

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

DESARROLLO DE MÉTODOS PARA EL MEJORAMIENTO DE LA CAPACIDAD DE IDENTIFICACIÓN DE PÉPTIDOS MEDIANTE LA ESPECTROMETRÍA DE MASAS POR TIEMPO DE VUELO

PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE FÍSICO

MIGUEL FRANCISCO GARCÍA VERA

miguel_garcia@whoever.com

Director: César Costa Vera, Ph.D.

cesar.costa@epn.edu.ec

Quito, Junio de 2011

DECLARACIÓN

Yo, Miguel Francisco García Vera, declaro que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional, puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Miguel Francisco García Vera

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Miguel Francisco García Vera, bajo mi supervisión.

César Costa Vera, Ph.D.
DIRECTOR DE PROYECTO

AGRADECIMIENTOS

En primer lugar agradezco a mis padres por haberme dado la oportunidad de estudiar y darme su apoyo a lo largo de mi vida. Por haber estado siempre conmigo en los momentos más difíciles y haber confiado en mí en todo momento.

A la Escuela Politécnica Nacional, por haberme dado la formación necesaria para convertirme en un profesional capacitado. A mis profesores durante estos largos años de estudio, a mis compañeros de clase y de laboratorio con quienes tuve la oportunidad de compartir interminables experiencias dentro y fuera de las aulas. De manera muy especial quisiera agradecer a mi director de tesis el Dr. César Costa Vera, quien hizo posible mi participación en este proyecto de investigación, por su amistad, consejos, y por el tiempo dedicado a la revisión y supervisión de este trabajo.

A la Unidad de Espectrometría de masas del Departamento de Parasitología del Centro Médico de la Universidad de Leiden (LUMC), por medio de su director Prof. Andree Deelder. A Kate, Rico, Bart, Simone, Tiziana y al resto de compañeros del laboratorio en Leiden, y de manera muy especial a mi supervisor Magnus Palmblad, por su constante apoyo y enseñanzas que hicieron posible la realización de gran parte de este trabajo.

Al Centro Internacional de Zoonosis (CIZ), a Richar y Elizabeth, con quienes tuve la oportunidad de trabajar durante el desarrollo de este proyecto.

A la Fundación Dr. P. C. Flu por el apoyo con los recursos económicos durante mi estadía en Leiden.

DEDICATORIA

Dedico este trabajo a mis padres, que han estado siempre a mi lado.

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS	I
LISTA DE TABLAS	III
RESUMEN	IV
ABSTRACT	V
DECLARACIÓN DE AUTORÍA	VI
1. INTRODUCCIÓN	1
1.1. OBJETIVOS	4
1.1.1. OBJETIVO GENERAL	4
1.1.2. OBJETIVOS ESPECÍFICOS	4
2. FUNDAMENTOS DE LA ESPECTROMETRÍA DE MASAS	5
2.1. CONCEPTOS BÁSICOS	6
2.2. FUENTES DE IONIZACIÓN	8
2.2.1. ELECTROSPRAY	10
2.3. ANALIZADORES DE MASA	12
2.3.1. ANALIZADORES QUADRUPOLES	15
2.3.2. ANALIZADORES DE TIEMPO DE VUELO	20
2.3.2.1. Extracción Ortogonal	21
2.3.2.2. Reflectrón	23
2.4. ESPECTROMETRÍA DE MASAS ITERADA (TANDEM)	25
2.4.1. ESPECTRÓMETRO DE MASAS QqTOF	27
2.5. TÉCNICAS DE FRAGMENTACIÓN	28
2.5.1. NOMENCLATURA	28
2.5.2. DISOCIACIÓN INDUCIDA POR COLISIÓN	31

2.5.3.	DISOCIACIÓN POR CAPTURA ELECTRÓNICA Y DISOCIACIÓN POR TRANSFERENCIA ELECTRÓNICA	36
3.	BASES DE LA PROTEÓMICA	38
3.1.	CONCEPTOS BÁSICOS	39
3.2.	SEPARACIÓN DE PÉPTIDOS Y PROTEÍNAS	43
3.2.1.	DIGESTIÓN ENZIMÁTICA DE PROTEÍNAS	44
3.2.1.1.	Tripsina	44
3.2.1.2.	Digestión <i>In Silico</i>	45
3.2.2.	CROMATOGRAFÍA LÍQUIDA	46
3.2.2.1.	Cromatografía de fase normal	47
3.2.2.2.	Cromatografía de fase reversa	47
3.2.2.3.	Cromatografía de intercambio iónico	48
3.2.2.4.	Cromatografía por afinidad molecular	48
3.2.2.5.	Cromatografía de exclusión molecular	48
3.3.	IDENTIFICACIÓN Y CARACTERIZACIÓN DE PÉPTIDOS Y PROTEÍNAS POR MEDIO DE MS	50
3.3.1.	TRATAMIENTO DE LOS ESPECTROS DE MASA	51
3.3.1.1.	Deisotopización y determinación del estado de carga	52
3.3.1.2.	Promedio de espectros de masa	52
3.3.1.3.	Sustracción del ruido de fondo (Background)	52
3.3.1.4.	Espectros de masa con centroides	53
3.3.2.	IDENTIFICACIÓN POR MEDIO DE MS (PMF: PEPTIDE MASS FINGERPRINTING)	53
3.3.2.1.	Scores en la identificación de proteínas	54
3.3.3.	ANÁLISIS POR MEDIO DE LA ESPECTROMETRÍA DE MASAS TANDEM	55
3.3.4.	BASES DE DATOS	56
3.3.5.	<i>DE NOVO</i> SEQUENCING	58
4.	METODOLOGÍA EXPERIMENTAL	60
4.1.	EXTRACCIÓN DE PROTEÍNAS EN UN CULTIVO DE <i>ESCHERICHIA COLI</i>	60
4.2.	SEPARACIÓN CROMATOGRÁFICA (HPLC)	61
4.3.	ESPECTROMETRÍA DE MASAS QqTOF	61
4.4.	FRAGMENTACIÓN CID A DIFERENTES ENERGÍAS	62
4.5.	PROCESAMIENTO DE DATOS	63
4.6.	IDENTIFICACIÓN POR MEDIO DE BASE DE DATOS	64

4.7. DE NOVO SEQUENCING	64
4.8. RECALIBRACIÓN DE LOS ESPECTROS DE MS Y MS/MS	66
4.8.1. ELECCIÓN DE LOS CALIBRANTES INTERNOS	67
4.8.2. CONVERSIÓN DE LOS VALORES DE m/z A UNIDADES DE TIEM- PO DE VUELO FICTICIAS	68
4.8.3. RECALIBRACIÓN	69
4.9. PROGRAMACIÓN DE RUTINAS	70
4.9.1. RECALIBRACIÓN	70
4.9.2. ANÁLISIS DE LOS ESPECTROS DE FRAGMENTACIÓN A DIFERENTES ENERGÍAS DE COLISIÓN	71
4.9.2.1. Conjunto de Aprendizaje	72
4.9.2.2. Espectros de fragmentación con 35 eV	73
4.9.2.3. Espectros de fragmentación con 55 eV	74
5. RESULTADOS Y DISCUSIÓN	76
5.1. RECALIBRACIÓN DE LOS ESPECTROS DE MS Y MS/MS	76
5.1.1. RELACIÓN ENTRE LOS PARÁMETROS DE LA ECUACIÓN DE RECALIBRACIÓN	80
5.2. INVESTIGACIÓN DE LOS ESPECTROS DE FRAGMENTACIÓN CID A DIFERENTES ENERGÍAS DE COLISIÓN	87
5.2.1. RESULTADO DEL CONJUNTO DE DATOS DE APRENDIZAJE	89
5.2.2. RESULTADO DEL RE-SCORING DE LAS IDENTIFICACIONES OBTENIDAS CON PEPNOVO	93
5.3. LIMITACIONES Y ALCANCE	94
6. CONCLUSIONES Y RECOMENDACIONES	96
6.1. CONCLUSIONES	96
6.2. RECOMENDACIONES	98
REFERENCIAS	100
ANEXOS	110

LISTA DE FIGURAS

2.1. Esquema de un espectrómetro de masas	6
2.2. Resolución de un pico en el espectro de masas	7
2.3. Esquema de un espectrómetro de masas tandem	9
2.4. Esquema de un sistema de ESI	10
2.5. Cono de Taylor en la ionización por ESI	11
2.6. Esquema de un analizador de masas cuadrupolar	15
2.7. Algunas funciones angulares de Mathieu	17
2.8. Diagrama de estabilidad en la ecuación de Mathieu	18
2.9. Diagramas de estabilidad en el espacio (U,V)	19
2.10. Esquema de un oa-TOF	22
2.11. Pulso de extracción en un oa-TOF	24
2.12. Esquema del reflectrón o espejo electrostático	25
2.13. Nomenclatura de los fragmentos de la columna del péptido	29
2.14. Fragmento interno R2R3	29
2.15. Estructura de un fragmento de ion imonio	29
2.16. Distribución de Energías Internas en CID	33
2.17. Ubicación del protón previa la fragmentación en la columna del péptido	35
3.1. Estructura de un péptido	40
3.2. Perfiles Isotópicos	42
3.3. Esquema de un sistema de HPLC	49
3.4. Ejemplo de un cromatograma	50
4.1. Ultimate 3000 nano LC system	61
4.2. maXis QqTOF	62
4.3. Esquema de la lista de potenciales calibrantes en el proceso de recalibración	68
4.4. Esquema de la lista de péptidos identificados por Pepnovo	72

5.1. Distribución de los errores antes de la recalibración (S1)	77
5.2. Histograma de los errores (luego de recalibrar) para los picos incluidos y excluidos en el proceso de recalibración	79
5.3. Distribución de los errores en función de la intensidad (S1)	81
5.4. Gráfico de los parámetros a_i de la Función 1 vs. Número de Scan. Set de datos S1.	82
5.5. Relación entre los parámetros a_i de la Función 1	83
5.6. Diagrama de flujo del proceso de recalibración	86
5.7. Diagrama de flujo del proceso de re-scoring	88
5.8. Espectro de Fragmentación a 15 eV, 35 eV y 55 eV	91
5.9. Histograma del número de iones immonio (identificados) por péptido	92
6.1. Diagrama de flujo para la identificación de péptidos mediante los métodos propuestos	99

LISTA DE TABLAS

2.1. Comparación de las características de algunos analizadores de masas . . .	14
2.2. Masa molecular de varios iones fragmentos	30
3.1. Masas moleculares de los 20 aminoácidos más comunes	41
3.2. Isótopos estables en los compuestos orgánicos	42
3.3. Especificidad de las proteasas más comunes	44
4.1. Conjuntos de datos utilizados	63
4.2. Parámetros de búsqueda en la base de datos	65
4.3. Parámetros para de novo sequencing en PepNovo	66
4.4. Iones immonio utilizados en el procedimiento de re-scoring	75
5.1. Errores antes de la recalibración	77
5.2. Media \pm Desviación Estándar (ppm) del error global para los datos, antes y después de la recalibración con las funciones utilizadas	78
5.3. Errores y estadísticas de la recalibración considerando la relación entre los parámetros a_i en la Función 1.	84
5.4. Resultados de <i>de novo</i> sequencing	87
5.5. Longitud y porcentaje de cobertura de la secuencia por parte de los frag- mentos observados	89
5.6. Longitud y porcentaje de cobertura de la secuencia por parte de los frag- mentos observados (2+ y 3+)	90
5.7. Parámetros α_t^{ch}	92
5.8. Resultados del procedimiento de <i>de novo</i> sequencing luego del re-scoring	94

RESUMEN

La metodología estándar para la identificación de péptidos y proteínas en una muestra de material biológico desconocido consiste en la utilización de técnicas de fragmentación molecular, como es el caso la disociación inducida por colisiones (CID), para identificar propiedades estructurales, tales como la secuencia primaria de aminoácidos, o las posibles modificaciones postraslacionales (PTMs). En este trabajo se introducen dos métodos basados en el uso conjunto de TOF-MS y CID para mejorar la tasa de identificación de péptidos y proteínas mediante la espectrometría de masas.

El primer método se desarrolló tomando en cuenta que la precisión en la determinación de la masa es un aspecto fundamental para la identificación de péptidos y proteínas. Hemos desarrollado un algoritmo computacional automático (no supervisado) para recalibrar conjuntos de datos de corridas completas de LC/TOF-MS adquiridos de muestras complejas como digestiones de cultivos de poblaciones enteras de células de algún organismo. Adicionalmente, comparamos varias ecuaciones genéricas de calibración de espectros TOF con distinto número de parámetros libres. El error promedio en la determinación de la masa se redujo de decenas de ppm a menos de un ppm. La dispersión del error mejoró hasta diez veces luego de la recalibración automática.

El segundo método involucra una mejora en la capacidad de identificación de péptidos mediante *de novo* sequencing. Se introduce un esquema de re-scoring de los péptidos identificados previamente mediante un programa comercial para realizar *de novo* sequencing, que utiliza la información presente en los espectros de fragmentación a diferentes energías de colisión. En esta tesis se consideraron tres conjuntos de datos adquiridos a 15 eV, 35 eV y 55 eV de energía de colisión en el sistema de centro de masa. Los datos adquiridos a energías de colisión de 35 eV y 55 eV fueron utilizados para obtener información de canales de fragmentación en fragmentos de tipo *a*, *c*, *x*, *z* y iones immonio. La tasa de aminoácidos y péptidos correctamente predichos luego del procedimiento de re-scoring se incrementó hasta en un 10 %.

Palabras clave: Espectrometría de masas por tiempo de vuelo, Disociación inducida por colisiones, Proteómica.

ABSTRACT

A standard methodology for peptide and protein identification in complex biological samples uses mass spectrometry together with fragmentation techniques as collision induced dissociation (CID) to obtain structural information such as the primary sequence of aminoacids, or possible post-translational modifications (PTMs). In this work, we introduce two methodologies based on the combination of TOF-MS and CID to improve the rate of successful peptide and protein identifications using mass spectrometry.

The first approach was developed by noticing the fact that mass accuracy in MS and MS/MS are essential for peptide and protein identification. We have developed an unsupervised (automatic) computer algorithm to recalibrate entire LC/TOF-MS datasets acquired from complex peptide mixtures, such as whole cell-digests. Additionally, we compare several TOF mass calibration equations with a varying number of free parameters. The average (systematic) mass measurement errors in the original datasets could be reduced from tens of ppm to less than one ppm. The spread of random errors or precision improved ten-fold after automatic recalibration.

The second method is intended to improved the rate of correct peptide identifications using *de novo* sequencing algorithms. We introduced a workflow for the re-scoring of peptides previously identified using a commercial program for *de novo* sequencing, that uses the additional information obtained from the spectra acquired at different collision energies. In this work, three datasets were acquired using collision energies of 15 eV, 35 eV and 55 eV in the center of mass system of reference. The spectra acquired at 35 eV and 55 eV were used to evaluate fragmentation of type *a*, *c*, *x*, *z* and the presence of immonium ions. The rate of peptides and aminoacids correctly identified after the re-scoring procedure was increased about a 10 %.

Keywords: Time of flight mass spectrometry, Collision-induced dissociation, Proteomics.

DECLARACIÓN DE AUTORÍA

En el presente trabajo se introducen dos métodos para mejorar la tasa de identificación de péptidos en muestras biológicas mediante disociación inducida por colisiones y espectrometría de masas de tiempo de vuelo. El primero consiste en una recalibración automática de espectros de tiempo de vuelo obtenidos en corridas de LC-MS/MS. El segundo es un esquema de re-scoring para mejorar la identificación de péptidos mediante *de novo* sequencing, que utiliza la información de fragmentación en varias energías de colisión CID.

Los capítulos 2 y 3 son una revisión bibliográfica de los aspectos de la espectrometría de masas y la proteómica que son relevantes para la presente investigación.

En el capítulo 4 se describe el set-up experimental utilizado para la adquisición de datos y el procesamiento de los mismos. Se utilizaron un total de seis sets de datos de LC-MS/MS, de los cuales cuatro fueron obtenidos experimentalmente por el autor como parte de este trabajo, y dos de ellos (S3 y S4) fueron una contribución de Magnus Palmblad (LUMC). Para el proceso de recalibración se utilizó como partida el trabajo de Palmblad et. al. [7] referente a recalibración de espectros de FTICR-MS, y el trabajo previo de Costa-Vera et. al. [8] referente a las ecuaciones de recalibración de espectros de TOF-MS. En base a estos estudios previos, el trabajo del autor fue modificarlos y adaptarlos para la recalibración automática de corridas completas de digestiones enzimáticas de muestras biológicas, procesadas por medio de LC-TOF-MS/MS. El esquema de re-scoring utilizando tres distintos niveles de energía, la definición de los scores utilizados y el desarrollo del esquema de trabajo es contribución del autor.

La programación de rutinas en *Mathematica* tiene como base scripts previos desarrollados por César Costa Vera (EPN). La programación en C++ fue realizada en su totalidad por el autor.

Todos los resultados presentados en el capítulo 5 fueron obtenidos a lo largo del desarrollo de esta investigación.

CAPÍTULO 1

INTRODUCCIÓN

El estudio de las proteínas expresadas en los organismos vivos es un problema que actualmente presenta gran interés no sólo en el área de la biología básica, sino principalmente por sus aplicaciones en áreas como la medicina, y la biotecnología [1]. La proteómica es el área que se encarga del estudio de este conjunto de proteínas expresadas en un organismo vivo (el proteoma). Para esto utiliza un conjunto de técnicas complementarias que permiten obtener información de la estructura, función y naturaleza de las proteínas estudiadas [1].

Durante las últimas décadas esta rama ha crecido enormemente gracias al apoyo de la espectrometría de masas. La espectrometría de masas brinda una plataforma para estudiar el proteoma que se caracteriza por su versatilidad, rapidez, y alta sensibilidad [2]. El enfoque predilecto utilizado con la espectrometría de masas se conoce como de abajo-arriba (bottom-up o shotgun proteomics en inglés), en el que se realiza primero una digestión proteolítica de las proteínas a estudiar y luego se analizan los péptidos resultantes mediante espectrometría de masas iterada (tándem) [3–5].

La identificación de péptidos y posteriormente de proteínas, normalmente se realiza por medio de la comparación de los espectros de masas experimentales y espectros teóricos obtenidos de una base de datos o del genoma del organismo (si se conoce) en donde se encuentran las proteínas que se pretende identificar. Este método si bien ha probado tener una alta eficiencia en el proceso de identificación y caracterización de muestras complejas, presenta la desventaja de que si la proteína no se encuentra en la base de datos, será imposible identificarla en la muestra. Una forma de enfrentar este problema es lo que se conoce como *de novo* sequencing y básicamente consiste en

obtener la secuencia de aminoácidos a partir de la información presente en los espectros de masa tándem sin contar con ninguna información adicional [6]. Un punto crítico para la aplicación del método es su gran dependencia de la calidad y resolución de los espectros de masa de fragmentación.

En este trabajo se introducen dos metodologías para mejorar la tasa de identificación de péptidos mediante espectrometría de masas de tiempo de vuelo. El primero consiste en un proceso de recalibración que permite aumentar la precisión en la determinación de la masa por medio del estudio de las ecuaciones de tiempo de vuelo. El procedimiento de recalibración que se propone tiene la ventaja de que no depende del conocimiento de las unidades originales de tiempo, a las cuales no siempre se tiene acceso en equipos comerciales. Este trabajo se basa en dos trabajos previos realizados por Palmblad et.al. [7] y Costa et. al. [8]. El segundo método consiste en un procedimiento de re-scoring de las identificaciones realizadas con un programa comercial de *de novo* sequencing (PepNovo®) al utilizar espectros de fragmentación con diferentes energías en la cámara de colisión (fragmentación por medio de disociación inducida por colisión o CID por sus siglas en inglés).

El trabajo experimental y la adquisición de datos fueron realizados en los laboratorios del Departamento de Parasitología, Unidad de Espectrometría de Masas del Centro Médico de la Universidad de Leiden (LUMC) en la ciudad de Leiden, Holanda, en el marco de una colaboración con el Departamento de Física de la Escuela Politécnica Nacional.

Para el procedimiento de recalibración se utilizaron cuatro conjuntos de datos de la digestión proteolítica de cultivos de *Escherichia Coli* obtenidos mediante dos espectrómetros comerciales distintos y con diferentes calidades de calibración originales. Se utilizó la calibración original del fabricante para hacer una identificación preliminar de los péptidos presentes en la muestra y utilizarlos como calibrantes internos para el proceso de recalibración. Para la recalibración se probaron distintas funciones matemáticas de calibración, y se compararon los errores con respecto a la calibración original. Los resultados obtenidos muestran que el algoritmo eficientemente produce una distribución de errores centrados alrededor de cero y reduce la dispersión en el error entre 2 y 20 veces dependiendo de la calidad de la calibración original. En el mejor de los casos se obtuvo una dispersión del error menor a 1 ppm luego de la recalibración.

Para el procedimiento de re-scoring con diferentes energías de colisión se utilizaron 3 energías de colisión para evaluar distintos tipos de iones fragmento. Los resultados se evaluaron utilizando un programa comercial de libre distribución para realizar

de novo sequencing (PepNovo®). El objetivo del procedimiento de re-scoring fue escoger el mejor candidato entre 5 secuencias predichas por PepNovo y determinar cual de ellas tiene la mayor probabilidad de ser el origen del espectro de fragmentación observado. De esta manera, los resultados originalmente rankeados según PepNovo fueron re-evaluados al utilizar un score que toma en cuenta la información adicional de la fragmentación a 35 eV y 55 eV. Las mejoras en el número de péptidos correctamente identificados ascendieron a valores superiores al 10 %. El número total de aminoácidos correctamente predicho también mostró incrementos cercanos al 10 %.

Este trabajo se organiza de la siguiente manera. En el Capítulo 1 se presenta una introducción general al trabajo. En el Capítulo 2 se describen los principios físicos e instrumentales básicos de la espectrometría de masas, con especial atención al tema de los analizadores cuadrupolares, los analizadores de tiempo de vuelo y la fragmentación por disociación inducida por colisión (CID). En el Capítulo 3 se describen los conceptos básicos de la proteómica basada en espectrometría de masas. Se presentan las técnicas de separación de péptidos y proteínas y se introduce a la identificación de péptidos mediante espectrometría de masas tándem (de este punto en adelante se utiliza la palabra inglesa *tandem*, sin tilde, de acuerdo con la convención internacional). En el Capítulo 4 se describe la metodología experimental desde la adquisición de datos hasta los métodos utilizados para la recalibración y para la definición del score considerando diferentes energías de colisión. En el Capítulo 5 se muestran y discuten los resultados obtenidos al utilizar la metodología descrita en el capítulo 4. Se presentan en primera instancia los resultados del procedimiento de recalibración, seguidos por los resultados de *de novo* sequencing antes y después del procedimiento de re-scoring. En el Capítulo 6 se presentan las conclusiones obtenidas en el presente trabajo, las recomendaciones y posibles trabajos futuros que surgen a partir de esta investigación.

Debajo se presenta un esquema de los objetivos que se plantearon al inicio de esta investigación.

1.1. OBJETIVOS

1.1.1. OBJETIVO GENERAL

Estudiar las bases físicas de la espectrometría de masas por tiempo de vuelo (TOF-MS) y los métodos de la proteómica para formular y aplicar técnicas de mejoramiento del procesos de identificación de péptidos generados mediante disociación inducida por colisión de haces moleculares en alto vacío.

1.1.2. OBJETIVOS ESPECÍFICOS

- Investigar las bases físicas y químicas de los métodos utilizados para identificar péptidos mediante TOF-MS.
- Estudiar experimental y teóricamente los fenómenos físicos relacionados con la fragmentación de iones de macromoléculas mediante la disociación inducida por colisión con moléculas de gas residual y su efecto en los espectros de masas.
- Usar la información espectral producto de la fragmentación molecular inducida por colisiones en niveles de energía crecientes para determinar la secuencia primaria de péptidos y proteínas con alta confiabilidad.
- Estudiar el problema de la calibración de los espectros de masas de los iones padre y de los fragmentos considerando los procesos físicos de la espectrometría de masas.
- Formular un esquema automático de recalibración de espectros de masa para mejorar la precisión en la determinación de la masa molecular y consecuentemente, la asignación de secuencias primarias de péptidos identificados por espectrometría de masas.
- Aplicar los métodos desarrollados para la identificación de péptidos y proteínas en una muestra de digestión enzimática de *Escherichia Coli* para evaluar su rendimiento y confiabilidad.

CAPÍTULO 2

FUNDAMENTOS DE LA ESPECTROMETRÍA DE MASAS

La espectrometría de masas incluye un conjunto de técnicas analíticas que permiten determinar de manera muy precisa la masa molecular de sustancias desconocidas; y en ciertos casos hacer análisis cuantitativos de los componentes presentes en una muestra [9, 10]. El funcionamiento de un espectrómetro de masas depende del tipo de espectrómetro pero en todos los casos se basa en principios físicos ligados con el movimiento de partículas cargadas en campos eléctrico - magnéticos. La interacción de los iones con los campos electromagnéticos producen distintos efectos físicos en los iones que están relacionados con la relación de masa (m) sobre carga (z), de manera que lo que se mide en un espectrómetro no es directamente la masa sino más bien la relación m/z .

Un espectrómetro de masas se compone de tres partes fundamentales:

1. Fuente de Ionización
2. Analizador de Masas
3. Detector

La fuente de ionización es la primera componente de un espectrómetro de masas y es fundamental para su funcionamiento pues los espectrómetros de masa miden únicamente la masa de moléculas ionizadas. Desde este punto de vista, se busca ionizar la mayor cantidad de moléculas posibles sin afectar la integridad de las mismas, es decir, sin producir fragmentación indeseada de las moléculas presentes en la muestra.

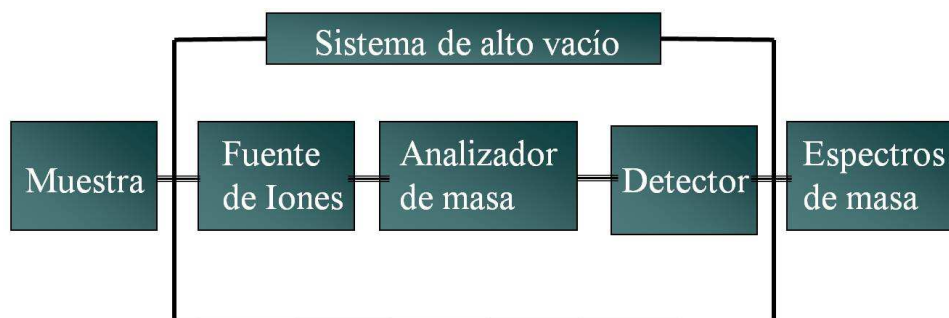


Figura 2.1: Esquema de un espectrómetro de masas

La segunda etapa de un espectrómetro de masas es el analizador de masas y es el corazón del espectrómetro en el sentido de que el analizador es el que separa los diferentes compuestos de acuerdo a su valor de m/z . Existen diferentes tipos de analizadores de masa que se diferencian en la forma de aplicar los campos electromagnéticos para separar a las especies con distinto m/z .

Finalmente, está la etapa de detección, en la cual se obtiene la información proveniente del analizador de masa y se registran los correspondientes espectros de masa.

2.1. CONCEPTOS BÁSICOS

Ion: Se denomina Ion a cualquier partícula o grupo de partículas, que presentan una carga eléctrica no nula.

Isótopo: Dos elementos se conocen como isótopos si tienen el mismo número atómico (es decir, el mismo número de protones en su núcleo) pero diferente número de masa (es decir, un número distinto de neutrones en su núcleo).

Masa Promedio: Masa de un compuesto calculada usando todos sus isótopos y su abundancia relativa.

Masa Monoisotópica: Masa exacta calculada el isótopo más abundante. Para el caso de moléculas se obtiene con la combinación de los isótopos más abundantes de los elementos presentes en la muestra.

Masa Nominal: Valor obtenido al sumar las masas de los isótopos más abundantes redondeadas al número entero más próximo. Es equivalente a la suma de los números

de masa de los elementos constituyentes.

Poder de Resolución: Se define como poder de resolución a la relación entre el valor de m/z correspondiente al centroide de un pico en el espectro y su ancho a media altura; es decir, el poder de resolución es igual a $m/\Delta m$, donde m debe entenderse como m/z . Ver [Figura 2.2](#).

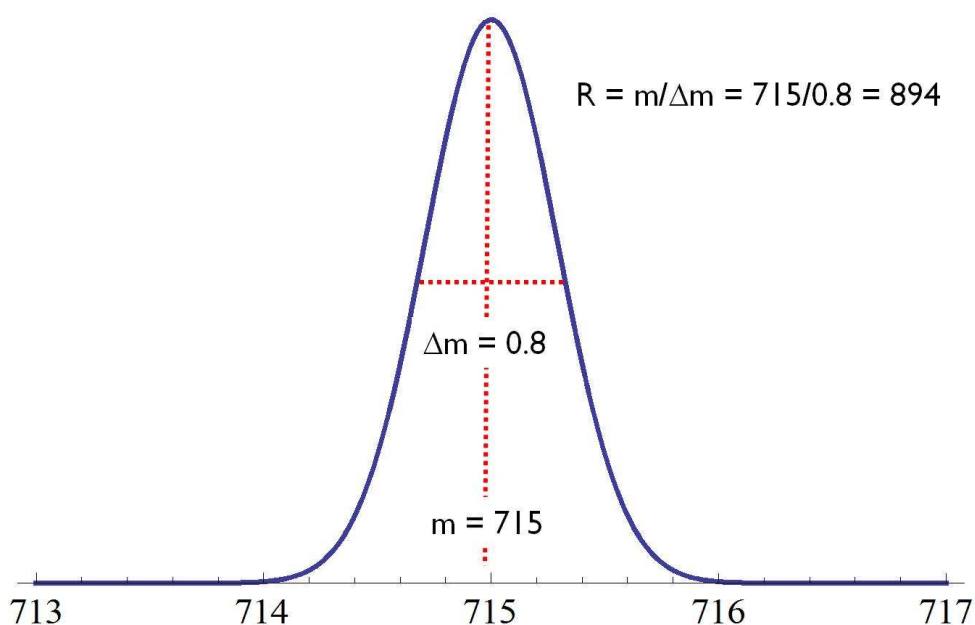


Figura 2.2: Resolución(R) de un pico en el espectro de masas.

Calibración: Calibrar un espectro de masas es una de las tareas más importantes cuando se trabaja con los datos obtenidos del analizador de masa. Calibrar un espectro significa transformar las magnitudes físicas directamente registradas por el sistema detector del espectrómetro de masas (por ejemplo el tiempo de vuelo en un analizador TOF o el espectro de frecuencias en un analizador de resonancia ciclotrónica) a su valor correspondiente de m/z . Dependiendo del tipo de analizador, la ecuación de calibración es diferente y muchas veces incorpora efectos térmicos del equipo, efectos por las condiciones iniciales de generación de iones, o efectos espaciales de carga, entre otros.

Ion precursor: Ion precursor es aquel que es seleccionado para reaccionar y producir iones producto. La reacción puede ser unimolecular (disociación, cambio del estado de carga, isomerización, etc), o reacciones ion/ion o ion/molécula. En espectrometría de masas mucha veces corresponde al ion de la especie molecular completa o entera, antes de ser fragmentada o rota.

Ion producto: Los iones producto son el resultado de la reacción que sufre el ion precursor. En el caso de disociación molecular los iones producto se conocen generalmente con el nombre de iones fragmento. En general corresponde a uno de los iones resultantes de la fragmentación del ion precursor.

Espectrometría de Masas Tandem (MS/MS y MSⁿ): Se conoce como espectrometría de masas tandem a la espectrometría de masas con múltiples etapas de análisis. La espectrometría de masas tandem más utilizada es aquella que se refiere a una doble etapa de análisis conocida como MS/MS en la que se mide la masa de un ion precursor, el cual luego es fragmentado y en una segunda etapa se mide la masa de los iones fragmento. En principio este procedimiento puede extenderse varias veces dando origen a la MSⁿ en donde n representa el número de análisis de masa realizados. La espectrometría de masas tandem permite obtener información acerca de la estructura molecular de los compuestos, y es una herramienta fundamental en el campo de la proteómica en la elucidación de la secuencia primaria de péptidos y proteínas. En la [Figura 2.3](#) se muestra el esquema de un espectrómetro de masas tandem.

Rango Dinámico: Se refiere al rango de m/z en el cual existe correlación entre la intensidad de la señal en el espectro de masa y la cantidad real de moléculas presentes [11]. Otra definición del rango dinámico es la relación entre el pico de más alta intensidad en el espectro y el pico de menor intensidad que no es considerado ruido [12].

2.2. FUENTES DE IONIZACIÓN

Existen distintos tipos de fuentes de ionización que se distinguen por el mecanismo físico-químico que produce la ionización y en ciertos casos la transformación de la muestra al estado gaseoso. Dependiendo de la volatilidad del compuesto o de sus características térmicas se pueden utilizar técnicas de ionización en fase gaseosa, en fase líquida o en fase sólida. Para el caso de sustancias muy volátiles se emplean métodos como el de la ionización electrónica [13], o la ionización química [14]. En el caso de sustancias poco volátiles se emplea la ionización desde la fase sólida o líquida respectivamente. La ionización en fase líquida generalmente se produce a presión atmosférica y se genera a partir de la nebulización de la muestra disuelta en estado líquido. Las técnicas más utilizadas para la ionización en estado líquido son la ionización por electrospray y la ionización química a presión atmosférica [15] o APCI por sus siglas en

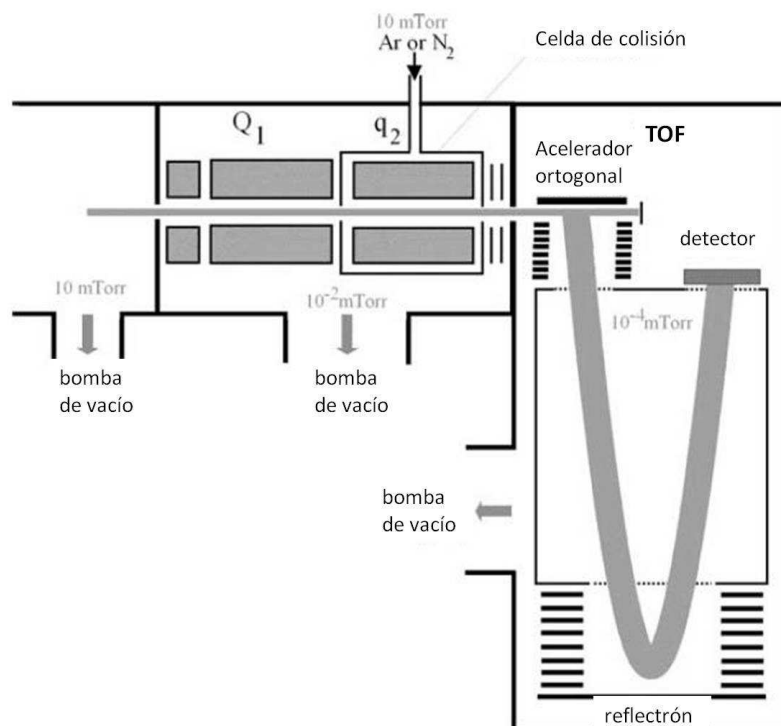


Figura 2.3: Esquema de un espectrómetro de masas tandem QqTOF (de doble cuadrupolo y tiempo de vuelo). En la primera etapa, los iones precursores son seleccionados mediante un espectrómetro cuadrupolar (Q_1); los iones son fragmentados en q_2 , y finalmente las masas de los iones fragmentos son determinadas en un espectrómetro de tiempo de vuelo (TOF). [11]

inglés. En el caso de la ionización en estado sólido, el enfoque utilizado en el campo de la proteómica consiste en estudiar la muestra condensada junto con una matriz, que es irradiada con partículas energéticas para producir la desorción y posterior ionización de la muestra. Los métodos más comunes para la ionización en estado sólido son la desorción/ionización por láser asistida por matriz o MALDI [16, 17] por sus siglas en inglés, la desorción en plasma, el bombardeo con átomos rápidos o FAB por sus siglas en inglés [18]; entre otros.

Como se mencionó anteriormente, la técnica de ionización que finalmente se emplee depende en gran medida del estado de la muestra y de la preparación previa de la misma. En este trabajo, se trabajó con una muestra disuelta en estado líquido y por tanto se empleó la técnica de ionización por electrospray que se describe a continuación.

2.2.1. ELECTROSPRAY

La ionización por electrospray o ESI por sus siglas en inglés fue introducida a la espectrometría de masas en los trabajos de Fenn et. al.[19, 20] en su trabajo sobre la ionización de biomoléculas y proteínas. Una de las grandes ventajas de esta técnica, es su capacidad de acoplarse fácilmente con un sistema de cromatografía líquida (HPLC o UPLC) y su característica de producir iones con baja energía interna. El principio de funcionamiento de la ESI consiste en aplicar un alto voltaje (2-6 kV) entre la punta de un capilar por el que circula la muestra disuelta en un solvente polar muy volátil, y un electrodo frontal. El flujo a través del capilar puede ser del orden de $\mu\text{L}/\text{min}$ o de nL/min en el caso de nano-ESI [21], por lo cual el acople con HPLC o UPLC es prácticamente directo. Una característica a recalcar de la ionización por ESI es que la sensibilidad no depende directamente de la cantidad de muestra utilizada sino más bien de la concentración de la muestra en el solvente [22], lo que hace posible trabajar con flujos a nivel de nano-ESI y conseguir alta sensibilidad utilizando cantidades mínimas de muestra. En este sentido, se ha demostrado niveles de detección en el rango de attomoles [23].

En la **Figura 2.4** se observa todas las partes de un sistema típico de electrospray. El flujo proveniente del capilar viaja hasta el electrodo de menor potencial y pasa a la zona de desolvatación, la misma que generalmente contiene un gas de átomos inertes de N_2 para ayudar a la evaporación del solvente. Como parte del sistema se usan colimadores para mejorar el enfoque del haz.

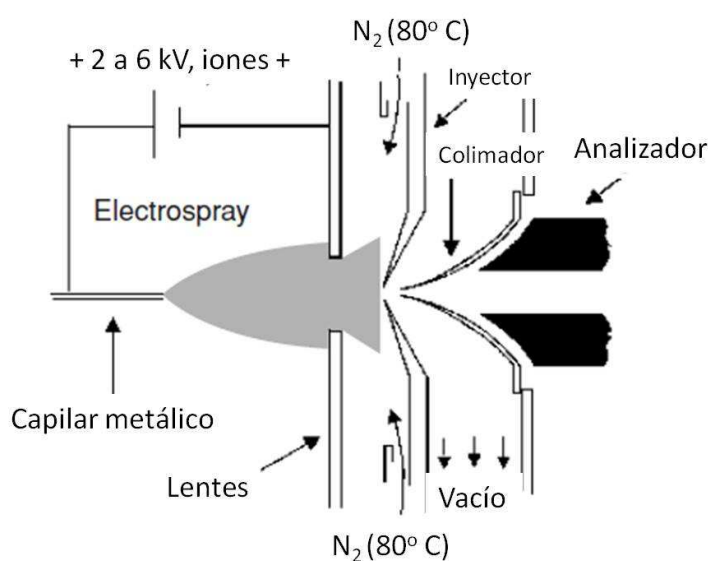


Figura 2.4: Esquema de un sistema de ESI. [24]

El alto voltaje en la punta del capilar produce campos eléctricos del orden de 10^6 V m^{-1} [12] y el fluido adquiere una forma conocida como Cono de Taylor que tiene la forma alargada que se muestra en la Figura 2.5. Del extremo del cono se liberan gotas cargadas (diámetro promedio de $1.5 \mu\text{m}$ y aproximadamente 50000 cargas elementales) que producen un spray del cual lentamente el solvente se evapora con la ayuda del baño de gas de N_2 [24] como se ve en la Figura 2.4.

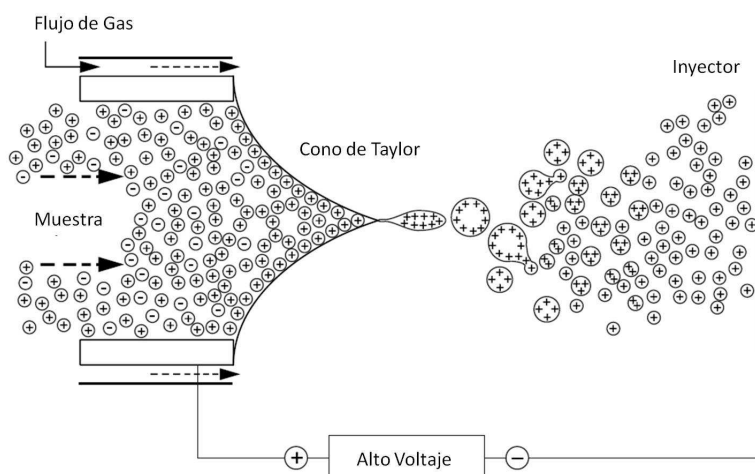


Figura 2.5: Cono de Taylor en la ionización por ESI. [12]

A medida que el solvente se evapora, la carga por unidad de volumen se incrementa en la gota, y al igual que en la salida del tubo capilar, la gota se deforma por el campo eléctrico para producir un nuevo Cono de Taylor. Éste se descompone nuevamente en gotas de menor volumen (diámetro promedio de $0.1 \mu\text{m}$ y aproximadamente 300 cargas elementales) [24, 25]. El proceso continua hasta llegar al punto en que la repulsión electrostática interna en las gotas produce la desorción de moléculas cargadas de su superficie. Esta expulsión continua de material se produce por la competencia entre la tensión superficial de la gota y la repulsión por la acumulación de cargas superficiales [25]. Este proceso fue estudiado por Rayleigh y en honor a él se define el límite de Rayleigh como aquel en que se produce la explosión coulombica que expulsa material de la gota precursora para producir las gotas de menor diámetro [26].

$$\gamma = \frac{(Ne)^2}{8\pi\epsilon_0 d_0^3} \quad (2.1)$$

Donde Ne es la carga neta, γ la tensión superficial del solvente, y d_0 el diámetro de la gota esférica. Según esto se han estimado tiempos de inestabilidad menores a los μs , lo que corresponde a tiempos mucho menores que la razón de evaporación del

solvente [27] por lo que la expulsión de material por medio de una explosión coulombica explica la producción de moléculas individuales antes de entrar a la zona del analizador de masa. Mientras la gota sigue en el proceso, llega un punto en que la relación entre la carga residual y la evaporación del solvente no permiten alcanzar el límite de Rayleigh y la carga residual es retenida en las moléculas del analito. Este modelo es conocido como el modelo de la carga residual o CRM por sus siglas en inglés [25].

Alternativamente, se ha propuesto un mecanismo distinto para la producción de iones por medio de ESI. El modelo de la evaporación de iones o IEM por sus siglas en inglés fue originalmente propuesto por Iribarne y Thomson [28], y a diferencia del modelo CRM propone una liberación individual de los iones a partir de la gota sin la necesidad de explosiones coulombicas. El modelo se basa en el hecho de que bajo ciertas condiciones la liberación de un ión (o cluster) de la superficie de la gota es termodinámicamente favorable antes de alcanzar el límite de Rayleigh.

En principio, el proceso de generación de iones puede pasar por una combinación de los dos mecanismos y en la actualidad no se cuenta con un modelo definitivo para explicar la ionización por medio de ESI.

2.3. ANALIZADORES DE MASA

Los analizadores de masa pueden considerarse como el corazón del espectrómetro por cuanto son la región en la cual actúan los campos electromagnéticos que permiten la separación de los compuestos de acuerdo a su relación de m/z . Existen analizadores de masa de distintos tipos con distintas características y desempeño (ver [Tabla 2.1](#)). De entre ellos, el analizador de masa por tiempo de vuelo es uno de los más utilizados, principalmente por su rapidez en la adquisición de datos, ilimitado rango de masas y alta precisión. Sin embargo, una de las limitaciones de este tipo de analizadores es su bajo límite de resolución cuando es comparado con otro tipo de analizadores como los ciclotrones de transformada de Fourier o FT-ICR [29]. A pesar de esto, la espectrometría de masas por tiempo de vuelo presenta un rendimiento muy bueno cuando se trabaja con extracción retardada [30], espejos electrostáticos [31] y/o extracción ortogonal [32]. En especial, el caso de la extracción ortogonal (oaTOF) permite no solo incrementar considerablemente la resolución sino también incorporar métodos de ionización continuos como el electrospray.

Otro tipo de analizadores muy utilizados en la práctica es el de tipo cuadrupolar, el cual no presenta una precisión tan alta en la determinación de la masa pero es muy útil para la interfaz entre fuentes de iones a presión atmosférica. A más de esto, los analizadores cuadrupolares son versátiles en el sentido de que pueden utilizarse como analizadores de masa o como una etapa intermedia en la transmisión de iones.

Una condición general que deben cumplir los analizadores de masa es eliminar las interacciones entre los iones a estudiar y cualquier otra molécula presente en el ambiente, ya que esto podría ocasionar choques indeseados que pueden producir cambios en las trayectorias o cambios en el estado de carga. Para asegurar esto, los analizadores de masas son operados en condiciones de alto vacío, ya que así el recorrido libre medio de los iones será lo suficientemente grande para evitar colisiones. Según la teoría cinética de los gases, el recorrido libre medio de una molécula viene dado por [24]:

$$\lambda = \frac{kT}{\sqrt{2}p\sigma} \quad (2.2)$$

Donde k es la constante de Boltzmann, T es la temperatura en Kelvin, p es la presión y σ es la sección eficaz de colisión.

De esta manera, a condiciones normales con $T = 300 \text{ K}$, $\sigma \approx 45 \times 10^{-20} \text{ m}^2$, y garantizando un recorrido libre medio λ de al menos 1 metro, se tiene que la presión de trabajo debe ser menor a 66 nbar. Bajo estas condiciones de presión se asegura que los iones no sufran colisiones durante su trayectoria. Más adelante discutiremos casos en los que se facilitan las colisiones para promover por ejemplo la disociación controlada de las moléculas en el analizador de masa.

Tabla 2.1: Comparación de las características de algunos analizadores de masas [12]

Analizador	TOF	Quadrupolo	FT-ICR	Trampa de Iones
Resolución	Alta	Baja	Muy alta	Baja - Media
Precisión	Alta	Baja	Muy alta	Baja - Media
Rango m/z	Muy alto	Bajo	Medio	Bajo - Medio
Sensitividad	Alta	Alta	Alta	Alta
Velocidad de Adquisición	Muy alta	Alta	Baja - Media	Baja - Media
Fuente de Iones	Pulsada/Continua	Continua	Pulsada/Continua	Pulsada/Continua
Costo	Medio - alto	Bajo	Alto	Medio

2.3.1. ANALIZADORES QUADRUPOLES

Los analizadores cuadrupolares basan su funcionamiento en la estabilidad de las trayectorias de iones con diferentes valores de m/z . Un analizador cuadrupolar se compone de 4 electrodos circulares (o idealmente hiperbólicos) paralelos conectados a un voltaje continuo U y otro alterno V (RF) de manera que el voltaje en cada uno de ellos viene dado por:

$$\Phi_0 = \pm (U - V \cos \omega t) \quad (2.3)$$

Los signos \pm dependen de la posición de los electrodos de manera que se encuentran emparejados como se observa en la [Figura 2.6](#)

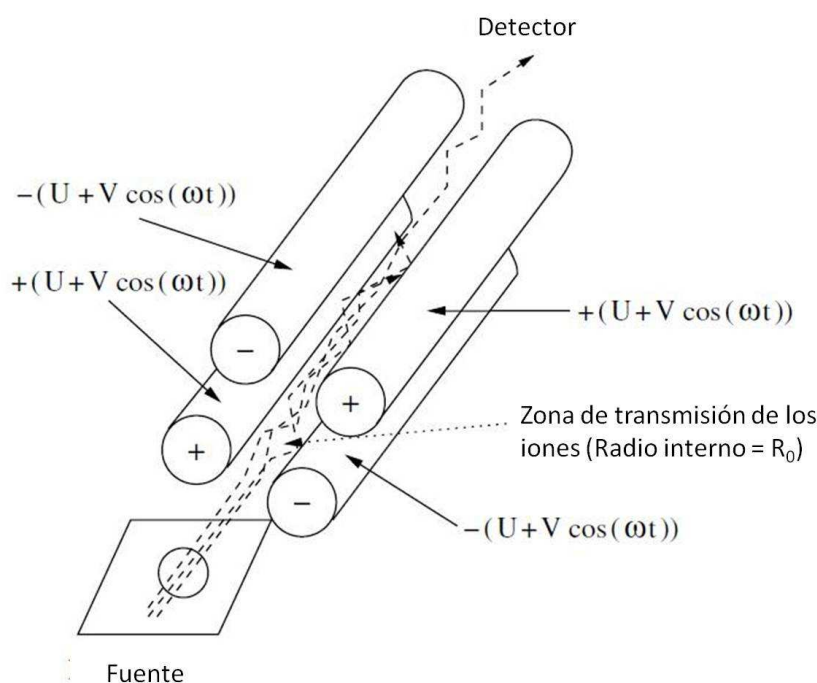


Figura 2.6: Esquema de un analizador de masas cuadrupolar. Los iones positivos atraviesan la región entre los electrodos solo para ciertos valores de m/z . Al escanear los valores de los voltajes U y V , iones de diferente m/z atraviesan esta región. [11]

La descripción de las ecuaciones y el funcionamiento del analizador de masa cuadrupolar presentada a continuación se basa en el capítulo 2 del texto de Ekman et. al. [12], en el capítulo 2 del texto de Hoffman et. al. [24] y en el capítulo 3 del texto de Kaltashov et. al. [25].

Para generar las diferentes trayectorias de los iones, se varían los valores de U y V . Las condiciones dinámicas que condicionan las trayectorias permitidas de los iones en

el cuadrupolo permiten transmitir secuencialmente distintos iones hasta construir el espectro de masa. El diseño del espectrómetro cuadrupolar hace que solo las especies con ciertos valores de m/z tengan trayectorias estables, es decir, que su movimiento en el plano XY se mantenga dentro de los bordes marcados por las barras, esto es, que $x^2 + y^2 < R_0^2$ (ver [Figura 2.6](#)). Para todas las demás especies, sus trayectorias tendrán una amplitud mayor que R_0^2 de manera que chocarán con los electrodos y los iones se perderán. Los cuadrupolos generalmente se operan a velocidades de escaneo que van hasta los 6 kDa/s. Mientras menor es la velocidad de adquisición, mayor es la resolución que se puede alcanzar.

La trayectoria de un ion en el cuadrupolo es obtenida a partir de las ecuaciones de movimiento que se derivan del campo generado al interior del cuadrupolo. El potencial en el interior del cuadrupolo se obtiene a partir del potencial de los electrodos según la [Ecuación 2.3](#) y es el siguiente:

$$\Phi = \frac{x^2 - y^2}{R_0^2} \Phi_0 = \frac{x^2 - y^2}{R_0^2} (U - V \cos wt) \quad (2.4)$$

Las ecuaciones dinámicas de movimiento se obtiene directamente como:

$$m \frac{d^2x}{dt^2} + \frac{2ze}{mR_0^2} (U - V \cos wt) x = 0 \quad (2.5)$$

$$m \frac{d^2y}{dt^2} - \frac{2ze}{mR_0^2} (U - V \cos wt) y = 0 \quad (2.6)$$

Las mismas que deben ser integradas para obtener la ecuación de la trayectoria. Sin embargo, no nos interesa la trayectoria real de los iones siempre y cuando cumplan con la condición de estabilidad mencionada anteriormente. Para estudiar esta condición hacemos el cambio de variable $\varepsilon = wt/2$ de manera que podemos escribir las ecuaciones de movimiento como:

$$\frac{d^2u}{d\varepsilon^2} + (a_u - 2q_u \cos 2\varepsilon) u = 0 \quad (2.7)$$

Donde u representa a x o y , y donde:

$$a_u = a_x = -a_y = \frac{8zeU}{mw^2R_0^2} \quad y \quad q_u = q_x = -q_y = \frac{4zeV}{mw^2R_0^2} \quad (2.8)$$

La **Ecuación 2.7** es conocida como la ecuación de Mathieu, y fue originalmente estudiada en la propagación de ondas en membranas [33]. La solución viene dada por las funciones de Mathieu. Las soluciones de la ecuación diferencial de Mathieu se conocen como funciones de Mathieu Angulares y funciones de Mathieu Radiales [34], las mismas que corresponden a la solución en el eje x y en el eje y respectivamente. Tanto las funciones radiales como las funciones angulares dependientes de los parámetros a y q pueden encontrarse en la literatura [34]. De manera general las funciones de Mathieu son funciones de tipo oscilatorio como puede verse con algunos ejemplos en la **Figura 2.7**.

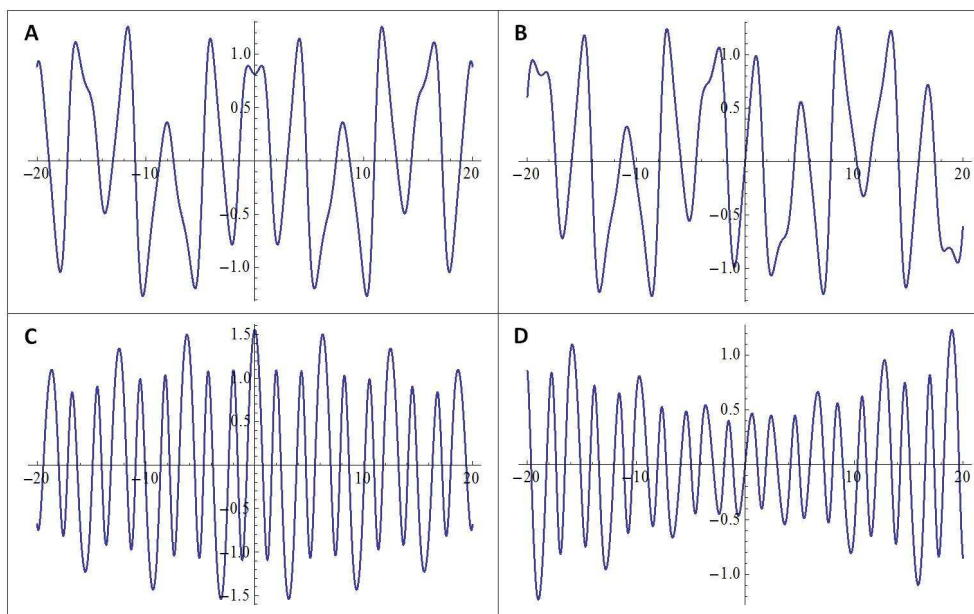


Figura 2.7: Algunas funciones angulares de Mathieu. (A) Función angular par con parámetros $a = 3$, $q = 2$. (B) Función angular impar con parámetros $a = 3$, $q = 2$. (C) Función angular par con parámetros $a = 10$, $q = 3$. (D) Función angular impar con parámetros $a = 10$, $q = 3$.

Sobre estas funciones, no estamos interesados en la solución exacta sino en la zona de estabilidad para los iones en el cuadrupolo. Imponiendo la condición de estabilidad se puede construir un diagrama en el espacio (a,q) que representa la zona de estabilidad. La **Figura 2.8** muestra esta zona gráficamente. Los cuadrupolos trabajan en la zona en que $a/q = const = 0.336$, que en el gráfico corresponde a la línea entrecortada

marcada como línea de scan. De esta manera al variar los valores de U y V manteniendo su relación constante, se tienen distintos valores de m/z que caen dentro de la zona de estabilidad. Adicionalmente, el valor de U/V se escoge de manera que la ventana de m/z permitidos sea lo suficientemente pequeña para garantizar que se escanea las especies de una en una, esto es, una buena resolución espectral.

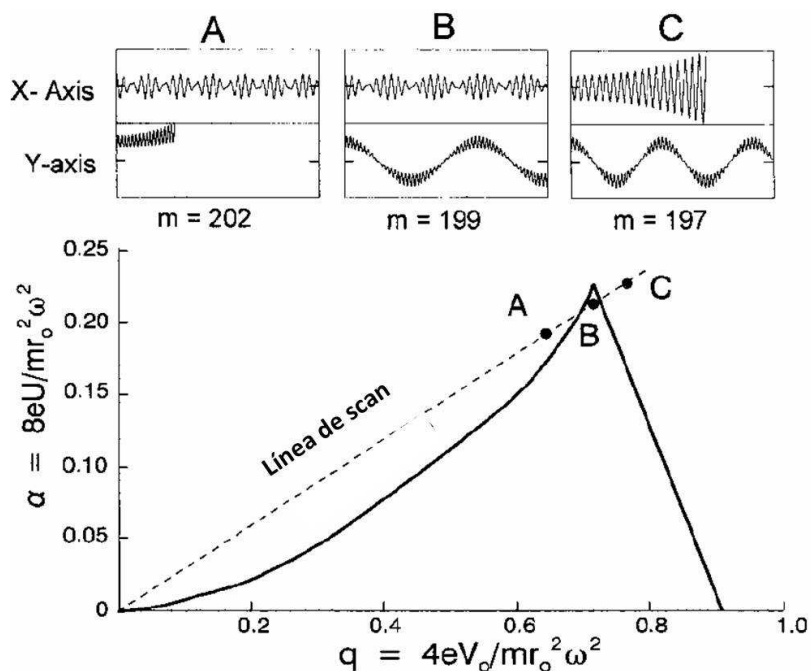


Figura 2.8: Diagrama de estabilidad en la ecuación de Mathieu. La región bajo la curva representa los valores de (a,q) en los que la trayectoria de los iones es estable. Los valores de U y V se escogen de manera que el escaneo se efectúe sobre la línea $a/q = 0.336$. En este caso solo el ion con m/z de 199 unidades atraviesa el cuadrupolo mientras que el resto de masas tienen trayectorias inestables. [12]

El diagrama mostrado en la [Figura 2.8](#) puede ser visto de otra manera si se transforma a un diagrama U vs. V ya que en ese caso la zona de estabilidad se escala según cambia el valor de m/z tal y como se muestra en la [Figura 2.9](#). En la [Figura 2.9](#) puede notarse que si $U = 0$ y V toma un valor fijo (V_1), entonces todos los valores de m/z superiores a un cierto límite inferior (m_1 en la gráfica) son transmitidos en el cuadrupolo. Este modo se conoce como modo RF del cuadrupolo y es muy utilizado especialmente como una interfaz en sistemas de ionización a presión atmosférica.

Los valores de energía con los que se trabajan a la entrada y salida del cuadrupolo son menores a los 100 eVs lo que hace posible trabajar con presiones no tan bajas ($\approx 10^{-5}$ Torr) [24]. La capacidad de transmisión de los cuadrupolos está limitada tanto en el límite inferior como en el superior por el voltaje alterno V. Para un voltaje V fijo,

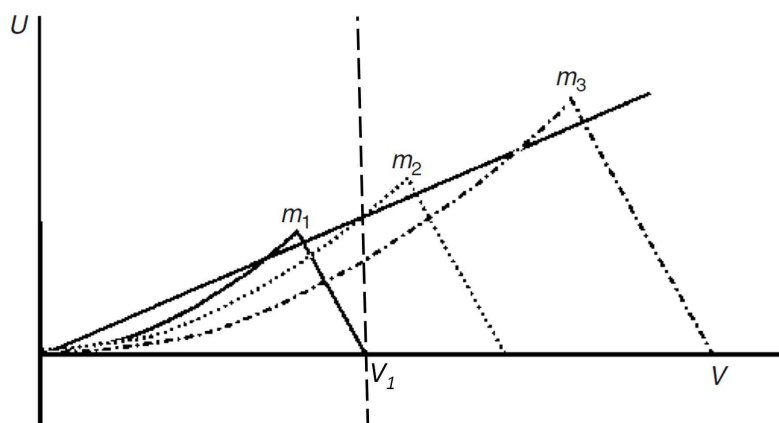


Figura 2.9: Diagramas de estabilidad en el espacio (U,V). La región de estabilidad se escala según el valor de m/z . Para el caso de un voltaje fijo V_1 , solo las masas mayores a m_1 como m_2 y m_3 atraviesan el cuadrupolo. [24]

existe un límite inferior en m/z dado por los diagramas de estabilidad como se observa en la [Figura 2.9](#); y en el caso de m/z altos, el enfoque es inversamente proporcional a m/z y por tanto los iones de m/z mayores a 4000 Da tienen bajas probabilidades de atravesar el cuadrupolo [12, 24]. Una forma de solucionar este problema de transmisión es la utilización de hexapolos u octapolos.

Los hexapolos u octapolos actúan de manera semejante a los cuadrupolos con cada electrodo consecutivo conectado a un voltaje RF de signo opuesto. Para estos casos el potencial dentro del n-polo es:

$$U(R) = \frac{(nzeV)^2}{4mw^2R_0^2} \left(\frac{R}{R_0} \right)^{2n-2} \quad (2.9)$$

La forma del potencial en la [Ecuación 2.9](#) indica que el hexapolo y octopolo tienen un potencial mayor en los electrodos cuando $R = R_0$ (proporcional a n^2), pero el potencial en el cuadrupolo crece más rápidamente cuando R es cercano a 0. De esta manera, el hexapolo y octopolo tienen una mejor transmisión en un rango más amplio de masas pero no son tan buenos para enfocar como el cuadrupolo. Por esta razón, generalmente se utilizan hexapolos y octopolos como guías de iones para transmisión y cuadrupolos como selectores de iones o analizadores de masa por su mayor resolución y capacidad de enfoque.

Comúnmente, los analizadores cuadrupolares se utilizan a resolución unitaria para la selección de precursores y el análisis posterior se puede hacer por medio de un analizador

como el de tiempo de vuelo que tiene una mejor resolución. Para la transmisión, se utilizan hexapolos u octapolos dando origen al espectrómetro tandem conocido como QqTOF (cuadrupolo-hexapolo/octopolo-TOF) y sobre él hablaremos más a detalle en una sección posterior.

2.3.2. ANALIZADORES DE TIEMPO DE VUELO

El principio de funcionamiento de los analizadores por tiempo de vuelo o analizadores TOF por sus siglas en inglés (Time-of-Flight) se basa en medir el tiempo que le toma a especies con distintos valores de m/z en llegar al detector del espectrómetro. Inicialmente, los iones son expuestos a campos eléctricos que les otorgan distintas velocidades dependiendo de la relación m/z , de tal manera que los tiempos de vuelo también dependen de esta relación. En principio, las especies con el mismo valor de m/z deberían alcanzar el detector al mismo tiempo, lo cual no ocurre en la práctica debido a las condiciones iniciales de generación de los iones y a defectos en la homogeneidad de los campos eléctricos. Sin embargo, el desarrollo de la espectrometría TOF ha producido métodos para reducir estos efectos y garantizar el menor ancho posible de los picos al alcanzar el detector (i.e. mayor resolución). Tal es el caso de la espectrometría TOF con extracción retardada [30], con espejos electrostáticos [31] o con extracción ortogonal [32]. Más adelante discutiremos algunos de estos conceptos. Con todas estas mejoras, los analizadores por tiempo de vuelo se encuentran en un punto medio en cuanto a su capacidad de resolución [12, 24], la cual es mayor que la de los analizadores cuadrupolares pero se ve limitada en comparación con el rendimiento de otros analizadores como el FT-ICR.

El esquema general de un analizador de tiempo de vuelo se compone de una zona de aceleración y una zona de vuelo libre (ver Figura 2.10). La zona de aceleración está compuesta por un grupo de electrodos conectados a alto voltaje que producen iones con la misma energía cinética o con la misma cantidad de movimiento [35]. El tipo de zona de aceleración más común es aquella en que los iones son sometidos a un campo eléctrico constante y por tanto adquieren todos la misma energía cinética. Si el voltaje aplicado entre los electrodos es V entonces se tiene que la energía cinética de los iones será:

$$K = \frac{1}{2}mv^2 = zeV \quad (2.10)$$

iones con diferentes razones de m/z tendrán diferentes velocidades pues el voltaje V es el mismo para todos los iones. Es así que los iones con m/z mayores tendrán velocidades menores y viceversa. La separación según m/z se produce al permitir a los iones viajar libremente por una zona de longitud L sin campos eléctricos en ella y ubicada más allá de los electrodos de aceleración; de manera que aquellos iones con mayor velocidad (menor m/z) llegan primero al detector ubicado al final del recorrido. Así, se registra un espectro en el dominio del tiempo. La relación entre el tiempo y la tasa m/z se obtiene al despejar el tiempo para cada uno de los iones según la siguiente ecuación:

$$t = \frac{L}{v} = L\sqrt{\frac{m/z}{2eV}} \quad (2.11)$$

O equivalentemente:

$$m/z = 2eV \left(\frac{t}{L}\right)^2 \quad (2.12)$$

De esta manera se registra un espectro en el dominio de m/z en donde la altura de los picos es proporcional al número de iones que llegan al detector para cada instante de tiempo (idealmente, y dentro del rango dinámico instrumental). La **Ecuación 2.11** muestra la relación más sencilla para calibrar un espectro de tiempo de vuelo e implica una proporcionalidad entre el tiempo de vuelo t y la raíz cuadrada de m/z .

Como mencionamos anteriormente, existen condiciones en la generación de iones que incrementan el ancho de los picos en el espectro. De estos efectos los más importantes son los que vienen dados por: (1) iones que empiezan en planos equipotenciales diferentes en la zona de aceleración; y (2) iones con distintas velocidades iniciales en la dirección de vuelo [36]. Existen distintos mecanismos para corregir en cierta medida estos efectos indeseados, y de entre ellos mencionamos a la extracción ortogonal y al reflectrón, pues son los que se utilizan en el esquema del espectrómetro tandem QqTOF.

2.3.2.1. Extracción Ortogonal

La espectrometría TOF con extracción ortogonal (oa-TOF) tiene dos ventajas fundamentales, y éstas son:

- Permite acoplar la espectrometría TOF con fuentes de ionización continuas [32].

- Mejora la resolución al disminuir la dispersión en la componente de la velocidad inicial paralela a la dirección de vuelo libre [37, 38].

La principal característica de la oa-TOF es que mediante un pulso ($\approx 10 \mu\text{s}$) de alto voltaje se extraen a los iones en una dirección ortogonal a la velocidad axial con la que entran a la zona de aceleración. En este analizador, los iones provenientes de una etapa previa de análisis ingresan de manera continua a la zona de aceleración hasta que se les aplica un campo eléctrico constante en una dirección ortogonal a la dirección de su ingreso. Al aplicar la extracción ortogonal, la velocidad axial del haz de iones original, que es la que mantiene con largura la mayor dispersión en los valores de las velocidades iniciales de los iones, no influye en el tiempo de vuelo de los iones, los mismos que describen una trayectoria como puede verse en la Figura 2.10. La Figura 2.10 muestra un oa-TOF con espejo electrostático o reflectrón, en el cual x representa la distancia desde el electrodo inicial y v_0 es la velocidad en la dirección de vuelo libre. La ventaja del oa-TOF es que disminuye la dispersión en v_0 pues la velocidad radial del haz entrante puede controlarse mediante hexapolos u octapolos antes de la entrada al oa-TOF produciendo un haz de iones axialmente simétrico para el que $\bar{v}_0 = 0$.

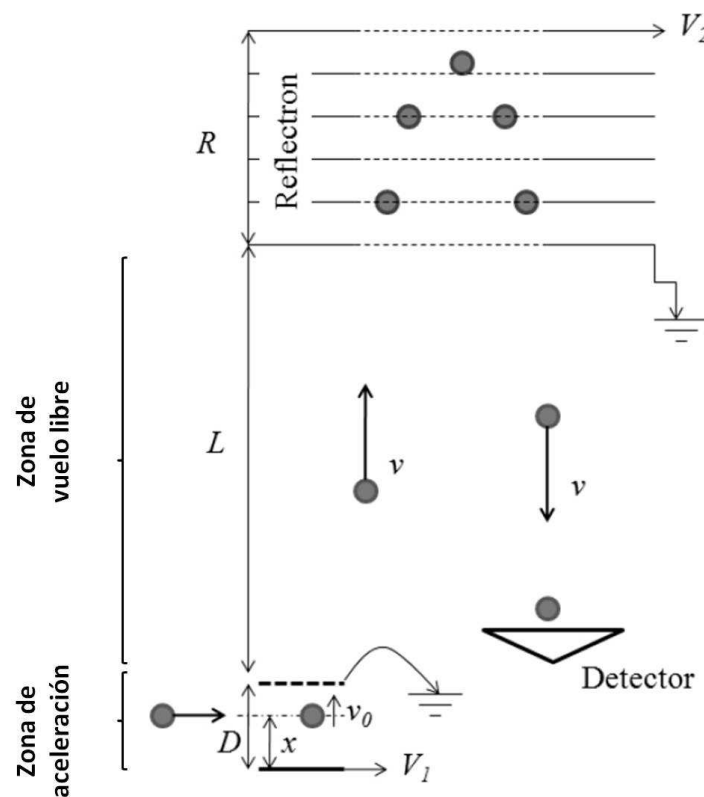


Figura 2.10: Esquema de un oa-TOF

Al considerar las variables x y v_0 , la ecuación exacta del tiempo de vuelo tiene una forma funcional complicada y se aleja de la **Ecuación 2.11**. La manera más sencilla de afrontar el problema es considerar aproximaciones adecuadas que tomen en cuenta que $x \ll D$ y que $w = \frac{1}{2}Mv_0^2/(eV_1 \frac{D-x}{D}) \ll 1$, donde w es el déficit relativo de energía inicial. Tomando en cuenta estos factores la expresión para el tiempo de vuelo total (tiempo de vuelo en la zona de aceleración + tiempo de vuelo en la zona de vuelo libre + tiempo de vuelo en el reflectrón) puede obtenerse como un desarrollo en series de potencias de w y por ende como una serie en los valores de m/z [39, 40].

$$t = t_0 + a_0 \left(\frac{m}{z}\right) + \sum_{k=1,3,5,\dots} a_k \left(\frac{m}{z}\right)^{\frac{k}{2}} \quad (2.13)$$

En donde los parámetros a_k dependen de los voltajes y geometría del espectrómetro [39, 40].

La velocidad axial del haz entrante no se modifica al entrar en la zona de aceleración, de manera que el detector puede ubicarse en el mismo plano que el acelerador ortogonal (ver **Figura 2.10**), y el ángulo Θ en que se desvía el haz depende de la relación entre los voltajes del haz a la entrada V_{in} y el voltaje de aceleración V_1 , de manera que $\Theta = \arctan(V_1/V_{in})^{1/2}$ [32]. Además, los voltajes y tiempo de espera del pulso pueden controlarse de manera que el tiempo de vuelo de los iones sea el mismo que el tiempo en que se acumulan los iones en la zona de aceleración. Con esto se incrementa la sensibilidad y el rendimiento del analizador.

La extracción de iones se realiza mediante un pulso de alto voltaje que usualmente tiene una duración del orden de $10 \mu\text{s}$, tiempo en el cual todos los iones abandonan la región de aceleración y se dirigen a la zona de vuelo libre. Un aspecto muy importante es que el tiempo de subida del pulso debe ser lo más pequeño posible (en el orde de 10 ns). Para obtener estos tiempos de subida, se usan switches de alto voltaje muy rápidos y voltajes que no sobrepasan los 2kV [32].

2.3.2.2. Reflectrón

El reflectrón o espejo electrostático consiste en un campo eléctrico opuesto a la dirección de la velocidad de los iones entrantes que se ubica luego de una primera zona de vuelo libre. Este dispositivo permite compensar el efecto de la distribución energética inicial de los iones al salir de la zona de aceleración. El concepto del reflectrón se basa

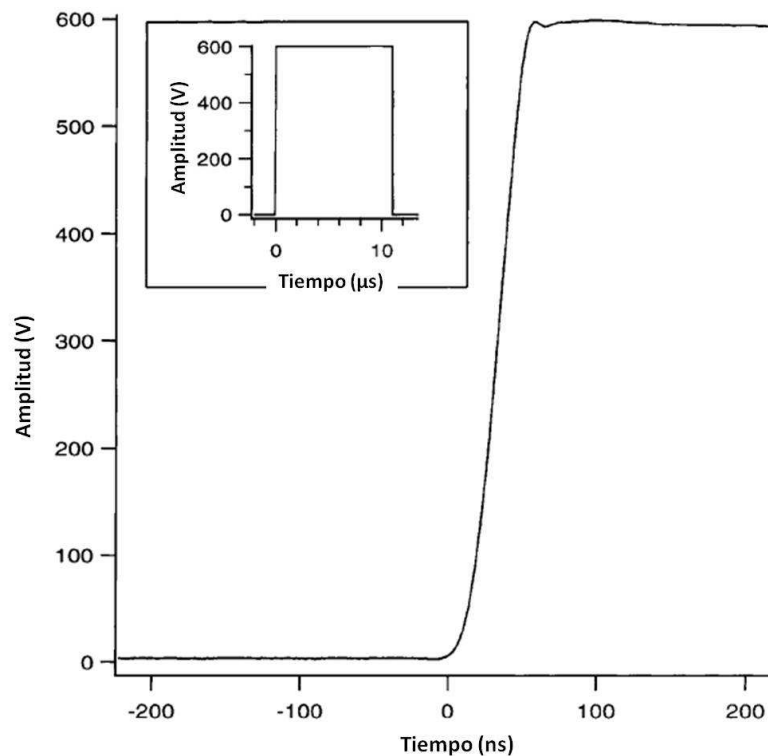


Figura 2.11: Pulso de extracción en un oa-TOF. La figura interna corresponde al total del pulso. [32]

en que iones con la misma m/z pero con mayores energías llegan antes al reflectrón pero tienen que penetrar más en virtud de su velocidad adicional, con el consiguiente incremento en su tiempo de vuelo. Bajo las condiciones de enfoque adecuadas es posible cancelar el efecto de la distribución de energías iniciales y conseguir que iones de iguales m/z lleguen al detector simultáneamente incrementando drásticamente la resolución del espectrómetro.

De esta manera, el campo electrostático producido por el reflectrón afecta únicamente a la componente paralela al eje de vuelo y se tiene una relación lineal entre el tiempo de permanencia (t_p) de un ión en el espejo electrostático y su velocidad de entrada (v_e) [41]

$$t_p = 2 \frac{v_e}{a_R}$$

Donde a_R es la aceleración producida por el campo constante del reflectrón y es igual a: zeV_R/mR (ver Figura 2.10).

De esta manera, los iones con una mayor velocidad de entrada tardarán más tiempo dentro de la zona del reflectrón, haciendo posible que los iones de menores velocidades

de entrada salgan antes del reflectrón y lleguen iguales al detector. Un esquema sencillo de este proceso puede observarse en la [Figura 2.12](#)

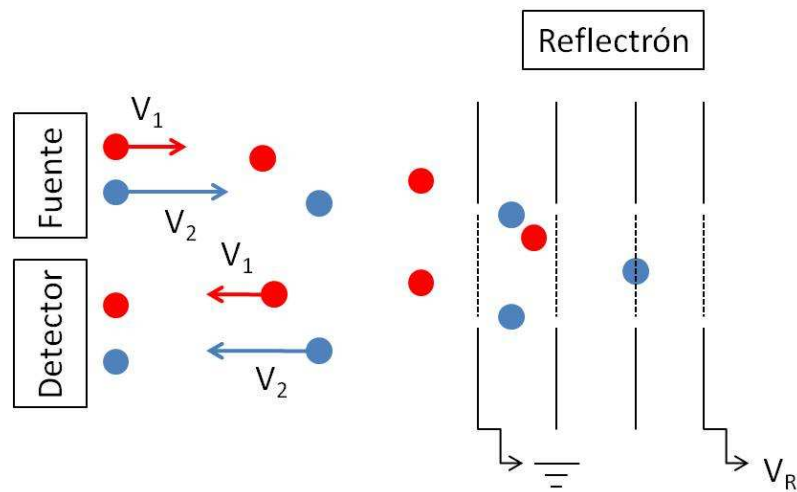


Figura 2.12: Esquema del reflectrón o espejo electrostático. El ión 1 (V_1) tiene una velocidad menor al ión 2 (V_2) a la salida de la fuente. Debido a esta diferencia de velocidades el ión 2 penetra una distancia mayor en el reflectrón incrementando su tiempo de vuelo de manera que ambos iones llegan simultáneamente al detector.

2.4. ESPECTROMETRÍA DE MASAS ITERADA (TANDEM)

La espectrometría de masas iterada, o espectrometría de masas tandem como será referida de ahora en adelante, consiste básicamente en el uso de dos o más espectrómetros de masas, uno después de otro, para obtener información adicional de los iones en estudio. Típicamente, los iones son seleccionados o filtrados en las primeras etapas y analizados posteriormente. En particular, se puede decir que la espectrometría de masas tandem es un grupo de técnicas que permiten la elucidación de las características estructurales de péptidos, proteínas, o la molécula que se pretende estudiar. La espectrometría de masas tandem puede realizarse en varios modos dependiendo de la información que se desea obtener. Los 4 métodos más utilizados son:

- Modo de Ion precursor
- Modo de Pérdidas Neutrales
- Modo de Monitoreo de Reacción Única o Modo SRM por sus siglas en inglés
- Modo de Ion-Producto o Modo MRM por sus siglas en inglés

La descripción de estos métodos puede encontrarse en la literatura [42]. De los 4 métodos, el modo MRM es utilizado cuando se trata de identificar la estructura de sustancias desconocidas. De manera simple, en este modo de operación se realiza una primera etapa de selección de ión precursor, se pasa luego a la zona de reacción y finalmente se obtiene el espectro de todos los iones fragmento provenientes del ión precursor seleccionado. Este proceso se puede realizar en múltiples etapas si en vez de obtener el espectro de los iones producto se repite el proceso y uno de los iones producto es seleccionado como ion precursor y se repite el ciclo.

La espectrometría de masas tandem también se divide en espectrometría por separación espacial o temporal [12]. La separación espacial se realiza en distintos espectrómetros ubicados de manera espacial consecutiva; se tienen por ejemplo los analizadores QqTOF o QqQ. En el caso de la separación temporal, los distintos análisis se realizan en el mismo analizador pero en instancias consecutivas de tiempo; ejemplos son el caso del FT-ICR o la Trampa de Iones. La desventaja de los analizadores espaciales es que en cada instancia de análisis la transmisión de iones se reduce drásticamente, haciendo impráctica la realización de más de 3 o 4 etapas de análisis [12]; además, al incluir más etapas de análisis se incrementa el número de analizadores y por consiguiente el costo del equipo. En el caso de la separación temporal, todos los análisis se realizan en el mismo espacio físico, y se consigue una mayor transmisión de iones en las diferentes etapas de análisis, haciendo posible llegar hasta 7 u 8 etapas consecutivas en analizadores como el FT-ICR [12].

Un aspecto importante de la espectrometría de masas tandem es que el resultado de las reacciones presentes entre cada etapa de análisis debe ser controlado y conocido, de manera que el espectro final pueda ser interpretado y obtener la información estructural del compuesto estudiado. Uno de los métodos más utilizados para producir fragmentación molecular es la colisión inducida por colisión o CID por sus siglas en inglés. Este método consiste en hacer colisionar al ion precursor con moléculas de un gas inerte a presión y energías controladas. Con esto se incrementa la energía interna del ion precursor produciendo la posterior disociación del mismo. En la [Sección 2.5.2](#) se discute a detalle este proceso.

2.4.1. ESPECTRÓMETRO DE MASAS QqTOF

Como se mencionó anteriormente, el espectrómetro de masas tandem QqTOF es un espectrómetro tandem por separación espacial [12]. La estructura de este espectrómetro consiste básicamente en dos cuadrupolos seguidos por un analizador de tiempo de vuelo (ver Figura 2.3). El primer cuadrupolo puede funcionar en modo RF o en modo scan, mientras que el segundo cuadrupolo/hexapolo/octopolo es utilizado en modo RF como cámara de colisión con algún gas como Ar o N₂. La sensibilidad y resolución de este tipo de espectrómetros es muy alta debido a la combinación de estos dos tipos de analizadores pues se beneficia de la alta resolución del espectrómetro oa-TOF con las capacidades de transmisión y selección del cuadrupolo [36, 43, 44]. Las principales limitaciones de este espectrómetro vienen dadas por el rango de masas que efectivamente se puede analizar en el cuadrupolo, el cual típicamente no supera los 4000 Da [24]. Este dispositivo es no obstante muy adecuado para su acoplamiento con la ionización ESI, ya que ésta típicamente genera iones múltiplemente cargados que incrementan el rango de masas efectivo del cuadrupolo.

El primer cuadrupolo actúa como un filtro de masas cuando se adquieren espectros MS/MS y permite la selección del precursor para trabajar en modo MRM. El segundo cuadrupolo/hexapolo se utiliza únicamente en modo RF y se utiliza como cámara de colisión para producir la fragmentación de los iones mediante el proceso de disociación inducida por colisión que será revisado más adelante. El primer cuadrupolo también puede operarse en modo RF para la transmisión de iones al momento de adquirir espectros MS. En un análisis LC-MS/MS completo, el modo de operación del primer cuadrupolo varía continuamente para adquirir de manera alternada espectros MS y MS/MS [43].

En los espectrómetros comerciales normalmente se utiliza un hexapolo adicional a la entrada al oa-TOF para condicionar el haz reduciendo su velocidad transversal mediante una cámara de enfoque por medio de colisiones con un gas residual [45]. El papel de este hexapolo final es crucial para obtener un poder de resolución de hasta 40000 unidades en espectrómetros comerciales de última generación como por ejemplo el MaXis QqTOF (Bruker Daltonics) y el 6540-UHD Q-TOF (Agilent Technologies).

El modo de operación en el QqTOF permite acoplar la resolución del oa-TOF con el sistema de inyección continua de la ionización por ESI. Esta ventaja hace del QqTOF una herramienta muy poderosa para el análisis por medio de LC-MS/MS. La rapidez de adquisición se ve limitada principalmente por la rapidez de escaneo en el primer

cuadrupolo y por el acople con el oa-TOF. Bajo condiciones normales de operación, la tasa de adquisición puede llegar hasta a 20 Hz manteniendo sensibilidad y resolución¹.

2.5. TÉCNICAS DE FRAGMENTACIÓN

Como se mencionó anteriormente, la espectrometría de masas tandem permite obtener información sobre la estructura de algunos tipos de moléculas. Para esto hace falta utilizar algún método de fragmentación que permita obtener fragmentos específicos de la especie molecular original. Luego, usando la información espectral correspondiente a los fragmentos y con ayuda de modelos de fragmentación teórica, se trata de reconstruir la composición de la especie molecular original. En particular, la espectrometría de masas tandem por medio de la fragmentación molecular ha demostrado ser útil en la elucidación de la secuencia primaria de péptidos y proteínas [2]. A continuación se describen las principales técnicas de fragmentación utilizadas en la actualidad y su relación con la fragmentación de péptidos.

2.5.1. NOMENCLATURA

La nomenclatura estándar utilizada para referirse a los fragmentos de péptidos fue originalmente propuesta por Roepstorff y Fohlman en 1984 [46]. Ésta fue posteriormente modificada por Johnson et. al. [47]. Finalmente, Biemann completó la nomenclatura para incluir fragmentos satélites e iones de immonio [48].

Para que un fragmento sea detectado este debe poseer carga eléctrica. Los fragmentos de la columna del péptido se clasifican en *a*, *b*, *c* si el fragmento contiene el terminal N y en fragmentos *x*, *y*, *z* si el fragmento contiene el terminal C (ver Sección 3.1). El número de aminoácidos en el fragmento se indica por medio de un subíndice. Por ejemplo, en la Figura 2.13 se observan los diferentes tipos de fragmentos producidos en un péptido compuesto por 4 residuos. En la parte superior se anotan los fragmentos que contienen el terminal C y en la parte inferior los fragmentos complementarios que contienen el terminal N.

Adicionalmente existen fragmentos internos provenientes de más de una fragmentación

¹<http://www.bdal.com/products/lc-ms/uhr-tof/maxis/overview.html>

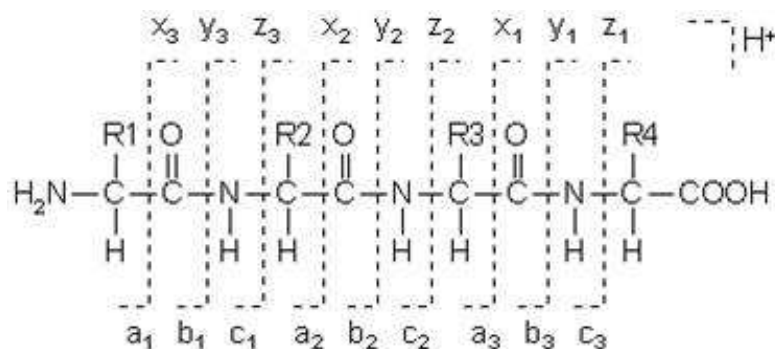


Figura 2.13: Nomenclatura de los fragmentos de la columna del péptido. Fuente: http://www.matrixscience.com/help/fragmentation_help.html

en la cadena del péptido. La nomenclatura de los fragmentos internos es la secuencia de aminoácidos con su código de una sola letra (ver [Tabla 3.1](#)) y normalmente se producen por la combinación de una fragmentación de tipo *b* y una de tipo *y*. Por ejemplo, en la [Figura 2.14](#) se observa el fragmento interno R2R3, donde R2 y R3 son las identificaciones correspondientes a los dos residuos presentes en el fragmento.

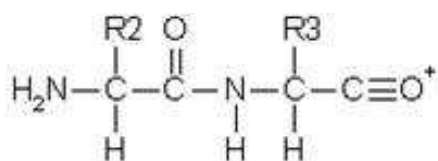


Figura 2.14: Fragmento interno R2R3.

Fuente: http://www.matrixscience.com/help/fragmentation_help.html

En el caso que el fragmento interno contenga un solo aminoácido, toma el nombre de *ion immonio*. Estos iones son generados por una fragmentación consecutiva de tipo *a* y tipo *y*. Su nomenclatura es la del código de una sola letra del aminoácido.

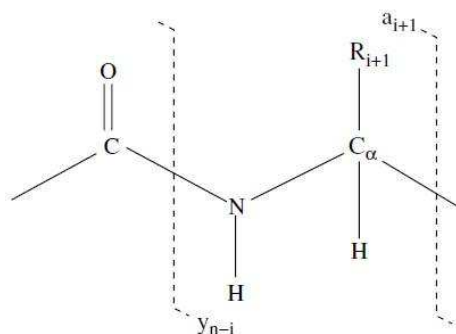


Figura 2.15: Estructura de un fragmento de ion immonio. [11]

Finalmente, los iones fragmentos satélites se producen por fragmentación adicional

lateral en los fragmentos de la columna del péptido. Los más comunes son los tipo *d* por fragmentación parcial de la cadena lateral en los fragmentos tipo *a*, los tipo *v* por fragmentación total de la cadena lateral en los fragmentos tipo *y*, y los tipo *w* por fragmentación parcial de la cadena lateral en los fragmentos tipo *z*.

Adicionalmente, existen fragmentos provenientes de la pérdida de una molécula de H₂O en cualquiera de los fragmentos tradicionales. La nomenclatura de los fragmentos es *b^o*, *y^o*, etc. En el caso de la pérdida de una molécula de amonía la nomenclatura de los fragmentos es de *b^{*}*, *y^{*}*, etc.

En la **Tabla 2.2** se describen estos fragmentos. Las masas corresponden a los fragmentos neutros.

Tabla 2.2: Masa molecular de varios iones fragmentos

Fragmento	Masa ^a
<i>a</i>	[N]+[M]-CHO
<i>a[*]</i>	<i>a</i> -NH ₃
<i>a^o</i>	<i>a</i> -H ₂₀
<i>b</i>	[N]+[M]-H
<i>b[*]</i>	<i>b</i> -NH ₃
<i>b^o</i>	<i>b</i> -H ₂₀
<i>c</i>	[N]+[M]+NH ₂
<i>d</i>	<i>a</i> -cadena parcial
<i>v</i>	<i>y</i> -cadena completa
<i>w</i>	<i>z</i> -cadena parcial
<i>x</i>	[C]+[M]+CO-H
<i>y</i>	[C]+[M]+H
<i>y[*]</i>	<i>y</i> -NH ₃
<i>y^o</i>	<i>y</i> -H ₂ O
<i>z</i>	[C]+[M]-NH ₂

^a N es la masa molecular del grupo terminal-N, C es la masa molecular del grupo terminal-C, M es la masa molecular de los residuos de los aminoácidos presentes en el fragmento. Fuente: http://www.matrixscience.com/help/fragmentation_help.html

2.5.2. DISOCIACIÓN INDUCIDA POR COLISIÓN

Uno de los métodos de fragmentación es el de disociación inducida por colisión o CID por sus siglas en inglés (Collision-Induced Dissociation). La fragmentación se produce luego de la colisión del ión a estudiar con moléculas neutras (N_2 , Ar, He). El efecto de la colisión es incrementar la energía interna de la molécula ocasionando la disociación de la misma. El proceso puede verse dividido en dos pasos [49]:

1. La colisión de la molécula a ser estudiada con el gas inerte y el correspondiente incremento de la energía interna de la misma. ($A^+ + X \rightarrow A^{+*} + X$)
2. La disociación (fragmentación) de la molécula. ($A^{+*} \rightarrow A_1 + A_2$)

Donde A^+ corresponde al ión que se desea fragmentar (A^{+*} es el ión en el estado excitado) y X a la molécula de gas neutro.

Los procesos de CID se dividen en CID en baja energía (< 100 eV) y CID en alta energía (> 100 eV) dependiendo de la energía con que se acelera la molécula antes de producir la colisión con el gas neutro [24, 49, 53]. Estos dos tipos de activación molecular producen distintos tipos de fragmentación asociados con el tiempo de la transformación de la energía cinética en energía interna en la molécula. En ambos casos, la colisión entre la molécula a estudiar y el gas inerte es una colisión inelástica de manera que parte de la energía cinética inicial es transformada en energía interna que permite la disociación de la molécula.

La manera más sencilla de estudiar esta colisión es hacerlo en el sistema centro de masa (CM) ya que en este sistema el momento total es igual a 0. En el sistema CM, las leyes de conservación de la energía y momento permiten calcular la máxima energía cinética disponible para su transformación en energía interna, la cual resulta ser igual a la energía cinética en el sistema CM a la cual notaremos como E_k^{CM} [24, 49]. De manera sencilla puede hallarse la relación entre E_k^{CM} y la energía cinética en el sistema de referencia laboratorio (E_k^L):

$$E_k^{CM} = E_k^L \frac{m_X}{m_A + m_X} \quad (2.14)$$

En la práctica, no toda esta energía disponible es transformada a energía interna ΔE_i , y la Ecuación 2.14 es una aproximación cuando ambas especies se consideran esferas

unimoleculares [49]. Ecuaciones más exactas toman en cuenta el ángulo de dispersión luego de la colisión para determinar la fracción de energía cinética efectivamente transformada a energía interna del ión precursor [49, 50]. Una ecuación más exacta que la Ecuación 2.14 expresa ΔE_i promedio transferida como:

$$\overline{\Delta E_i} = \frac{1}{2} \chi E_k^L \quad (2.15)$$

Donde

$$\chi = 4 \frac{m_a m_X}{(m_a + m_X)^2} \quad (2.16)$$

La Ecuación 2.15 es válida cuando la masa del ión proyectil m_A es mucho mayor que la masa de la molécula objetivo m_X , por lo que la colisión efectiva se produce únicamente entre la molécula objetivo y una sección del ión proyectil de masa m_a ($m_a \ll m_A$); lo cual es el caso de la fragmentación de péptidos en donde m_A está normalmente sobre los 600 Da. Si bien la Ecuación 2.15 ya no considera al ión proyectil como un cuerpo rígido, el problema real es mucho más complicado y la mayoría de resultados relacionados con la transferencia de energía interna se basan en resultados experimentales obtenidos con técnicas como la espectrometría de masas por resolución angular (ARMS) [51–53] o la espectroscopia de energía traslacional (TES) [54, 55]. Además, la transferencia de energía hacia los grados internos de libertad de la molécula es un problema mecano-cuántico, como se indicará más adelante. La Figura 2.16 muestra los espectros de masa y distribuciones de energía interna para diferentes valores de E_k^{CM} .

A más de las consideraciones sobre la energía interna depositada por colisión, se debe tomar en cuenta las propiedades estadísticas del ambiente en el que estas colisiones se desarrollan. Así, por ejemplo, la presión afecta el número de colisiones por molécula así como el número de colisiones totales, por lo cual se busca regular la presión para alcanzar la máxima intensidad en los iones fragmento o por ejemplo reducir la intensidad del ión precursor al 30 %. En trabajos previos se ha analizado la correlación entre el número de colisiones y el porcentaje de reducción en la intensidad del ión precursor [53]; en particular para un trayecto de colisión de 1 cm una reducción en la intensidad superior al 70 % indica claramente la ocurrencia de múltiples colisiones. En el caso de celdas de colisión cuadrupolares como en un espectrómetro de tipo QqTOF se garantizan múltiples colisiones lo que aumenta la energía interna efectivamente adquirida por el ión proyectil [49]. La reducción en la intensidad tiene que ver con un mayor número de reacciones

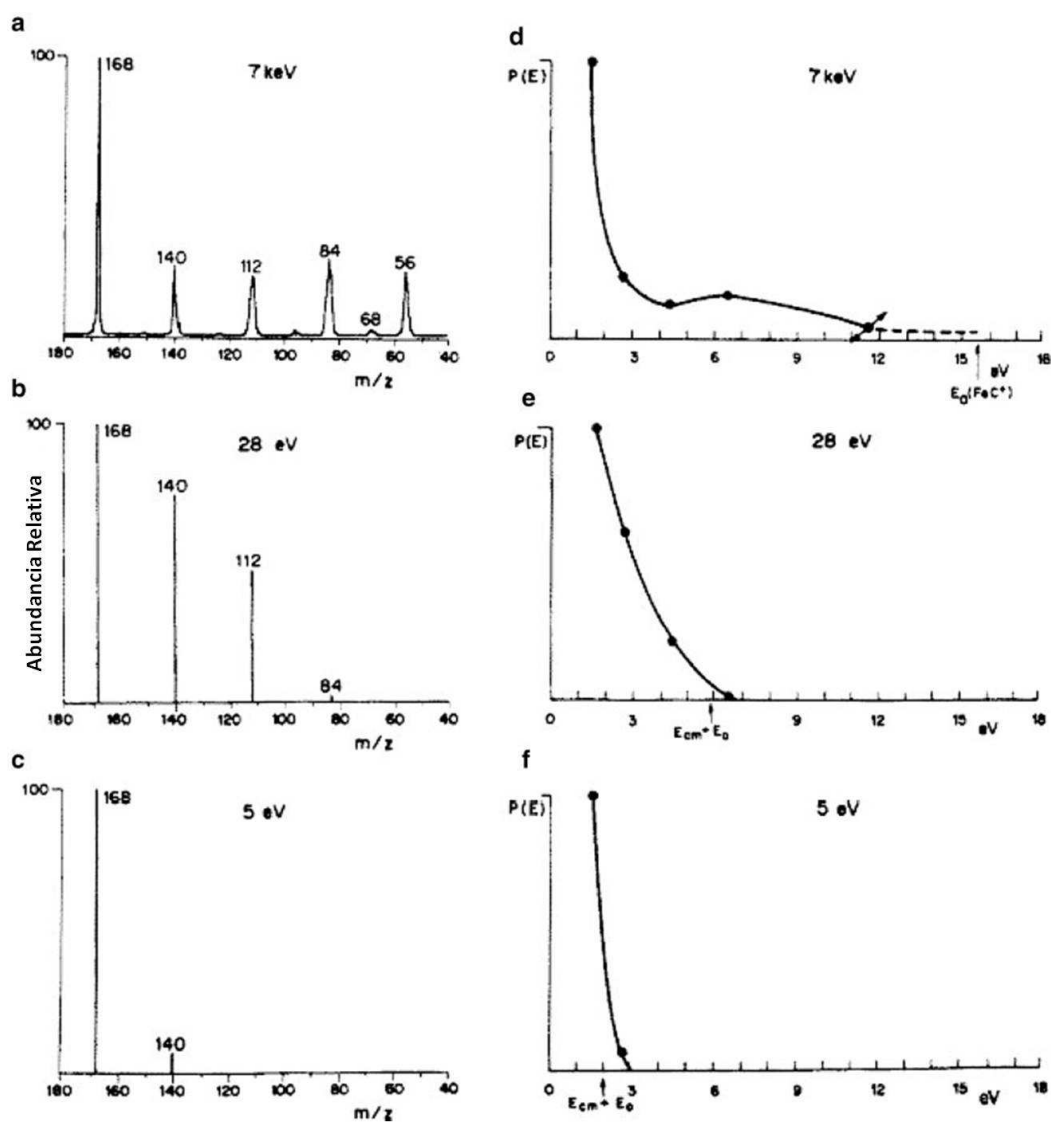


Figura 2.16: Distribución de Energías Internas en CID. Espectros de masas de fragmentación CID y sus correspondientes distribuciones de energía (a,d) 7 keV, (b,e) 28 eV, (c,f) 5 eV para $\text{Fe}(\text{CO})_5^-$. [56]

de disociación molecular, es decir, un mayor número de iones precursores se fragmenta al atravesar la cámara de colisión. Esto implica que una disminución en la intensidad del ión precursor se ve acompañada por un incremento en la intensidad de los iones fragmento observados.

Las tasas de decaimiento en el proceso de fragmentación molecular, es decir, la probabilidad de una reacción del tipo $A^{+*} \rightarrow A_1 + A_2$ a partir del incremento en la energía interna ha sido descrita en términos de la teoría de cuasi-equilibrio de Rice-Ramsperger-Kassel-Marcus (RRKM/QET) [50, 57]. La teoría se basa en la premisa de que las tasas en la reacción unimolecular dependen de los modos vibracionales cuánticos de la molécula, y no en el modo como la molécula fue activada. La teoría se basa en 4 suposiciones básicas [24]:

1. El tiempo del proceso de disociación es largo comparado con el tiempo de excitación de la molécula.
2. La tasa de disociación es baja comparada con la tasa de redistribución de la energía en los modos internos de la molécula.
3. Antes de la disociación el ión llega a una condición de equilibrio en que la energía se distribuye con igual probabilidad en todos los modos internos ($3N - 6$ modos vibracionales para una cadena lineal de átomos).
4. Los productos de la disociación son el resultado de un conjunto de reacciones que compiten entre sí.

La teoría considera a la molécula como un conjunto de s osciladores armónicos idénticos, de los cuales uno de ellos es denominado el oscilador crítico y se asocia con la coordenada correspondiente a la reacción. La teoría permite aproximar la constante de decaimiento $k(E, \nu)$ como [50, 57]:

$$k(E, \nu) = \nu \left(\frac{n - m}{n} \right)^{s-1} = \nu \left(\frac{E - E_0}{E} \right)^{s-1} \quad (2.17)$$

Donde $E = nh\nu$ y $E_0 = mh\nu$ son las energías del estado base y del estado excitado del oscilador crítico.

La **Ecuación 2.17** es una aproximación que permite entender de manera cualitativa el comportamiento de la constante de reacción. La ecuación más exacta es de la forma:

$$k(E, \nu) = \frac{\sigma \Omega(E - E_0)}{h \rho(E)} \quad (2.18)$$

Donde σ es la degeneración de la reacción, $\Omega(E - E_0)$ es el número de estados con energía menor a $E - E_0$ y $\rho(E)$ es la densidad de estados para una energía E .

Existen algunos métodos para calcular los valores de Ω y ρ , cuya discusión excede los propósitos de este trabajo. Algunos de estos métodos pueden revisarse en [57].

La teoría RRKM/QET permite calcular exactamente las constantes de reacción para sistemas simples y así predecir la abundancia de los diferentes picos en el espectro de fragmentación. Sin embargo para el caso de moléculas más complejas como los péptidos con los que se tiene que lidiar en la proteómica, los cálculos exactos se vuelven complejos y se trabaja con resultados de tipo cualitativo [58]. Estos resultados sugieren un papel crucial del protón adicional (protón obtenido durante la etapa de ionización) en la fragmentación del péptido, por la cual el modelo toma el nombre de *Modelo del Protón Móvil* [59]. Según este modelo, la fragmentación de los péptidos es mediada por la presencia del protón, el cual juega un papel crucial en la zona de ruptura del enlace, lo cual se observa por la relación entre la energía necesaria para la disociación y la basicidad del aminoácido más básico presente en el péptido [58, 59]. Esta evidencia indica que la probabilidad de que el protón adicional se ubique en el oxígeno del grupo carboxilo o en el nitrógeno del grupo amino [59] (ver en la **Figura 2.17**) aumenta cuando el protón se encuentra menos atraído por el aminoácido más básico del péptido. Esta ubicación del protón facilita la ruptura del enlace para producir iones de tipo *b* y tipo *y* según se encuentra descrito en la literatura [59–61].



Figura 2.17: Ubicación del protón previa la fragmentación en la columna del péptido. [59]

Hay que recalcar que esta forma particular de fragmentación tiene prevalencia en CID ya que el extendido tiempo de la reacción hace posible la excitación de los modos vibracionales de las moléculas según se describe en la teoría RRKM. En este sentido, la fragmentación en iones tipo *b* y iones tipo *y* se produce gracias a tiempos de transferencia de energía de aproximadamente 10^{-13} s en el caso de CID a bajas energías y 10^{-14} s en el caso de CID a altas energías [49, 53]. Estos tiempos de interacción son del orden de la frecuencia natural de las vibraciones moleculares en el caso de bajas energías, y al periodo de Bohr en los electrones de valencia de una molécula poliatómica [49, 53]. Todo esto implica que las colisiones logran inducir excitaciones electrónicas en el caso de colisiones a altas energías y excitaciones directas de los modos vibracionales en colisiones a bajas energías. En ambos casos, la relajación del sistema se produce por medio de un proceso ergódico que conduce a la ruptura de los enlaces más débiles según las teorías de RRKM/QET y los postulados presentados anteriormente.

2.5.3. DISOCIACIÓN POR CAPTURA ELECTRÓNICA Y DISOCIACIÓN POR TRANSFERENCIA ELECTRÓNICA

Otros métodos a más de CID se han desarrollado para la disociación de moléculas en estado gaseoso. De estos métodos recalcamos el caso de la disociación por captura electrónica o ECD [62] por sus siglas en inglés (Electron Capture Dissociation), y la disociación por transferencia electrónica o ETD [63] por sus siglas en inglés (Electron Transfer Dissociation). Estos dos métodos son de especial importancia porque producen patrones de fragmentación que son complementarios a aquellos obtenidos por medio de CID. Esto presenta grandes ventajas cuando se busca determinar la secuencia primaria de un péptido a través de la información de los espectros de fragmentación pues ayuda a eliminar posibles sitios de fragmentación que fueron pasados por alto en alguno de los dos métodos. Resultados previos sobre los patrones de fragmentación utilizando CID o ECD/ETD muestran una correlación cercana a cero cuando se analiza la preferencia en aminoácidos al momento de producirse la fragmentación molecular [64].

La fragmentación por medio de ECD ocurre cuando un ión precursor captura un electrón de baja energía ($<1\text{eV}$) formando una especie radical excitada que rápidamente se disocia. La energía de recombinación de 5 - 7 eV produce el incremento en la energía interna del ión precursor lo que provoca la rápida disociación del mismo. En este caso, el proceso de activación ocurre en tiempos mucho menores a los 10^{-14} s por lo que el proceso de disociación se considera como no ergódico y no necesariamente favorece

la ruptura de los enlaces más débiles como ocurre con la fragmentación por CID. En el caso de péptidos es el enlace C α -N el que se rompe durante la disociación produciendo mayormente iones fragmento de tipo *c* y tipo *z*. Además, ECD facilita la fragmentación de péptidos de masas altas o proteínas pues no depende del número de modos normales de la molécula.

Debido a los tiempos de reacción necesarios para obtener una buena relación entre señal y ruido (10^{-3} s) y a la baja relación de *m/z* para los electrones, la técnica de fragmentación por ECD estaba originalmente limitada a analizadores de tipo FT-ICR [65]. Sin embargo, el estudio de reacciones ion - ion en fase gaseosa permitió el desarrollo de la técnica de fragmentación ETD en las que un anión (de baja afinidad electrónica, generalmente aniones de antraceno) juega el papel del electrón utilizado en ECD. Al utilizar aniones mucho más masivos que el electrón la técnica se aplica no sólo para el caso de analizadores de tipo FT-ICR pero también para analizadores del tipo trampas de iones y cuadrupolares. En la actualidad la técnica de ETD en trampas de iones o cuadrupolos es utilizada rutinariamente [66, 67].

CAPÍTULO 3

BASES DE LA PROTEÓMICA

Las proteínas son componentes básicos de los seres vivos, puesto que cumplen un sin número de funciones, ya sea como catalizadoras de reacciones bioquímicas, como componentes estructurales en las células, como transportadores inter o intra celulares, entre otros. El área que se encarga del estudio de las proteínas expresadas en un organismo vivo se conoce como proteómica, término que se hereda del área de la genómica en la cual se estudia el mapa de genes característico de un organismo. La relación entre el genoma y el proteoma es estrecha pues todas y cada una de las proteínas que forman parte de un ser vivos están codificadas por genes específicos, los mismos que proveen con la receta básica para la construcción de las mismas [68]. Las proteínas a su vez pueden ser modificadas en su constitución por otros genes o por procesos celulares dependiendo de su funcionalidad o el rol fisiológico particular que las proteínas deben cumplir. Tanto la estructura primaria como las estructuras de orden superior son fundamentales en la definición de esta funcionalidad. En particular, la estructura terciaria, esto es la forma tridimensional geométrica, que adquieren las proteínas define en gran medida su actividad y funcionalidad [25, 68]. Por esta razón la caracterización de las proteínas así como las diferentes modificaciones a las que son sometidas, en los diferentes niveles de su estructura es uno de los problemas más básicos de la biología.

El dogma fundamental de la biología establece que el ADN de los seres vivos, o genoma, constituye el esquema o manual de construcción de los seres vivos [12, 69]. El genoma está constituido por partes que codifican instrucciones y partes que no lo hacen (la parte a veces llamada “junk DNA” en inglés, o ADN basura). La parte codificadora está constituida por los llamados genes. Mucha de la información del proteoma (o mapa completo de proteínas) de un organismo puede ser obtenida a partir de la información

contenida en el genoma [12, 69]. Sin embargo, la proteómica es un área mucho más compleja en el sentido de que el proteoma de un organismo varía con las condiciones en que éste se encuentra y puede cambiar a lo largo del tiempo. Es así que en muchos casos el punto de interés no es la proteína codificada en el genoma sino las modificaciones posteriores que esta sufre al interactuar con el entorno [25]. Gracias a avances tecnológicos recientes, en especial en el campo de la espectrometría de masas, el área de la proteómica ha tomado gran interés en los últimos años [2].

Los objetivos fundamentales de la proteómica son identificar, caracterizar, y cuantificar las proteínas presentes en una muestra. Para llevar a cabo estas tareas se apoya en la espectrometría de masas. En este ámbito, las mayores ventajas de la espectrometría de masas son su gran rapidez, y su alta sensibilidad, llegando a límites de detección de picomoles o femtomoles [70, 71]. En este capítulo discutimos la metodología utilizada para identificar y caracterizar péptidos y proteínas mediante la espectrometría de masas y las técnicas complementarias que brindan apoyo a la espectrometría de masas con este fin. En particular, la cromatografía líquida acoplada al análisis por espectrometría de masas tandem (LC-MS/MS) [3–5, 72] es uno de los métodos más utilizados y será descrito a mayor detalle en secciones posteriores.

3.1. CONCEPTOS BÁSICOS

Proteína: Para los fines concernientes, una proteína es una cadena consecutiva de polipéptidos, donde se entiende por polipéptido a una macro-cadena de aminoácidos. De manera general, los complejos de polipéptidos conforman lo que se conoce como complejos proteínicos, y entenderemos por proteína a alguna de las subunidades que conforman el complejo proteínico.

Péptidos y Polipéptidos: Un péptido o polipéptido es una cadena lineal de aminoácidos unidos consecutivamente mediante enlaces peptídicos (≤ 10 aminoácidos = Péptidos, ≤ 100 aminoácidos = Polipéptidos). La estructura general de un péptido puede verse en la [Figura 3.1](#). Todos los aminoácidos tienen la misma estructura y se componen de un grupo amino (terminal-N), un grupo carboxilo (terminal-C) y un residuo R que es diferente para cada uno de los 20 aminoácidos más comunes (ver [Tabla 3.1](#)). Cuando un aminoácido se une a la cadena de un péptido, se forma un enlace del tipo $N-C\alpha-C-N$ y se pierde una molécula de H_2O . A la cadena de enlaces entre distintos aminoácidos

se la conoce como la columna del péptido [11].

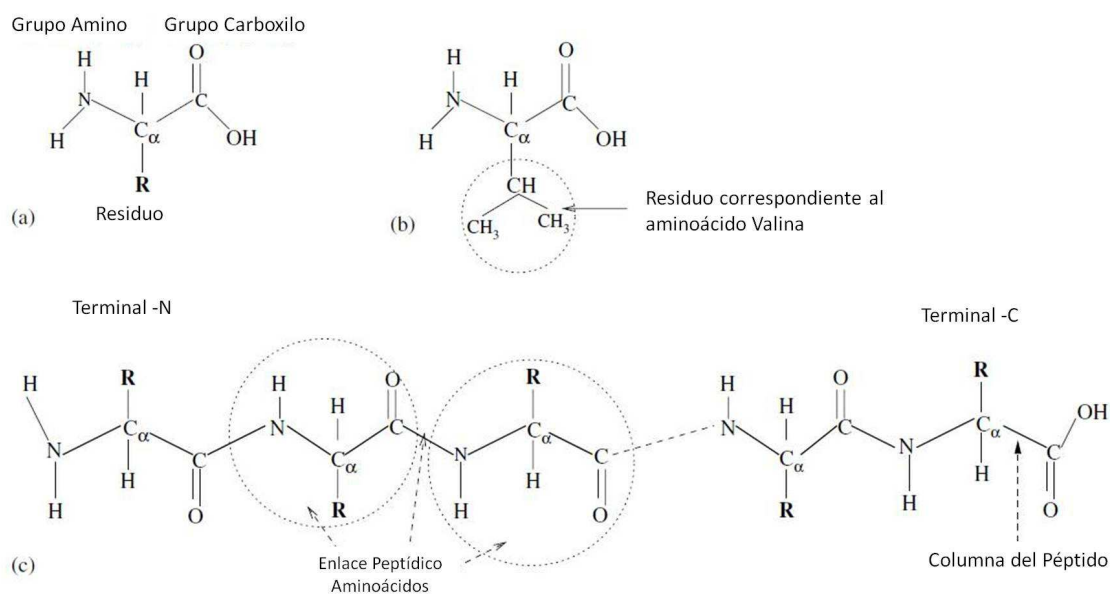


Figura 3.1: Estructura de un péptido. Un péptido se forma por la unión consecutiva de aminoácidos, los cuáles presentan la misma estructura salvo el residuo R que los diferencia. (a) Estructura general de un aminoácido. (b) Detalle del residuo correspondiente a la Valina. (c) Estructura general de un péptido. [11]

Perfil Isotópico: El perfil isotópico de una especie corresponde a todos los picos observados en el espectro que corresponden a distintas combinaciones de los isótopos existentes compatibles con su composición. La **Figura 3.2** muestra el perfil isotópico característico de un péptido. En el caso de péptidos o sustancias orgánicas, los perfiles isotópicos aparecen por la presencia de isótopos en los elementos H, C, O, N (ver **Tabla 3.2**). Dependiendo de la masa molecular del péptido, el perfil isotópico es diferente y la intensidad relativa entre los distintos picos isotópicos cambia.

Deisotopización: La deisotopización consiste en transformar todo el perfil isotópico de un pico en el espectro a un sólo pico correspondiente a la masa monoisotópica o a la masa promedio de la especie.

Deconvolución del estado de carga: En la mayoría de casos, la información proporcionada por el espectro de masas debe ser lo más simple posible e incluir sólo picos representativos de las especies presentes en la muestra. En el caso de que exista un mismo ión con diferentes estados de carga, la deconvolución del estado de carga consiste en transformar todos los picos del espectro a uno sólo con un único estado de

Tabla 3.1: Masas moleculares de los 20 aminoácidos más comunes [11]

Aminoácido			Masa Monoisotópica	Masa Promedio	% de ocurrencia ^a
Alanina	Ala	A	71.04	71.08	7.87
Arginina	Arg	R	156.10	156.19	5.42
Asparagina	Asn	N	114.04	114.10	4.13
Ácido Aspártico	Asp	D	115.03	115.09	5.34
Cisteína	Cys	C	113.01	113.15	1.50
Glutamina	Gln	Q	128.06	128.13	3.96
Ácido Glutámico	Glu	E	129.04	129.12	6.66
Glicina	Gly	G	57.02	57.05	6.95
Histidina	His	H	137.06	137.14	2.29
Isoleucina	Ile	I	113.08	113.16	5.91
Leucina	Leu	L	113.08	113.16	9.65
Lisina	Lys	K	128.10	128.17	5.92
Methionina	Met	M	131.04	131.20	2.39
Fenilalanina	Phe	F	147.07	147.18	3.95
Prolina	Pro	P	97.05	97.12	4.82
Serina	Ser	S	87.03	87.08	6.84
Threonina	Thr	T	101.05	101.11	5.41
Triptofano	Trp	W	186.08	186.21	1.13
Tirosina	Tyr	Y	163.03	163.18	3.02
Valina	Val	V	99.07	99.13	6.73

^a El % de ocurrencia es el promedio de apariciones de cada aminoácido en las secuencias de la base de datos Swiss-Prot (versión 5.21)

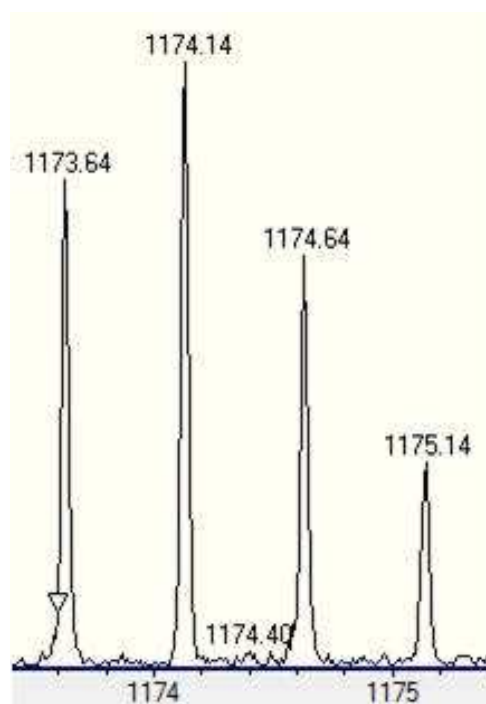


Figura 3.2: Perfiles Isotópicos. Un péptido o cualquier sustancia orgánica presenta un perfil isotópico en el espectro de masas producto de los isótopos de los elementos H, C, O, N. En este caso los picos isotópicos se encuentran separados por 0.5 unidades en m/z lo que permite identificar la carga de la especie como +2.

Tabla 3.2: Isótopos estables en los compuestos orgánicos [73]

Elemento	Masa (Da)	% Abundancia Relativa
H	1.0078	99.9885
	2.0141	0.0115
C	12.0000	98.93
	13.0034	1.07
N	14.0031	99.632
	15.0001	0.368
O	15.9949	99.757
	16.9991	0.038
	17.9992	0.205

carga, el cual por facilidad es casi siempre elegido como el estado de carga con $z = 1$. Para un pico de m/z en el espectro, su deconvolución de carga lo convierte en un pico equivalente en la posición $((m/z - 1)z) + 1$. De esta manera, todos los picos que corresponden a una misma especie son sumados en intensidad en la posición del pico monocargado.

3.2. SEPARACIÓN DE PÉPTIDOS Y PROTEÍNAS

Como se mencionó anteriormente, dos de los objetivos de la proteómica son identificar y caracterizar las proteínas que conforman el proteoma de un organismo. Ninguna de estas dos tareas es sencilla, y entre algunos de los problemas se pueden mencionar la complejidad intrínseca de las muestras, así como las limitaciones propias de la técnica. En el caso de la espectrometría de masas, la resolución del espectrómetro es generalmente dependiente de m/z y valores muy altos presentan una menor resolución; además, los límites de detección de algunos tipos de analizadores no sobrepasan las decenas de kDa, así que el estudio de proteínas con masas mucho mayores a estos valores se ve limitado por las características propias del espectrómetro. Por otra parte, las muestras con que se trabaja en el campo de la proteómica son generalmente muestras complejas que incluyen cientos o miles de proteínas.

En cierta manera, para limitar estos problemas no se analiza directamente la muestra en el espectrómetro de masas; se sigue, más bien, una serie de pasos previos que permiten obtener espectros de masas mucho más sencillos y fáciles de analizar. En la **Sección 3.3** se discuten a detalle las dificultades que existen en la interpretación y análisis de los espectros de masas. Las etapas previas al análisis por medio de MS generalmente se encargan de separar los péptidos y proteínas según ciertas características, como pueden ser su masa o su punto de enfoque isoelectrico. Esta separación previa, a más de brindar información útil en el proceso de caracterización, permite analizar gradualmente la muestra en el espectrómetro en lugar de analizar todas las especies en un solo paso. Por su parte, para asegurar una buena resolución y mantenerse dentro del límite de detección del espectrómetro, por lo común: (1) las proteínas son digeridas enzimáticamente de manera controlada, y (2) los polipéptidos resultantes son analizados en el espectrómetro.

Comúnmente para estos pasos de separación previa se trabaja con separación en gel

por masa molecular), por punto de enfoque isoelectrico o por una combinación de ambas (separación en 2 dimensiones) [11, 74]. Adicionalmente, las técnicas de cromatografía líquida o cromatografía en gas permiten una separación de los péptidos de acuerdo a la polaridad de los mismos, y en particular la cromatografía líquida permite el acople directo con la ionización por ESI.

3.2.1. DIGESTIÓN ENZIMÁTICA DE PROTEÍNAS

La digestión enzimática es el proceso por el cual una proteína es degradada por el efecto de una enzima proteolítica, también llamada proteasa. En las aplicaciones de la MS a la proteómica se trabaja con enzimas que se anclan en ciertas zonas específicas de las proteínas y rompen los enlaces únicamente en esas zonas. La proteasa más utilizada para la caracterización de proteínas por medio de MS es la tripsina, sin embargo, existen otros tipos de proteasas que producen la ruptura de la cadena de aminoácidos en zonas diferentes, y pueden ser utilizadas para obtener péptidos complementarios a aquellos obtenidos por medio de la tripsina. Entre ellas se destacan las proteasas GluC, LysC, AspN y Chymotripsina [75].

Tabla 3.3: Especificidad de las proteasas más comunes [75]

Proteasa	Especificidad
Tripsina	Terminal C de los residuos R y K Si no están seguidos por el residuo P
GluC	Terminal C del residuo Q Si no están seguidos por el residuo P
LysC	Terminal C del residuo K
AspN	Terminal N del residuo D
Chymotripsina	Terminal C de los residuos Y, F, W y L Si no están seguidos por el residuo P

3.2.1.1. Tripsina

Existen varias razones por las cuales la tripsina es la proteasa de elección:

- Tiene una alta especificidad (pocas anclajes no realizados) y casi nunca se ancla en zonas distintas a las predichas.
- Los péptidos resultantes tienen un promedio de 11 aminoácidos en longitud, i.e., una masa molecular adecuada para el análisis por medio de MS.
- Es particularmente útil en la espectrometría de masas con iones positivos al asegurar la presencia de residuos básicos (R y K) en los péptidos.
- Es relativamente fácil de obtener.

3.2.1.2. Digestión *In Silico*

Se conoce como digestión *In Silico* a la digestión teórica de una proteína. Como se explicará más adelante, la identificación de péptidos y proteínas por medio de MS se basa en la comparación de espectros de masa experimentales con espectros de masa teóricos correspondientes a la proteína o péptido correspondiente. En este sentido, la digestión *in silico* es el conjunto de masas (y por tanto el espectro) que se espera obtener al realizar la digestión enzimática de alguna proteína. Para realizar una digestión *in silico* se debe tomar en cuenta la proteasa elegida, posibles modificaciones en los péptidos y anclajes no realizados. Los anclajes no realizados son raros cuando se utiliza tripsina, sin embargo, pueden suceder y por tanto es preferible considerar de 1 a 2 anclajes no realizados en la digestión *in silico*. Si por ejemplo la digestión teórica de una proteína produce n péptidos y se consideran k anclajes no realizados, el número de total de péptidos a considerar en la digestión *in silico* será de $\frac{1}{2}(2n + 2nk - k^2 + k + 2)$ [11]. Por su parte, las modificaciones en los péptidos pueden ser modificaciones fijas o modificaciones variables.

- Las modificaciones fijas son fáciles de simular ya que simplemente cambian la masa de uno o varios aminoácidos en toda la muestra.
- Las modificaciones variables son aquellas que pueden o no aparecer en la muestra y son las que aumentan en gran medida la complejidad de la digestión *in silico*.

En el caso que se observen n posibles modificaciones variables en la secuencia de un péptido, se deberían considerar 2^n masas moleculares distintas en la digestión *in silico*, lo cual incrementa el número de posibles identificaciones erradas cuando se compara un espectro experimental con la información de la digestión *in silico*.

Dentro del grupo de posibles modificaciones existen aquellas que se dan dentro de la célula y que son realizadas por la maquinaria celular. Por ejemplo, hay grupos que se añaden a la proteína durante el proceso de su fabricación y que la dotan de propiedades especiales. Un ejemplo de esto es la glicosilación [76], donde, enzimas especiales, atan azúcares a la proteína a medida que ésta es producida por los ribosomas. Otros ejemplos, incluyen la fosforilación [77], entre otras. La caracterización de estas llamadas modificaciones postranslacionales es fundamental en el estudio proteómico.

3.2.2. CROMATOGRAFÍA LÍQUIDA

La cromatografía es una técnica originalmente propuesta por Michael Tswett [78, 79] a comienzos del siglo 20 y en la actualidad es uno de los métodos más utilizados en la proteómica para la separación de péptidos y proteínas. De manera sencilla, la cromatografía se basa en la existencia de una fase móvil que transporta la muestra a través de una fase estacionaria fija en la columna de separación cromatográfica [12]. La muestra atraviesa la columna, y dependiendo de la relación entre su interacción con la fase estacionaria y la fase móvil le tomará mayor o menor tiempo abandonar la columna. El tiempo de elución de un componente a través de la columna cromatográfica se conoce como **tiempo de retención**. Generalmente, la fase estacionaria se encuentra dentro de una columna de un diámetro y longitud determinadas, y la separación se produce por los distintos tiempos de retención al atravesar la columna.

Los tipos de cromatografía más utilizados son aquellos en que la fase móvil se presenta en estado líquido o gaseoso, dando origen a la cromatografía en gas o GC por sus siglas en inglés, y la cromatografía líquida o LC por sus siglas en inglés. De las dos, la LC es la más ampliamente utilizada en proteómica pues los péptidos y proteínas al ser compuestos no volátiles son difíciles de acoplar con un sistema de cromatografía en fase gaseosa [12].

En la cromatografía líquida, la muestra es transportada en el estado líquido a través de la fase estacionaria, la cual consiste de un ligando adherido a la columna cromatográfica por medio de partículas esféricas de sílica. El ligando depende del tipo de interacción que se busque y del tipo de cromatografía que se esté empleando. Existen básicamente 5 tipos de LC que son los más utilizados, sin embargo, la cromatografía de fase reversa es la más utilizada en la separación de péptidos y proteínas. Para propósitos analíticos discutimos sus características con mayor detalle:

La descripción de las técnicas de los distintos mecanismos de cromatografía está basada en el capítulo 4 del texto de Eidhamer et. al. [11], en el capítulo 5 del texto de Hoffman et. al. [24], y en el capítulo 4 del texto de Ekman et. al. [12].

3.2.2.1. Cromatografía de fase normal

En la cromatografía de fase normal, la fase estacionaria es polar y la fase móvil es relativamente no polar. En este tipo de cromatografía los compuestos son separados en base a su polaridad de manera que los compuestos más polares tendrán un tiempo de retención mayor que aquellos menos polares. Al aumentar la polaridad de la fase móvil se logran tiempos de retención menores. Se utiliza principalmente cuando se trata de separar compuestos polares.

3.2.2.2. Cromatografía de fase reversa

La cromatografía de fase reversa (RE-LC) es largamente la más utilizada en las aplicaciones a MS, y al igual que la cromatografía en fase normal, se basa en la polaridad de los compuestos, pero en este caso la fase estacionaria es menos polar que la fase móvil. En la RE-LC se utilizan columnas de sílica con cadenas de átomos de carbono como ligandos (C_4 , C_8 , C_{18}). La ventaja de este método respecto a la cromatografía en fase normal es que muestra mayor reproducibilidad, y el tiempo de vida útil de las columnas es mayor. En la RE-LC los menores tiempos de retención corresponden a los compuestos más hidrófobos, y al igual que en la cromatografía en fase normal, los tiempos de retención pueden variarse al cambiar la polaridad de la fase móvil.

La fase móvil generalmente está compuesta de una mezcla de dos soluciones, notadas como A y B. La solución A es una solución polar, generalmente H_2O (grado analítico) con 0.1 % de TFA (ácido trifluoroacético), y la solución B suele ser una solución menos polar, como por ejemplo 90 % acetonitrilo acuoso con 0.1 % de ácido fórmico. En un principio, la muestra disuelta en A es introducida en la columna, y por afinidad los compuestos presentes en la muestra (péptidos) son retenidos en la fase estacionaria. Gradualmente, se incrementa el porcentaje de B (este incremento se conoce como gradiente en la columna cromatográfica) de manera que la interacción compuesto - fase móvil se incrementa y los compuestos son paulatinamente liberados de la fase estacionaria. Normalmente, en un 60 % de B todos los péptidos han sido eluidos de la columna cromatográfica.

3.2.2.3. Cromatografía de intercambio iónico

La cromatografía de intercambio iónico (IEC) se basa en las interacciones electrostáticas entre los grupos funcionales de los ligandos en la fase estacionaria y los compuestos presentes en la muestra. Existen dos tipos de IEC y dependen de la carga en los grupos funcionales de los ligandos en la fase estacionaria, la cual puede ser negativa (Cromatografía de intercambio aniónico), o positiva (Cromatografía de intercambio catiónico). El tiempo de elución a través de la columna se puede controlar variando el PH de la fase móvil para afectar la fuerza de la interacción electrostática.

3.2.2.4. Cromatografía por afinidad molecular

La cromatografía por afinidad molecular se basa en las interacciones biológicas que pueden existir entre los ligandos en la fase estacionaria (receptores, anticuerpos, metales, etc) y grupos funcionales de los compuestos en la muestra (péptidos o proteínas). Generalmente se emplea para purificar muestras de interés o para separar formas funcionales de un mismo biopolímero.

3.2.2.5. Cromatografía de exclusión molecular

La cromatografía de exclusión molecular o también conocida como cromatografía de exclusión en gel separa los compuestos en base a su tamaño mientras recorren la columna cromatográfica formada por un entramado de biopolímeros. La fase estacionaria consiste en un entramado de polímeros en estado de gel de manera que se forman poros y caminos a través de la columna, lo que hace que los compuestos de diferentes tamaños tengan distintos tiempos de retención en la columna.

Como dijimos, el método más utilizado es el de la cromatografía de fase reversa, y generalmente cuando se habla de LC se hace referencia a este tipo de cromatografía. Originalmente, la elución de la muestra a través de la columna cromatográfica era llevada a cabo únicamente bajo la acción de la gravedad, sin embargo, en la actualidad se utilizan bombas de alta presión para mejorar el flujo a través de la columna, con el consiguiente incremento en la eficiencia de la separación. Es así que se habla de HPLC (cromatografía líquida de alto rendimiento o cromatografía líquida de altas presiones) o

UPLC (cromatografía líquida de ultra rendimiento o cromatografía líquida de ultra altas presiones) las cuáles hacen referencia a la LC con altas presiones.

Las columnas cromatográficas más utilizadas en las aplicaciones a la proteómica tienen una longitud de entre 15 y 20 cm, y el diámetro interno depende de la aplicación específica. En procesos preliminares se utilizan las micro columnas con diámetros de unos pocos milímetros, mientras que para la separación de péptidos se utilizan columnas capilares con diámetros menores a los 300 μm o nano columnas con diámetros de entre 50 a 100 μm [11].

En la **Figura 3.3** se muestra el esquema básico de un sistema de HPLC. El resultado de la separación por medio de HPLC, además de las fracciones de proteína colectadas, es lo que se conoce como cromatograma, en el cual se observa el tiempo de retención en el eje x y la intensidad de la señal en el eje y. La obtención del cromatograma puede hacerse por medio de espectroscopia de absorción a 280 nm (aminoácidos aromáticos) o a 210-220 nm (enlace peptídico), o en el caso de acople directo a un sistema de MS midiendo el número total de iones en el tiempo. El objetivo de un buen sistema de HPLC debe ser producir un cromatograma con el menor ancho de los picos cromatográficos (picos simétricos) y con una buena separación de los distintos compuestos de manera que no existe un traslape de los picos en el cromatograma.

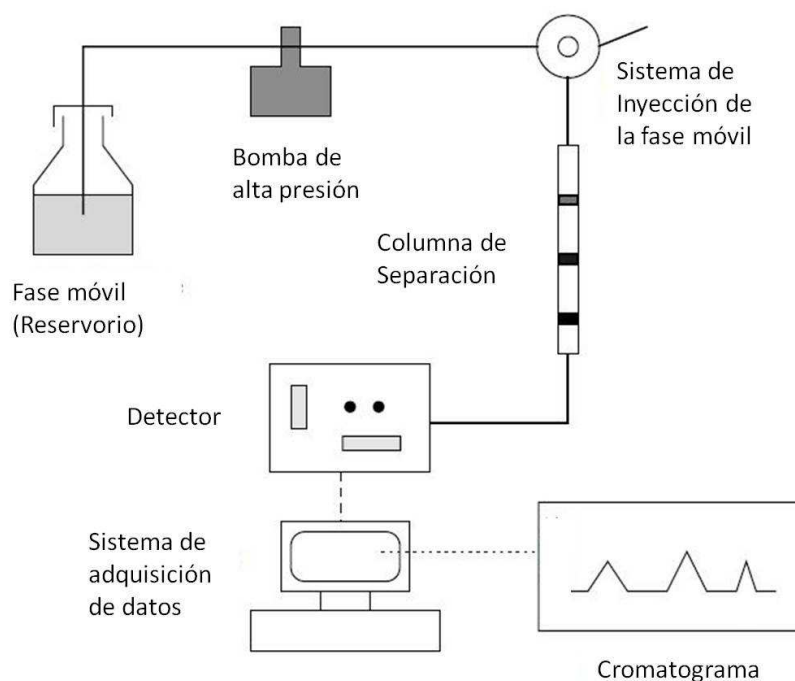


Figura 3.3: Esquema de un sistema de HPLC. [11]

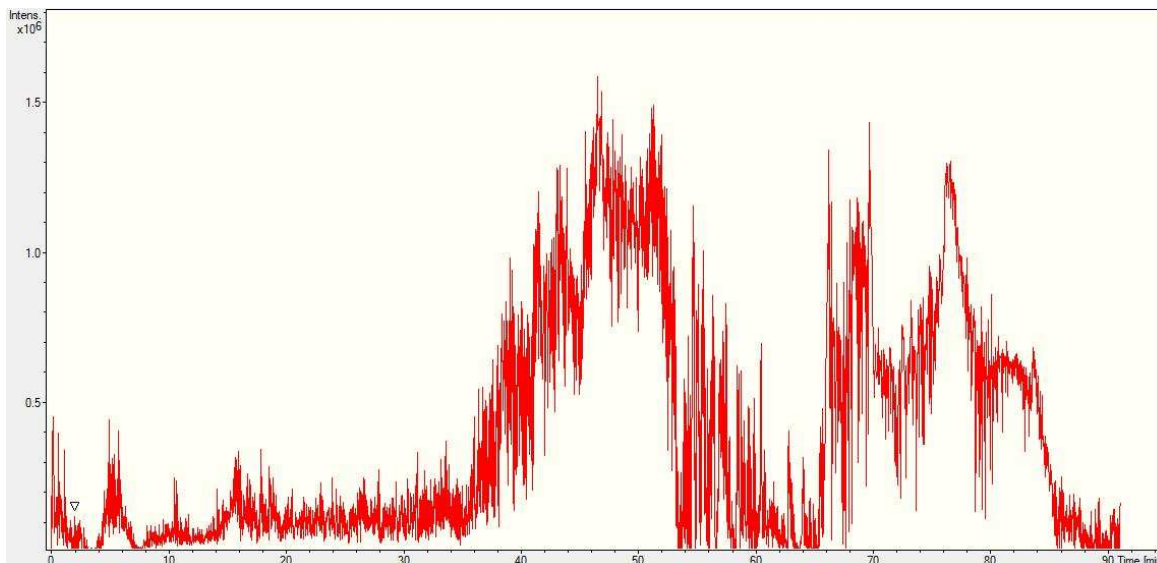


Figura 3.4: Ejemplo de un cromatograma. En el eje x se tiene el tiempo de retención y en el eje y la intensidad de la señal. El cromatograma corresponde a la digestión enzimática de una muestra de *E.Coli* en una columna de separación de 90 min. Cada pico corresponde a un péptido presente en la muestra

3.3. IDENTIFICACIÓN Y CARACTERIZACIÓN DE PÉPTIDOS Y PROTEÍNAS POR MEDIO DE MS

Como se ha mencionado, la espectrometría de masas permite determinar de manera muy precisa la masa molecular, y en esta sección respondemos a la pregunta de cómo esta información puede llevarnos a determinar la secuencia primaria de péptidos y proteínas, lo cual es uno de los objetivos de la proteómica. Los péptidos y proteínas se observan en los espectros de masas como iones producidos por la ganancia o pérdida de un protón, por lo cual un péptido de masa molecular m y carga z tendrá un valor de $(m + z)/z$ en el espectro de masa, donde por simplicidad hemos asumido la masa del protón igual a 1 Da.

En un espectro de masas se puede observar la misma especie con distintos estados de carga, en donde la predominancia de ciertos estados sobre otros depende principalmente del método de ionización empleado, y de la longitud de los péptidos. En el caso de MALDI normalmente la carga predominante será $z = 1$, pero si se utiliza ESI, la probabilidad de encontrar moléculas con estados de cargas superiores es mayor, y los iones con $z = 2, 3$ y 4 son especialmente observados en los espectros de masa. Por su parte, al trabajar con polipéptidos de mayor longitud, se incrementa la

probabilidad de tener más de un aminoácido con gran afinidad protónica en la cadena, y por consiguiente, estados de mayor estado de carga aumentan su probabilidad de ocurrencia.

A más de los diferentes estados de carga, los espectros de masas pueden contener el efecto del ruido de fondo o de ruido electrónico que afectan la claridad en la interpretación del espectro. Para sobrellevar todas estas dificultades existen un sin número de programas computacionales que realiza tareas necesarias antes de empezar con la interpretación de los espectros. Estas tareas referentes al tratamiento de los espectros de masa son brevemente descritas más adelante.

Como puede esperarse, la identificación de una proteína no puede obtenerse únicamente conociendo su masa molecular, por lo que existen dos enfoques distintos para obtener esta identificación. Los dos enfoques se conocen como top-down [62, 63, 80] y bottom-up [5, 72], y los dos enfrentan el problema de la identificación de proteínas desde dos puntos de vista contrapuestos. El enfoque bottom-up o también conocido como shotgun proteomics parte de la digestión de la proteína y su identificación por medio de la identificación primaria de los péptidos resultantes de la digestión enzimática. Por su parte en el enfoque top-down se analiza la proteína entera por medio de MS y se utilizan técnicas de fragmentación como ECD o PSD para analizar los péptidos resultantes de la fragmentación [62]. El enfoque más utilizado es el bottom-up por cuanto la resolución de los espectrómetros es reducida cuando se trabaja con masas de proteínas mayores a algunas decenas de kDa, lo que hace más complicado trabajar en el régimen del enfoque top-down. Sin embargo, los analizadores de tipo FT-ICR permiten hacer este tipo de análisis debido a su alta resolución en rangos de masas altos. En este trabajo nos centramos en el enfoque bottom-up y en particular en la información obtenida de la espectrometría tandem para la identificación de péptidos que será discutida en la [Sección 3.3.3](#).

Debe aclararse que en realidad, una estrategia más comprehensiva para la identificación exitosa de péptidos y proteínas típicamente consiste de una combinación de métodos top-down y bottom-up [81].

3.3.1. TRATAMIENTO DE LOS ESPECTROS DE MASA

Para facilitar la interpretación de los espectros de masa normalmente se siguen una serie de pasos que simplifican el espectro y ayudan por ejemplo a la discriminación

entre picos de señal y picos de ruido. La mayoría de estos procedimientos se realizan mediante softwares especializados, pues la cantidad de datos con la que se trabaja actualmente hace imposible una revisión y tratamiento manual de los espectros. Muchos de estos programas computacionales son parte integral del espectrómetro y sus licencias deben ser adquiridas directamente del fabricante del equipo.

3.3.1.1. Deisotopización y determinación del estado de carga

La determinación del estado de carga de un polipéptido es sencilla si se toma en cuenta la existencia del perfil isotópico. De esta manera, el estado de carga puede obtenerse al observar que los picos isotópicos distan entre sí una distancia de aproximadamente $1/z$ en el espectro de masa. Normalmente, el proceso de identificación de la carga se hace automáticamente mediante programas computacionales que identifican un perfil isotópico y realizan el proceso de deisotopización o deconvolución de la carga. (Ver [Figura 3.2](#)).

3.3.1.2. Promedio de espectros de masa

El promedio de espectros de masa consiste en un promedio simple de varios espectros de masa en donde las intensidades son sumadas y luego divididas para el total de espectros considerados en el promedio. Este procedimiento es muy utilizado cuando se trabaja con LC-MS ya que normalmente un pico cromatográfico que corresponde a un compuesto es adquirido en varios espectros de masa.

3.3.1.3. Sustracción del ruido de fondo (Background)

La mayoría de espectros de masa presentan una gran cantidad de ruido, lo que hace que la línea base no sea uniforme y aparezca una gran cantidad de señal indeseada. La sustracción del ruido de fondo se hace escogiendo algunos espectros a la derecha y a la izquierda de un pico cromatográfico (siempre y cuando la señal en el cromatograma sea baja en esa zona) y sustrayéndolos de los espectros obtenidos dentro de la zona del pico cromatográfico. Generalmente, primero se promedian los espectros correspondientes al ruido de fondo y el promedio es sustraído de los espectros con señal.

3.3.1.4. Espectros de masa con centroides

Los espectros de masa son en principio medidos de manera que los picos tienen un ancho característico, lo que origina el concepto de resolución espectral. Sin embargo, esta información no puede ser utilizada por ejemplo en búsquedas en bases de datos o *de novo* sequencing, de manera que los programas especializados se encargan de transformar estos picos a parejas de datos (m/z , Intensidad). Para esto, se determina una ventana en m/z que represente unívocamente al pico en el espectro de masas y suma las intensidades dentro de esa ventana. Finalmente, el centroide del pico determina el valor de m/z que representa el pico y la intensidad se determina como se mencionó anteriormente.

En principio la determinación del centroide debe hacerse en todos los espectros de masa para conocer la masa exacta del compuesto que se está observando, sin embargo, la determinación de la intensidad como se mencionó anteriormente solo es especialmente útil cuando se trabaja con espectrómetros de alta resolución y permite obtener información cuantitativa sobre la presencia del compuesto en la muestra. En otros casos, la intensidad del pico puede ser simplemente calculada como la altura del pico en su centroide.

3.3.2. IDENTIFICACIÓN POR MEDIO DE MS (PMF: PEPTIDE MASS FINGERPRINTING)

La digestión enzimática de las proteínas produce un conjunto de péptidos con masas moleculares características para la proteína que ha sido digerida, y en ese sentido, el espectro de masas de este conjunto de péptidos puede ser catalogado como una huella digital de la proteína a la que representa. Este es el principio en el que se basa la identificación por medio de PMF [82].

El procedimiento consiste en la digestión enzimática de la proteína y en la adquisición del correspondiente espectro de masa de los péptidos resultantes. Luego este espectro es comparado con espectros teóricos obtenidos mediante la digestión *in silico* de varias secuencias que se encuentran en una base de datos. Los factores que deben considerarse al realizar esta comparación son los siguientes [11]:

1. La especificidad de la proteasa utilizada para realizar la digestión *in silico*.

2. Posibles anclajes no realizados.
3. Modificaciones fijas y variables.
4. La tolerancia en la determinación de la masa.

3.3.2.1. Scores en la identificación de proteínas

Una vez que se han adquirido y analizado los datos de la MS, normalmente se cuenta no solo con una, sino con varias proteínas que pueden encajar con el espectro de masa de los péptidos obtenido. La asignación de scores a los posibles candidatos permite determinar aquellos que tienen una mayor probabilidad de ser los responsables del espectro de masas experimental observado, de manera que las identificaciones correctas siempre deberían tener un score más alto.

El score para una identificación puede descomponerse en los scores a nivel de péptidos y a nivel de proteínas [11]:

Score a nivel de péptidos: Considera diferencias en masa molecular teórica y experimental, el número de anclajes no realizados, el número de modificaciones, la intensidad de los picos en el espectro experimental, y las predicciones teóricas de los péptidos e intensidades que se esperan en el espectro de masas experimental.

Score a nivel de proteínas: Considera el número de péptidos identificados, el porcentaje de la secuencia cubierta experimentalmente, las diferencias en masa molecular teórica y experimental, las diferencias en el valor de pI teórico y experimental, la variación en los errores al nivel del péptidos.

Estos y otros factores son tomados en cuenta para elaborar el score global de una proteína candidata. El problema es determinar la mezcla adecuada de estos factores que provea un score global que favorezca a las identificaciones correctas sobre los falsos positivos. Existen distintos esquemas de scores propuestos y de manera general pueden clasificarse como scores probabilísticos y no probabilísticos. Los scores no probabilísticos combinan de alguna manera los factores mencionados anteriormente, mientras que los scores probabilísticos expresan la probabilidad de que una secuencia dada sea el origen de los espectros de masa observados.

Algunos de estos scores se construyen a través de un proceso de aprendizaje, es decir que los parámetros del esquema del score son obtenidos a través de identificaciones

previas y algoritmos de aprendizaje [83].

Valor P

El valor P es utilizado en la mayoría de esquemas de asignación de scores y expresa la significancia estadística de una identificación [11]. Si una secuencia de aminoácidos obtuvo un score S, entonces el valor P es la probabilidad de que se encuentre un score de valor mayor o igual a S por casualidad. Los algoritmos de identificación comúnmente presentan el score final como el valor de $-C\text{Log}P$, donde C es una constante.

El cálculo teórico de P dependerá del esquema particular de asignación de scores a las proteínas, y generalmente su cálculo es muy complicado. En este sentido, los programas aprenden la distribución de probabilidades para los scores S generando espectros aleatorios y calculando sus scores con respecto a la base de datos escogida.

3.3.3. ANÁLISIS POR MEDIO DE LA ESPECTROMETRÍA DE MASAS TANDEM

A pesar de que en principio se podría identificar proteínas por medio de PMF sin necesidad de realizar espectrometría de masas tandem, muchas veces en proteómica se trabaja con muestras que incluyen cientos de diferentes proteínas y es impráctico realizar procesos para separar las proteínas individualmente y hacer una identificación por medio de PMF. Además, por medio de la espectrometría tandem se pueden conseguir identificaciones de proteínas con coberturas de la secuencia de aminoácidos menores al 20 % [24], y esto debido a que a diferencia de PMF, cuando se trabaja con espectrometría tandem se busca determinar la secuencia primaria de los péptidos producidos en la digestión enzimática, lo que brinda mucha mayor información que solo la masa molecular de los péptidos producto.

En la [Sección 2.5](#) vimos los distintos mecanismos para producir la fragmentación controlada de péptidos, y en especial revisamos el caso de la disociación inducida por colisión. Como vimos, los patrones de fragmentación en péptidos entregan gran información sobre la secuencia de aminoácidos en especial si se consiguen escaleras de fragmentación en la columna del péptido. En particular el caso de la fragmentación por CID produce iones de tipo *b* y tipo *y* que permiten dilucidar la secuencia primaria del péptido.

3.3.4. BASES DE DATOS

Al igual que en el caso de la identificación por medio de PMF, la identificación por medio de espectrometría de masas tandem se basa principalmente en la comparación de espectros obtenidos experimentalmente con supuestos candidatos que pueden obtenerse a partir de una base de datos adecuada. Cuando hablemos sobre *de novo* sequencing discutimos el caso en el que no se cuenta con espectros teóricos obtenidos de una base de datos.

En el caso de la espectrometría de masas tandem, la información con que se cuenta incluye los espectros de fragmentación de los péptidos presentes en la muestra, por lo cual la identificación de las proteínas se puede realizar a partir de la información de la secuencia primaria de los péptidos obtenidos luego de la digestión enzimática. De manera sencilla el problema se expresa de la siguiente manera: dado un conjunto de péptidos obtenidos de la digestión *in silico* de las proteínas en la base de datos y sus correspondientes espectros de fragmentación teóricos, se debe encontrar aquellos que se asemejen en mayor grado a los espectros MS/MS experimentales.

En primera instancia, una comparación completa con todos los espectros producidos a partir de la base de datos tomaría demasiado tiempo, así que el procedimiento de comparación se realiza luego de aplicar filtros en la búsqueda en la base de datos. Los principales filtros son la masa del péptido precursor y posibles etiquetas de secuencia o sequence tags por su nombre en inglés (pequeñas secuencias de aminoácidos, generalmente de 2 ó 3 que pueden ser inferidos fácilmente del espectro). Al aplicar el filtro de masa debe considerarse la precisión en la determinación experimental de la masa, la cual es una característica intrínseca del instrumento. En este sentido, mientras mayor sea la precisión en la determinación de la masa, menor será el número de péptidos candidatos y por tanto el proceso de comparación es más ágil y toma menos tiempo. Además de esto, la búsqueda en la base de datos debe considerar modificaciones fijas y variables (ver [Sección 3.2.1.2](#)), la precisión en la determinación de la masa en espectros MS/MS, anclajes no realizados, y especificidad de la proteasa utilizada. En ciertos casos se puede contar con información adicional (polaridad, función de la proteína, etc.) que ayude a incrementar los filtros y mejorar la identificación acertada de los péptidos en la muestra.

Existen varios motores de búsqueda que utilizan diferentes algoritmos para filtrar la base de datos y asignar scores a las asignaciones de espectros experimentales. La asignación

nación de scores es semejante a la descrita en la [Sección 3.3.2.1](#).

Los principales motores de búsqueda son los siguientes:

- MASCOT¹
- SEQUEST²
- X!Tandem³

Cada uno de estos programas utiliza un diferente esquema para asignar scores a los espectros considerados [84–86]. De manera general, los esquemas de asignación de scores se basan en la presencia de picos predichos por la digestión *in silico* y en sus intensidades. La comparación de intensidades puede hacerse en distintos niveles dependiendo de la complejidad del algoritmo de comparación; es así que al crear un espectro teórico se lo puede realizar en 3 niveles:

1. Todos los picos en el espectro de fragmentación tienen la misma intensidad
2. Las intensidades de los picos son asignadas de acuerdo al tipo de ión que son: tipo *b*, tipo *y*, etc.
3. La intensidad de los picos es asignada de acuerdo al tipo de ión, longitud del péptido, secuencia, etc.

En base a esto, los esquemas de comparación pueden basarse en una búsqueda de picos correspondientes o en una división del espectro en intervalos de m/z y en una comparación de las intensidades en cada intervalo.

Como se mencionó anteriormente, existen algoritmos de asignación de scores probabilísticos y no probabilísticos, a partir de los cuales se obtiene una lista rankeada de las proteínas y péptidos que pueden estar presentes en la muestra. Si bien esta lista es generalmente acertada, éste no es siempre el caso y es recomendable validar las identificaciones obtenidas por medio de los programas comerciales [87].

¹<http://www.matrixscience.com/>

²<http://fields.scripps.edu/sequest/>

³<http://www.thegpm.org/tandem/>

3.3.5. DE NOVO SEQUENCING

Existen casos en los que la secuencia que se busca no se encuentra en la base de datos, y en estos casos el objetivo es encontrar la secuencia primaria de aminoácidos a partir de los espectros de fragmentación sin contar con ninguna información de partida. Este procedimiento es lo que se conoce como *de novo* sequencing.

La identificación por medio de *de novo* sequencing no es una tarea sencilla, y es por esta razón que sólo se utiliza con seguridad cuando se cuenta con espectros de alta precisión y alta resolución. Desde este punto de vista, el desarrollo de mejores espectrómetros en la actualidad le ha dado una mayor importancia a la identificación por medio de *de novo* sequencing. El problema en cuanto a la precisión en la determinación de la masa y la resolución viene dado principalmente por la semejanza en la masa de algunos aminoácidos o combinaciones de los mismos, considérese el caso de los aminoácidos:

Gln = 128.06 y Lys = 128.10

Val + Val = 198.14 y Pro + Thr = 198.10

Es así que si no se cuenta con una buena precisión y una buena resolución en la determinación de la masa, estos aminoácidos no pueden ser diferenciados a partir de la información en el espectro de masas.

La identificación mediante *de novo* sequencing se basa en la búsqueda de una escalera de iones del mismo tipo para determinar la secuencia de aminoácidos. Como se discutió en la [Sección 2.5](#), existen varios métodos para fragmentar los péptidos y dependiendo del método utilizado se obtienen patrones distintos de fragmentación. En el caso de la fragmentación por CID a bajas energías, los principales fragmentos que se observan son del tipo *b* y tipo *y*, por lo que al realizar *de novo* sequencing en CID-MS/MS se busca encontrar una escalera de iones tipo *b* o tipo *y* que permita elucidar la estructura primaria del péptido. En este sentido, la identificación por medio de *de novo* sequencing busca encontrar picos con masas correspondientes a iones de tipo *b* (*y*) (ver [Tabla 2.2](#)) que idealmente estarán acompañados por iones tipo *y* (*b*) complementarios y por picos $b(y)^o$ o $b(y)^*$ que ratifiquen la identificación del pico tipo *b* (*y*). En el caso ideal, la fragmentación generará toda una escalera completa de iones de un sólo tipo, con lo cual el problema se reduce a medir la diferencia entre los picos consecutivos y determinar el aminoácido correspondiente. Sin embargo, en la práctica los espectros de fragmentación se alejan mucho de la idealidad, y en la práctica muchas

veces se utilizan varias técnicas de fragmentación para mejorar la cobertura completa del péptido y obtener información complementaria que permita dilucidar la identidad del péptidos representado en el espectro de masas. Las principales combinaciones son CID con ECD/ETD [88] y CID en baja energía con CID en alta energía [89]. En el caso de CID con ECD/ETD se tiene la ventaja de contar con iones complementarios pues CID promueve fragmentos del tipo b y tipo y mientras que ECD/ETD promueve fragmentos del tipo a y tipo x . Por su parte la fragmentación CID a altas energías produce muchos más fragmentos que la fragmentación CID a bajas energías, por lo que produce no sólo fragmentos de tipo b o tipo y sino también fragmentos internos, fragmentos satelitales e iones de imonio que permiten ratificar la presencia de ciertos aminoácidos.

En resumen, el problema del secuenciamiento puede expresarse como:

Dado un espectro de masas $S = \{s_1, \dots, s_n\}$, la masa m del péptido, un conjunto de tipo de iones δ , y un esquema de asignación de scores Γ , encontrar la secuencia de aminoácidos P que maximice Γ dado S .

La solución para hallar P sería evaluar Γ para todos los péptidos cuya masa sea igual a m , sin embargo, un algoritmo eficiente debería utilizar la información contenida en el espectro para reducir el espacio de búsqueda. Estos algoritmos son utilizados en varios programas comerciales como PEAKS [90, 91], Lutefisk [92, 93], PepNovo [94], MSNovo [95], Sherenga [6], NovoHMM [96]. La mayoría de los scores consideran iones de tipo b e iones de tipo y junto con sus radicales ($b\text{-H}_2\text{O}$, $b\text{-NH}_3$, ...) para fragmentación CID. A más de esto, los scores generalmente consideran penalizaciones si alguno de los iones predichos no se encuentra en el espectro. Una descripción de los algoritmos utilizados en cada uno de estos programas puede encontrarse en las referencias correspondientes.

CAPÍTULO 4

METODOLOGÍA EXPERIMENTAL

En este capítulo se describen los métodos y técnicas utilizadas para la validación de los métodos. Se describe todo el proceso desde la muestra de material biológico hasta el planteamiento de los modelos computacionales para el análisis de datos.

4.1. EXTRACCIÓN DE PROTEÍNAS EN UN CULTIVO DE *ESCHERICHIA COLI*

La muestra biológica utilizada consistió en una digestión enzimática de los extractos de una población de células de la bacteria *Escherichia Coli*. Este organismo ha sido ampliamente estudiado y las bases de datos contienen la información completa sobre el genoma y proteoma de esta bacteria [97]. La digestión enzimática se realizó utilizando tripsina (Promega, Madison, WI, USA) y siguiendo el protocolo estándar descrito en el Kit de Pierce “Pierce In - Solution Tryptic Digestion Kit”¹. Parte del proceso de digestión incluye la alquilación de la Cysteina por lo que las modificaciones fijas consideradas incluyeron la modificación de tipo Carbamidomethyl Cysteina.

¹<http://www.piercenet.com/browse.cfm?fldID=973D9FA3-3C0B-452D-9633-AC1A8168D21D>

4.2. SEPARACIÓN CROMATOGRÁFICA (HPLC)

La separación cromatográfica en modo reverso (HPLC) fue realizada utilizando columnas de separación PepMap100 C18 (Dionex, Sunnyvale, CA, USA) de 15 cm de longitud y un diámetro interno de 300 μm . Los ligandos dentro de las columnas son cadenas de átomos de C_{18} con un tamaño medio de partícula de 3 μm y un tamaño de poro de 100 Å. Las columnas fueron montadas en el cromatógrafo de marca Ultimate 3000 nano LC system (Dionex).

La separación de los péptidos fue realizada en un gradiente de 10-90 % solvente B (80/20 ACN/ H_2O + 0.04 % TFA) en el solvente A (H_2O + 0.05 % TFA) durante 60 minutos, seguido por un lavado durante 5 min en 90 % solvente B y una re-equilibración de la columna en 10 % solvente B por 25 min.



Figura 4.1: Ultimate 3000 nano LC system (Dionex)

4.3. ESPECTROMETRÍA DE MASAS QqTOF

El análisis por medio de espectrometría de masas fue realizado en un espectrómetro de tipo QqTOF de última generación marca Bruker Daltonics, modelo maXis (Bruker Daltonics, Bremen, Germany). Los espectros de masas de fragmentación fueron adquiridos luego de la selección del ion precursor en el cuadrupolo y la posterior colisión inducida por disociación seguida del análisis en el espectrómetro TOF ortogonal (adquisición en modo MRM) a una velocidad de 1Hz. La selección de precursores se realizó a partir de los 3 iones de mayor intensidad observados en cada espectro (m/z 50 - 3000), dando

preferencia a aquellas con carga +2 y +3. La ventana de selección en el cuadrupolo se fijó en 1.0 m/z . Para evitar la selección de los mismos precursores se activó la opción de exclusión activa la misma que evita que los mismos precursores sean seleccionados en un periodo de 1 min después de haber sido seleccionados por primera vez.

Adicionalmente, para la validación del método de recalibración que será presentado más adelante, se realizó la adquisición de datos en un espectrómetro de tipo QqTOF de marca Bruker Daltonics, modelo micrOTOFQ (menor resolución que el espectrómetro maXis). Los parámetros de operación fueron los mismos que se utilizaron con el espectrómetro maXis.



Figura 4.2: maXis QqTOF (Bruker Daltonics)

Se adquirieron 6 conjuntos de datos como se describe en la [Tabla 4.1](#):

4.4. FRAGMENTACIÓN CID A DIFERENTES ENERGÍAS

Para el análisis de la fragmentación a diferentes energías se utilizó la muestra de *E. coli* y se cambiaron los parámetros de la energía de colisión en el instrumento maXis. Los valores de energías de colisión escogidos fueron de 15 eV, 35 eV y 55 eV en el

Tabla 4.1: Conjuntos de datos utilizados

Datos	Espectrómetro ^a	Energía de Colisión	Calibración Original
S1	maXis	Estándar ^b	Externa ^c
S2	maXis	15 eV	Externa ^d
S3	microTOFQ	Estándar ^b	Externa ^d
S4	microTOFQ	Estándar ^b	Externa ^d
S5	maXis	35 eV	Externa ^d
S6	maXis	55 eV	Externa ^d

^aBruker Daltonics, ^b 10 - 20 eV, ^c La ecuación de calibración fue intencionalmente elegida erróneamente para producir un conjunto de datos con mayores errores, ^d Calibración estándar.

sistema centro de masa. Además de cambiar los valores de la energía de colisión, se cambiaron algunos parámetros en la zona de transmisión luego de la celda de colisión para asegurar la transmisión de los iones y mejorar la señal en la zona de masas bajas.

Estas modificaciones no afectan la cinética de reacción de fragmentación ya que se aplican posterior a la celda de colisión, y solamente mejoran la calidad de los espectros obtenidos.

4.5. PROCESAMIENTO DE DATOS

El procesamiento de los datos brutos fue llevado a cabo en el programa Compass Data-Analysis 4.0 (Bruker Daltonics). Los procedimientos básicos que se realizaron fueron programados en un script conteniendo los siguientes elementos: deconvolución de la carga, deisotopización y determinación de centroides en el espectro de masa como se describió en la **Sección 3.3.1**. Adicionalmente en el script se definieron dos formatos para exportar los datos: Mascot Generic Format² (.mgf) y mzXML³ (.mzXML). El primero de ellos es el formato estándar para realizar una búsqueda de datos de tipo MS/MS en el motor de búsqueda Mascot y básicamente contiene la masa y carga del ión precursor

²http://www.matrixscience.com/help/data_file_help.html#GEN

³<http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML>

junto con una lista de los valores de m/z y sus correspondientes intensidades (opcional estado de carga) para los picos observados en el espectro de fragmentación. Por su parte, el formato mzXML es uno de los principales formatos abiertos para presentar información de proteómica y espectrometría de masas; es decir, un archivo .mzXML puede contener información de MS, MS/MS, MS^n , LC-MS/MS, etc. basada en el lenguaje XML. Los archivos .mgf y .mzXML fueron utilizados para el resto de etapas del análisis.

El análisis posterior se dividió en 3 partes:

1. Análisis por medio de base de datos
2. *De novo* sequencing
3. Fragmentación a diferentes energías de colisión

4.6. IDENTIFICACIÓN POR MEDIO DE BASE DE DATOS

Como primer paso y para poder validar los resultados se realizó un procedimiento estándar de búsqueda en base de datos para identificar los péptidos presentes en la muestra. La identificación se realizó utilizando el programa comercial MASCOT según los parámetros de la [Tabla 4.2](#). La identificación por medio de bases de datos tiene 2 objetivos dentro de este trabajo:

- Probar que el método de identificación por medio de *de novo* sequencing utilizando fragmentación a distintas energías permite mejorar la identificación acertada de péptidos.
- Validar el método de recalibración.

4.7. DE NOVO SEQUENCING

Se utilizó el programa comercial de libre distribución PepNovo+ (University of California, 2010) [98] para identificar los péptidos presentes en el set de datos S2. PepNovo+ es la última versión del programa PepNovo desarrollado en el 2005 [94]. Éste es un programa

Tabla 4.2: Parámetros de búsqueda en la base de datos

Parámetro	Valor
Base de Datos:	NCBIInr ¹
Enzima:	Tripsina
Anclajes no realizados:	1
Taxonomía:	Escherichia coli
Modificaciones Fijas:	Carbamidomethyl (C)
Modificaciones Variables:	Oxidación (M)
Tolerancia Precursor:	50 ppm ^a , 25 ppm ^b , 100 ppm ^c , 100 ppm ^d
Tolerancia Fragmentos:	100 mmu ^a , 50 mmu ^b , 100 mmu ^c , 100 mmu ^d
Estado de carga:	+2, +3
Instrumento:	ESI-QUAD-TOF

¹<http://www.ncbi.nlm.nih.gov/guide/>, ^a S1, ^b S2, ^c S3, ^d S4

El nivel de significancia estadística para las identificaciones se fijó en un 95 %.

que corre bajo línea de comandos dados los parámetros requeridos (ver [Tabla 4.3](#)). Para tener consistencia con la búsqueda en base de datos se utilizó el archivo .mgf para esta etapa, ya que este fue el mismo archivo utilizado anteriormente con Mascot.

Para medir la eficiencia del método se definieron dos medidas semejantes a las utilizadas previamente por Fischer et. al. [[96](#)].

$$Precisión = \frac{\text{número de aminoácidos correctamente identificados}}{\text{número de aminoácidos predichos}} \quad (4.1)$$

$$Eficiencia = \frac{\text{número de aminoácidos correctamente identificados}}{\text{número de aminoácidos en el conjunto de datos}} \quad (4.2)$$

Precisión se define como la relación entre los aminoácidos acertados y los aminoácidos

Tabla 4.3: Parámetros para de novo sequencing en PepNovo

Parámetro	Valor
Enzima:	Tripsina
Modificaciones Fijas:	Carbamidomethyl (C)
Modificaciones Variables:	Oxidación (M)
Tolerancia Precursor:	0.3 Da
Tolerancia Fragmentos:	0.05 Da
Número de Secuencias ^a :	5

^aNúmero de posibles secuencias (rankeadas) por espectro.

predichos por el algoritmo, mientras que *Eficiencia* se define como la relación de los aminoácidos predichos y los aminoácidos reales presentes en la muestra.

Estos índices se miden comparando sólo la secuencia con mayor score de las 5 predichas por PepNovo.

Los resultados generados por PepNovo se presentan a manera de una secuencia de aminoácidos en la que pueden haber faltantes en el terminal-N o en el terminal-C; por ejemplo, para el péptido GYRPQFYFR (tiempo de retención 32.4 min, MS/MS scan 1898), el candidato propuesto es de la forma [220.098]RPQFYFR; donde [220.098] es la diferencia en masa del terminal-N que no pudo ser identificada a partir del espectro de masas. En este caso el número de aminoácidos predicho es 7, el número de aminoácidos reales es 9, y el número de aminoácidos correctamente identificados es 7, por tanto se tendría que Precisión = 1, Eficiencia = 0.78.

4.8. RECALIBRACIÓN DE LOS ESPECTROS DE MS Y MS/MS

La recalibración de espectros de masa se realizó utilizando los péptidos identificados por medio de la búsqueda en la base de datos. Estos péptidos fueron utilizados como calibrantes internos según el proceso que se describe a continuación.

4.8.1. ELECCIÓN DE LOS CALIBRANTES INTERNOS

Calibrantes internos son aquellas especies identificadas en un espectro de masa que son utilizadas para determinar los parámetros en una ecuación de calibración adecuada. Los calibrantes internos pueden ser especies conocidas presentes en la muestra o especies añadidas externamente en la muestra. El primer enfoque es mucho más práctico pues la adición de elementos en la muestra puede afectar la señal original. Parte central de este trabajo ha sido proponer una forma de implementar una calibración usando calibrantes internos que no fueran añadidos artificialmente a la muestra. El método propuesto consiste en dos etapas. La primera etapa consiste en la identificación preliminar de péptidos mediante búsqueda en base de datos o *de novo* sequencing según la calibración original del instrumento. La segunda etapa es la recalibración de los espectros usando los péptidos identificados en la primera etapa como calibrantes internos.

El proceso de recalibración se realizó a partir de los péptidos identificados mediante la búsqueda en base de datos como se describe en la [Sección 4.6](#), usando los datos de la calibración original. Los conjuntos de datos S1, S2, S3 y S4 fueron utilizados con este propósito. La elección de los potenciales calibrantes internos se realizó de manera semejante a la descrita por Palmblad et. al. para el caso de un espectrómetro FT-ICR [7]. Una lista de candidatos fue construida a partir de los péptidos identificados por medio de MS/MS. En esta lista se incluyeron todos los picos correspondientes al ión precursor independientemente de su estado de carga (+1, +2, +3, +4), observados en un intervalo entre 15 espectros antes y 50 espectros después del espectro MS/MS en que el péptido fue identificado. La ventana de selección asimétrica indicada se justifica con la observación de que los péptidos abundantes tienden a mostrar una cola en el pico cromatográfico y pueden aparecer por largo tiempo después de haber sido detectados por primera vez [7]. Para aceptar un pico como posible calibrante se fijó una tolerancia de error de 50 ppm para S1, 5 ppm para S2, 20 ppm para S3 y 20 ppm para S4. En los casos en que dos picos fueron considerados aceptables dentro de la tolerancia de error para un mismo espectro, se eligió como potencial calibrante a aquel con mayor intensidad.

Por este procedimiento se obtuvo una lista de valores de m/z reducida, consistente de calibrantes con mayor probabilidad de haber sido correctamente identificados, reportados por el m/z medido, el m/z teórico asignado, y su intensidad (ver [Figura 4.3](#)). Esta lista fue utilizada en las etapas posteriores de la recalibración.

# scan	Péptido	Carga	m/z teórico	m/z experimental	Intensidad
426	AVESGDKKPLR	2	600.3413178	600.3385376	14064
430	AVESGDKKPLR	2	600.3413178	600.3386597	11177
442	SRVCQVTGK	2	517.7747566	517.7830566	12243
454	SRVCQVTGK	2	517.7747566	517.7582764	22386

Figura 4.3: Esquema de la lista de potenciales calibrantes en el proceso de recalibración

4.8.2. CONVERSIÓN DE LOS VALORES DE m/z A UNIDADES DE TIEMPO DE VUELO FICTICIAS

El siguiente paso fue transformar en la lista antes indicada, los valores de m/z a unidades tiempo de vuelo ficticias para poder realizar la recalibración como se describe previamente en el trabajo de Costa-Vera et.al. [8]. El objetivo en este punto es prescindir de las ecuaciones de calibración originales utilizadas por el fabricante, ya que en muchos casos estas no son accesibles al usuario. Como se indicó anteriormente, las ecuaciones putativas utilizadas para la recalibración se derivan de la ecuación de tiempo de vuelo [Ecuación 2.13](#). La evaluación de estas funciones en términos de su mejor rendimiento frente al proceso de recalibración es otra parte importante de este trabajo.

$$\text{Función TOF1: } \quad tof = a_0 \left(\frac{m}{z}\right)^{1/2} + a_1 \quad (4.3)$$

$$\text{Función TOF2: } \quad tof = a_0 \left(\frac{m}{z}\right) + a_1 \left(\frac{m}{z}\right)^{1/2} + a_2 \quad (4.4)$$

$$\text{Función TOF3: } \quad tof = a_0 \left(\frac{m}{z}\right)^{3/2} + a_1 \left(\frac{m}{z}\right)^{1/2} + a_2 \quad (4.5)$$

$$\text{Función TOF4: } \quad tof = a_0 \left(\frac{m}{z}\right)^{3/2} + a_1 \left(\frac{m}{z}\right) + a_2 \left(\frac{m}{z}\right)^{1/2} + a_3 \quad (4.6)$$

En primer término, para construir una lista de tiempos de vuelo ficticios asociados con las masas teóricas de los calibrantes identificados, se usaron estas ecuaciones con los coeficientes a_i igualados a 1. Este simple procedimiento solamente tiene la intención de transformar las unidades a tiempo para ser utilizadas internamente en el algoritmo de recalibración.

4.8.3. RECALIBRACIÓN

La recalibración a partir de las unidades de tiempo de vuelo ficticias (tof) se realizó escogiendo las funciones inversas de las ecuaciones 4.3 - 4.6. Para los primeros dos casos las funciones inversas son sencillas; en los casos restantes las soluciones analíticas de las funciones inversas fueron halladas con la ayuda de *Mathematica*® (Wolfram Research, Champaign, IL).

$$\text{Función 1: } \frac{m}{z} = a_0(\text{tof})^2 + a_1(\text{tof}) + a_2 \quad (4.7)$$

$$\text{Función 2: } \frac{m}{z} = \frac{2a_0(\text{tof}) + a_1^2 - 2a_0a_2 - a_1\sqrt{4a_0(\text{tof}) + a_1^2 - 4a_0a_2}}{2a_0^2} \quad (4.8)$$

Las funciones 3 y 4 tienen formas algebraicas más complicadas y pueden hallarse en los anexos.

La función TOF 1 corresponde al caso más simple de la ecuación de recalibración y su inversa es de la forma $m/z = (\text{tof} - a_1)^2/a_0^2$, consecuentemente, para definir completamente con esta función la relación putativa entre m/z y tiempo de vuelo, se necesita únicamente encontrar el valor de dos parámetros. Sin embargo, para mayor generalidad se escogió la **Ecuación 4.7**. Para el resto de casos se utilizó simplemente la función inversa.

El número mínimo de calibrantes internos por espectro fue definido como el número de parámetros en la ecuación de recalibración más 1 para lograr un ajuste que sea estadísticamente significativo. En algún caso, por ejemplo, se usaron 6 parámetros para calibrar una ecuación con cuatro parámetros desconocidos. En la mayoría de espectros el número de calibrantes disponibles es mucho mayor que el número de parámetros de la ecuación de recalibración. En esos casos realizamos múltiples recalibraciones con diferentes conjuntos de calibrantes para probar la robustez del método.

Para cada espectro se realizó la determinación de los parámetros a_i por medio del método de mínimos cuadrados y una vez obtenida la ecuación, se la utilizó para volver a unidades de m/z . A continuación se determinó el error respecto a las masas teóricas de los péptidos identificados. En este punto se decidió utilizar un nuevo filtro para la selección de calibrantes, pues podría ocurrir que algunos de los péptidos identificados preliminarmente mediante la búsqueda en base de datos no hayan sido correctamente asignados. Si utilizamos estos péptidos como calibrantes internos podrían haber

mayores errores luego de la recalibración.

El filtro planteado consiste en un nivel de tolerancia máximo para el error residual luego de recalibración. Este valor se eligió de manera que se obtenga un bajo error luego de la recalibración sin reducir considerablemente el número de calibrantes aceptados. La tolerancia luego de recalibración fue fijado en 3 ppm (S1), 1.5 ppm (S2), 10 ppm (S3, S4) para cada uno de los sets de datos considerados.

El filtro descrito funciona de tal manera que en caso de que un calibrante potencial, admitido previamente, tenga un error superior al mencionado anteriormente, el calibrante es desechado y la ecuación de recalibración es obtenida nuevamente utilizando los calibrantes restantes. El proceso se repite hasta que todos los calibrantes cumplan la condición o hasta que el número de calibrantes en el espectro sea menor que el mínimo requerido, en cuyo caso el espectro es rechazado pues no es posible su recalibración.

Todo el proceso de recalibración fue programado en una rutina escrita en *Mathematica*. En el mismo programa, los parámetros del ajuste fueron obtenidos minimizando el error cuadrático medio con el método de Levenberg-Marquardt [99] disponible en el programa.

4.9. PROGRAMACIÓN DE RUTINAS

4.9.1. RECALIBRACIÓN

El proceso de recalibración fue realizado por medio de rutinas escritas en *Mathematica* y en C++. Los calibrantes fueron seleccionados mediante una rutina escrita en C++ (Microsoft Visual C++ 2010 Express) siguiendo los pasos descritos previamente. Los archivos .mzxml fueron leídos utilizando la librería mzxml.lib⁴ y los archivos de texto conteniendo las identificaciones obtenidas de la búsqueda en base de datos fueron leídos en la implementación en C++. Los resultados (potenciales calibrantes) fueron exportados en un archivo de texto conteniendo las columnas que pueden verse en la **Figura 4.3**.

El de texto fue importado en *Mathematica* y en base a éste se realizó la recalibración, i.e., la determinación de los parámetros a_i y el análisis de los resultados.

⁴en desarrollo, Bruin, J., Leiden University Medical Center

4.9.2. ANÁLISIS DE LOS ESPECTROS DE FRAGMENTACIÓN A DIFERENTES ENERGÍAS DE COLISIÓN

Con el objetivo de mejorar la capacidad de identificación de péptidos mediante *de novo* sequencing, se ha explorado el uso de diferentes energías (regímenes) de fragmentación. Estos resultados se describen a continuación.

Se realizó una búsqueda en base de datos de los resultados del conjunto de datos S2 con el objetivo de tener una lista de péptidos identificados para comparar con las identificaciones obtenidas por medio de *de novo* sequencing. El objetivo es utilizar la información adicional contenida en los conjuntos de datos S5 y S6 para aumentar el número de identificaciones obtenidas originalmente por PepNovo en S2. Para esto se realizó un procedimiento de re-scoring de los 5 péptidos predichos por PepNovo según el esquema discutido más adelante.

Del total de péptidos identificados en S2 mediante base de datos se tomó aleatoriamente un 20 % de los mismos como un conjunto de aprendizaje (cuya funcionalidad se explica en la [Sección 4.9.2.1](#)) y las pruebas se realizaron con el restante 80 % para validar el método. El total de péptidos identificados en S2 mediante energías de colisión estándar (15 eV) fue de 545 (316 con carga 2+, 216 con carga 3+, 12 con carga 4+ y 1 con carga 5+). De estos 545 péptidos 487 son péptidos únicos.

El análisis de los espectros de fragmentación a diferentes energías de colisión se realizó mediante una rutina escrita en C++. Los resultados obtenidos del proceso de *de novo* sequencing fueron exportados a un archivo de texto para cada identificación exitosa, esto es, para cada secuencia putativa generada por PepNovo. El formato del archivo de texto se muestra en la [Figura 4.4](#). La selección de la secuencia más probable entre éstas por medio de un procedimiento de re-scoring es precisamente el objetivo del análisis de fragmentación con diferentes energías.

Los valores presentados en el archivo de texto corresponden a las identificaciones realizadas utilizando la energía de colisión de 15 eV y por tanto el número de scan y el tiempo de retención no tienen porque ser exactamente los mismos que aquellos observados en los casos de energías de 35 eV y 55 eV. Este problema es lo que se conoce como el problema de alineamiento en LC-MS y consiste en encontrar los espectros correspondientes en dos o más conjuntos de datos adquiridos para la misma muestra. Para lograr este alineamiento de los datos de LC-MS se han desarrollado métodos y al-

# scan	Carga	m/z precursor	Tiempo de retención
1245	2	388.71156	21.3
YGNFFK	}	Secuencias predichas por PepNovo	
HTPFFK			
YGNFMK			
YFFGNK			
YFFNGK			

Figura 4.4: Esquema de la lista de péptidos identificados por PepNovo

goritmos que toman en cuenta las fluctuaciones esperadas en los tiempos de retención y en la adquisición de los espectros de MS, tanto en caso de LC [100], como en el caso de CE [101], entre otros.

En este trabajo decidimos no centrarnos en este aspecto y utilizamos un esquema sencillo para correlacionar los péptidos observados en 15, 35 y 55 eV. En principio, puede asumirse que los péptidos seleccionados para fragmentación se encuentran en tiempos de retención similares en los 3 sets de datos. La identificación del espectro MS/MS correspondiente al mismo péptido se basa entonces en una comparación de la masa del precursor dentro de un rango de 1.5 min antes y después del tiempo de retención determinado en el set de datos de 15 eV. El valor de 1.5 min fue escogido tomando en cuenta la opción de exclusión activa (1 min) y dejando un margen de 0.5 min por posibles diferencias en los tiempos de retención. La tolerancia en masa para la identificación del ión precursor se fijó en 0.1 Da. Además, para evitar seleccionar picos de ruido se tomaron solo los 50 picos más intensos del espectro para realizar la identificación.

4.9.2.1. Conjunto de Aprendizaje

El conjunto de aprendizaje fue utilizado para determinar el tipo de iones a considerar en la validación por medio de las energías de colisión de 35 y 55 eV. Se escogieron 109 péptidos identificados por medio de bases de datos con cargas 2+ (55 péptidos) y 3+ (54 péptidos). Se dejaron de lado los casos de cargas 4+ y 5+ pues son mucho menos frecuentes. Esta separación en carga tiene como objetivo diferenciar la fragmentación para diferentes estados de carga.

Básicamente, el procedimiento consiste en generar el conjunto de picos teóricos correspondientes a los fragmentos tipo a, tipo c, tipo x, tipo z y iones de immonio para la

secuencia de aminoácidos identificada. Debido a que PepNovo ya utiliza los fragmentos tipo *b* y tipo *y* para la identificación primaria, este tipo de iones se excluyó del proceso de validación para enfocar el método en aquellos fragmentos que no fueron considerados en la primera etapa con PepNovo. Esta lista de *m/z* artificialmente construida se comparó con el espectro de masas real (archivo .mzxml) y se determinó las coincidencias usando una tolerancia de 0.05 Da. S5 fue utilizado para identificar los fragmentos generados por la ruptura de la columna peptídica, mientras que S6 se utilizó para buscar iones immonio.

4.9.2.2. Espectros de fragmentación con 35 eV

Los espectros de fragmentación con 35 eV fueron utilizados para verificar la presencia de iones tipo *a*, *c*, *x*, y *z*. La lista de fragmentos para cada una de las 5 secuencias más probables predichas por PepNovo se comparó de la manera indicada con el espectro utilizando una tolerancia de 0.05 Da. El sistema de re-scoring que se propone en este trabajo considera un score positivo por cada pico presente en el espectro y un score negativo si el pico no se encuentra en el espectro. Tanto el score positivo como el score negativo dependen del tipo de fragmento considerado. Para cada tipo de fragmento se definen los factores $0 \leq \alpha_t^{ch}, \beta_t^{ch} \leq 1$ donde $t = a, c, x, z$ según el tipo de fragmento correspondiente, y el superíndice *ch* corresponde al estado de carga del precursor. El score para cada fragmento *i* es entonces:

$$\hat{S}_{c_i} = \alpha_t^{ch} I_i - \beta_t^{ch} \frac{I_M}{2} \quad (4.9)$$

Donde I_M corresponde a la intensidad del ión más intenso en el espectro. El score final para el set de datos de 35 eV es la suma del score de los picos identificados más el score negativo de los picos no identificados dividido para $\frac{I_M}{2}$. Es decir, si P es el conjunto de picos tipo *a*, *c*, *z* predichos y E el conjunto de picos efectivamente identificados en el espectro se tiene que:

$$S_{c_1} = \frac{\sum_{t \in (a,c,z)} \sum_{i \in E} \alpha_t^{ch} I_i}{|E| (I_M/2)} - \sum_{t \in (a,c,z)} \sum_{P-E} \beta_t^{ch} \quad (4.10)$$

El proceso se realiza para los 5 péptidos candidatos identificados por PepNovo.

4.9.2.3. Espectros de fragmentación con 55 eV

Para los espectros de fragmentación con 55 eV se realizó un procedimiento semejante al realizado con 35 eV, con la diferencia de que estos espectros se utilizaron para verificar la aparición de iones immonio. La observación de iones immonio es un tema de gran relevancia y en general es pasado por alto. Uno de los trabajos más importantes en el área corresponde al realizado por Falick et. al. [102] al proponer un modelo para determinar la presencia de un aminoácido basados en los iones immonio observados en los espectros con altas energías de colisión. Más recientemente, Hohmann et. al. [103] han estudiado la relación entre la presencia de aminoácidos y la intensidad de los iones immonio observados mediante un espectrómetro QqTOF. Más estudios en esta área son necesarios para incorporar la información de los iones immonio en programas de *de novo* sequencing o en programas de identificación por medio de base de datos. En este trabajo se utilizaron los iones de immonio para re-evaluar las identificaciones de PepNovo al incorporarlos en el esquema de re-scoring por medio del score S_{c_2} que se define como:

$$S_{c_2} = \frac{\text{número de iones immonio observados}}{\text{número de iones immonio teóricos}} \quad (4.11)$$

La [Tabla 4.4](#) resume los iones immonio utilizados en el procedimiento de re-scoring.

El score total S_c debe ser una combinación de S_{c_1} y S_{c_2} de manera que:

$$S_c = \lambda_1 S_{c_1} + \lambda_2 S_{c_2} \quad (4.12)$$

Las constantes λ_1 y λ_2 fueron determinadas utilizando los resultados del conjunto de aprendizaje. Los valores de λ_1 y λ_2 se variaron entre 0.04 y 1 con pasos de 0.04 y para cada caso se determinó el score para las 190 secuencias de prueba. El valor óptimo de (λ_1, λ_2) se eligió de manera que se maximice la suma de los scores individuales.

Tabla 4.4: Iones immonio utilizados en el procedimiento de re-scoring [102]

Aminoácido	Masa Nominal de los Iones Immonio
Ala	44
Arg	129
Asp	88
Asn	87
Cys	76
Gly	30
Gln	101
Glu	102
His	110
Ile/Leu	86
Lys	101
Met	104
Phe	120
Pro	70
Ser	60
Thr	74
Trp	159
Tyr	136
Val	72

CAPÍTULO 5

RESULTADOS Y DISCUSIÓN

5.1. RECALIBRACIÓN DE LOS ESPECTROS DE MS Y MS/MS

El procedimiento descrito en la [Sección 4.8](#) fue utilizado para probar el método de recalibración en los conjuntos de datos S1, S2, S3 y S4. Para cada uno de ellos, se determinó el error original en los picos identificados en el espectro como la diferencia relativa entre el valor m/z medido y el valor m/z teórico determinado a partir de la secuencia de aminoácidos obtenida de la identificación en base de datos. Cada set de datos se caracterizó por el error promedio y la dispersión de valores de error correspondientes, por medio de la desviación estándar del error globales (ver [Tabla 5.1](#)). Nótese que el mejor caso concebible corresponde a un error y a una dispersión nulos. El set de datos proveniente del espectrómetro maXis con calibración estándar (S2) presenta un bajo error medio y una desviación estándar del error menor a 2 ppm, lo cual ya de por sí son excelentes resultados. La [Figura 5.1](#) muestra un histograma que caracteriza la distribución de los errores medidos para el conjunto de datos S1.

Para validar la robustez de los resultados de la recalibración, se eligieron dos miembros de la lista de posibles calibrantes por espectro, pero que no habían sido usados para determinar los parámetros de las ecuaciones de recalibración. Los valores de m/z fueron determinados tanto para los calibrantes incluidos como para aquellos que fueron excluidos y los resultados se muestran en la [Tabla 5.2](#). Como puede verse, la Función 3 obtuvo en general los mejores resultados en cuanto a la disminución de la dispersión del

Tabla 5.1: Errores antes de la recalibración

	Error Medio (ppm)	Desviación Estándar (ppm)
S1	12.531	24.28
S2	-7.395	1.29
S3	8.440	16.27
S4	12.192	11.93

error. Sin embargo, la Función 1 más simple obtuvo resultados casi igual de buenos y aún superiores en el caso del conjunto de datos S2. Como podía esperarse, esto muestra que la función cuadrática simple es bastante buena para calibrar espectros de masa obtenidos por medio de un oa-TOF [104].

Además, se observa que la adición de muchos términos puede ser perjudicial como se observa para los resultados de la Función 4. Debido a estos resultados, y tomando en cuenta la complejidad computacional proveniente de utilizar funciones de recalibración muy complicadas, el resto del análisis se realizó utilizando la Función 1. En ese caso, la disminución marginal en la reducción del error es compensada por un menor esfuerzo computacional. La diferencia entre los resultados para S1 y S2 indican que la calibración inicial tiene efecto en el proceso de recalibración pero que independientemente de la misma, el procedimiento descrito disminuye los errores aún para espectros originalmente mal calibrados.

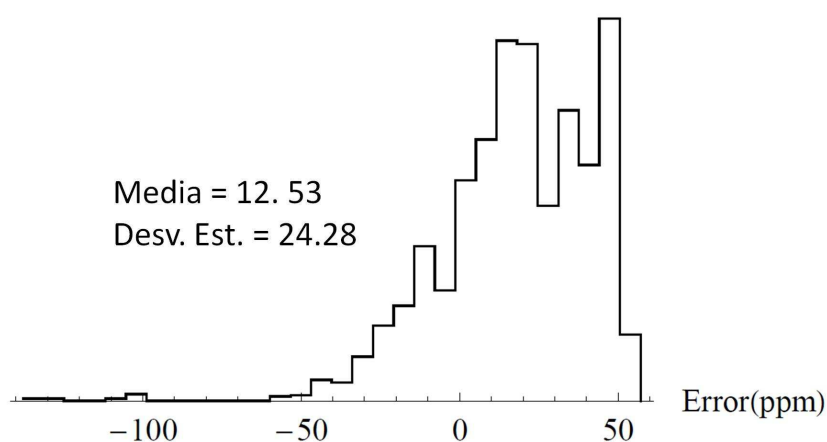


Figura 5.1: Distribución de los errores en la determinación de la masa para los calibrantes antes de la recalibración (conjunto de datos S1).

Otro aspecto que debe tomarse en cuenta es que el proceso de recalibración utiliza péptidos identificados en la muestra y los espectros que pueden ser recalibrados son aquellos con el número mínimo de calibrantes necesarios según la ecuación de recalibración utilizada. Es así que mientras menos parámetros sean necesarios en la ecuación de recalibración, mayor será el número de espectros efectivamente recalibrados.

Tabla 5.2: Media \pm Desviación Estándar (ppm) del error global para los datos, antes y después de la recalibración con las funciones utilizadas

	S1	S2	S3	S4
Original	12.531 \pm 24.28	-7.395 \pm 1.29	8.440 \pm 16.27	12.192 \pm 11.93
Función 1	0.005 \pm 1.23	0.004 \pm 0.65	-0.044 \pm 3.89	-0.035 \pm 3.83
Función 2	-0.015 \pm 1.24	-0.014 \pm 0.66	0.001 \pm 3.89	0.037 \pm 3.89
Función 3	-0.002 \pm 1.22	-0.005 \pm 0.66	-0.038 \pm 3.89	-0.031 \pm 3.80
Función 4	-0.063 \pm 2.88	0.894 \pm 2.70	-0.377 \pm 8.99	-0.526 \pm 10.16
Validación	-0.071 \pm 1.80	-0.124 \pm 0.95	-0.214 \pm 5.63	-0.270 \pm 6.41
%Scans recalibrados ^a	355/446 (79.6 %)	166/207 (80.2 %)	268/335 (80 %)	221/284 (77.8 %)

^a % Scans recalibrados, indica el número de scans recalibrados relativo al número Scans con al menos 6 calibrantes, en porcentaje.

La [Figura 5.2](#) muestra un histograma de los errores para los calibrantes incluidos y excluidos en la determinación de los parámetros a_i en la Función 1 para el set de datos S1. El total de calibrantes excluidos representa un tercio del total de calibrantes incluidos por lo que el histograma en la [Figura 5.2](#) se encuentra escalado para mostrar ambos conjuntos de datos en el mismo gráfico. Como podía esperarse, los calibrantes excluidos muestran una dispersión del error superior (1.80 ppm) a aquella de los calibrantes incluidos (0.96 ppm). El error cuando todos los calibrantes son puestos juntos se encuentra en el intermedio de estos dos valores. La [Tabla 5.2](#) muestra también los resultados de la recalibración para los calibrantes excluidos (Validación).

Un aspecto importante fue notar que el proceso produjo buenos resultados solo cuando los calibrantes cubren todo el rango de masas recalibrado; es decir, que cuando uno de los 2 calibrantes excluidos se encontraba fuera del rango cubierto por los calibrantes incluidos, el error para el calibrante excluido luego de la recalibración era más alto que

el original. En este sentido, el método es aplicable para el rango de masas cubierto por los calibrantes escogidos. O inversamente, se podría definir que el método de recalibración escoja calibrantes capaces de encerrar todo el rango de masas a calibrar, eventualmente.

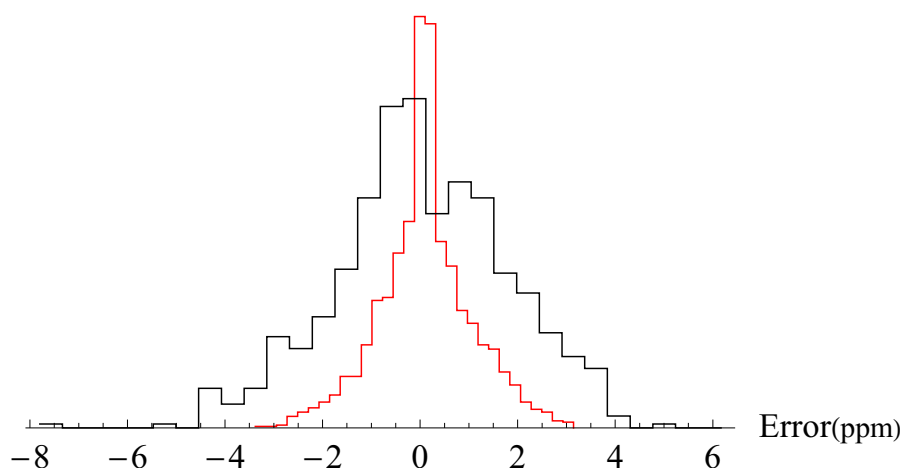


Figura 5.2: Histograma de los errores (luego de recalibrar) para los picos incluidos (rojo) y excluidos (negro) en el proceso de recalibración. Los datos corresponden a S1.

Para S1, el total de espectros recalibrados alcanza el 41 % de los espectros MS en la corrida, lo que equivale a un 51 % de los espectros en que al menos un péptido fue identificado. Como se esperaba, los primeros y últimos espectros en la corrida no fueron recalibrados porque no hubieron péptidos identificados en dichas regiones. El primer espectro en que al menos un calibrante fue encontrado es el scan número 472, y el último fue el scan número 3849; así que de un total de 701 scans (uno de cada 4 scans es un espectro MS, los otros tres corresponden a espectros MS/MS), 355 fueron exitosamente recalibrados. Los porcentajes para el resto de sets de datos pueden verse en la [Tabla 5.2](#).

Para evaluar más allá el potencial del método, se decidió investigar el mejoramiento en la precisión en la determinación de la masa de los péptidos que inicialmente fueron rechazados y dejados fuera del proceso de recalibración. Este test más exigente permite evaluar la bondad del método para valores que estuvieron inicialmente mal condicionados. Así, la ecuación de recalibración fue probada en el conjunto de potenciales calibrantes que fueron rechazados por el filtro de tolerancia en el error interno. Los resultados para los picos rechazados en S1 produjeron un error promedio de -2.63 ppm con una desviación estándar de 24.88 ppm, lo cual comparado con el error original para este subconjunto (1.83 ± 33.60 ppm) muestra una mejora en la dispersión del error que

es mucho menor que la mejora observada para aquellos picos que pasaron el filtro de tolerancia en el error interno.

Para un procedimiento de recalibración como el descrito aquí, es importante asegurarse de que no se introduce un error sistemático debido al rechazo de un número grande de picos de baja intensidad en el espectro de masa. Si esto ocurriera habría el peligro de sesgar de alguna manera la interpretación de los resultados hacia los picos de alta intensidad. La **Figura 5.3** muestra la distribución de los errores originales y de los errores después de la recalibración como una función de la intensidad de los picos. Como referencia debe mencionarse que en terminos de intensidad, se fijó un límite mínimo de intensidad de 10 000 cuentas; además, los picos de mayor intensidad observados alcanzan alrededor de 780 000 cuentas (relación aproximada de 80:1). De la figura se ve claramente que los errores más altos en la determinación de la masa se presentan naturalmente, antes y después de la recalibración, para los picos de menor intensidad. Sin embargo, como puede verse en la distribución de errores luego de la recalibración, la distribución de errores originalmente no-simétrica se vuelve simétrica alrededor del valor central de 0 ppm. Adicionalmente, la magnitud del error para los iones de baja intensidad se reduce en la misma medida que para los iones de alta intensidad. Esto indica que el procedimiento de recalibración no introduce un error sistemático adicional en contra de los iones de baja intensidad, sino que más bien, disminuye la extensión del error, y que por lo tanto puede ser utilizado para recalibrar todos los picos en el espectro con la misma eficiencia.

5.1.1. RELACIÓN ENTRE LOS PARÁMETROS DE LA ECUACIÓN DE RECALIBRACIÓN

Una inquietud natural que aparece cuando se realiza un procedimiento como el realizado para la recalibración es si los diferentes parámetros calculados en los ajustes son totalmente independientes entre sí. Esta no es únicamente una inquietud de orden matemático, sino físico, ya que en los parámetros de la ecuación de tiempo de vuelo se establecen relaciones varias entre las dimensiones del instrumento, los voltajes usados, etc.

Por esta razón, los parámetros a_i de la Función 1 fueron investigados y la **Figura 5.4** muestra los valores de a_0 , a_1 y a_2 para los espectros efectivamente recalibrados del conjunto de datos S1. Se puede ver que los valores no se alejan de los valores de 1,

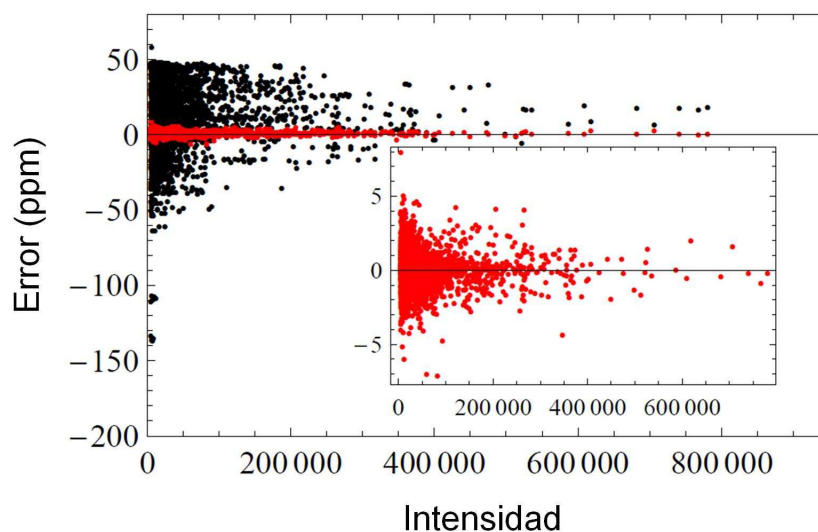


Figura 5.3: Distribución de los errores en función de la intensidad para los datos originales (negro) y para los datos recalibrados (rojo). El sub-gráfico presenta un acercamiento a la distribución de errores luego de la recalibración. El error fue disminuido consistentemente en todo el dominio de intensidades.

0 y 0 para a_0 , a_1 , y a_2 , respectivamente; pero las pequeñas diferencias son claramente responsables del éxito de la recalibración. Una mirada rápida a la [Figura 5.5](#) sugiere una relación lineal entre los parámetros, por lo que se realizó una regresión lineal para verificar esta tendencia. Los valores de R^2 obtenidos de la regresión lineal (>0.99) confirmaron esta suposición.

Esta relación lineal implica que ambos coeficientes a_0 y a_1 pueden ser expresados en términos de a_2 , el cual es el único parámetro independiente en la ecuación de recalibración. Por tanto, el número de parámetros que deben ser calculados en la Función 1 pasa a ser sólo 1. Este hecho hace posible expresar la Función 1 como una nueva función cuadrática dependiente de un sólo parámetro a_2 y por tanto el número de espectros que pueden ser recalibrados se incrementa ya que el número mínimo de calibrantes requerido disminuye de 4 a tan sólo 2. Desde el punto de vista físico esto sería razonable si la recalibración solamente está contribuyendo a eliminar la influencia de pequeños efectos, tales como la dispersión de velocidades iniciales transversales de los iones en la parte de la entrada del oa-TOF.

La nueva ecuación fue utilizada para recalibrar los espectros que anteriormente fueron rechazados por no tener el mínimo número de calibrantes requeridos (6 calibrantes = 4 incluidos + 2 excluidos). En este caso, la nueva condición es tener al menos 4 calibrantes (2 incluidos + 2 excluidos) por espectro. Esta modificación del conjunto de scans

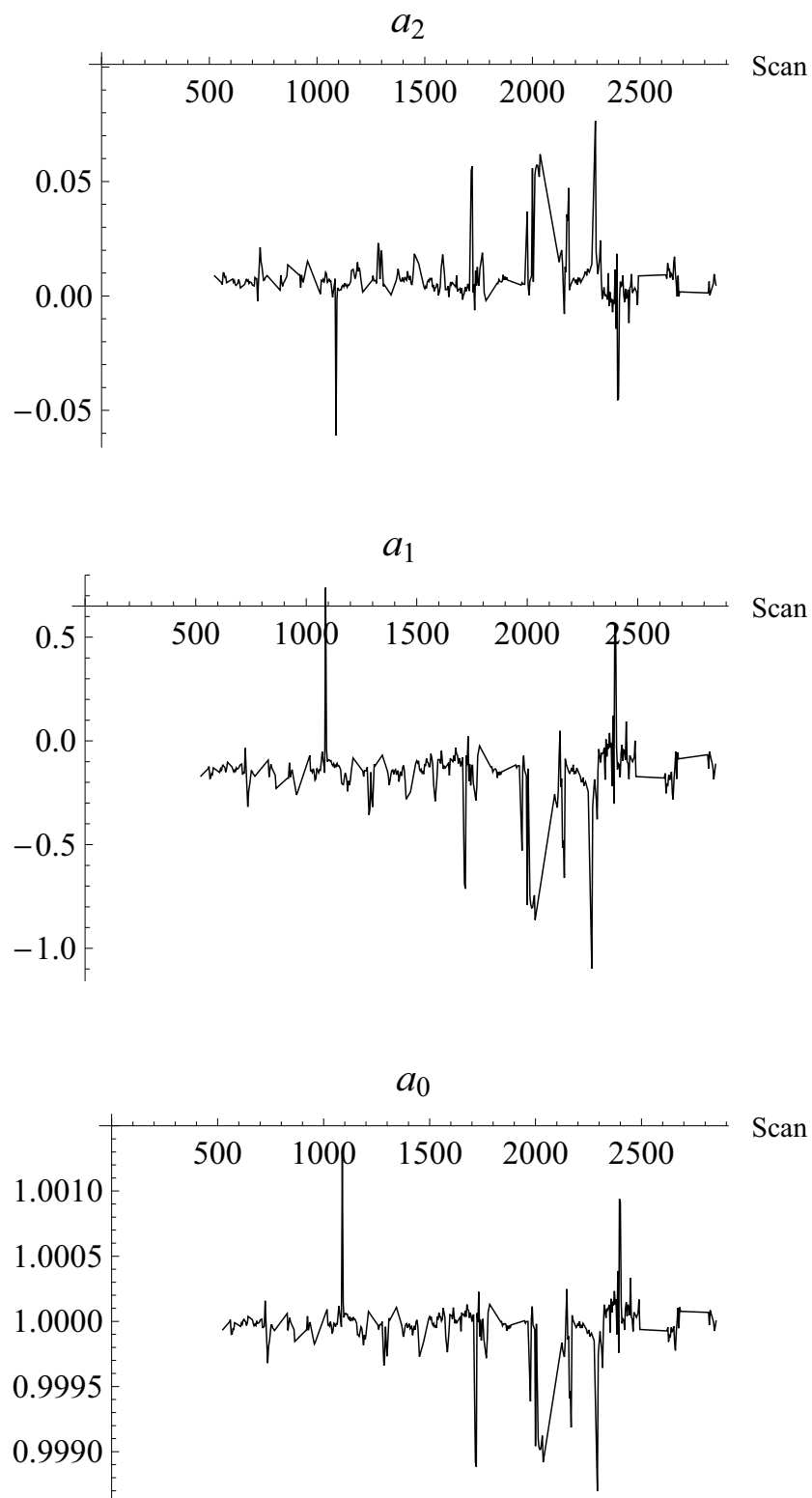


Figura 5.4: Gráfico de los parámetros a_i de la Función 1 vs. Número de Scan. Set de datos S1.

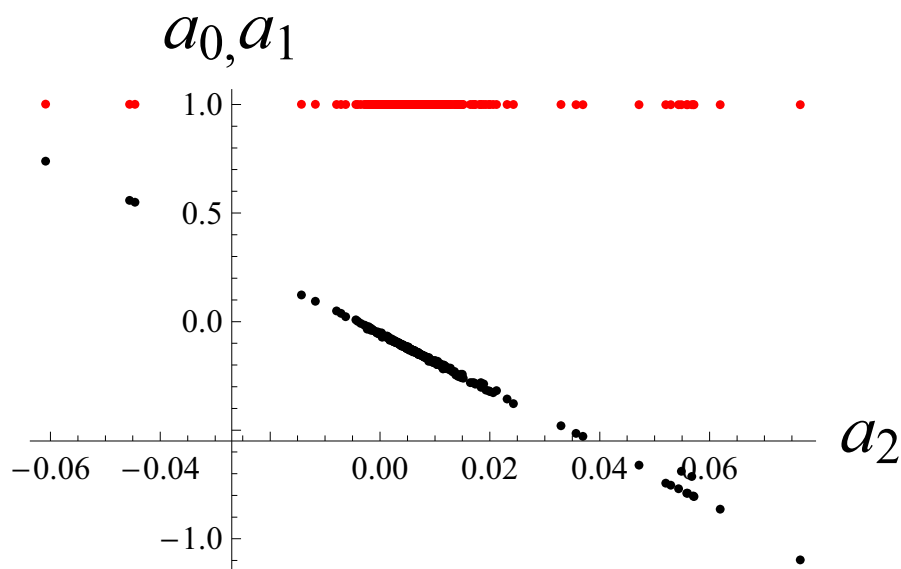


Figura 5.5: Relación entre los parámetros a_i de la Función 1 para el set de datos S1. El eje x representa al parámetro a_2 , mientras que en el eje y se presentan a_0 (rojo) y a_1 (negro). La figura sugiere claramente una relación lineal entre los parámetros.

calibrables eleva su número sustancialmente, de 446 a 582 (30 % de incremento) para S1. De igual manera, para este conjunto aumentado, el número de scans efectivamente recalibrados también se incrementa de 355 a 461. Para el caso de S2, S3 y S4, los incrementos en el número de scans considerados es todavía mayor llegando al 85%, 39%, 49% respectivamente. Así mismo, el porcentaje de scans efectivamente recalibrados se incrementó para todos los sets de datos excepto para S1 (ver [Tabla 5.3](#)), lo cual puede ser atribuido a la mala calibración original de este set de datos. En la [Tabla 5.3](#) se muestran los resultados para los nuevos picos recalibrados (Recalibración Reducida) y el resultado global de la recalibración (Recalibración Global) que incluye los picos calibrados por la nueva Función 1 con un sólo parámetro, y aquellos picos que fueron recalibrados anteriormente utilizando la Función 1 original. Aún cuando los errores utilizando la Función 1 con un sólo parámetro son mayores que los errores utilizando la Función 1 con tres parámetros, la disminución respecto al error original se mantiene y el número total de espectros efectivamente recalibrados es mucho mayor al utilizar la ecuación con un solo parámetro.

Tabla 5.3: Errores y estadísticas de la recalibración considerando la relación entre los parámetros a_i en la Función 1.

	S1	S2	S3	S4
Original	13.305±23.61	-7.423±1.28	8.843±16.24	12.842±11.43
RR ^a	-0.017±1.83	-0.066±0.85	-0.035±5.61	0.223±5.31
RG ^b	0.008±1.33	-0.026±0.74	-0.042±4.39	0.046±4.34
%Scans	431/582	336/383	411/464	386/424
Recalibrados	(74 %)	(87.7 %)	(88.6 %)	(91 %)
Incremento en % Scans Recalibrados	-6 %	8 %	9 %	13 %
Incremento en el Número de Scans Considerados	30 %	85 %	39 %	49 %

^aRR (Recalibración Reducida) corresponde a la calibración de los scans que previamente no fueron recalibrados (i.e., scans con un número de calibrantes internos menor a 6) y que ahora son recalibrados utilizando la relación interna entre los parámetros a_i . ^bRG (Recalibración Global) incluye tanto los scans previamente recalibrados como aquellos recalibrados en Calibración Reducida.

La condición de que en cada proceso de recalibración se exija tener dos péptidos identificados y potencialmente utilizables como calibrantes, pero que se excluyen a propósito del proceso de ajuste, se debe a la necesidad de tener en cada caso picos que se puedan usar para evaluar la mejora eventual en la precisión de la determinación de la masa del espectro en general. Estos dos péptidos corresponden a testigos en ese sentido. En condiciones normales, cuando no se requieren estos testigos, el número de espectros efectivamente calibrables se incrementaría correspondientemente, pues se deben tomar en cuenta espectros con al menos 2 calibrantes en lugar de 4. De esta manera, en S1 se pueden considerar 671 en lugar de 582 con al menos 4 calibrantes, lo cual equivale a un 96 % del número de espectros con al menos un péptido identificado por medio de la base de datos y un 78 % del total de scans MS en la corrida, lo que implica una cobertura potencial muy amplia del conjunto total de escanes registrada.

En base a lo discutido anteriormente, el procedimiento de recalibración puede resumirse en los siguientes pasos (ver [Figura 5.6](#)):

1. Transformar los valores m/z originales (según la calibración del instrumento) a unidades de tiempo de vuelo utilizando la ecuación TOF1.
2. Escoger aquellos espectros con un número de calibrantes mayor o igual a 4.
3. Recalibrar los espectros que pasen las condiciones previamente descritas utilizando la Función 1 (ver [Figura 5.6](#)).
4. A partir de las recalibraciones exitosas, hallar la relación lineal entre los parámetros a_0 , a_1 y a_2 y reescribir la Función 1 en términos de a_2 .
5. Utilizar la ecuación derivada en el paso 4 para recalibrar los espectros que no fueron considerados en el paso 2 con al menos 2 calibrantes y aquellos que fueron rechazados en el paso 3.
6. Repetir el procedimiento de identificación de péptidos.

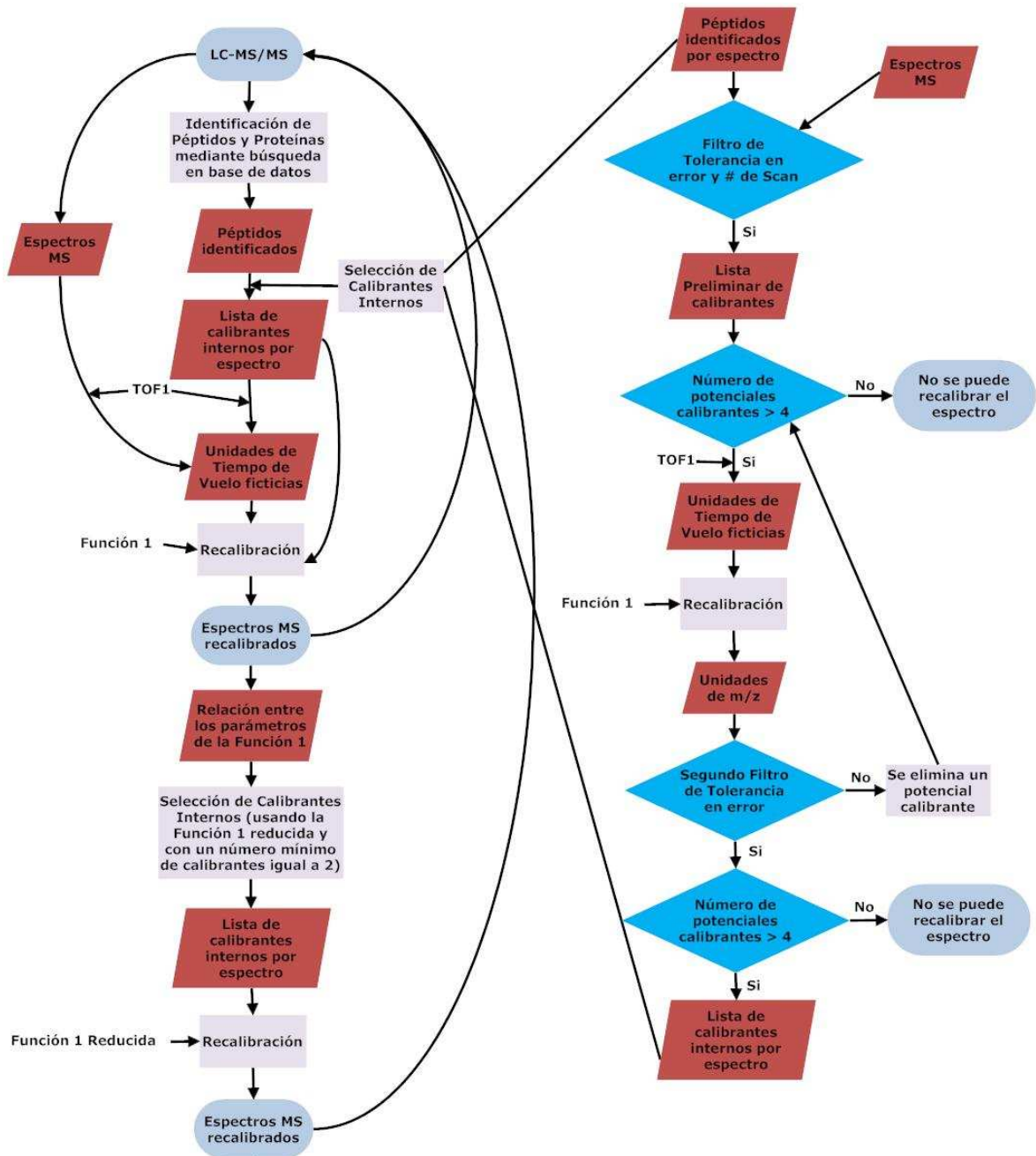


Figura 5.6: Diagrama de flujo del proceso de recalibración

5.2. INVESTIGACIÓN DE LOS ESPECTROS DE FRAGMENTACIÓN CID A DIFERENTES ENERGÍAS DE COLISIÓN

Para incrementar aun más la tasa de identificación de péptidos con la metodología usada, es necesario obtener información adicional independiente. Para esto, se propone cambiar las condiciones de fragmentación CID de los péptidos, en particular cambiando la energía de colisión. Con esto es de esperarse que los patrones de fragmentación reflejen las diferentes condiciones y consecuentemente generen información adicional. Esta información puede provenir de la estimulación de otras rutas de fragmentación, esto es, fragmentos diferentes a los anteriormente generados. Los espectros de fragmentación con diferentes energías de colisión fueron utilizados para mejorar la tasa de identificaciones por medio de *de novo* sequencing.

Los resultados obtenidos con PepNovo para bajas energías se presentan en la [Tabla 5.4](#). Estos resultados no distinguen entre los aminoácidos *I* y *L*, ni entre los aminoácidos *Q* y *K*. Tanto el valor de la *Precisión* como de la *Eficiencia*, como se definieron en la [Sección 4.7](#), se encuentran dentro del rango demostrado en estudios previos [96, 105, 106]. La longitud promedio de los péptidos es de 13.10 mientras que la predicción de PepNovo es de 10.89 aminoácidos por péptido. La similitud entre el parámetro predicho y el real, puede tomarse como una figura de mérito para el programa PepNovo. En nuestro caso, el resultado es bastante bueno.

Tabla 5.4: Resultados de *de novo* sequencing

Precisión	72.6 %
Eficiencia	63.7 %
Longitud Promedio	10.89

El uso de las mediciones obtenidas con mayores valores de energía de colisión, sirven en este caso para mejorar estos resultados. Esto es así, porque la información adicional obtenida con estas mediciones permite escoger con mejor criterio dentro de la lista de péptidos mejor ranqueados por PepNovo, como explicamos a continuación.

La mayoría de programas comerciales para realizar *de novo* sequencing arrojan como

resultado un número dado de potenciales candidatos según el requerimiento del usuario. Esta lista de potenciales candidatos se encuentra rankeada según algún mecanismo de asignación de scores; sin embargo, esto no asegura que la primera secuencia de la lista sea efectivamente la correcta. En este sentido, un re-scoring utilizando un número mayor de fragmentos permite determinar con mayor certeza el péptido efectivamente presente en la muestra siempre y cuando haya sido identificado inicialmente por el programa de *de novo* sequencing.

En este trabajo utilizamos dos niveles distintos de energías para definir un score que considere iones fragmento de la columna peptídica e iones immonio según el diagrama de la [Figura 5.7](#). Los métodos de score fueron descritos en la [Sección 4.9.2](#). A continuación se muestran los resultados del conjunto de datos de aprendizaje utilizado para determinar los parámetros usados en los esquemas de scoring.

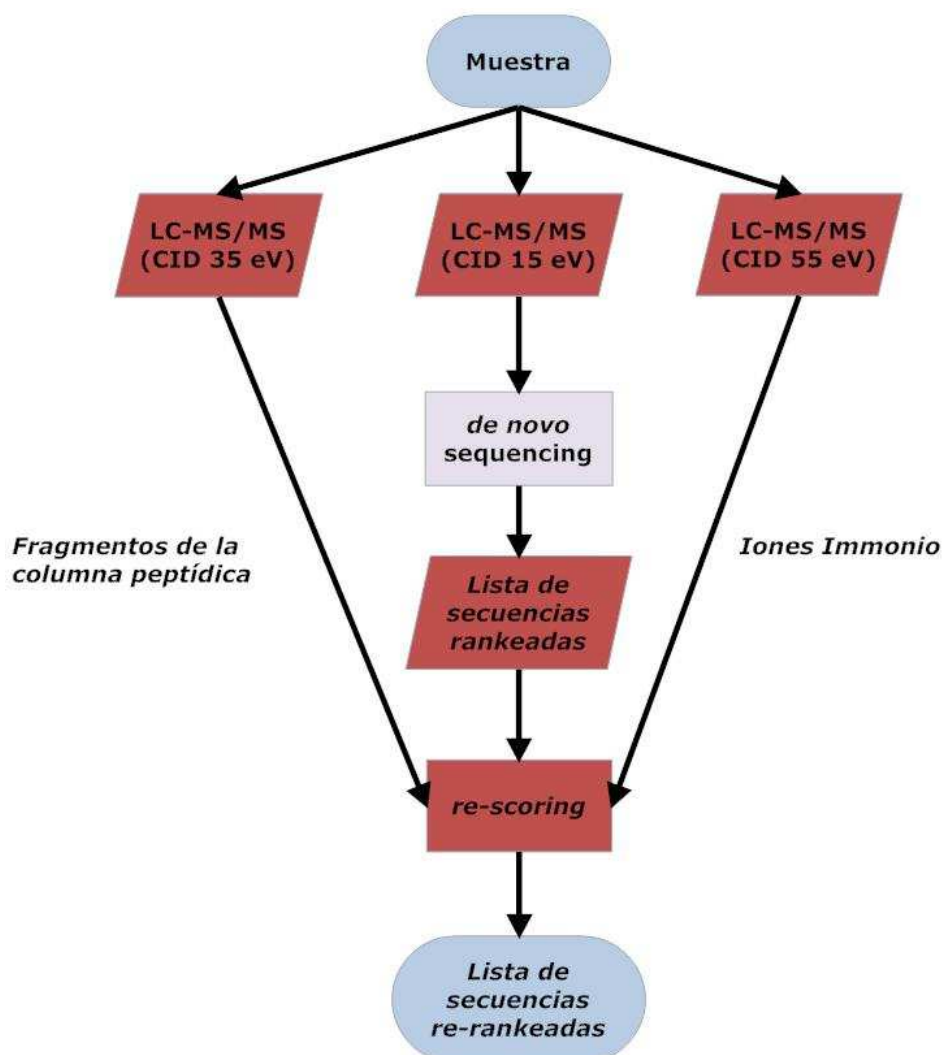


Figura 5.7: Diagrama de flujo del proceso de re-scoring

5.2.1. RESULTADO DEL CONJUNTO DE DATOS DE APRENDIZAJE

Bajo las condiciones normales de adquisición, el total de péptidos identificados en el conjunto de datos S2 fue de 545. Para las mediciones con diferentes energías de colisión, de este total se tomaron 109 (20 %) scans para el conjunto de aprendizaje. Para cada uno de ellos se identificó el correspondiente espectro en S5 y en S6 (se “alinearon” los espectros respectivos). En S5 se realizó una búsqueda de fragmentos de la columna el péptido según se describió en la sección experimental. La **Tabla 5.5** muestra la longitud promedio de los péptidos en el conjunto de aprendizaje y el número promedio de picos observados para cada tipo de fragmento. Idealmente, si se produjese la disociación en todos los enlaces de la columna del péptido, el número de picos observados para cualquier tipo de ión de la columna debería ser igual a la longitud de la secuencia de aminoácidos menos 1. Desde este punto de vista, la razón entre el número de picos observados de un fragmento en particular y la longitud de la cadena de aminoácidos puede convertirse en una medida porcentual de cobertura de la secuencia por parte del tipo de ión en particular, y es así como deberá entenderse de aquí en adelante (ver **Tabla 5.5**). Idealmente se trata de lograr los máximos porcentajes de cobertura en cada caso.

Tabla 5.5: Longitud y porcentaje de cobertura de la secuencia por parte de los fragmentos observados

Fragmento	Longitud Promedio	Porcentaje de cobertura promedio
Secuencia completa	13.65	100 %
Tipo <i>a</i>	2.47	12.2 %
Tipo <i>c</i>	3.19	17.3 %
Tipo <i>x</i>	2.15	15.8 %
Tipo <i>z</i>	3.56	23.1 %

Los porcentajes de cobertura observados para fragmentos de tipo *a*, *c*, *x*, y *z* resultan bajos cuando se comparan con el caso de los iones tipo *b* y tipo *y* que pueden alcanzar niveles del 70 % o superiores [107], presumiblemente debido a que la energía interna entregada a los iones no es la suficiente para inducir procesos esencialmente distintos a los de CID a 15 eV y quizás el proceso sigue siendo principalmente de carácter ergódi-

co (Sección 2.5.2). Sin embargo, la identificación de estos fragmentos puede ayudar a mejorar la identificación de los péptidos correspondientes mediante *de novo* sequencing. La Tabla 5.6 muestra los resultados para los dos subconjuntos de datos, es decir, para aquellos péptidos con carga 2+ y 3+. Puede verse que los resultados no son muy diferentes a los resultados globales de la Tabla 5.5 y la diferencia entre los estados de carga 2+ y 3+ parecen indicar una mayor fragmentación en los estados de mayor carga. Esto puede explicarse desde el punto de vista del modelo del protón móvil ya que los protones adicionales facilitan la ruptura de los enlaces en la columna del péptido. Hay que señalar que si bien los patrones de fragmentación cambian para diferentes energías de colisión (ver Figura 5.8), y que podría ser deseable adicionar información de espectros de fragmentación a energías mucho más altas, las características del cuadrupolo hacen imposible obtener espectros de fragmentación CID en altas energías (varios cientos o miles de voltios) como en espectrómetros de tipo TOF-TOF.

Tabla 5.6: Longitud y porcentaje de cobertura de la secuencia por parte de los fragmentos observados (2+ y 3+)

Fragmento	Longitud Promedio	Porcentaje de cobertura promedio
Secuencia completa	10.02 ^a	100 % ^a
	17.35 ^b	100 % ^b
Tipo a	2.53 ^a	12.7 % ^a
	2.85 ^b	15.0 % ^b
Tipo c	2.17 ^a	16.3 % ^a
	3.25 ^b	19.3 % ^b
Tipo x	2.08 ^a	15.4 % ^a
	3.01 ^b	16.3 % ^b
Tipo z	3.43 ^a	20.6 % ^a
	4.57 ^b	24.2 % ^b

^a Estado de carga: 2+

^b Estado de carga: 3+

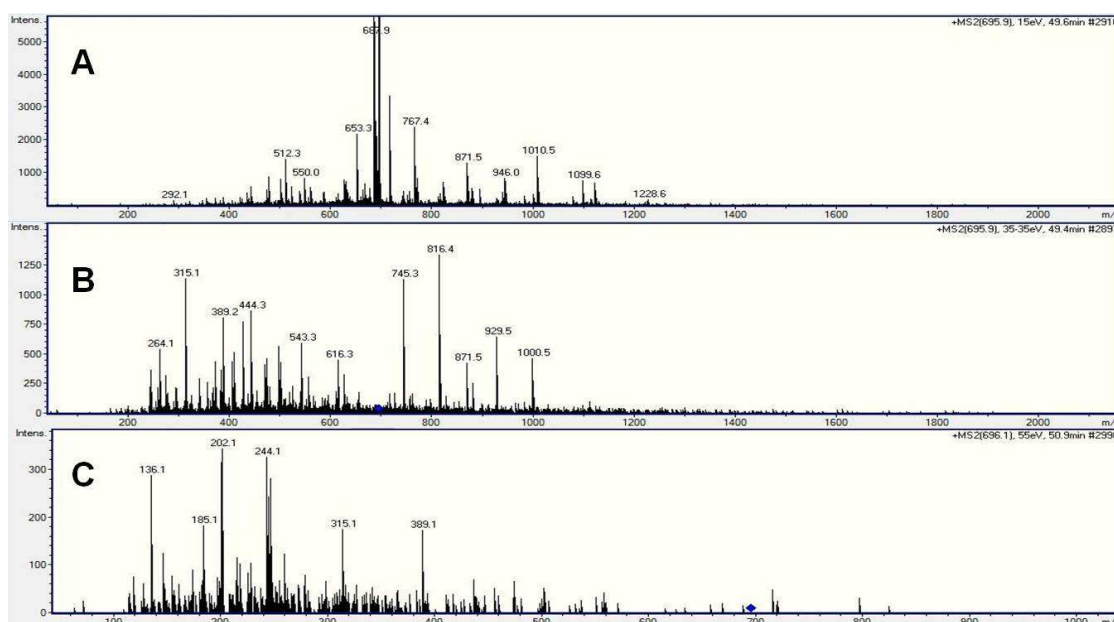


Figura 5.8: (A) Espectro de fragmentación en S2 (Energía de colisión = 15 eV, $m/z = 695.9$ y tiempo de retención = 49.6 min). (B) Espectro de fragmentación en S5 (Energía de colisión = 35 eV, $m/z = 695.9$ y tiempo de retención = 49.4 min). (C) Espectro de fragmentación en S6 (Energía de colisión = 55 eV, $m/z = 696.1$ y tiempo de retención = 50.9 min). Se puede observar que el incremento de energía de colisión provoca mayor fragmentación. Mientras mayor la energía, además, mayor el número de fragmentos livianos.

A partir de los resultados de la [Tabla 5.6](#) se definieron los parámetros α_t^{ch} y β_t^{ch} . Debido a los bajos porcentajes de cobertura de los fragmentos, se decidió tomar $\beta_t^{ch}=0$ y considerar solo la contribución positiva por la identificación de alguno de los fragmentos en el espectro. Los valores para los parámetros α_t^{ch} se observan en la [Tabla 5.7](#). Los parámetros fueron escogidos simplemente como las fracciones promedio de cobertura de la secuencia según los resultados para el conjunto de datos de aprendizaje.

Tabla 5.7: Parámetros α_t^{ch}

α_t^{ch}	Tipo a	Tipo c	Tipo x	Tipo z
2+	0.127	0.163	0.154	0.206
3+	0.150	0.193	0.163	0.242

El set de datos S6 se utilizó para verificar la presencia de iones immonio (ver [Tabla 4.4](#)). La [Figura 5.9](#) muestra un histograma con el número de iones immonio observados por péptido. Según esto, el número medio de iones immonio observados por péptido es de 6.50 con un mínimo de 1 y un máximo de 13. Debido a que la longitud promedio de los péptidos es de 13.65 el valor promedio de S_{c_2} es de 0.476.

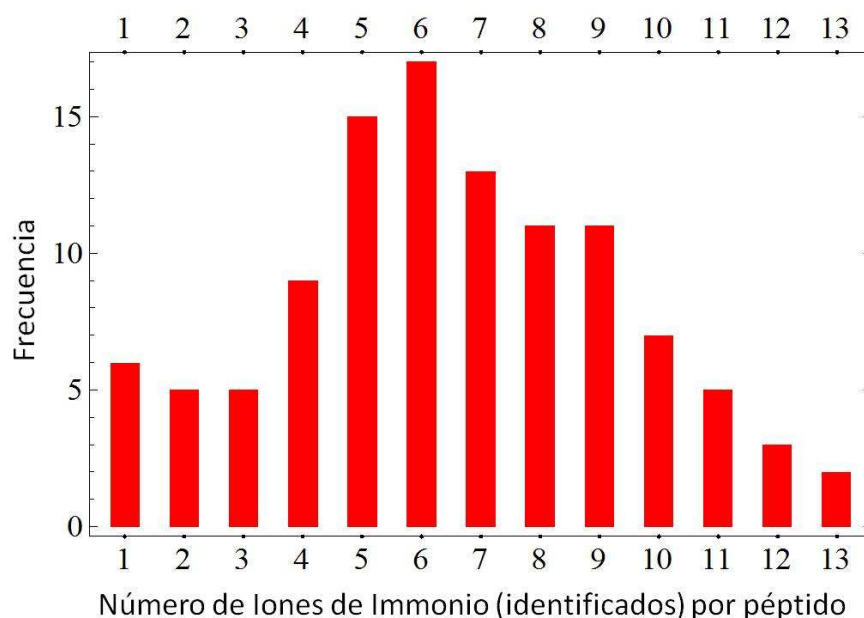


Figura 5.9: Histograma del número de iones immonio (identificados) por péptido.

Finalmente, se calcularon los parámetros λ_1 y λ_2 de la [Ecuación 4.12](#). Para cada combinación de valores de los parámetros, se evaluaron los scores para las 190 secuencias

y su suma total. Los valores que maximizaron esta suma total fueron $\lambda_1=0.84$ y $\lambda_2=0.32$ de manera que el esquema de score se escribe como:

$$Sc = 0,84Sc_1 + 0,32Sc_2 \quad (5.1)$$

5.2.2. RESULTADO DEL RE-SCORING DE LAS IDENTIFICACIONES OBTENIDAS CON PEPNOVO

El procedimiento de re-scoring de las identificaciones se realizó utilizando la información de los espectros de fragmentación obtenidos con energías de colisión de 35 y 55 eV según el score definido en la [Ecuación 5.1](#). Para esto, se utilizaron las 476 secuencias que no formaron parte del conjunto de aprendizaje. Los resultados de PepNovo arrojaron 5 candidatos rankeados según el score del programa para cada espectro de fragmentación. En este trabajo se cálculo un score adicional según la [Ecuación 5.1](#). El uso de ambos scores permitió elevar sustancialmente el nivel de certeza en la identificación del péptido correspondiente. Esto es así, por que no necesariamente el primer candidato en la lista de PepNovo es el péptido real.

Se siguió el procedimiento descrito en la [Sección 4.9.2](#) para validar los fragmentos de la columna del péptido utilizando el set de datos S5, y la presencia de iones immonio mediante el conjunto de datos S6. Como resultado se reportó la secuencia con el mayor score según el proceso de validación. Con esta nueva lista de principales candidatos se volvieron a evaluar los índices de *Precisión* y *Eficiencia* para conocer la mejora obtenida por la utilización del método.

En la [Tabla 5.8](#) se muestran los resultados antes de la validación ([Tabla 5.4](#)) y los resultados después de realizar el proceso de validación. Puede verse que ambos índices han mejorado (7 y 14 % respectivamente), lo cual implica una mejor capacidad de identificar correctamente las secuencias de aminoácidos a partir de los espectros de fragmentación.

Evidentemente, el procedimiento de re-scoring aumentó el número total de aminoácidos correctamente predichos dentro de las secuencias peptídicas. La mejora en la identificación se da a nivel global y no necesariamente a nivel local; es decir, que algunos de los péptidos correctamente identificados inicialmente se perdieron en el nivel de validación. Por ejemplo, el péptido AGENVGVLLR (tiempo de retención 33.9 min, MS/MS

Tabla 5.8: Resultados del procedimiento de *de novo* sequencing luego del re-scoring

	Resultados Originales	Luego del re-scoring
Precisión	72.6 %	77.4 %
Eficiencia	63.7 %	72.8 %

scan 1990) fue correctamente identificado por PepNovo; sin embargo, después de la validación, el péptido AADNVGVLLR, inicialmente rankeado segundo obtuvo el score más alto según nuestra definición. El opuesto también es cierto (es decir que péptidos de bajo ranking pasan a ser la primera opción), y ocurre con más frecuencia (como era de esperarse). El número de identificaciones correctas a nivel de péptidos (Eficiencia = 1) según PepNovo fue de 128 péptidos, y después del procedimiento de re-scoring, este número se incrementó a 143. Por su parte, las identificaciones correctas parciales, es decir aquellas en que todos los aminoácidos predichos por PepNovo son correctos pero la cobertura de la secuencia no es completa (Precisión = 1, Eficiencia \neq 1) pasó de 187 a 211 después de la validación. En ambos casos el incremento fue superior al 10 %.

5.3. LIMITACIONES Y ALCANCE

En este trabajo se presentaron dos métodos que permiten mejorar la identificación de péptidos por medio de 2 mecanismos distintos. En el primer caso se propuso un método de recalibración automático que utiliza la información de los picos presentes en la muestra y permite realizar una calibración interna sin la necesidad de adicionar elementos extraños a la muestra. Este método no depende del acceso a las unidades de tiempo de vuelo originales, por lo cual es un método independiente del tipo de espectrómetro utilizado y puede aplicarse al caso de espectrómetros comerciales en que las ecuaciones de calibración originales son desconocidas. Una de las limitaciones del método es su dependencia del número de calibrantes disponibles por scan; sin embargo, la relación interna entre los parámetros a_i de las ecuaciones de recalibración hace que el método sea aplicable a un porcentaje muy alto de los espectros que se observan en una corrida típica de LC-MS/MS. La limitación principal tiene que ver con una característica intrínseca de la calibración de espectros de tiempo de vuelo, que es que la recalibración sólo puede ser realizada con éxito para aquellos picos dentro de rango cubierto por los

calibrantes internos, es decir, no es posible extrapolar los resultados dentro de un espectro. La recalibración presenta gran potencial para realizar una segunda búsqueda en base de datos con los espectros recalibrados, para lo cual debería volver a escribirse los archivos .mzxml con los nuevos valores de m/z y realizar una segunda búsqueda en base de datos o un proceso de *de novo* sequencing tomando en cuenta la mejora en la precisión de la determinación de la masa. Idealmente, el procedimiento aquí introducido puede iterarse varias veces hasta alcanzar el máximo número de identificaciones positivas. Por supuesto, en este caso será necesario llegar a un compromiso entre la tasa de mejora en cada paso de la iteración y el tiempo de cómputo involucrado.

En segundo lugar, se introdujo un método de re-aseguramiento y re-scoring de las identificaciones positivas usando información adicional de espectros de fragmentación con energías diferentes. Este método puede ser acoplado a cualquier tipo de análisis por medio de LC-MS/MS en un espectrómetro de tipo QqTOF. Una gran ventaja de este procedimiento, es que el análisis consecutivo de los distintos sets de datos producidos para sendas energías puede automatizarse con una rutina de programación adecuada, sin supervisión. Para esto es necesario un trabajo previo para determinar los parámetros de transmisión que se utilizarán en el espectrómetro. La limitación principal de este método es el factor tiempo pues requiere el triple del tiempo para adquirir los tres sets de datos necesarios.

CAPÍTULO 6

CONCLUSIONES Y RECOMENDACIONES

6.1. CONCLUSIONES

Se ha utilizado una muestra de material biológico proveniente de un extracto de células completa de *Escherichia Coli* para realizar procedimientos de recalibración que permitan mejorar la precisión en la determinación de la masa mediante espectrometría de masas de tiempo de vuelo, y analizar la fragmentación a diferentes energías de colisión para obtener información complementaria en la identificación mediante *de novo* sequencing. Para esto se realizaron procedimientos estándar de análisis en el marco de la proteómica incluyendo separación por HPLC, espectrometría de masas tándem e identificación por medio de bases de datos y *de novo* sequencing.

A partir del trabajo realizado se pueden extraer las siguientes conclusiones:

Recalibración

- Se ha mostrado la validez de un método de recalibración automático independiente del instrumento utilizado en la adquisición de datos. El método se basa en las ecuaciones de tiempo de vuelo y en la utilización de péptidos identificados en un primer procedimiento de búsqueda en base de datos como calibrantes internos para determinar los parámetros en varias ecuaciones de recalibración.
- Se determinó que la ecuación simple representada por la Función 1 produce re-

sultados comparables con el resto de funciones con más parámetros; y aunque la Función 3 obtuvo los mejores resultados, la Función 1 presenta ventajas computacionales al facilitar el cálculo de los parámetros a_i .

- El hecho de que la Función 1 obtuviera buenos resultados implica que la ecuación cuadrática simple es suficiente para corregir los errores producto de la dispersión de velocidad y energía inicial en el caso de espectrómetros de tipo oa-TOF.
- Para todos los conjuntos de datos se mostró que el procedimiento de recalibración exitosamente centra la distribución de errores alrededor de cero y la dispersión estándar del error se reduce entre 2 y 20 veces dependiendo de la calibración original del instrumento. En el mejor de los casos, es decir, cuando se trabajó con los datos del espectrómetro maXis con una buena calibración original, la dispersión del error obtenida mediante el procedimiento de recalibración llegó a ser menor a 1 ppm.
- La relación interna de los parámetros a_i permitió desarrollar un procedimiento de recalibración de 2 etapas, al final de las cuales se consiguió recalibrar entre un 74 % y un 91 % de los espectros dependiendo de la calibración original del instrumento.

Re-scoring utilizando diferentes energías de colisión en CID

- La utilización de energías de colisión adicionales de 35 eV y 55 eV en el sistema de centro de masas, mostró patrones de fragmentación distintos a los observados en el caso de energías de colisión de 15 eV. Ver [Figura 5.8](#).
- Al utilizar la información de fragmentación adicional incluida en los conjuntos de datos de 35 eV y 55 eV, se diseñó un sistema de re-scoring que considera fragmentos de la columna del péptido de tipo *a*, *c*, *x*, *z*, y la presencia de iones immonio en el caso de energía de fragmentación de 55 eV. Este sistema de re-scoring permitió re-evaluar las identificaciones obtenidas mediante el programa comercial PepNovo, e incrementar el número de péptidos correctamente identificados en valores superiores al 10 %. Las figuras de mérito utilizadas para medir la eficacia del método mostraron también incrementos alrededor del 10 %.
- Aún cuando los porcentajes de cobertura de los iones fragmento considerados no son altos (alrededor del 30 %), se mostró que la información adicional es suficiente

para desarrollar un sistema que mejore las identificaciones obtenidas por medio de programas computacionales comerciales como PepNovo.

6.2. RECOMENDACIONES

En base a la experiencia adquirida durante el desarrollo de esta investigación se plantean las siguientes recomendaciones y trabajos futuros:

- Una vez conocida la mejor ecuación para realizar el procedimiento de recalibración (Función 1), se debería mostrar su validez dentro de un proceso iterativo de recalibración e identificación por medio de base de datos o *de novo* sequencing. Es decir, debería recalibrarse un conjunto de datos y re-escribir el archivo .mzxml con los nuevos valores de m/z para realizar una nueva identificación de péptidos y comparar la tasa de identificaciones con la original. Luego se podría utilizar los nuevos péptidos identificados para repetir todo el proceso hasta que no se alcance una mejora sustancial en el número de péptidos identificados en la muestra.
- Los resultados obtenidos con el esquema de re-scoring sencillo propuesto en este trabajo son muy alentadores y debería realizarse un procedimiento semejante considerando funciones de re-scoring más complejas que tomen en cuenta modelos teóricos de predicción de fragmentos. En el caso particular de la identificación de iones immonio podría utilizarse un esquema probabilístico para la asignación de scores basado en el trabajo de Falick et. al. [102].
- En base a los dos métodos introducidos en este trabajo, se podría elaborar un esquema de trabajo que incorpore ambos métodos de manera automática en un algoritmo de identificación de péptidos mediante espectrometría de masas de tiempo de vuelo según el esquema que se muestra a continuación:

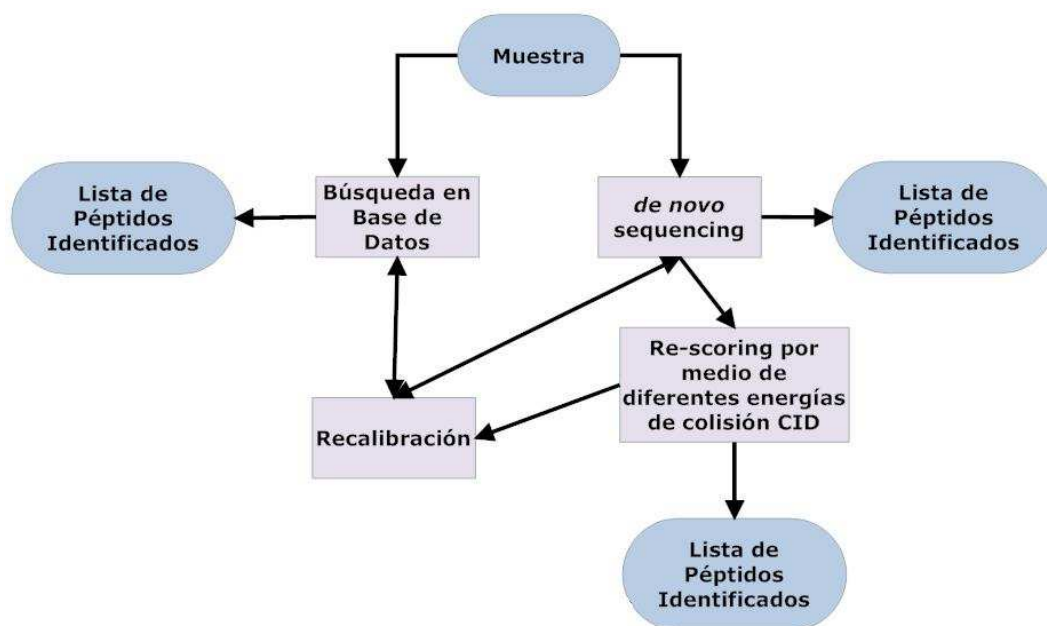


Figura 6.1: Diagrama de flujo para la identificación de péptidos mediante los métodos propuestos.

REFERENCIAS

- [1] Abhilash, M.; Tech, B., Applications Of Proteomics, The Internet Journal of Genomics and Proteomics. 2009, 4(1).
- [2] Aebersold, R.; Mann M., Mass spectrometry-based proteomics. Nature. 2003, 422, 198 - 207.
- [3] Bronstrup, M. Absolute quantification strategies in proteomics based on mass spectrometry, Expert Rev Proteomics, 2004, 1(4), 503-512.
- [4] Conrads, T.P.; Anderson, G.A.; Veenstra, T.D.;Pasa-Tolic, L.; Smith R.D., Utility of accurate mass tags for proteome-wide protein identification, Anal Chem. 2000, 72(14), 3349-3354.
- [5] Mitulovic, G.; Mechtler, K., HPLC techniques for proteomics analysis—a short overview of latest developments. Brief Funct Genomic Proteomic. 2006, 5 (4), 249-260.
- [6] Dancik, V. et al., De novo protide sequencing via tandem mass-spectrometry, J. Comp. Biology. 1999, 6, 327 - 342.
- [7] Palmblad, M.; Bindschedler, L. V.; Gibson, T. M.; Cramer, R., Automatic internal calibration in liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry of protein digests, Rapid Commun Mass Spectrom. 2006, 20 (20), 3076-3080.
- [8] Costa-Vera C. et. al. (2009). A comparison of Calibration Equations for Improving Mass Accuracy by Internal Recalibration of TOF MS/MS Data Sets From Whole-cell Digests. 57th ASMS Conference, Philadelphia.
- [9] Zhang, G.; Ueberheide, G.; Waldemarson, S.; Myung, S.; Molloy, K.; Eriksson, J.; Chait, B.; Neubert, T.; Fenyo, D., Protein Quantitation Using Mass Spectrometry, Methods Mol Biol. 2010, 673, 211 - 222.

- [10] Kline, K.; Sussman, M., Protein Quantitation Using Isotope-Assisted Mass Spectrometry, *Annual Rev Biophysics*. 2010, 39, 291 - 308.
- [11] Eidhammer I, Flikka K, Martens L, Mikalsen S-O., 2007. *Computational Methods for Mass Spectrometry Proteomics*. 1 ed. Chichester: Wiley.
- [12] Ekman, R.; Silberring, J.; Westman-Brinkmalm, A.; Kraj, A., 2009. *MASS SPECTROMETRY: Instrumentation, Interpretation and Applications*. 1 ed. New Jersey:Wiley.
- [13] Bleakney, W., A New Method of Positive Ray Analysis and Its Application to the Measurement of Ionization Potentials in Mercury Vapor, *Phys. Rev.* 1929, 34(1), 157 - 160.
- [14] Harrison, A., 1983. *Chemical Ionisation Mass Spectrometry*, CRC Press, Boca Raton, FL.
- [15] Horning, E.; Horning, M.;Carroll, D.; Dzidic, I.; Stillwell, R., New Picogram Detection System Based on a Mass Spectrometer with an External Ionization Source at Atmospheric Pressure. *Anal. Chem.*, 1973, 45, 936 -943.
- [16] Karas, M.;Bachmann, D.;Bahr, U.; Hillenkamp, F., Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds, *Int. J. Mass Spectrom. Ion Proc.* 1987, 78, 53 - 68.
- [17] Karas, M.; Bahr, U.; Hillenkamp, F., UV Laser Desorption/Ionization Mass Spectrometry of Proteins in the 100,000 Dalton Range, *Int. J. Mass Spectrom. Ion Proc.* 1989, 92, 231 - 242.
- [18] Barber, M.; Bordoli, R.; Sedgewick, R.; Tyler, A., Fast Atom Bombardment of Solids (F. A. B.): A New Ion Source for Mass Spectrometry, *J. Chem. Soc. Chem. Commun.* 1981, 7, 325 - 327.
- [19] Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S.F.; Whitehouse, C.M., Electrospray ionization for mass spectrometry of large biomolecules, *Science*. 1989, 246 (4926), 64-71.
- [20] Mann, M.; Meng, C.K.; Fenn, J.B., Interpreting mass spectra of multiply charged ions, *Anal. Chem.* 1989, 61, 1702 - 1708.
- [21] El-Faramawy, A.; Siu, K.W.; Thomson, B.A., Efficiency of Nano Electrospray Ionization, *J Am Soc Mass Spectrom.* 2005, 16, 1702-1707.

- [22] Covey, T., Analytical characteristics of the electrospray ionization process, in Biochemical and Biotechnological applications of electrospray ionization mass spectrometry. 1996, ed. A.P. Snyder, ACS Symposium Series 619, American Chemical Society: Washington.
- [23] Wahl, J.; Goodlett, D.; Udseth, H.; Smith, R., Attomole level capillary electrophoresis-mass spectrometric protein analysis using 5 μm i.d. capillaries, *Anal Chem.* 1992, 64(24), 3194 - 3196.
- [24] Hoffman, E.; Stroobant, V., 2007. *Mass Spectrometry: Principles and Applications.* 3 ed. Chichester:Wiley.
- [25] Kaltashov, I.; Eyles, S., 2005. *Mass Spectrometry in Biophysics.* 1 ed. New Jersey: Wiley.
- [26] Rayleigh, J., On the equilibrium of liquid conducting masses charged with electricity. *Philos. Mag.* 1882, 14, 184 - 186.
- [27] Belonozhko, D.; Grigorev, A., Characteristic time for the evolution of instability of a droplet charged to the Rayleigh limit. *Tech. Phys. Lett.* 1999, 25, 610 - 611.
- [28] Iribarne, J.; Thomson, B., On the evaporation of small ions from charged droplets. *J. Chem. Phys.* 1976, 64, 2287 - 2294.
- [29] Marshall, A.; Hendrickson, C.; Jackson, G., Fourier Transform Ion Cyclotron Resonance mass spectrometry: a primer, *Mass Spectrom Rev.* 1998, 17, 1 - 35.
- [30] Wiley, W.; McLaren, I., Time-of-flight Mass Spectrometer with Improved Resolution, *Rev Sci Instr.* 1955, 26, 1150 - 1157.
- [31] Mamyrin, B.; Karataev, V.; Schmikk, D.; Zagulin, V., *Sov Phys. JETP.* 1973, 37, 4.
- [32] Guilhaus, M.; Selby, D.; Mlynski, V., Orthogonal acceleration time-of-flight mass spectrometry, *Mass Spectrom Rev.* 2000, 19 (2), 65-107.
- [33] Mathieu, E., Le mouvement vibratoire d'une membrane de forme elliptique, *J. Math. Pures Appl.* 1868, 13, 137 - 203.
- [34] Gutiérrez-Vega, C.; Rodríguez, R.; Meneses, M.; Chávez, S., Mathieu functions, a visual approach, *Am. J. Phys.* 2003, 71(3), 233 - 242.
- [35] Santacruz, C.; Hakansson, P.; Barofsky, D.; Piyadasa, C., A Constant-Momentum/Energy-Selector Time-of-Flight Mass Spectrometer, *J. Am. Soc. Mass Spectrom.* 2007, 18, 92 - 101.

- [36] Chernushevich, I.; Loboda, A.; Thomson, B., An introduction to quadrupole-time-of-flight mass spectrometry, *J Mass Spectrom.* 2001, 36 (8), 849 - 865.
- [37] Wolff, J.; Fuentes, T.; Taylor, J., Investigations into the accuracy and precision obtainable on accurate mass measurements on a quadrupole orthogonal acceleration time-of-flight mass spectrometer using liquid chromatography as sample introduction, *Rapid Commun Mass Spectrom.* 2003, 17 (11), 1216-1219.
- [38] Weaver, P.; Laures, A.; Wolff, J., Investigation of the advanced functionalities of a hybrid quadrupole orthogonal acceleration time-of-flight mass spectrometer, *Rapid Commun Mass Spectrom.* 2007, 21 (15), 2415-2421.
- [39] Vestal, M.; Juhasz, P., Resolution and Mass Accuracy in Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight, *J American Society for Mass Spectrometry.* 1998, 9, 892-911.
- [40] Costa, C.; Zubarev, R.; Ehring, H.; Hakansson, P.; Sunqvist, U. R., A Three-point Calibration Procedure for Matrix-assisted Laser Desorption/Ionization Mass Spectrometry Utilizing Multiply Charged Ions and Their Mean Initial Velocities, *Rapid Commun Mass Spectrom.* 1996, 10, 1429-1432.
- [41] Cotter, R., 1997. Time of flight mass spectrometry: instrumentation and applications in biological research. 1 ed. Washington:Am Chem Soc.
- [42] Schwartz, J.; Wade, A.; Enke, C.; Cooks, R.; Systematic delineation of scan modes in multidimensional mass spectrometry. *Anal. Chem.* 1990, 62, 1809 - 1818.
- [43] Ens, W.; Standing, K., Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules, *Methods Enzymol.* 2005, 402, 49-78.
- [44] Kristensen, D.; Imamura, K.; Miyamoto, Y.; Yoshizato, K., Mass spectrometric approaches for the characterization of proteins on a hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer, *Electrophoresis.* 2000, 21 (2), 430-439.
- [45] Krutchinsky, A.; Chernushevich, IV.; Spicer, V.; Ens, W.; Standing, K., Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer, *J Am Soc Mass Spectrom.* 1998, 9, 569 - 579.
- [46] Roepstorff, P.; Fohlman, J., Letter to the editors. *Biological Mass Spectrometry.* 1984, 11: 601

- [47] Johnson, R.; Martin, S.; Biemann, K.; Stults, J.; Watson, J.; Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine, *Analytical Chemistry*. 1987, 59 (21), 2621 - 2625.
- [48] Biemann, K., Nomenclature for Peptide Fragment Ions, In *Methods in Enzymology*. 1990, 193, 886 - 888.
- [49] Mayer, P.; Poon, C., The Mechanism of collisional activation of ions in mass spectrometry, *Mass Spec. Rev.* 2009, 28, 608 - 639.
- [50] Cooks, R.; Beynon, J.; Caprioli, R.; Lester, G., 1973. *Metastable Ions*. 1 ed. Amsterdam: Elsevier.
- [51] Laramee, J.; Carmody, J.; Cooks, R., Angle-resolved mass spectrometry, *Int J Mass Spectrom Ion Phys.* 1979, 31, 333 - 343.
- [52] Hubik, A.; Hemberger, P.; Laramee, J.; Cooks, R., Control of energy deposition by impact parameter in polyatomic ion collisions, *J Am Chem Soc.* 1980, 102, 3997 - 4000.
- [53] Hayes, R.; Gross, M., Collision-Induced Dissociation, *Methods in Enzymology*. 1990, 193, 237 - 263
- [54] Fournier, P.; VandeRunstraat, C.; Govers, T.; Schopman, J.; deHeer, F.; Los, J., Collision-induced dissociation of 10 KeV N₂ ions. Evidence for predissociation of the C₂S state, *Chem Phys Lett.* 1971, 9, 426-428.
- [55] Fuentes, B.; Martinez, H., Translational spectroscopy of N produced by collision induced dissociation of N₂ ion He and Ar, *Int J Mass Spectrom.* 2004, 238, 55-58.
- [56] Wysocki, V.; Kenttamaa, H.; Cooks, R., Internal Energy Distributions of Ions Activated by Various Methods, *Int. J. Mass Spectrom. Ion Processes.* 1987, 75, 181-208.
- [57] Baer, T.; Mayer, P., Statistical Rice-Ramsperberg-Kassel-Marcus Quasiequilibrium Theory Calculations in Mass Spectrometry, *J Am Soc Mass Spectrom.* 1997, 8, 103-115.
- [58] Dongre, A.; Jones J.; Somogyi, A.; Wysocki, V., Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model, *J Am Chem Soc.* 1996, 118, 8365-8374.

- [59] Wysocki, V.; Tsaprailis, G.; Smith, L.; Brechi, L., Mobile and localized protons: a framework for understanding peptide dissociation, *J Mass Spectrom.* 2000, 35, 1399-1406.
- [60] Schlosser, A.; Lehmann, W., Five-membered ring formation in unimolecular reactions of peptides: a key structural element controlling low energy collision induced dissociation of peptides *J. Mass Spectrom.* 2000, 35, 1382-1390.
- [61] Polce, M.; Ren, D.; Wesdemiotis, C., Dissociation of the peptide bond in protonated peptides, *Mass Spectrom.* 2000, 35, 1391-1398.
- [62] Zubarev, R.; Kelleher, N.; McLafferty, F., Electron Capture Dissociation of Multiply Charged Protein Cations. A Non-ergodic Process, *J. Am. Chem. Soc.* 1998, 120, 3265-3266.
- [63] Syka, J.; Coon, J.; Schroeder, M.; Shabanowitz, J.; Hunt, D., Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry, *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 9528 - 9533.
- [64] Zubarev, R.; Zubarev, A.; Savitski, M., Electron Capture/Transfer versus Collisionally Activated/Induced Dissociations: Solo or Duet?, *J Am Soc Mass Spectrom.* 2008, 19, 753 - 761.
- [65] Haselmann, K.; Budnik, B.; Olsen, J.; Nielsen, M.; Reis, C.; Clausen, H.; Johnson, A.; Zubarev, R., Advantages of External Accumulation for Electron Capture Dissociation in Fourier Transform Mass Spectrometry, *Anal. Chem.* 2001, 73, 2998 - 3005.
- [66] Appella, E.; Anderson, C., New Prospects for Proteomics: Electron-Capture (ECD) and Electron-Transfer Dissociation (ETD) Fragmentation Techniques and Combined Fractional Diagonal Chromatography (COFRADIC). *FEBS J.* 2007, 274, 6255 - 6255.
- [67] Molina, H.; Horn, D.; Tang, N.; Mathivanan, S.; Pandey, A., Global Proteomic Profiling of Phosphopeptides Using Electron Transfer Dissociation Tandem Mass Spectrometry, *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 2199 - 2204.
- [68] Brenner, S., The molecular evolution of genes and proteins: a tale of two serines, *Nature.* 1988, 334(6182), 528 - 530.
- [69] Lesk, A., 2007. *Introduction to Genomics.* 1 ed. New Jersey: Oxford Univ. Press.

- [70] Valaskovic, G.A.; Kelleher, N.L.; McLafferty, F.W., Attomole protein characterization by capillary electrophoresis mass spectrometry, *Science*. 1996, 273 (5279), 1199 - 1202.
- [71] Martin, S.E.; Shabanowitz, J.; Hunt, Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* 2000, 72, 4266 - 4274.
- [72] Wolters, D. A.; Washburn, M. P.; Yates, J. R. 3rd, An automated multidimensional protein identification technology for shotgun proteomics, *Anal Chem.* 2001, 73 (23), 5683-5690.
- [73] Winter, M., WebElements: the periodic table on the WWW, [consultada: 29 de mayo de 2011], <http://webelements.com/>
- [74] Steven, P.; Corthals, G.; Zhang, Y.; Rochon, Y.; Aebersold, R., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci.* 2000, 97(17), 9390 - 9395.
- [75] Thermo Fischer Scientific Inc., Proteases for Mass Spectrometry: High-purity individual proteases for powerful analysis by mass spectrometry, [consultada: 2 de Junio de 2011], <http://www.piercenet.com/browse.cfm?fldID=70778998-5056-8A76-4E98-B25CE62646AD>
- [76] Mariño, K.; Bones, J.; Kattla, J.; Rudd, P., A systematic approach to protein glycosylation analysis: a path through the maze, *Nature Chem. Biol.* 2010, 6, 713 - 723.
- [77] Weise, C.; Lenz, C., Identification of protein phosphorylation sites by advanced LC-ESI-MS/MS methods, *Methods Mol Biol.* 2008, 446, 33 - 46.
- [78] Tswett, M., Physikalische-chemische Studien u ber das Chlorophyll. Die Adsorptionen, *Deutschen Botan. Ges.* 1906, 24, 316-323.
- [79] Ettre, M. S., Tswett and the Invention of Chromatography, *LCGC N. Amer.* 2003, 21, 458-467.
- [80] Durbin, K.; Tran, J.; Zambdborg, L.; Sweet, S.; Catherman, A.; Ji, E.; Mingxi, L.; Kellie, J.; Kelleher, N., Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale, *Proteomics.* 2010, 10(20), 3589-3597.

- [81] VerBerkmoes, N.; Bundy, J.; Hauser, L.; Asano, K.; Razumovskaya, J.; Larimer, F.; Hettich, R.; Stephenson, J., Integrating Top-Down and Bottom-Up Mass Spectrometric Approaches for Proteomic Analysis of *Shewanella oneidensis*, *Journal of Proteome Research*. 2002, 1(3), 239-252.
- [82] Henzel, W.; Billeci, T.; Stults, J.; Wong, S., *Proc. Natl. Acad. Sci.* 1993, 90, 5011 - 5015.
- [83] Kall, L.; Canterbury, J.D.; Weston, J.; Noble, W.S.; MacCoss, M.J., Semi-supervised learning for peptide identification from shotgun proteomics datasets, *Nature*. 2007, 4 (11), 923-925.
- [84] Eng, J.; McCormack A.; Yates J. R. 3rd, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *J Am Soc Mass Spectrom.* 1994, 5, 976-989.
- [85] Perkins, D.; Pappin, D.; Creasy D.; Cottrell, J., Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis*. 1999, 20, 3551-3567.
- [86] Craig, R.; Beavis, R., TANDEM: matching proteins with tandem mass spectra, *Bioinformatics*. 2004, 20, 1466-1467.
- [87] Johnson, R.; Davis, M.; Tayler J.; Patterson S., Informatics for protein identification by mass spectrometry, *Methods*. 2005, 35, 223-236
- [88] Bertsch, A. et al, De novo peptide sequencing by tandem MS using complementary CID and electron transfer dissociation, *Electrophoresis*. 2009, 30, 3736-3747
- [89] Yates, J.; Griffin, P.; Hood, L., 1991. Editor, *Techniques in Protein Chemistry Vol. II*, Academic Press:San Diego, 477-489.
- [90] Ma, B. et al, PEAKS: powerful software for peptide de novo sequencing by MS/MS, *Rapid Commun. Mass Spectrom.* 2003, 17, 2337-2342.
- [91] Ma, B. et al, An effective algorithm for peptide de novo sequencing from MS/MS spectra, *Journal of Computer and System Sciences*. 2005, 70, 418-430.
- [92] Taylor, J.; Johnson, R., Sequence database searches via de novo peptide sequencing by tandem mass spectrometry, *Rapid Commun. Mass Spectrom.* 1997, 11, 1067-1075.

- [93] Taylor, J.; Johnson, R., Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry, *Anal. Chem.* 2001, 73, 2594-2604.
- [94] Frank, A.; Pevzner, P., Pepnovo: De novo peptide sequencing via probabilistic network modeling, *Anal. Chem.* 2005, 77, 964-973.
- [95] Mo, L.; Dutta, D.; Wan, Y.; Chen, T., MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing Via Tandem Mass Spectrometry. *Anal Chem.* 2007, 79, 4870-4878.
- [96] Fischer, B. et al., NovoHMM: A Hidden Markov Model for de novo peptide sequencing. *Anal. Chem.* 2005, 77, 7265-7273.
- [97] Blattner, F.; Plunket, G.; Bloch, C.; Perna, N.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.; Rode, C.; Mayhew, F.; Gregor, J.; Davis, N.; Kirkpatrick, H.; Goeden, M.; Rose, D.; Mau, B.; Shao, Y., The Complete Genome Sequence of *Escherichia coli* K-12, *Science*. 1997, 277(5331), 1453-1462.
- [98] Frank, A., A Ranking-Based Scoring Function for Peptide-Spectrum Matches, *J. Proteome Research*. 2009, 8, 2241-2252.
- [99] Moré, J., The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical Analysis*, Watson, G., Ed.; Springer: Berlin / Heidelberg, 1978; Vol. 630, 105-116.
- [100] Wang, P.; Tang, H.; Fittzgibbon, M.; McIntosh, M., A statistical method for chromatographic alignment of LC-MS data, *Biostatistics*. 2007, 8(2), 357-367.
- [101] Nevedomskaya, E.; Derks, R.; Deelder, A.; Mayboroda, O.; Palmblad, M., Alignment of capillary electrophoresis-mass spectrometry datasets using accurate mass information, *Analytical and Bioanalytical Chem.* 2009, 395(8), 2527-2533.
- [102] Falick, A.; Hines, W.; Medzihradzky, K.; Baldwin, M.; Gibson, B., Low-Mass Ions produced from peptides by High-Energy Collision-Induced Dissociation in Tandem Mass Spectrometry, *J Am Soc Mass Spectrom.* 1993, 4, 882-893.
- [103] Johmann, L.; Eng, J.; Gemmill, A.; Klimek, J.; Vitek, O.; Reid, G.; Martin D., Quantification of the compositional information provided by Immonium Ions on a Quadrupole-Time-of-Flight Mass Spectrometer, *Anal Chem.* 2008, 80, 5596-5606.
- [104] Ekroos, K.; Shevchenko, A., Simple two-point calibration of hybrid quadrupole time-of-flight instruments using a synthetic lipid standard, *Rapid Commun Mass Spectrom.* 2002, 16 (12), 1254-1255.

- [105] Pevtsov, S.; Fedulova, I.; Mirzaei, H.; Buck, C.; Zhang, X., Performance Evaluation of Existing De Novo Sequencing Algorithms, *J Prot Res.* 2006, 5, 3018-3028.
- [106] Mo, L.; Dutta, D.; Wan, Y.; Chen, T., MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing Via Tandem Mass Spectrometry, *Anal Chem.* 2007, 79, 4870-4878.
- [107] Lu, B.; Chen, T., A Suboptimal Algorithm for De Novo Peptide Sequencing via Tandem Mass Spectrometry, *J Comp Biology.* 2003, 10(1), 1-12.

ANEXOS

ANEXO A - FUNCIONES 3 Y 4

Función 3:

$$\frac{m}{z} = -\frac{4a_1}{6a_0} + \frac{2^{4/3}a_1^2}{6 \left(2a_0^3a_1^3 + 27a_0^4(tof - a_2)^2 + 3^{3/2}\sqrt{a_0^7(4a_1^3 + 27a_0(tof - a_2)^2)(tof - a_2)^2} \right)^{1/3}}$$
$$+ \frac{2^{2/3} \left(2a_0^3a_1^3 + 27a_0^4(tof - a_2)^2 + 3^{3/2}\sqrt{a_0^7(4a_1^3 + 27a_0(tof - a_2)^2)(tof - a_2)^2} \right)^{1/3}}{6a_0^2}$$

Función 4:

$$\begin{aligned}
\frac{m}{z} = & \frac{2(a_1^2 - 2a_0a_2)}{6a_0^2} + (2^{4/3} (a_1^4 - 4a_0a_1^2a_2 + a_0^2(a_2^2 - 6a_1(tof - a_3)))) / \\
& (6a_0^2) \left(27(tof)^2a_0^4 - 18(tof)a_0^2a_1^3 + 36(tof)a_0^3a_1a_2 - 12a_0a_1^4a_2 + 15a_0^2a_1^2a_2^2 + 2a_0^3a_2^3 \right. \\
& + 3^{3/2} (a_0^4(a_1a_2 + a_0(tof - a_3))^2 (-a_1^2(a_2^2 + 4a_1(tof - a_3)) + 2a_0a_2(2a_2^2 + 9a_1(tof - a_3))) \\
& + 27a_0^2(tof - a_3)^2)^{1/2} - 54(tof)a_0^4a_3 + 18a_0^2a_1^3a_3 - 36a_0^3a_1a_2a_3 + 27a_0^4a_3^2)^{1/3} \\
& + \left(2^{2/3} (27(tof)^2a_0^4 - 18(tof)a_0^2a_1^3 + 2a_1^6 + 36(tof)a_0^3a_1a_2 - 12a_0a_1^4a_2 + 15a_0^2a_1^2a_2^2 + 2a_0^3a_2^3 \right. \\
& + 3^{3/2} (a_0^4(a_1a_2 + a_0(tof - a_3))^2 (-a_1^2(a_2^2 + 4a_1(tof - a_3)) + 2a_0a_2(2a_2^2 + 9a_1(tof - a_3))) \\
& + 27a_0^2(tof - a_3)^2)^{1/2} - 54(tof)a_0^4a_3 + 18a_0^2a_1^3a_3 - 36a_0^3a_1a_2a_3 + 27a_0^4a_3^2)^{1/3} \left. \right) / (6a_0^2)
\end{aligned}$$