



**Universitat de les
Illes Balears**

Facultat de Ciències

Memòria del Treball de Fi de Grau

Iniciación a la experimentación en proteómica. Análisis de un caso práctico.

Elena Fiol Diez

Grau de Biología

Any acadèmic 2013-14

DNI de l'alumne: 43195064Y

Treball tutelat per: Francisca Barceló Mairata

Departament de Biología

Paraules clau del treball:

proteómica, Gammapatía monoclonal de significado incierto (GMSI), biomarcadores, espectrometría de masas MALDI-TOF, R-studio, paquete MALDIquant

Abstract

Actualmente los estudios de proteómica desempeñan un papel importante en la identificación de biomarcadores en fluidos biológicos asociados a patologías. La espectrometría de masas MALDI-TOF es una técnica esencial en el contexto de la proteómica debido a su alta capacidad de análisis, su sensibilidad y su precisión en la determinación de masas moleculares proteicas.

El objetivo de este estudio es analizar posibles biomarcadores en seroproteínas asociados a la patología Gammapatía monoclonal de significado incierto (GMSI) usando la técnica de espectrometría de masas MALDI-TOF y el software R-studio.

A partir del análisis de 4 muestras de seroproteínas de pacientes diagnosticados como GMSI y 5 controles de seroproteínas de individuos sanos, mediante el software R-studio con la aplicación MALDIquant, se han identificado 4 posibles biomarcadores de seroproteínas asociados a GMSI.

Los resultados obtenidos no son válidos estadísticamente debido al número limitado de muestras que se han analizado. Para que el estudio sea significativo se debería analizar un número elevado de muestras.

ÍNDICE

1. INTRODUCCIÓN	4
1.1 ¿QUE ES LA PROTEOMICA?	4
1.2 LA ESPECTOMETRÍA DE MASAS COMO UNA TÉCNICA DE ESTUDIO DE PROTEÓMICA.....	4
1.3 CONCEPTO DE LA ENFERMEDAD GAMMAPATÍA MONOCLONAL DE SIGNIFICADO INCIERTO (GMSI).....	6
2. PROTOCOLO GENERAL EN LA EXPERIMENTACION CON ESPECTOMETRIA DE MASAS MALDI-TOF	6
2.1 PREPARACIÓN DE LAS MUESTRAS	6
2.2 ANÁLISIS POR ESPECTROMETRÍA DE MASAS. TÉCNICA MALDI-TOF	7
2.3 ANÁLISIS DEL EXPECTRO DE MASAS.....	9
3. OBJETIVOS	10
4. CASO PRÁCTICO	11
4.1 METODOLOGÍA	11
5. RESULTADOS	17
6. CONCLUSIONES.....	26
7. BIBLIOGRAFÍA.....	27

1. INTRODUCCIÓN

1.1 ¿QUE ES LA PROTEOMICA?

El término "Proteoma" fue usado por vez primera en 1995 para describir el conjunto de PROTEÍNAS de un genOMA, en una célula o un tejido. El proteoma a diferencia del genoma es dinámico y puede variar dependiendo de numerosos factores internos y externos (por ejemplo, el estado fisiológico, el comportamiento alimentario, el estrés, las enfermedades, los tratamientos médicos...) (1, 2). La palabra proteoma dio lugar a una nueva disciplina, la "Proteómica.

La proteómica constituye una línea prioritaria de investigación en el área de la Biología, siendo definida como el conjunto de técnicas y estrategias utilizadas para el estudio del componente proteico total de una célula, tejido u organismo.

El auge actual de la proteómica se debe a tres razones fundamentales: La secuenciación de genomas a gran escala y el desarrollo de bases de datos de proteínas, al desarrollo de técnicas de espectrometría de masas para analizar proteínas y péptidos y a los avances realizados en la separación de proteínas mediante la electroforesis 2D-PAGE y la cromatografía líquida en fase reversa C18 (1,2).

La proteómica clínica se centra en la identificación y la comprensión de los cambios en el proteoma relacionados con diferentes enfermedades y su seguimiento durante el tratamiento médico. En este campo es importante el estudio de biomarcadores asociados a una determinada patología.

1.2 LA ESPECTOMETRÍA DE MASAS COMO UNA TÉCNICA DE ESTUDIO DE PROTEÓMICA

La espectrometría de masas (MS) es una técnica analítica que tiene una alta capacidad de análisis, sensibilidad y precisión en la determinación de masas moleculares proteicas (3).

La técnica de espectrometría de masas MALDI-TOF es una técnica de ionización suave. Se denomina MALDI por sus siglas en inglés Matrix-Assisted Laser Desorption/Ionization (desorción/ionización láser asistida por matriz) y TOF por el detector de iones que se acopla al MALDI y cuyo nombre procede también de sus siglas en inglés Time-Of-Flight. El MALDI-TOF es un instrumento muy utilizado en proteómica para la identificación de proteínas (4) y el desarrollo de biomarcadores.

En Biología, las aplicaciones de la espectrometría de masas MALDI-TOF en fluidos biológicos, se centran principalmente en tres campos.

A. Identificación de proteínas por MALDI-TOF

La técnica utilizada para la identificación de proteínas es la huella peptídica. Utilizando ésta técnica la proteína desconocida se hidroliza por acción de proteasas específicas (generalmente tripsina) en pequeños péptidos cuyas masas absolutas pueden determinarse mediante un espectrómetro de masas MALDI-TOF (5). Las masas obtenidas constituyen un puzzle que podemos comparar con unas bases de datos de proteínas para determinar la secuencia de la proteína problema e identificarla.

B. Identificación de microorganismos por MALDI-TOF

El desarrollo de la tecnología MALDI-TOF ha permitido la utilización de la espectrometría de masas en el diagnóstico microbiológico de rutina (6, 7). Permite la identificación de microorganismos a partir del análisis de proteínas, principalmente ribosomales de una cepa o colonia bacteriana.

Cada microorganismo tiene una composición proteica característica. Por consiguiente el espectro de masas MALDI-TOF de un microorganismo permite identificarlo a partir de una base de datos utilizando un software adecuado como es el MALDI-Biotyper.

C. Análisis de biomarcadores en seroproteínas

De acuerdo con el National Institute of Health (NIH) un biomarcador es una característica que es medida y evaluada de forma objetiva como un indicador de procesos biológicos normales o patológicos (8). Por lo tanto, un biomarcador puede emplearse para diagnosticar una enfermedad y para seguir su evolución durante el tratamiento médico.

El desarrollo de la proteómica ha abierto grandes expectativas para la identificación de biomarcadores, ya que la espectrometría de masas permite identificar proteínas en muy baja concentración (9) y puede realizarse un análisis sistemático de cientos o miles de proteínas en una muestra clínica, sin requerir del conocimiento de su secuencia de aminoácidos. La principal fuente de biomarcadores son los fluidos biológicos humanos (sangre (suero y plasma), orina y saliva, entre otros). Nuestro trabajo se centra principalmente en el análisis de las proteínas del suero sanguíneo. El suero presenta muchas cualidades beneficiosas para la investigación proteómica ya que tiene un alto contenido de proteínas (es decir, 60 a 80 mg / ml) (10). Un inconveniente importante es que hay ciertas proteínas en elevada concentración y alto peso molecular, como es la albumina, que interfiere con las proteínas de baja concentración en el análisis MALDI-TOF, por lo que deberán eliminarse antes del análisis de espectrometría de masas MALDI-TOF.

C-1. Análisis de un caso práctico

En el caso práctico que se aborda, se han estudiado potenciales biomarcadores de Gammapatías monoclonales de significado incierto (GMSI) a partir del análisis de seroproteínas en humanos.

Este trabajo ha sido aceptado y desarrollado con los siguientes condicionantes:

- 1- He participado como observadora en todo el protocolo experimental de preparación de muestra y análisis por espectrometría de masas MALDI-TOF de muestras de seroproteínas de individuos sanos decodificadas que forman parte de un proyecto evaluado.
- 2- El trabajo experimental que he desarrollado se ha centrado en el análisis bioinformático de los datos experimentales que me han proporcionado, con el objetivo único de adquirir conocimientos básicos en la experimentación en proteómica.

1.3 CONCEPTO DE LA ENFERMEDAD GAMMAPATÍA MONOCLONAL DE SIGNIFICADO INCIERTO (GMSI)

Las gammapatías monoclonales constituyen un grupo heterogéneo de trastornos que van desde enfermedades neoplásicas malignas, como el Mieloma Múltiple que requieren tratamiento activo, hasta cuadros benignos como las gammapatías monoclonales de significado incierto (GMSI) que generalmente no requieren de ningún tratamiento. Se caracterizan por la proliferación estable o progresiva de un clon de células plasmáticas que producen inmunoglobulinas (Ig) monoclonales (11). La patología GMSI constituye una etapa precancerosa asintomática caracterizada por la ausencia de lesiones óseas y se asocia con un riesgo de por vida de la progresión al estado de cáncer de mieloma múltiple (MM). GMSI representa la más frecuente de todas las gammapatías monoclonales detectadas en el laboratorio.

Su prevalencia en la población general es de 0,7 – 1,7 %, pero en los individuos mayores de 70 años alcanza hasta el 3% (12).

2. PROTOCOLO GENERAL EN LA EXPERIMENTACION CON ESPECTOMETRIA DE MASAS MALDI-TOF

2.1 PREPARACIÓN DE LAS MUESTRAS

Para el análisis de proteínas mediante la técnica de MALDI-TOF, las muestras biológicas tienen que ser purificadas mediante técnicas de electroforesis 2D-PAGE o de cromatografía.

Con la electroforesis bidimensional 2D-PAGE se separan proteínas en función de su peso molecular y del punto isoeléctrico (13). Posteriormente la proteína de interés se recupera del gel cortando la banda manualmente, aunque existen también sistemas automatizados.

En cromatografía se utiliza la cromatografía en fase reversa C18 y la instrumentación HPLC. La instrumentación nano-HPLC emplea columnas capilares con un diámetro interior (ID) más pequeño que la cromatografía habitual (HPLC) para permitir separar una cantidad de muestra muy pequeña.

Las proteínas purificadas deben ser digeridas con una proteasa específica para obtener fragmentos peptídicos pequeños para el posterior análisis MALDI-TOF. Como proteasa de digestión se utiliza la tripsina, que específicamente rompe los enlaces peptídicos de las proteínas mediante hidrólisis para formar péptidos de menor tamaño. Es una enzima específica que rompe en las posiciones del grupo carboxilo de los residuos de Arginina (Arg) o Lisina (Lys).

Para determinar los posibles biomarcadores a partir de las seroproteínas nos interesa obtener las proteínas del suero de bajo peso molecular. Para eliminar la albumina, que es una proteína que se encuentra en una concentración bastante elevada y de alto peso molecular que interfiere con el resto de proteínas se utiliza una cromatografía en fase reversa C18 Zip-Tip.

Después de purificar la proteína o el extracto proteico las muestras están listas para ser analizadas por MS

2.2 ANÁLISIS POR ESPECTROMETRÍA DE MASAS. TÉCNICA MALDI-TOF

EL espectrómetro de masas se compone de los siguientes elementos básicos

- 1- Sistema de ionización de muestra MALDI
- 2- Analizador de masas (TOF)
- 3- Detector de iones para la obtención del espectro de masas MALDI-TOF
- 4- Software de adquisición y análisis de datos

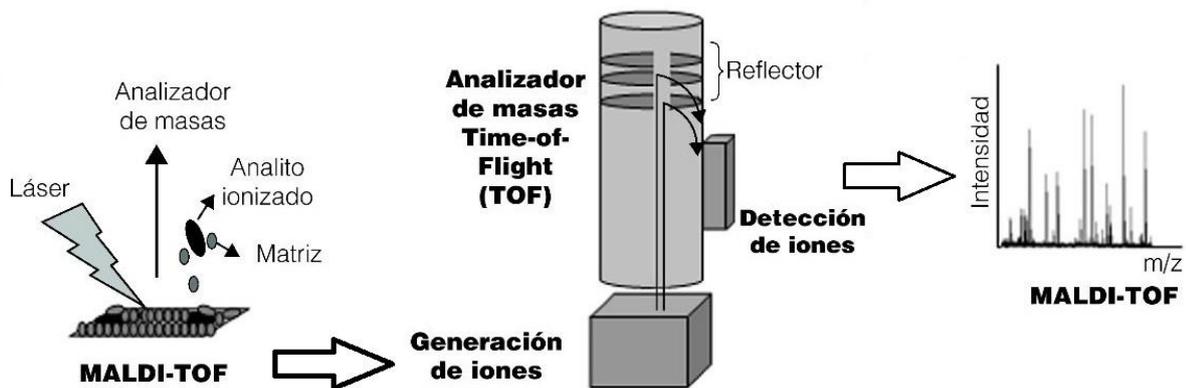


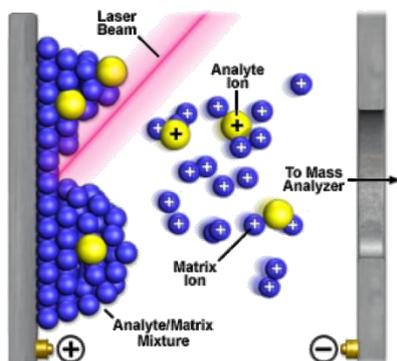
Figura 1A: Esquema del espectrómetro de masas MALDI-TOF

Los espectrómetros de masa con analizadores de tiempo de vuelo acoplados a sistemas de ionización MALDI se denominan espectrómetros MALDI-TOF

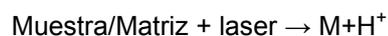
1. Sistema de ionización de muestra MALDI

La primera etapa del análisis MALDI-TOF es la ionización de la muestra. Para la ionización se utiliza el sistema (MALDI, *Matrix-Assisted Laser Desorption/Ionization*) Ionización/desorción con láser asistida con una matriz para muestras en estado sólido. Sobre una placa MALDI-TOF se mezcla la muestra proteica con una matriz. Las matrices más habituales son: el ácido α -ciano-4-hidroxicinamínico (CHCA) que se utiliza para analizar péptidos, el ácido di-hidroxi benzoico (DHB) y el ácido sinapínico (SA) que se utiliza para proteínas grandes.

En esta técnica la muestra proteica se mezcla con la matriz en exceso sobre una superficie de metal de tal forma que ambas cocrystalizan cuando se evapora el solvente. Esta preparación es sometida a pulsos cortos de láser en alto vacío durante algunos nanosegundos. La matriz es capaz



de absorber la energía a la longitud de onda del láser. Esta energía es convertida en energía de excitación que luego cede a las moléculas de la muestra de una forma controlada. El resultado es que las moléculas proteicas quedan como iones moleculares protonados volatilizadas (9).



Los iones moleculares producidos son posteriormente acelerados hasta el analizador.

Figura 1B: Esquema de la ionización de la muestra (14)

2. Analizador de masas

En el analizador de masas, los iones son separados de acuerdo a su relación masa/carga (m/z) en el vacío. El analizador más utilizado es el *TOF* (*Analizador de tiempo de vuelo*).

El analizador de masas TOF mide la relación m/z de un ion en función del tiempo requerido para recorrer la longitud de un tubo de vuelo desde que el ion abandona la fuente de iones hasta que llega al detector.

Algunos analizadores de masa TOF incluyen un reflector al final de la zona de vuelo que desvía la trayectoria de los iones hasta encontrar un segundo detector. Este artilugio permite obtener información estructural de la molécula y se emplea para identificar proteínas.

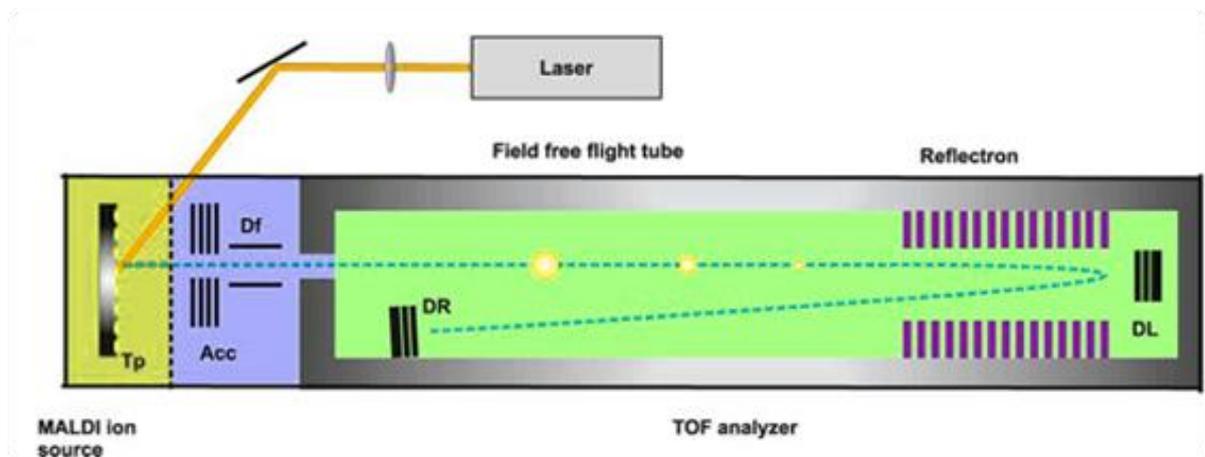


Figura 1C: Esquema del espectrómetro MALDI-TOF

(http://nutricionpersonalizada.wordpress.com/2010/10/26/peptidomica_analisis_peptidos_bioactivos/2_tmbs_23b/)

3. Detector de iones

Los detectores proporcionan información sobre el flujo de iones y la abundancia de iones que salen del analizador de masas. Convierten el chorro de iones en una señal eléctrica que finalmente registramos como el espectro de masas MALDI-TOF. Debido a la cantidad de información que se genera en un simple experimento, se necesita un sistema informático para el control del equipo, así como un software adecuado para la adquisición, almacenamiento y análisis de los espectros.

2.3 ANÁLISIS DEL ESPECTRO DE MASAS

La representación gráfica de los datos generados constituye el espectro de masas (Figura 2), donde se representa la abundancia de los iones detectados en función del ratio masa/carga (m/z)

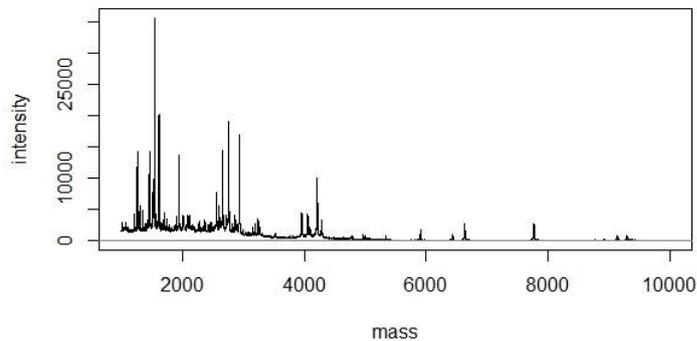


Figura 2: Espectro de masas obtenido en el análisis MALDI-TOF de una mezcla de seroproteínas.

Para el análisis de los datos experimentales se utilizan softwares específicos de acuerdo al objetivo del experimento. En este estudio, se utiliza el software gratuito R-studio con la aplicación MALDIquant para determinar posibles biomarcadores en muestras biológicas, comparando los espectros de masas MALDI-TOF de individuos sanos y pacientes con MGS1.

3. OBJETIVOS

El objetivo de este trabajo es identificar posibles biomarcadores en seroproteínas asociados a la enfermedad de Gammapatía monoclonal de significado incierto (GMSI) mediante la técnica de espectrometría de masas MALDI-TOF y el programa de análisis gratuito R-studio con la aplicación MALDIquant.

4. CASO PRÁCTICO

4.1 METODOLOGÍA

4.1.1 Preparación de muestras para el espectrómetro de masas MALDI-TOF

Se pretende obtener una solución de seroproteínas de bajo peso molecular. Para ello se emplea una cromatografía en fase reversa C 18. Se utiliza una punta de pipeta ZipTip C18 de 10 microL. Con ello se eliminan las proteínas más abundantes como la albumina de alto peso molecular y las sales, y se concentra la muestra.

En nuestro experimento se han analizado 5 controles sanos y 4 muestras MGSI, por lo que se ha repetido este protocolo 9 veces.

En la tabla 1 se describe el protocolo seguido de la preparación de la muestra para ser analizada por MALDI-TOF.

Solution	ZipTip_{C18}	Pipette Tips
Wetting solution	100% acetonitrile (ACN)	
Sample preparation	Adjust sample to 0.1% trifluoroacetic acid (TFA); final sample pH <4	
Equilibration solution	0.1% TFA in Milli-Q grade water	
Wash solution	0.1% TFA in Milli-Q grade water	
Elution solution	0.1% TFA/50% ACN with or without matrix	

Tabla 1: Procedimientos para concentrar y purificar las seroproteínas mediante cromatografía en fase reversa C18, mediante el micrométodo de Zip-Tip C18 (User guide for Reversed-Phase, MILLIPORE)

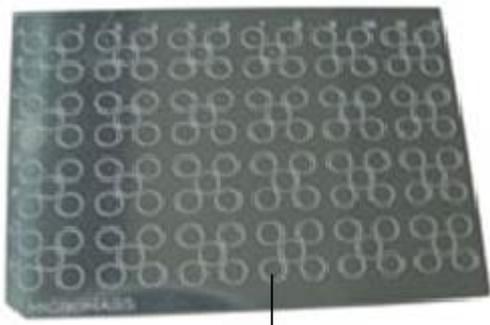
En la primera etapa se prepara la muestra de seroproteínas mezclando 5 microL de suero sanguíneo con 5 microL de TFA al 0.1%. La presencia de TFA permite la unión máxima de las proteínas a la punta de la pipeta Zip-Tip. Hay que asegurarse de que la solución este a pH <4.

En la segunda etapa se tiene que equilibrar la punta de la pipeta Zip-Tip. Para ello se pipetea tres veces en acetonitrilo y una vez en una mezcla de 0.1% de TFA con agua Milli-Q. A continuación, se aplica la solución de proteínas preparada a la pipeta Zip-Tip.

En la tercera etapa se aplica la solución de lavado que consiste en una mezcla de 0.1% de TFA con agua Milli-Q.

En la cuarta etapa se eluyen los péptidos y/o proteínas que ha quedado retenidas dentro de la Zip-Tip utilizando una mezcla de 0.1% de TFA con el 50% de ACN (acetonitrilo)

Finalmente se mezcla la muestra purificada con la Zip-Tip C18 con la matriz α -ciano-4-hidroxycinamínico y se deposita directamente sobre una placa MALDI-TOF MS (Figura 3). Se dejar secar al aire libre unos minutos y seguidamente se introduce en el espectrómetro de masas MALDI-TOF.



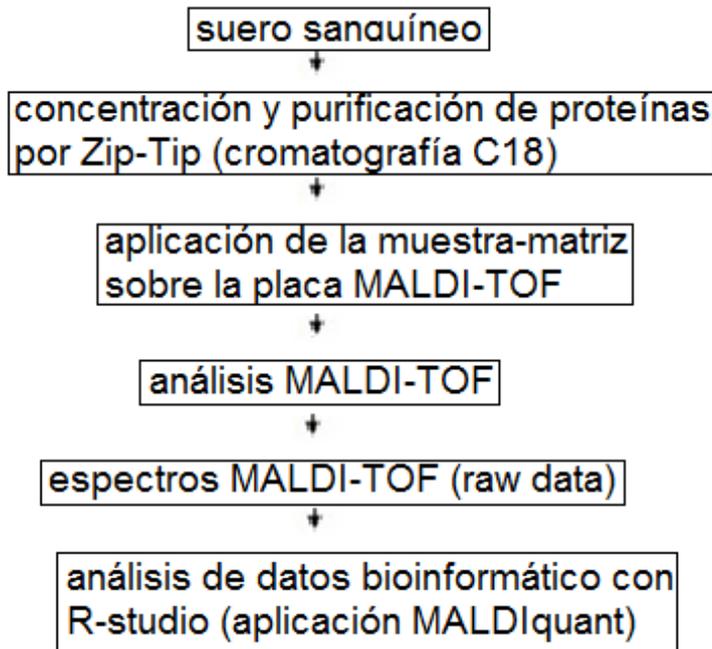
pocillo donde se
deposita la mezcla
muestra-matriz



Figura 3: Placa MALDI-TOF y sistema de introducción de la placa en el espectrómetro de masas MALDI-TOF

4.1.2 Análisis de muestras mediante el espectrómetro de masas MALDI-TOF

El proceso experimental que se sigue para analizar la muestra se presenta en el esquema 1:



Esquema 1: Protocolo general de análisis

Se ha trabajado con un espectrómetro de masas Autoflex III MALDI-TOF/TOF (Bruker Daltonics, Leipzig, Alemania) equipado con un láser de 200 Hz Smartbeam (Figura 3B). Los espectros se registraron en el modo lineal LP-ClinProt a una frecuencia de 200 Hz en el intervalo de masa desde 500 hasta 20000 Da. Para cada espectro se lanzaron 1000 disparos de láser desde diferentes posiciones para obtener un espectro de la muestra.



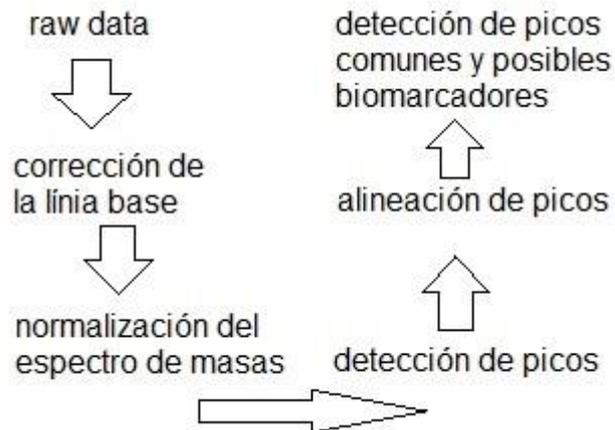
En cada experimento se hace un calibrado externamente del aparato utilizando una mezcla de proteínas de calibración estándar de Bruker.

Figura 3B: Espectrómetro de masas Autoflex III MALDI-TOF (<http://www.directindustry.es/prod/bruker-daltonics-inc/espectrometros-masas-maldi-tof-automatizados-30029-259986.html>)

4.1.3 Análisis bioinformático

Los datos del espectrómetro de masas obtenidos se analizaron utilizando el software gratuito R-studio. Este software dispone de un paquete de libre acceso que se denomina MALDIquant, que se ha utilizado en el análisis de las muestras.

El protocolo que se sigue consta de los siguientes pasos (Esquema 2).

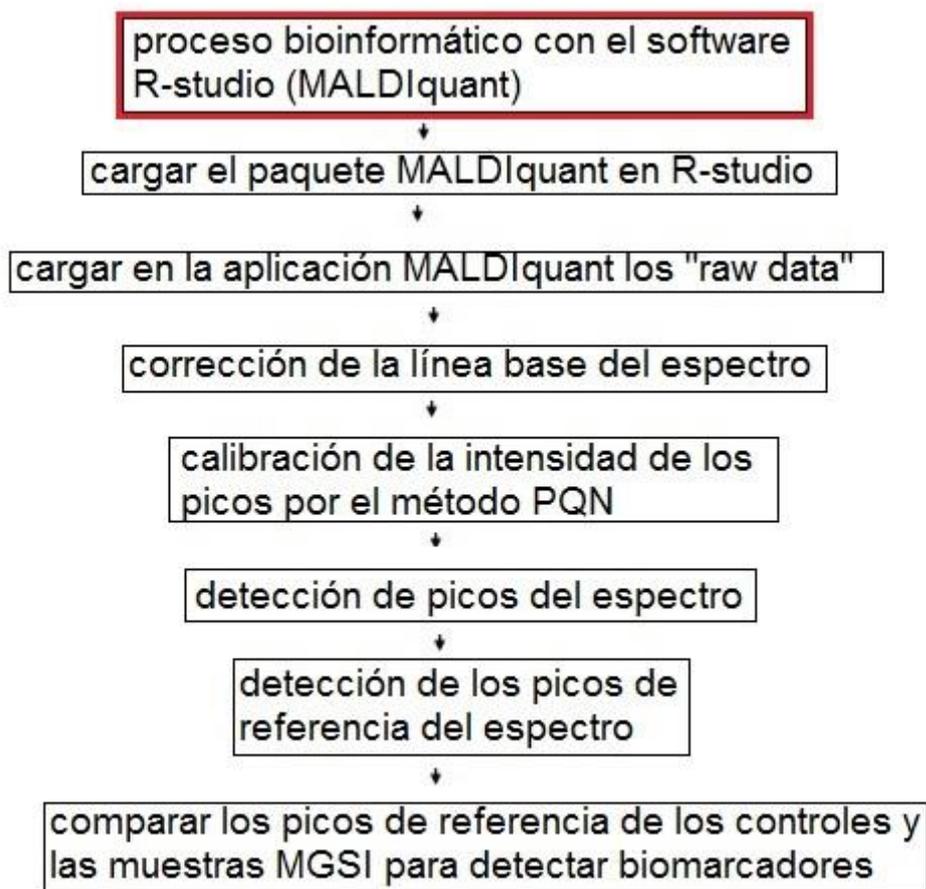


Esquema 2: Diagrama básico para el análisis con R-studio

4.1.3.1 Análisis de espectros de mass MALDI-TOF mediante el software R-studio con el paquete de libre acceso MALDIquant.

En la primera etapa se ordenaron los datos en tres carpetas diferentes. En una carpeta se introducen las 4 muestras MGSI, en otra carpeta los 5 controles sanos y en otra carpeta todas las muestras controles sanos + muestras MGSI. Cada espectro de cada muestra se ha obtenido por triplicado

El protocolo general que se sigue se ilustra en el esquema 3.



Esquema 3: Esquema del análisis de datos bioinformático mediante el software R-studio con el paquete de libre acceso MALDIquant

A. Iniciación del programa R-studio y carga de los datos experimentales

El primer paso en el análisis bioinformático es iniciar todos los paquetes necesarios para el programa de investigación R-studio. Para iniciar el paquete MALDIquant es necesario cargar el paquete "readBrukerFlexData" que le permite acceder directamente a los datos experimentales en bruto ("raw data") obtenidos del análisis MALDI-TOF. Después de la iniciación del programa, se tienen que cargar los datos experimentales obtenidos de las muestras MGS1 y los controles sanos en su forma "raw data", incluyendo los triplicados de cada muestra.

B. Transformación de la intensidad y corrección de la línea base de cada espectro

El programa en primer lugar realizó una transformación de los datos experimentales para estabilizar y alisar las intensidades. A continuación se aplicó la función "remove base line", que corrige la línea base del espectro

C. Calibración de la intensidad de los picos de cada espectro

Después de la corrección de la línea base se observa que hay picos muy altos y otros mucho más pequeños. Para poder hacer una mejor comparación se igualan las intensidades utilizando el método PQN que calibra los espectros utilizando la función "TIC" (total ion current).

La "total ion current (TIC)" representa la suma de las intensidades de todo el espectro. Así cada punto del espectro es dividido por la suma total de intensidades.

D. Detección de picos en el espectro

A continuación se aplica la función "detection peaks" para detectar los picos de cada espectro de cada muestra.

Las operaciones A-D se repiten para cada triplicado de cada muestra MGSI o control sano.

E. Picos de referencia

La función "reference peaks" se utiliza para detectar los picos de referencia de todas las muestras MGSI, de todos los controles sanos y de todas las muestras controles sanos + MGSI.

Esta función detecta todos los picos que son comunes a cada grupo de muestras analizado. Además, asegura que los picos que no existen en todo el grupo de espectros analizados son eliminados y el espectro de picos de referencia final solo consta de los picos comunes a todo el grupo de espectros analizados en muestras MGSI, en controles sanos y en todas las muestras controles sanos + MGSI.

G. Identificación de biomarcadores

Cuando se han obtenido los picos de referencia de las muestras MGSI, de los controles sanos y de todas las muestras controles sanos + GMSI, se procede a analizar los posibles biomarcadores asociados a GMSI.

Para ello se comparan los picos de referencia de todas las muestras controles sanos + MGSI con los picos de referencia de las muestras MGSI. Se eliminan los picos comunes y los picos que quedan en los espectros de las muestras GMSI constituyen posibles biomarcadores.

5. RESULTADOS

En este estudio se han analizado 5 muestras de seroproteínas de individuos sanos y 4 muestras de individuos clínicamente caracterizados como pacientes MGSI. El trabajo de este proyecto se ha centrado principalmente en el análisis bioinformático de los espectros MALDI-TOF de las muestras anteriormente citadas.

El experimento se ha realizado de acuerdo al protocolo descrito en el apartado 4.1.1. Para analizar los espectros MALDI-TOF obtenidos se ha aplicado el protocolo bioinformático del software R-studio utilizando el paquete MALDIquant descrito en el apartado 4.1.3.1. Los datos experimentales obtenidos constituyen los datos en bruto llamados "raw data".

La primera etapa del análisis tiene por objetivo detectar los picos de cada espectro MALDI-TOF obtenido experimentalmente. Los datos experimentales "raw data" son tratados mediante tres etapas consecutivas (Figuras 5 y 6):

- 1- En la primera etapa se hace la corrección de la línea base (panel B)
- 2- En la segunda etapa se realiza el calibrado del espectro mediante la función PQN (panel C).
- 3- En la última etapa se obtienen los picos detectados para cada espectro (panel D).

Se observa en la figura 5 y 6 que hay diferencias entre los espectros de seroproteínas de los controles sanos y las muestras MGSI.

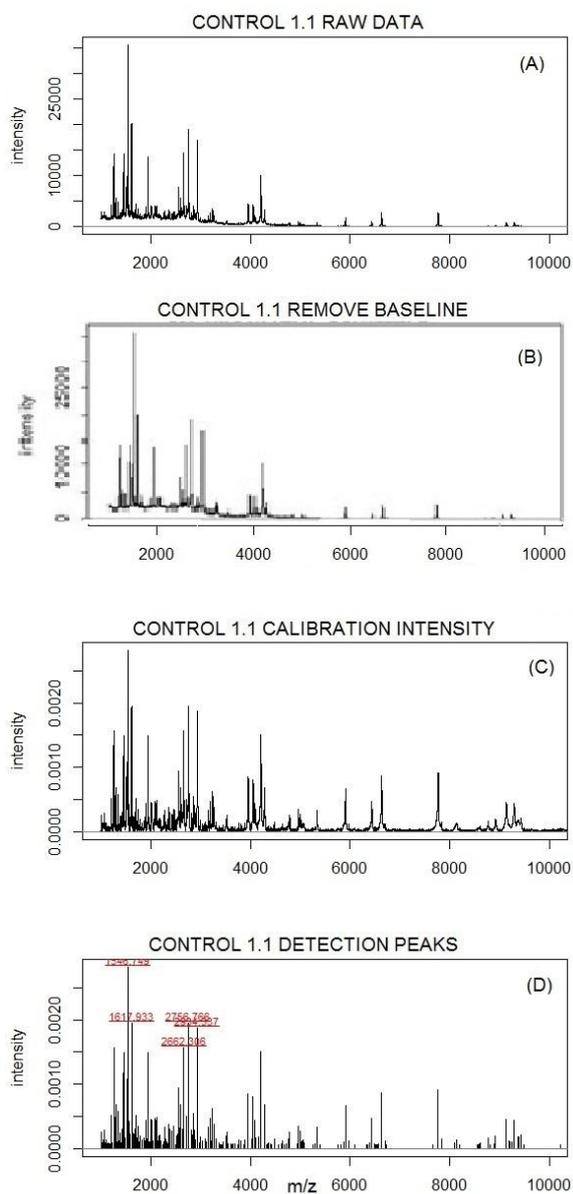


Figura 5: Protocolo de análisis de los datos obtenidos de un control en el análisis MALDI-TOF mediante el software R-studio con el paquete de libre acceso MALDIquant. Datos del control 1.1, primera replica

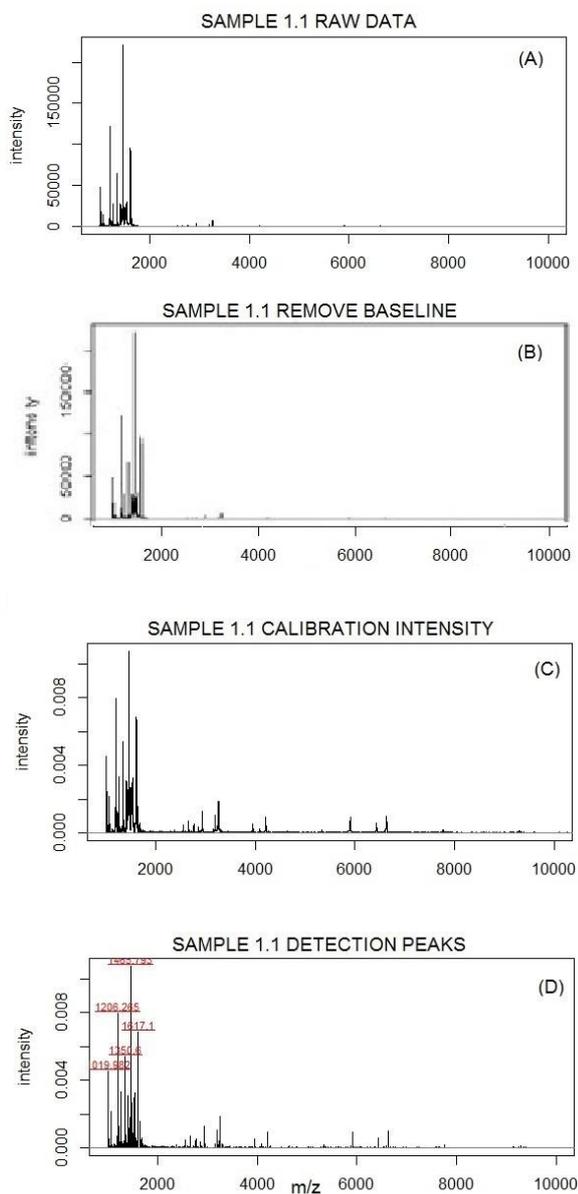


Figura 6: Protocolo de análisis de los datos obtenidos de una muestra MGS1 en el análisis MALDI-TOF mediante el software R-studio con el paquete de libre acceso MALDIquant. Datos de la muestra MGS1 1.1, primera replica.

En las figuras 7 y 8 se representan los espectros en un rango de masas (m/z) de 1000-10000 obtenidos en el análisis de los 5 controles sanos y 4 muestras MGSI.

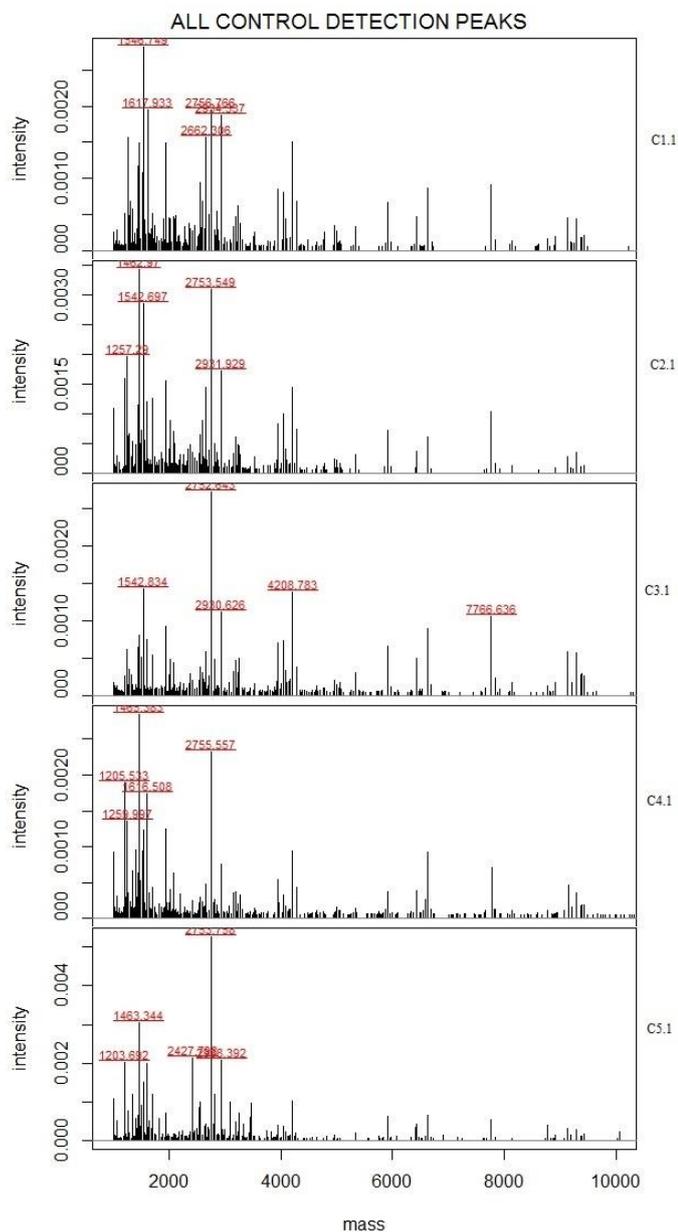


Figura 7: Espectros de los picos detectados en el análisis de los controles mediante el software R-studio con el paquete de libre acceso MALDIquant

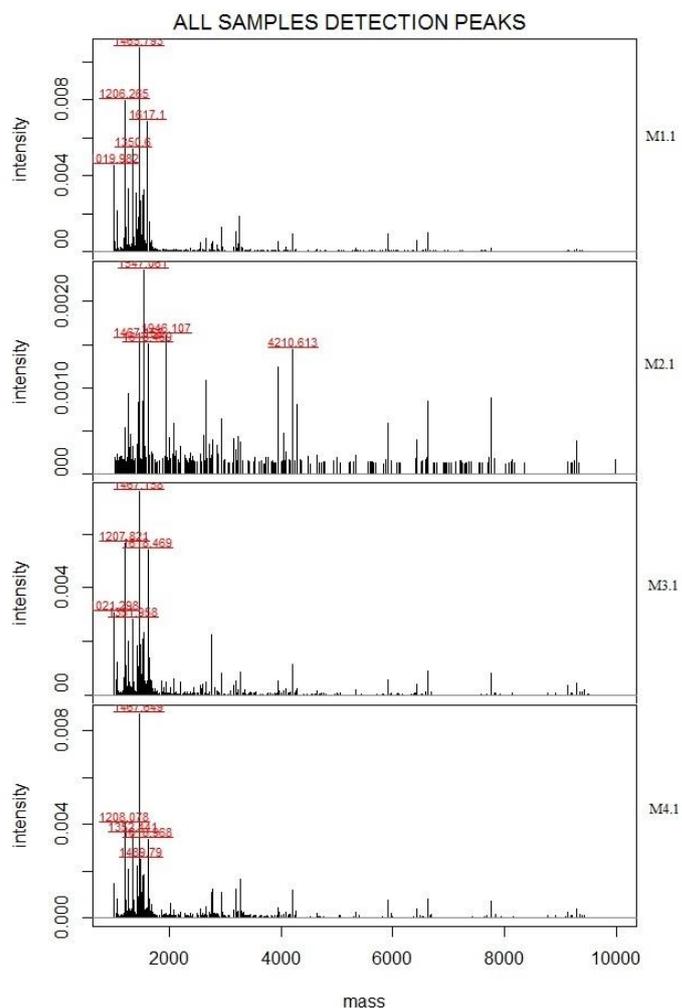


Figura 8: Espectros de los picos detectados en el análisis de las muestras MGSI mediante el software R-studio con el paquete de libre acceso MALDIquant

Para analizar estos espectros no se consideran las masas inferiores a 2000 Dalton, dado que es una zona del espectro de masas que puede verse afectada por cualquier tipo de contaminación que tengamos durante el proceso experimental y puede dar lugar a falsos picos biomarcadores. Además, en el análisis MALDI-TOF no se ha dispuesto de un calibrador en el rango de m/z inferior a 2500, por ello el análisis realizado en la identificación de picos comunes se realizara en el rango de masa (m/z) 3000-10000.

En las figuras 7B y 8B se representan los mismos espectros que en las figuras 7 y 8 pero en el rango de masas de 2000-10000Dalton.

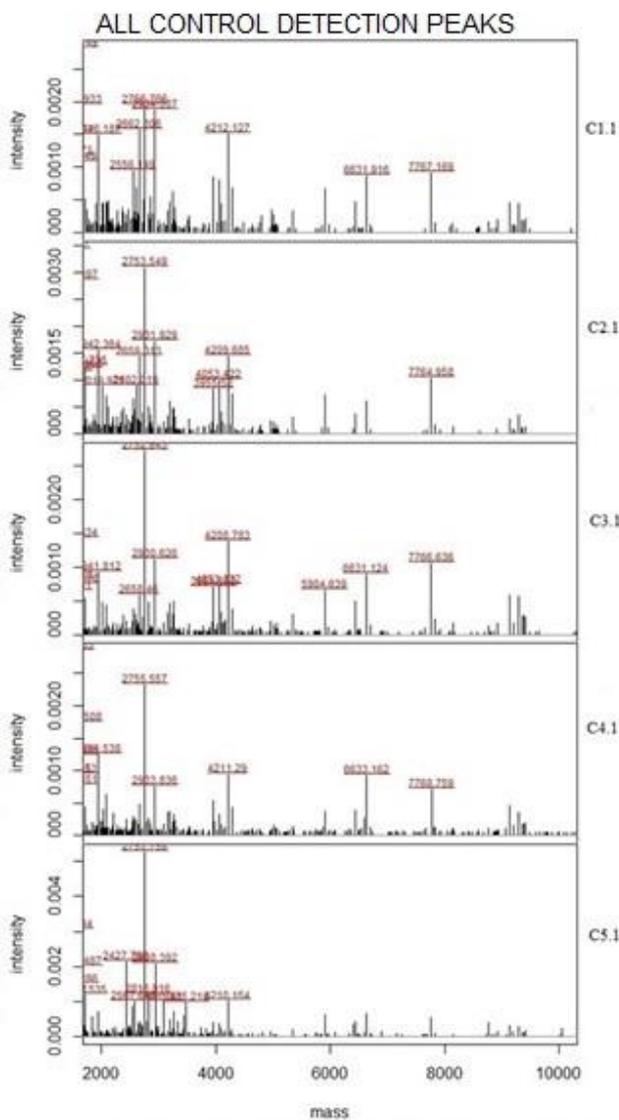


Figura 7B: Espectros de los picos detectados en el análisis de los controles mediante el software R-studio con el paquete de libre acceso MALDIquant

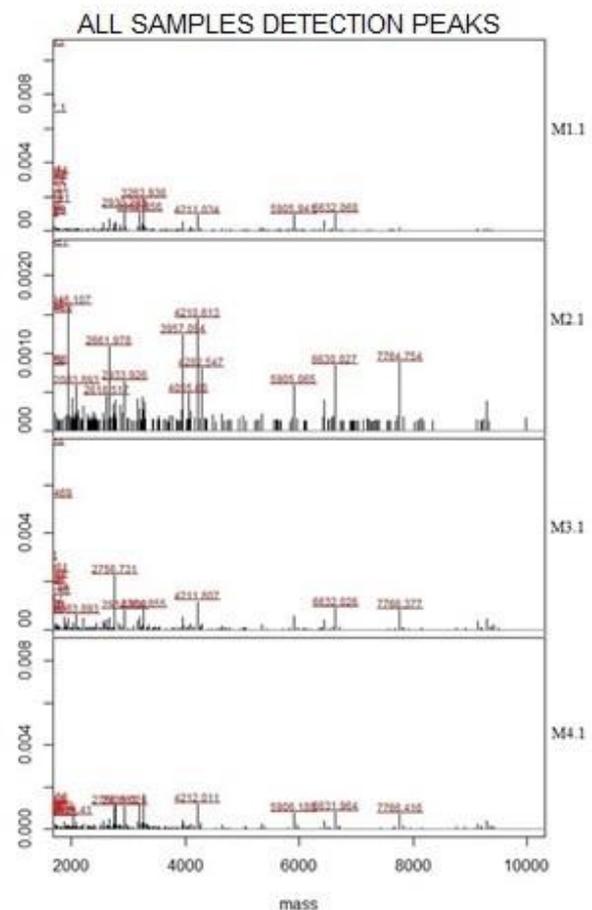


Figura 8B: Espectros de los picos detectados en el análisis de las muestras mediante el software R-studio con el paquete de libre acceso MALDIquant

Se puede observar en las figuras 7B y 8B que hay una gran variabilidad en los espectros de las distintas muestras biológicas, especialmente en las muestras de los pacientes MGS1.

En la siguiente etapa se determinan los picos comunes de referencia de los controles y los picos comunes de referencia de las muestras MGSi (Figura 9).

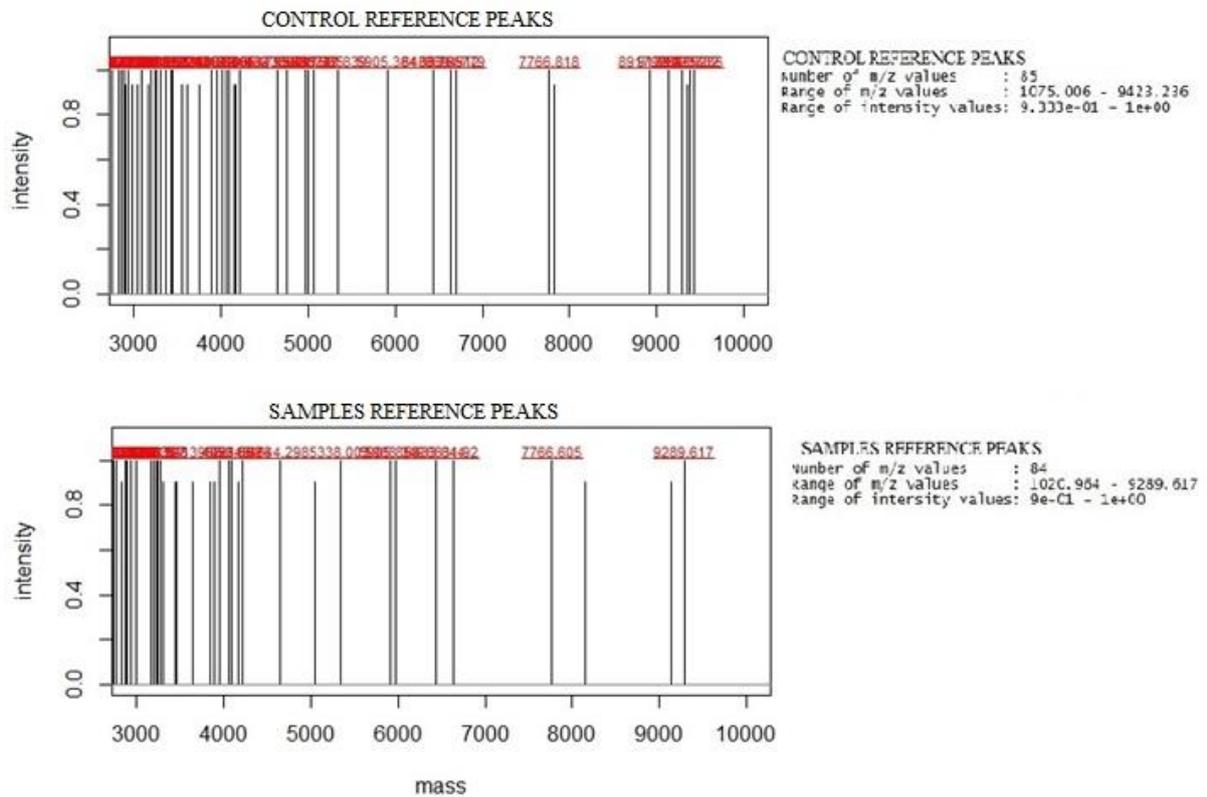


Figura 9: Picos de referencia de los controles y muestras alineados

En este análisis se han detectado 85 picos de referencia en los controles sanos (control reference peaks) y 84 picos de referencia en las muestra MGSi (sample reference peaks).

La última etapa es la identificación de picos en las muestras MGSi que pueden ser posibles biomarcadores.

Para la identificación de biomarcadores se comparan los picos de referencia de todas las muestras controles sanos + muestras MGSi (all reference peaks) con los picos de referencia de las muestras MGSi (samples reference peaks). En el análisis de los picos de referencia, un problema que se ha presentado ha sido determinar cuándo un pico que aparece en controles y en muestras puede ser identificado como un único pico de referencia. Este problema se ha abordado determinando el error asociado a la masa de un pico. A continuación se explica como se ha hecho.

La tabla 2 resume todos los picos de referencia detectados en todas las muestras controles sanos + muestras MGSi, controles sanos y muestras MGSi. Se puede observar por ejemplo que los picos de referencia 1205.669 y 3159.350 tienen un valor diferente pero aproximado en los grupos de todas las muestras controles sanos + MGSi (all sample reference peaks), de los controles sanos (control reference peaks) y de las muestras MGSi (sample reference peaks), por lo tanto, no se sabe si se trata del mismo pico o no. Para determinar si los tres valores detectados pueden ser identificados como un único pico calculamos el error manualmente que se puede admitir en las diferencias de masa, para aceptar dos picos diferentes como un único pico.

Para ver el error de masas en espectrometría de masas se aplica la siguiente fórmula:
$$\frac{(\text{pico mayor} - \text{pico pequeño})}{\text{pico mayor}} \times 10^6.$$

A continuación se aplica la fórmula a los picos mencionados a modo de ejemplo:

$$\frac{(1207.516 - 1204.437)}{1207.516} \times 10^6 = 2549.740$$

$$\frac{(3160.350 - 3158.636)}{3160.350} \times 10^6 = 542.436$$

Se puede observar que en el valor de masas 1205.669 el error es muy elevado (2549.740ppm). En cambio en el valor de masas 3159.350 el error es bastante más pequeño (542.436ppm)

Por eso, en nuestras condiciones experimentales se da por válido que dos picos diferentes son identificados como un único pico cuando el error experimental es igual o inferior a 600ppm.

Como los picos 1205.6691, 1204.4376 y 1207.5164 tienen un error mayor a 600ppm los trataremos como si fueran picos diferentes. En cambio los picos 3159.350, 3158.636 y 3160.350 tienen un error inferior a 600ppm por lo que los trataremos como el mismo pico.

Dado que todos los picos detectados inferiores a 3000Dalton tienen un error superior a 600ppm se aborda el análisis de biomarcadores en el rango de masas entre 3000 y 10000 daltons.

All peaks	Control peaks	Samples peaks	Error
		1021	
		1043	
	1075	1078	2965.6645
1098	1097		796.600644
		1153	
	1192		
		1173	
1205.6691	1204.4376	1207.5164	2549.7405
		1230	
		1252	
	1259	1264	3804.44243
		1286	
	1300		
		1333	
	1348	1352	2735.21999
		1374	
	1387	1390	2516.73148
		1406	
		1420	
		1450	
1465	1464	1467	1968.37019
1488		1489	
1519	1518	1521	2025.51212
1545	1544	1547	1959.24288
	1566	1569	1984.31803
1616	1615	1618	1901.22659
1639	1637	1640	1900.58906
		1662	
1679	1678	1681	1705.7353
		1703	
	1739	1742	2147.84198
		1764	
	1774		
		1790	
1831	1830	1833	1607.31071
		1888	
1897	1896	1898	1322.12214
	1927		
1944	1943	1946	1510.8589
	1967		
	1988	1988	-94.1216865
		2045	
2082	2081	2084	1256.83488
2116	2115	2117	1241.38201
2165	2164	2166	888.627161
		2182	
	2191		
		2195	
2211		2212	
		2295	
2306	2305		-398.324372
2325	2325	2325	-5.41482907
	2355		
2379	2378	2381	1113.36284
	2429		

All peaks	Control peaks	Samples peaks	Error
		2500	
	2529		
		2556	
	2568		
2604	2603		321.147551
		2606	
2642		2643	560.084484
2661	2660		342.048246
		2662	
	2681		
	2709		
		2726	
2755	2754		348.125091
		2757	
2818	2817		369.847443
		2820	
		2864	
2863	2862		286.163786
2884	2883		632.84993
		2886	
	2900		
2934.166	2933.945	2934.496	187.567713
	2973		
		2993	
	3041		
	3085		
3159.350	3158.636	3160.350	542.436
3193	3192	3194	501.289138
		3218	
3242	3241	3243	633.915292
3264	3263	3265	561.23392
3306.3653	3305.6615	3307.5384	567.4824
	3364.512		
3432.903	3431.380		443.762
		3436.088	
3448.597	3447.671	3450.139	715.453
	3542.713		
		3640.987	
	3609.468		
	3749.261		
		3844.838	
3883.962	3883.673	3885.115	371.155
3955.466	3955.545	3955.348	-49.858
	4018.449		
4055	4055	4056	358.617738
4092	4092	4093	148.711357
	4154		
4169	4168	4170	486.562844
4211	4210	4212	295.528306
4643.7488	4643.3828	4644.2978	197.003568
4756.4101	4755.4906		
	4963.7527		
	5004.7426		
		5049.3565	
5065.8674	5065.3645		99.2686
5337.9046	5337.8389	5338.0032	30.7664
5905.5719	5905.3842	5905.8534	79.4532
5968.4641		5968.3489	-19.3094
6434	6434	6434	-18.8075868
6632	6632	6632	1.20144361
6695	6695		-21.1139644
7766.733	7766.818	7766.605	-27.4450827
	7829.873		
		8141.373	
	8917.292		
9133.704	9133.475	9134.086	66.8867323
9289.477	9289.383	9289.617	25.1477243
	9357.788		
	9379.702		
	9423.236		

Tabla 2: Picos detectados en todas las muestras control sanos + MGSI, control sanos y muestras MGSI

Finalmente, la aplicación MALDIquant permite obtener los posibles biomarcadores una vez definido el rango de análisis de 3000-10000 (figura 10).

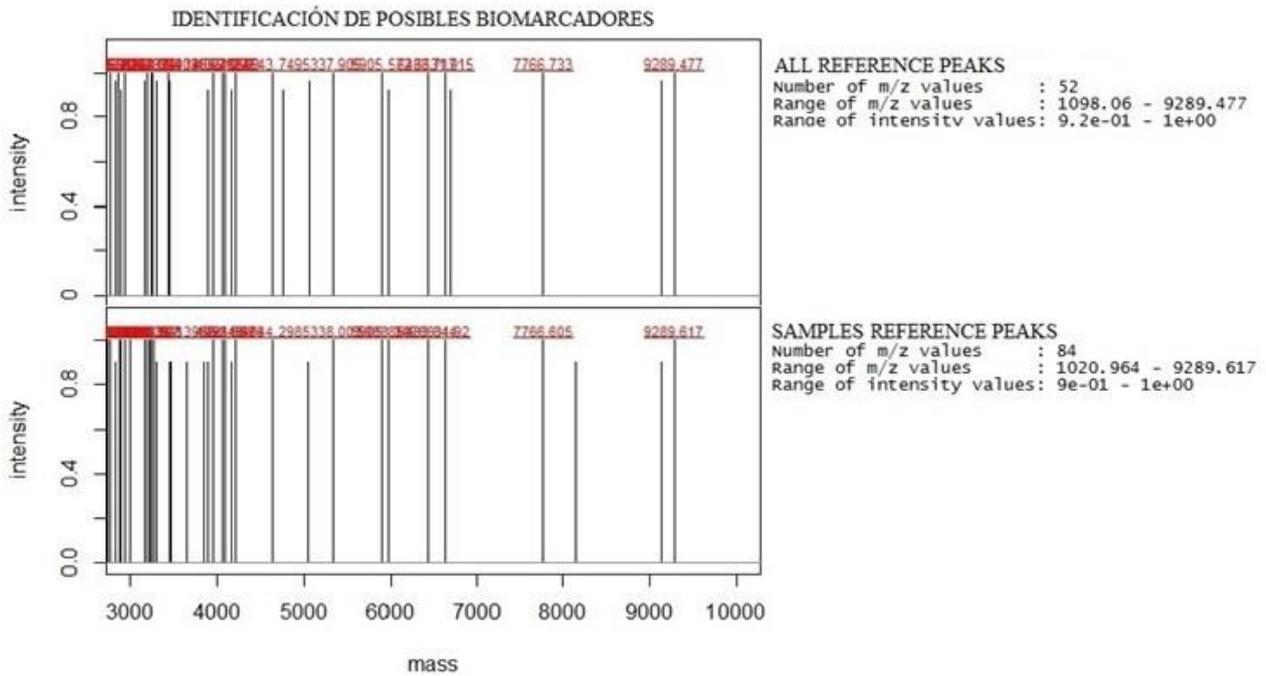


Figura 10: Picos de referencia alineados de todas las muestras control sanos + MGSI y muestras MGSI

Observando esta figura vemos que es difícil diferenciar los posibles biomarcadores debido a la amplia escala de masas. Por ello se ha desglosado esta figura en rangos de m/z: 2500-3500, 3500-4500, 4500-7000, 7000-10000. (figuras 10A, 10B, 10C, 10D)

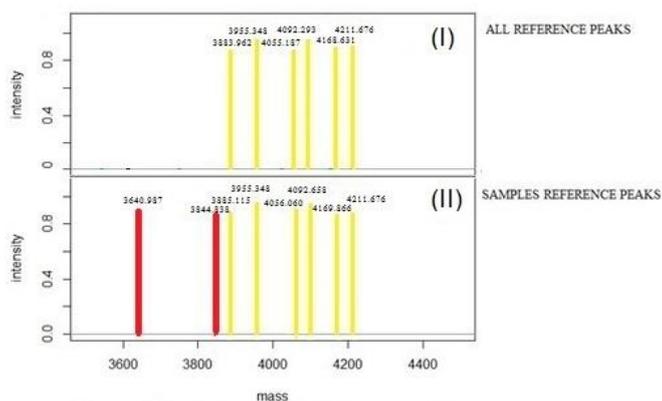


Figura 10A: Alineamiento de los picos de referencia en la zona 3500-4500m/z

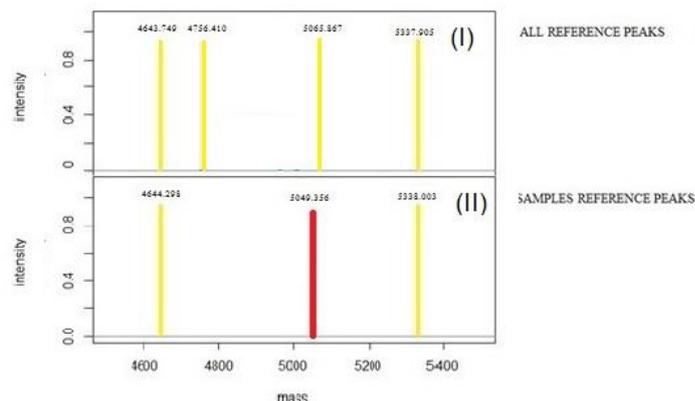


Figura 10B: Alineamiento de los picos de referencia en la zona 4500-5500m/z

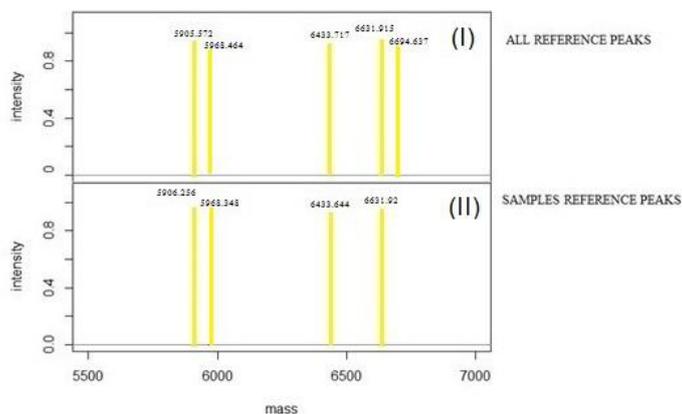


Figura 10C: Alineamiento de los picos de referencia en la zona 5500-7000m/z

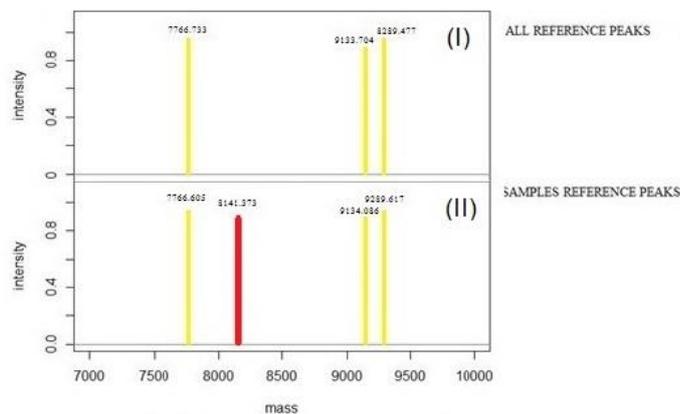


Figura 10D: Alineamiento de los picos de referencia en la zona 7000-10000m/z

Se puede observar que los picos de referencia 4756.410 y 5063.867 en la figura 10B y el pico de referencia 6694.637 en la figura 10C se encuentran en la grafía I, pero no se encuentran en la gráfica II. Esto es debido a la poca variabilidad de las muestras analizadas y al criterio que utiliza el software R-studio para asignar picos comunes.

Por ejemplo: Si el pico A esta en 2 de las 4 muestras MGSi analizadas, significa que esta presente en el 50% de las muestras MGSi, por lo que el programa no puede identificar al pico A como pico común de muestras MGSi. En cambio, si el pico A esta presente en 3 de los 5 controles sanos analizados, significa que esta presente en el 60% de los controles sanos, por lo que el software R-studio identifica el pico A como un pico común de controles sanos.

Cuando analizamos el conjunto de controles sanos + muestras MGSi, el pico A estaría presente en 5 de las 9 muestras totales analizadas, lo que supone que el pico A es común en el 55% de las muestras totales. Aplicando esta logica, el programa identifica el pico A como pico común de las muestras totales controles sanos + MGSi, mientras que no lo identifica como pico común en las muestras MGSi.

Los posibles biomarcadores se deducen eliminando los picos de referencia comunes de todas las muestras controles sanos + MGSi y de las muestras MGSi (graficas I y II). Los picos marcados en rojo que quedan solo en las muestras MGSi (samples reference peaks) son los posibles biomarcadores que podrían asociarse de forma específica al suero de individuos MGSi.

Se han identificado 4 posibles biomarcadores asociados a la patología GMSI en este simple análisis, de un total de 5 controles sanos y 4 muestras MGSi . (Tabla 3).

m/z
3640.987
3844.838
5049.356
8141.373

Tabla 3: Posibles biomarcadores asociados a la patología MGSi a partir del análisis de seroproteínas

Estos posibles biomarcadores no representan picos de proteínas, dado el fundamento de la técnica. Son picos de fragmentos peptídicos obtenidos a partir de las proteínas bombardeadas por el láser en presencia de la matriz.

El resultado de estos 4 posibles biomarcadores no son válidos estadísticamente ya que sólo se han analizado 5 controles sanos y 4 muestras MGSi. Para que este estudio pudiera ser concluyente se deberían analizar como mínimo 70 individuos, según indican los estudios bibliográficos consultados (15).

6. CONCLUSIONES

1. Se han identificado 4 posibles biomarcadores de seroproteínas para GMSI mediante la técnica proteómica de espectrometría de masas MALDI-TOF
2. Los resultados obtenidos no son válidos estadísticamente debido al número limitado de muestras que se han analizado. Para que el estudio sea significativo se debería analizar un número elevado de muestras.

7. BIBLIOGRAFÍA

- 1- Jesús Vázquez ; Presente y futuro de la proteómica: CSIC-UAM; ISSN 1695-0623, http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDAQFjAA&url=http%3A%2F%2Fwww2.cbm.uam.es%2Fjvazquez%2FPDFs%2FProteomica.pdf&ei=bXt_U-_OHoPV0QW6xoHIDA&usg=AFQjCNHwrcFhJrtgaGdYliMe_HlxTRRFdg
- 2- Daniela Morales Sánchez, Lilí Esmeralda Gallo Ramírez; Plataforma de proteómica; http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDYQFjAB&url=http%3A%2F%2Fwww.ibt.unam.mx%2Fcomputo%2Fpdfs%2Fmet%2Fplataformas_de_proteomica.pdf&ei=ydKAU7-zHquX0QWR1YGYCQ&usg=AFQjCNGa1f0NWDeB0y-oQI7_-w13waXGWg
- 3- Identificación de proteínas por espectrometría de masas. Equipamiento https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDAQFjAA&url=https%3A%2F%2Fwww.uco.es%2Fservicios%2Fscai%2Fpdf%2Fidentificacion_de_proteinas.pdf&ei=4nl_U8q0J8-z0QXU54CYBg&usg=AFQjCNEyNtjr000pd9kumz-W3oD8FwpUqA
- 4- <http://es.wikipedia.org/wiki/MALDI-TOF>
- 5- http://es.wikipedia.org/wiki/Huella_pept%C3%ADdica
- 6- Juan Luis Muñoz Bellido, Silvia Vega Castaño, Laura Ferreira, Fernando Sánchez Juanes y José Manuel González Buitrago; Revisión: Aplicaciones de la proteómica en el laboratorio de Microbiología Clínica: *Enferm Infecc Microbiol Clin.* 2012;30(7):383–393
- 7- Paulette Legarraga, Marcela Moraga, Marusella Lam, Enrique Geoffroy, Cecilia Zumarán y Patricia García; Impacto de la espectrometría de masas por MALDI-TOF MS en la identificación rápida de bacterias aeróbicas y anaeróbicas de importancia clínica; *Rev Chilena Infectol* 2013; 30 (2): 140-146
- 8- M. Martín-Carrasco; Biomarcadores en la enfermedad de Alzheimer: definición, significación diagnóstica y utilidad clínica; *Psicogeriatría* 2009; 1 (2): 101-114
- 9- M.^a Carmen Martín Gómez; Milagros Ballesteros González. Espectrometría de masas y análisis de biomarcadores; <http://www.analesranf.com/index.php/mono/article/view/1066/1063>
- 10- Radhakrishna S. Tirumalai, King C. Chan, DaRue A. Prieto, Haleem J. Issaq, Thomas P. Conrads, and Timothy D. Veenstra; Characterization of the Low Molecular Weight Human; *Mol Cell Proteomics* 2003; 2: 1096-1103
- 11- Qing- Tao Wang, Yong-Zhe Li, Yu-Fang Liang, Chao-Jun Hu, Yu-Hua Zhai, Guan-Fei Zhao, Jian Zhang, Ning Li, An-Ping Ni, Wen-Ming Chen, and Yang Xu; Construction of A Multiple Myeloma Diagnostic Model by Magnetic Bead-Based MALDI-TOF Mass Spectrometry of Serum and Pattern Recognition Software; *The anatomical record* 2009; 292: 604–610
- 12- <http://gammapatiasmonoclonales.blogspot.com.es/>
- 13- Yaniv Bledi, Alex Inberg and Michal Linial; A proteomic method for analysing plasma membrane proteins in living mammalian cells; *Briefing in functional genomics and proteomics* 09-2003; 2(3): 254–265
- 14- Elena Jordana-Lluch, Elisa Martró Català, Vicente Ausina Ruiz; La espectrometría de masas en el laboratorio de microbiología clínica; *Enferm Infecc Microbiol Clin.* 2012;30:635-44

15- María del Monte Moreno Jiménez; Estudio retrospectivo de pacientes con gammapatía monoclonal de significado incierto en el área de salud 6 de Valencia; 2010; http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCsQFjAA&url=http%3A%2F%2Fwww.recercat.net%2Fbitstream%2Fhandle%2F2072%2F199064%2FTR-MorenoJim%25C3%25A9nez.pdf%3Fsequence%3D1&ei=AlePU7_5GOSp0QXXhoH4CA&usg=AFQjCNF0_bx3wYF_Z8Z6A3dgkmVAaNyShw

AGRADECIMIENTOS

Quiero expresar mis agradecimientos al laboratorio científico Técnico de la Universidad de las Islas Baleares y concretamente a la Dra. Rosa Gomila como responsable del instrumento de espectrometría de masas MALDI-TOF, por la ayuda y atención que me ha prestado que ha hecho posible la realización de este trabajo de final de grado.

También deseo expresar mis agradecimientos a mi tutora la Dra. Francisca Barceló Mairata por su ayuda y guía durante todo el trabajo de final de grado.