Poster E-7

Sequence-structure-flexibility relationship in protein evolution.

Authors:
Sandra Maguid (*Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes*)
Gustavo Parisi (*Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes*)
Sebastian Fernandez-Alberti (*Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes*)
Julian Echave (*Instituto Nacional de Investigaciones Fisicoquimicas Teoricas y Aplicadas, Universidad Nacional de L*)

Short Abstract: We study the evolutionary divergence of protein backbone dynamics by comparing the Ca flexibility profiles for a large dataset of homologous proteins classified into families and superfamilies. Proteins were structurally aligned and for each alignment we calculated sequence similarity (Id%), structural dissimilarity (RMSd) and different measures of backbone flexibility similarity.

Long Abstract:
Internal protein dynamics is essential for biological function. During evolution, protein divergence is functionally constrained: properties more relevant for function vary more slowly than less important properties. Thus, if protein dynamics is relevant for function, it should be evolutionary conserved. In contrast with the well studied evolution of protein structure, the evolutionary divergence of protein dynamics has not been addressed systematically before, apart from a few case studies. X-ray diffraction analysis gives information not only on protein structure but also B-factors, which characterize the flexibility that results from protein dynamics. Here, we study the evolutionary divergence of protein backbone dynamics by comparing the Cα flexibility (B-factor) profiles for a large dataset of homologous proteins classified into families and superfamilies. A dataset of pair alignments of non-homologous protein pairs was used as reference. Proteins were structurally aligned and for each alignment we calculated sequence similarity (Id%), structural dissimilarity (RMSd) and two measures of backbone flexibility similarity: the Spearman rank-order correlation coefficient between the aligned B-factor profiles, ρB , and the Pearson linear correlation coefficient rB. Furthermore, the vibrational dynamics of each protein was studied by normal modes analysis (NMA). The low-frequency normal modes describe collective movements that are closely related to the protein´s biological function [Berendsen 2000]. The NMA was performed using the Gaussian Network Model (GNM) [Haliloglu 1997] [Bahar 1997]. This model is shown to be a simple and efficient computational method to study the collective motions of large proteins. In order to explore the common dynamics of homologous proteins, a new procedure was developed [Maguid 2005]. The method allows the comparison of patterns of vibrational motions obtained by GNM and involves the alignment, rearrangement, and Singular Value Decomposition of the normal modes of homologous proteins. We were able to study the common dynamics within a family and superfamily by the identification of collective coordinates that were conserved during the evolution.
We show that Cα flexibility profiles diverge slowly, so that they are conserved at family and superfamily levels, even for pairs of proteins with non-significant sequence similarity.

We also analyse and discuss the correlations between the divergence of flexibility, sequence, and structure [Maguid 2006]. The present work has gone beyond the comparison of native structures, by comparing backbone flexibility profiles. These contain information on the equilibrium distribution of protein conformations. Thus, the present results imply the conservation not only of the average structure but also of the dispersion of the equilibrium distribution around the average structure. Such distribution is determined by the intramolecular dynamics of the protein. Thus, the observed conservation of flexibility profiles provides indirect evidence of the conservation of protein dynamics. Furthermore, the comparative analysis of the normal modes in homologous proteins reveals different levels of evolutionary conservation at family and superfamily levels for the first 50 modes ordered by increasing frequency values. The differences between the vibrational dynamics of the homologous proteins rapidly increase with the average frequency of their corresponding equivalent modes. The high degree of conservation observed in the few lowest modes indicates that mutations in proteins are constrained to produce limited changes at the tertiary level of structure that guarantee a unique pattern of relative flexibilities that provide protein functionalities. Therefore, we show that mutations within a fold family not only conserve a certain degree of protein structure but also features of their dynamics.

References

Berendsen, J. C., and S. Hayward (2000). Curr. Opin. Struct. Biol. 10:165–169.
Haliloglu, T., I. Bahar, and B. Erman (1997). Phys. Rev. Lett. 79:3090–3093.
Bahar, I., A. R. Atilgan, and B. Erman (1997). Fold. Des. 2:173–181.
Maguid, S., S. Fernandez-Alberti, L. Ferrelli, and J. Echave (2005). Biophys. J. 89: 3-13.
Maguid, S., S. Fernandez-Alberti, G. Parisi, and J. Echave (2006). J. Mol. Evol. (in press).