

**An ENSEMBLE machine learning approach for the prediction of all-alpha membrane proteins -
(session: Novel Algorithms for Bioinformatics)**

Pier Luigi Martelli, Piero Fariselli, and Rita Casadio

Laboratory of Biocomputing, CIRB/Department of Biology, University of Bologna, via Imerio 42, 40126 Bologna, Italy

ABSTRACT

Motivation: All-alpha membrane proteins constitute a functionally relevant subset of the whole proteome. Their content ranges from about 10 to 30% of the cell proteins, based on sequence comparison and specific predictive methods. Due to the paucity of membrane proteins solved with atomic resolution, the training/testing sets of predictive methods for protein topography and topology routinely include very few well-solved structures mixed with a hundred proteins known with low resolution. Moreover, available predictors fail in predicting recently crystallised membrane proteins (Chen et al., 2002). Presently the number of well-solved membrane proteins comprises some 59 chains of low sequence homology. It is therefore possible to train/test predictors only with the set of proteins known with atomic resolution and evaluate more thoroughly the performance of different methods.

Results: We implement a cascade-neural network (NN), two different hidden Markov models (HMM), and their ensemble (ENSEMBLE) as a new method. We train and test in cross validation the three methods and ENSEMBLE on the 59 well resolved membrane proteins. ENSEMBLE scores with a per-protein accuracy of 90% for topography and 71% for topology, outperforming the best single method of 7 and 5 percentage points, respectively. When tested on a low resolution set of 151 proteins, with no homology with the 59 proteins, the per-protein accuracy of ENSEMBLE is 76% for topography and 68% for topology. Our results also indicate that the performance of ENSEMBLE is higher than that of the best predictors presently available on the Web.

Keywords: all-a membrane proteins; Neural networks; HMM; protein structure prediction; membrane protein topology.

Contact: gigi@biocomp.unibo.it, www.biocomp.unibo.it