Evaluation of protein models quality using neural networks. Application to the ETHE1 protein

Mereghetti P (1), Papaleo E (1), Fantucci P (1), Tiranti V (2), Zeviani M (2), Mineri R (2), De Gioia L (1)

(1) Department of Biotechnology and Bioscience, University of Milano-Bicocca, Milano
(2) Unit of Molecular Neurogenetics, Institute ``Carlo Besta'', Milano

Motivation

The development of reliable and accurate evaluation tools to check the quality of protein models is crucial for the improvement of useful prediction methods. Several energy functions and scoring algorithms for evaluating protein structures have been proposed and can be divided into different categories depending on the physical principles and on the structural features of the models considered in the evaluation. In the present contribution, we apply a method based on a neural network, to discriminate among correct and incorrect protein models of the protein encoded by the ETHE1 gene, which have been recently identified in mutated forms in patients affected by ethylmalonic encephalopathy.

Methods

A database of protein decoys with known three-dimensional structure was build combining models generated by different prediction methods. On each model we have computed several structural parameters, such as secondary structure content, solvent accessible surface, radius of gyration, stereochemical parameters and the backbone root mean square deviation between the model and the respective native structure, as measure of structural accuracy. The parameters dataset obtained was submitted to a principal component analysis to reduce redundacy and noise, the reduced dataset was then used to develop a neural network ables to describe the relationship between the principal components space and the model accuracy. The neural network is a three-layers feed-forward network with 11 sigmoid neurons in the first layer, 8 sigmoid neurons in the second layer and 1 linear neuron as output. Distinct neural networks were trained on different training-set partitions, obtaining an ensemble of 50 neural networks. As prediction results we consider the median over the 50 networks predictions. The models of "ETHE1" were generated starting form different sequence alignements, obtained with various fold recognition and homology modelling server.

Results

The comparison of different ETHE1 models and the evaluation of their reliability, as obtained by the procedure outlined in Methods, has allowed to disclose structure-function relationship that complement available experimental data.

Contact email: paolo.mereghetti@unimib.it