

MS2MODELS: probing protein interaction networks by MS-based proteomics and structural data integration

Postic, G.,^{*1,2} Marcoux, J.,³ Andreani, J.,⁴ Guérois, R.,⁴ Mouton-Barbosa, E.,³ Bousquet, M.P.,³ Rey, J.,² Vandenbrouck, Y.,⁵ Cianferani, S.,⁶ Schiltz, O.,³ Labesse, G.,⁷ Tuffery, P.,²

*lead presenter

¹guillaume.postic@univ-paris-diderot.fr

² *Université de Paris, CNRS UMR 8251, INSERM ERL U1133, Paris, France*

³ *CNRS UMR5089 Institut de Pharmacologie et de Biologie Structurale, 31077 Toulouse, France*

⁴ *Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France*

⁵ *CEA, BIG, Biologie à Grande Echelle, F-38054 Grenoble, France*

⁶ *Laboratoire de Spectrométrie de Masse Bio-Organique (LSMBO), IPHC, UMR 7178, Université de Strasbourg, CNRS, Strasbourg, France*

⁷ *CBS, Univ Montpellier, CNRS, INSERM, 34090 Montpellier, France*

Mass spectrometry (MS) has become essential for characterizing molecular species and their interactions. Most of the time, proteomic studies stop at listing the interacting proteins, without performing the analysis of the identified sequences. This is a wasted opportunity when considering the fact that structural and evolutionary aspects provide a powerful analysis framework for biologists: e.g. for interpreting patients mutations that interfere with assemblies, setting up directed mutagenesis and functional dissection experiments, or virtual screening. The MS2MODELS proteomics pipeline integrates structural biology to MS data, in order to enhance the analysis of the protein-protein interaction networks. The homology-based detection of relevant structures from the Protein Data Bank (PDB) is carried out with HHsearch¹. Annotations of homomultimeric complexes, as well as interaction data from BioGRID and the eukaryotic linear motifs (ELM) resource are also integrated into the analysis. We have used MS2MODELS on several MS datasets containing up to hundreds of proteins. Thanks to the integration of structural information, the pipeline is able (i) to identify true positives in MS data by validating interactions within the input list of proteins, and (ii) to find additional partners that are either below the MS detection threshold (false negatives) or not detected at all. Moreover, MS2MODELS indicates the potential involvement of each input protein into a homomultimeric complex. The pipeline comes with an easy-to-use web interface. Thus, the protein-protein interaction networks can be conveniently visualized in a web browser. Although MS can detect protein complexes, it cannot identify the protein residues involved in the interactions. This is why MS2MODELS offers the possibility to visualize the 3D structure of each partner within the context of its complex. The structure of the latter is either experimental or predicted, depending on its availability in the PDB or the Swiss Model Repository, respectively.

Bibliography :

[1] Söding, J. (2004). Protein homology detection by HMM–HMM comparison. *Bioinformatics*, 21(7), 951-960.