Title: **Establishment of databases and evaluation of data processing tools for the identification of glycans by mass spectrometry**

Student: *Ruben Guerrero Roman*

Date: January 2019

Supervisor/s: Dra. Estela Giménez López

Departament of Chemical Engineering and Analytical Chemistry

Glycoproteins are known to play a crucial role in many important of biological processes, such as cell signaling, cell-to-cell recognition and adhesion in cellular surfaces, which are regulated and controlled by linked oligosaccharides attached to these proteins. These oligosaccharide structures (also called glycans) are involved in many diseases such as hereditary disorders, immune deficiencies, chronic inflammation and cancer.

Some studies carried out recently have demonstrated the importance to identify correctly glycans and their isomers, as a certain isomer can berelated with some pathology. Lately, some authors have proposed N-glycans analysis by liquid chromatography coupled with mass spectrometry (LC-MS). These analyses present a huge and complex quantity of data.

For this reason, since that search is so time-consuming and tedious, in the last years, it has incremented the search of new software for data processing.

The main purpose of this work is to build database, with specialized software, that would permit to do an automatic search of glycan that can be present in different types of proteins. For this study, human alpha-1-acid glycoprotein (hAGP) will be use as model glycoprotein, because it has different types of comple type N-glycans such as bi, tri and tetra antennary glycans.

For this study, a commercial tool for data processing provided by Agilent Technologies, MassHunter Qualitative will be used. This software have three different data tools, depending if you want targeted or a non-targeted search (named "Find by Molecular Feature" , "Find by Molecular Feature (Bioconfirm extension)" and "Find by Formula"). These tools have been tested with standard parameters. "Find by Formula" was chosen as the best option, parameters were optimized for having more extensive and detailed search. An finally, the optimized data processing method was used to identify glycans from other type of glycoproteins such as EPO and OVA, for observing if these parameters are valid for different type and nature of glycans.

Keywords: Glycoprotein, glycans, Find by Formula, database, hAGP, EPO, OVA.