] is *ab initio* method and uses a molecular mechanics for the analysis of standard conformations in proteins. A model for prediction of α-helical regions in amino acid sequences has been tested on the mainly-α protein structure class. The modeling represents the construction of a continuous hypothetical α-helical conformation for the whole protein chain, and was performed using molecular mechanics program ICM [1] that represents macromolecules in internal coordinate system. The results of the modeling clearly demonstrated that the profile of energy along the model α-helical protein reveals minima corresponding to real α-helical segments in the native protein.

2 Method and results.

In spite of good performance for α-helices the method is rather computer time consuming (about 6h for protein of 100 residues on a regular PC). To move to the fast statistical prediction based on molecular modelling we have undertaken an analysis of an energy distribution of all possible short oligopeptides, natural and non-natural, involved in α-helical pattern. For oligopeptide in length of 4 we modeled 20^4 = 160 000 structures. In order to reveal a contribution of every tetrapeptide to forming α-helix we constructed and optimized polyalanine α-helices with analyzed tetrapeptide built into the central area of such model – $Ala_{10}\mathbf{X_4}Ala_{10}$. The energy profiles of all possible combinations of tetrapeptides were calculated (Fig. 1). The central, 11th point of the energy profiles corresponds to value of energy of a tested combination of four amino acids. Analysis of profiles results in some conclusions:

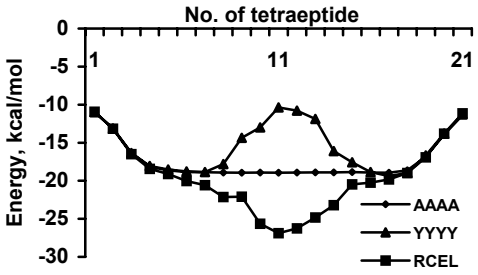

Figure 1: Energy profiles of optimized α-helical models of polyalanine chains with insertions of AlaAlaAlaAla, TyrTyrTyrTyr or ArgCysGluLeu at the 11th position.

[1] Microbiology and Tumor Biology Center, Karolinska Institute, Stockholm, Sweden. E-mail: Gelena.Kilosanidze@mtc.ki.se
[2] Microbiology and Tumor Biology Center, Karolinska Institute, Stockholm, Sweden. E-mail: Alexey.Kutsenko@mtc.ki.se

(1) The fact that end-regions of profiles coincide at different models after optimization of their structure, suggests that the central sites of these profiles can be compared with each other.
(2) Energy of polyalanine tetrapeptide fragments on the middle part of a chain, not subjected to edge effects (tetrapeptides 5th to 17th), is -18,9 kcal/mol and does not change.
(3) Energy of the other tetrapeptide fragments in the middle part of a chain can accept values above or below the polyalanine level and unambiguously characterize given tetrapeptide.

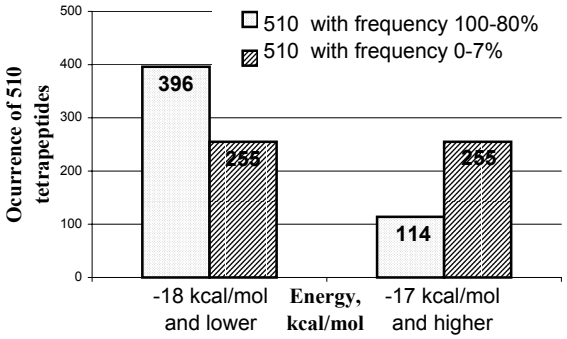

Figure 2: Energy distribution of modeled in α-helix tetrapeptides from two groups (of 510 combinations in each) that differ by occurrence in the nature.

This approach promises great benefit – exploring non-natural peptides. The estimated number of tetrapeptides that occur in native proteins is about 9000 [4], which is significantly less than the all-possible combinations of oligopeptide in length of four. Some of the native tetrapeptide fragments always prefer to be in α-helical conformation, others never adopt that type of secondary structure. Energies of model tetrapeptides will result in bias of given oligopeptides to the specific type of secondary structure. After the energy analysis of all oligopeptides one could suggest such amino acid combination that have not yet been found in nature but can adopt some kind of secondary structure with high efficiency, which will be of interest for design of *de novo* proteins.
We performed the exhaustive and statistical analyses of oligopeptide occurrence of in different conformations in PDB database [4]. That distribution was compared with calculated energy distribution.
The analysis shows that tetrapeptide fragments occurring with probability of 80 to 100% in the nature in α-helices preferably have energy values of –18 kcal/mol and lower in the model (Fig.2). Thus, this energy level serves as a cutoff benchmark to divide tetrapeptide fragments into favorable and unfavorable for α-helices formation.

References

[1] Abagyan, R.A. and Totrov, M. 1999. *Ab Initio* Folding of Peptides by the Optimal-Bias Monte Carlo Minimization Procedure. *J. Comp.Phys.*, 151:402-421.

[2] Kilosanidze, G.T., Kutsenko, A.S., Esipova, N.G. and Tumanyan, V.G. 2002. Use of molecular mechanics for secondary structure prediction. It is possible to reveal α-helix? *FEBS Lett.* 510:13-16.

[3] Kilosanidze, G.T., Kutsenko, A.S., Esipova, N.G. and Tumanyan, V.G. 2003. Analysis of forces leading the helix forming in α-proteins. *Protein Science* (in press).

[4] Vlasov, P.K., Kilosanidze, G.T., Ukrainskii, D.L., Kuz'min, A.V., Tumanian, V.G. and Esipova, N.G. 2001. Left-handed conformation of poly-L-proline type II in globular proteins. *Biofizika*. 46:573-576.