How many membrane proteins in the Human Genome?

Casadio R.(1), Fariselli P.(1), Martelli P.L.(1)

(1) Biocomputing group, University of Bologna

Motivation

Within the Biosapiens network of excellence (EC Framework VI), the Biocomputing Group of the Bologna University installed a DAS server in a pipeline connected to the EBI. Our task in collaboration with Gunnar von Hejne (Stockholm Bioinformatics Center, SCFAB, Stockholm University, Sweden), Gert Vriend (CMBI University of Nijmegen, the Netherlands) and David Jones (Bioinformatics Unit, University College London, United Kingdom) is the large scale screening of the human genome in order to annotate membrane proteins based on topology prediction of chains.

Methods

The DAS server in Bologna presently implements three top-scoring predictors: 1. TMHMM2.0 which is the predictor of transmembrane helices in proteins based on hidden Markov models developed by A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer (Journal of Molecular Biology, 305:567-580, 2001). TMHMM2.0 has the great advantage of being very fast, being based on only single sequence information. 2. MEMSAT is the new version of the predictor of transmembrane helices in proteins developed by Jones, D.T., Taylor, W.R. and Thornton, J. M. (Biochemistry 33:3038-3049, 1994). This new version takes advantage of the evolutionary information derived by multiple sequence alignment. 3. ENSEMBLE is the predictor of transmembrane helices in proteins developed by Martelli et al. (Bioinformatics 19:I205-I211, 2003). It is an ensemble of two hidden Markov models and one neural network. ENSEMBLE takes advantage of the evolutionary information derived by multiple sequence alignment.

Results

We found that out of 32001 unique sequences of the Ensemble release 35a1 (December 2004), 32.3 and 31.2 % are predicted as membrane proteins by MEMSAT and ENSEMBLE, respectively, (25% are predicted by both predictors). 19.6% of the sequences are predicted as membrane proteins by TMHMM2.0 (single sequenced-based), and 19% are predicted by all predictors. These results set the lower and upper bound for the membrane protein content of the human genome and allow a list of putative membrane proteins for further applications.

Contact email: casadio@alma.unibo.it
URL: http://www.biocomp.unibo.it/