

**Título:** APPLICATION OF DATA MINING AND ARTIFICIAL INTELLIGENCE TECHNIQUES TO MASS SPECTROMETRY DATA FOR KNOWLEDGE DISCOVERY

**Nombre:** López Fernández, Hugo

**Universidad:** Universidad de Vigo

**Departamento:** Escuela Superior de Ingeniería Informática

**Fecha de lectura:** 29/03/2016

**Mención a doctor europeo:** concedido

**Programa de doctorado:** Programa Oficial de Doctorado en Sistemas Software Inteligentes y Adaptables

**Dirección:**

> **Director:** MIGUEL REBOIRO JATO

> **Director:** Daniel González Peña

**Tribunal:**

> **presidente:** Gilberto Paulo Peixoto Igrejas

> **secretario:** FERNANDO DÍAZ GÓMEZ

> **vocal:** Anália María García Lourenço

**Descriptores:**

> INFORMATICA

> INTELIGENCIA ARTIFICIAL

**El fichero de tesis** no ha sido incorporado al sistema.

**Resumen:** Mass spectrometry using matrix assisted laser desorption ionization coupled to time of flight analyzers (MALDI-TOF MS) has become popular during the last decade due to its high speed, sensitivity and robustness for detecting proteins and peptides. This allows quickly analyzing large sets of samples in one single batch and doing high-throughput proteomics. In this scenario, bioinformatics methods and computational tools play a key role in MALDI-TOF data analysis, as they are able to handle the large amounts of raw data generated in order to extract new knowledge and useful conclusions.

A typical MALDI-TOF MS data analysis workflow has three main stages: data acquisition, preprocessing and analysis. Although the most popular use of this technology is to identify proteins through their peptides, analyses that make use of artificial intelligence (AI), machine learning (ML), and statistical methods can be also carried out in order to perform biomarker discovery, automatic diagnosis, and knowledge discovery.

In this research work, this workflow is deeply explored and new solutions based on the application of AI, ML, and statistical methods are proposed. In addition, an integrated software platform that supports the full MALDI-TOF MS data analysis workflow that facilitates the work of proteomics researchers without advanced bioinformatics skills has been developed and released to the scientific community.

