Poster C-23

Conformational diversity and evolutionary sequence divergence of proteins

Authors:
Maria Silvina Fornasari (*Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes*)
Gustavo Parisi (*Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes*)

Short Abstract: Using the Structurally Constrained Protein Evolution model, we found that the consideration of the quaternary structure and the multiple conformations that a protein could adopt, enhance the description of the sequence divergence. We used the protein triosephosphate isomerase from the archaeal Thermoproteus tenax which adopts a dimer/tetramer equilibrium.

Long Abstract:

A major constraint in protein sequence divergence is the conservation of protein structure. In recent years, there is an emerging picture which describes the native structure of a protein as an ensemble of different conformations. This concept was described earlier in the well known Monod-Wyman-Changeux model[1] for allosteric proteins or the so called pre-existing-equilibrium hypothesis. In this work we study the modelling of protein divergence as a consequence of the occurrence of multiple conformations for a protein native state.

It has been demonstrated that the explicit incorporation of structural information, as protein fold, stability and/or foldability enhances the description of sequence divergence during protein evolution[2;4;5]. To study how protein structure conservation modulates sequence divergence, we developed the Structurally Constrained Protein Evolution model(SCPE, [5]). The SCPE model simulates evolution by introducing random mutations and selecting them for structural conservation. Here, we first analyse SCPE performance when quaternary structural information is incorporated. We found that the consideration of the quaternary structure in the SCPE enhance the model performance when is compared with a monomeric description of the corresponding protein. Furthermore, we also show that the consideration of the different conformations a protein could adopt in our model also enhances the description of its evolution.

For this study we use as test system the protein triosephosphate isomerase from the hyperthermophilic archaeal Thermoproteus tenax (TtxTIM)[6]. This protein exists in a dimer/tetramer equilibrium. Although the dimer is inactive, the equilibrium is shifted to the active tetrameric form through a specific interaction with glycerol-1-phosphate dehydrogenase[6].

A sequence similarity search using TtxTIM as input was done using BLASTP with default parameters. The retrieved sequences (with E-values below 110-4) were aligned with CLUSTALX. Using this alignment a Neighbor-Joining tree was estimated using maximum likelihood distances calculated with the JTT+F model with gamma rate variation using the program HYPHY[]. From this tree the subtree containing TtxTIM and close homologous proteins was identified. SCPE simulations were performed with the monomeric, dimeric and tetrameric forms of TtxTIM to obtain a set of site-specific substitution matrices as described

previously[3]. Also using HYPHY, with these substitution matrices we calculated the maximum likelihood using the fixed topology (subtree) and the corresponding alignment mentioned above. Likelihood-ratio[7] test was used to study the statistical significance of the different models.

Our results indicate that the best conformation describing the divergence of TtxTIM is the tetrameric form in terms of the maximum likelihood analysis. We observed that the monomer performed worse than the dimer, and the dimer worse than the tetramer. In a profile analysis of the logarithm of the likelihood per position we found that the main difference between the tetramer and the dimer are located at the tetrameric interfaces. It is important to note that TtxTIM does not form stable tetramers but this is the active form of the protein. Then, the tetramer is the conformation subjected to selective pressure to conserve the biological activity, observation also supported by the maximum likelihood calculations.

References

[1]Monod, J., Wyman, J., Changeux, JP. 1965. On the nature of allosteric transitions: a plausible model. J. Mol. Biol. 12, 88-118.
[2]Bastolla, U., Roman, H.E., and Vendruscolo, M. 1999. Neutral evolution of model proteins: diffusion in sequence space overdispersion. J Theor Biol. 200:49-64
[3]Fornasari, M.S., Parisi, G., and Echave, J. 2002. Site-specific amino acid replacement matrices from structurally constrained protein evolution simulations. Mol Biol Evol. 19:352-6

[4]Govindarajan, S. and Goldstein, R.A. 1997. Evolution of model proteins on a foldability landscape. Proteins. 29:461-6
[5]Parisi, G. and Echave, J. 2001. Structural constraints and emergence of sequence patterns in protein evolution. Mol Biol Evol. 18:750-6
[6]Walden, H., Taylor, G., Lilie, H., Knura, T., and Hensel, R. 2004. Triosephosphate isomerase of the hyperthermophile Thermoproteus tenax: thermostability is not everything. Biochem Soc Trans. 32:305
[7]Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36, 182–198.

[8]Pond SL, Frost, SD and Muse, SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21(5):676-9