In Silico Screening of Peptides from *Escherichia coli* with Similarity to Eukaryotes' Proteins for Function Prediction

Porto, W.F.[1], Neto, S.M.[1], Candido, F.A.S.[1], Franco, O.L.[1]

[1]Centro de Análises Proteômicas e Bioquímicas, Universidade Católica de Brasília, Brasília, Brazil.

Through novel genome sequencing techniques, world data production is faster than relative analysis. In this view, thousands of proteins are currently annotated without a more detailed structural analysis or function evaluated. For these reasons, molecular modeling becomes each day an important tool to predict protein function, since three dimensional structures gives several evidences of a probable activity. In this work, we suggested a new approach for function's prediction of poorly characterized small proteins; which was based on several parameters appropriated to selection of interesting peptides. Data mining were started with 62,376 *Escherichia coli*'s proteins from the NCBI's non-redundant database, where only small proteins (in a range of 50-100 amino acids) were selected, remaining only 6,371 proteins. Moreover peptides that show transmembrane segments were also discarded, remaining 5,616. In a third step proteins with described function, were also removed remaining 483 proteins. Finally similarity with eukaryotes' proteins, considering all the BLAST's results, and absence of three-dimensional structure elucidated were also analyzed. After all these filters, only 75 proteins showed desired profile. For these process were used the phobius predictor (http://phobius.cbr.su.se/), the Embank files, the NCBI BLAST and the non-redundant protein databases of NCBI (ftp://ftp.ncbi.nih.gov/blast/db/nr.00.tar.gz; ftp://ftp.ncbi.nih.gov/blast/db/nr.01.tar.gz and ftp://ftp.ncbi.nih.gov/blast/db/nr.02.tar.gz) and PDB (http://dunbrack.fccc.edu/Guoli/culledpdb/pdbaanr.gz). We develop all *in silico* filters with PERL and PHP scripts. The next step is marked by study of relationship between structure and function of these proteins. This new method described here can be a relevant tool to generate information about the study or economic potential of these unsolved peptides functions.

Keywords: *ab initio* modeling, molecular modeling, protein function prediction, small proteins, structure/function relationship