PANALYZER: A SOFTWARE TOOL TO TACKLE THE PROTEIN INFERENCE PROBLEM

G. Prieto [(1)], K. Aloria [(1)], A. Fullaondo [(1)], J.M. Arizmendi [(1)].

[(1)] University of the Basque Country

Protein identification by shotgun proteomics is a peptide-centric approach where there is no straightforward way to restore the link between proteolytic products and their protein antecedents. The same peptide sequence can belong to multiple protein sequences, therefore ambiguities in determining the identity of the proteins present in the sample can occur. This issue becomes especially problematic in the case of complex samples of higher eukaryotes where large protein families and multiple protein isoforms are present. At the moment, most search engines do not provide effective tools to filter and sort identified proteins in large datasets and in many publications many protein identifications with no real evidence are reported.

In order to address this problem we have developed PAnalyzer, a tool that classifies protein identification results according to the proposal of Nesvizhkii and Aebersold [1] and proteins are defined as conclusive proteins, indistinguishable proteins, protein group and no conclusive proteins. The software tool has been developed in C# using the .NET platform [2], thus providing cross-platform interoperability. The peptide identification data is read from XML files, each one corresponding to a different replicate run, and the resulting protein classification data can be browsed from the GUI (Graphical User Interface) or saved to a CSV (Comma-Separated Values) file.

[1] Nesvizhskii, A.I. and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics*.**4:**1419-1440.
[2] ECMA. *ECMA-335: Common Language Infrastructure (CLI)*. ECMA (European Association for Standardizing Information and Communication Systems), 5th edition, December 2010.