

Temática: Ómicas

Xion-C: un software para facilitar la identificación de los sitios de conjugación de vacunas conjugadas por análisis LC-MS/MS

Xion-C: a software to facilitate the identification of conjugation sites of conjugated vaccines by LC-MS/MS analysis

Pablo Enmanuel Ramos Bermúdez ^{1*}, Luis Javier González López ¹, Satomy Pousa Ramírez ¹, Jorge Fernández de Cossío ¹

¹ Centro de Ingeniería Genética y Biotecnología. Ave. 31 No. 15802 e/ 158 y 190, Cubanacán, Playa, PO Box 6162. pabloenmanuel99@gmail.com, luis.javier@cigb.edu.cu, satomy.pousa@cigb.edu.cu, jorge.cossio@cigb.edu.cu

* Autor para correspondencia: pabloenmanuel99@gmail.com

Resumen

La espectrometría de masas es una herramienta analítica clave en la caracterización de vacunas conjugadas, por lo que las autoridades reguladoras solicitan la identificación de sitios de conjugación como atributo de calidad, información que se obtiene mediante la identificación de los péptidos tipo 2 en las corridas de LC-MS/MS. Este proceso incluye una etapa de validación manual que consume un tiempo considerable y no está exenta de subjetividades. Por ello, la presente investigación tiene como objetivo desarrollar una herramienta para: (1) aumentar la confiabilidad en la asignación de los sitios de conjugación; y (2) obtener una huella dactilar del conjugado sintetizado. La herramienta se desarrolló empleando los lenguajes de programación Java y Python v3.11, mediante este último, se realiza el enlace con archivos de datos sin procesar de las corridas de LC-MS/MS mediante la librería pyMSFileReader. Además, Xion-C analiza los archivos con los péptidos tipo 2 identificados y extrae de la Corriente Iónica Total (TIC), la contribución de cada péptido, el Cromatograma Iónico Extraído (XIC), el cual contiene información valiosa que no se debe de desechar o ignorar. La herramienta superpone los XICs de los péptidos lineales y los de tipo 2 de manera independiente para obtener dos perfiles cromatográficos que pueden utilizarse para evaluar la reproducibilidad de la síntesis de los conjugados. En más del 85% de los casos, los XICs fueron útiles para respaldar inequívocamente las asignaciones. El software simplifica considerablemente el análisis, alivia los procesos de validación manual y aumenta la confiabilidad de este proceso.

Palabras clave: espectrometría de masas, péptidos tipo 2, sitios de conjugación, vacuna conjugada, XIC

Abstract

Mass spectrometry is a key analytical tool in the characterization of conjugate vaccines, and regulatory authorities request the identification of conjugation sites as a quality attribute, information that is obtained by identifying type 2 peptides in LC-MS/MS runs. This process includes a manual validation step that is time-consuming and not free of subjectivities. Therefore, the present research aims to develop a tool to: (1) increase the reliability of conjugation site assignment; and (2) obtain a fingerprint of the synthesized conjugate. The tool was developed using Java and Python v3.11 programming languages, the latter of which links to raw data files from LC-MS/MS runs using the pyMSFileReader library. In addition, Xion-C analyses the files with the identified type 2 peptides and extracts from the Total Ion Current (TIC), the contribution of each peptide, the Extracted Ion Chromatogram (XIC), which contains valuable information that should not be discarded or ignored. The tool overlays the XICs of linear and type 2 peptides independently to obtain two chromatographic profiles that can be used to assess the reproducibility of conjugate synthesis. In more than 85% of cases, XICs were useful to unequivocally support the assignments. Xion-C considerably simplifies the analysis, alleviates manual validation processes and increases the reliability of this process.

Keywords: mass spectrometry, type 2 peptides, conjugation sites, conjugate vaccine, XIC

Introducción

Las vacunas conjugadas son de gran importancia para la salud, tanto humana como animal, su estructura se encuentra enlazada de manera covalente mediante el empleo de reactivos bifuncionales de entrecruzamiento molecular, dos unidades estructurales diferentes (Hermanson, 2013), la proteína portadora y el antígeno. Las posiciones en las que se enlazan la proteína portadora y el antígeno se conocen como sitios de conjugación. Estos son considerados como atributos de calidad, y su determinación resulta imprescindible en la caracterización físico-química de estos productos.

La industria biotecnológica cubana cuenta con varios productos biotecnológicos basados en vacunas conjugadas que han tenido impacto en la salud humana, como, por ejemplo, CIMAvax-EGF® (Saavedra & Crombet, 2017), producida por el Centro de Inmunología Molecular (CIM), Soberana 02 (Toledo-Romaní, y otros, 2022) y QuimiHib (Fernández-Santana, et al., 2004) (Fernández-Santana, et al., 2004) (Díaz, et al., 2013), producidas el Instituto Finlay de Vacunas. Además, el Centro de Ingeniería Genética y Biotecnología (CIGB) cuenta con varios candidatos vacunales contra la infestación por garrapatas, basados en la conjugación química del péptido P0 a varias proteínas portadoras entre las que se encuentran la p64K de *Neisseria meningitidis* (González, y otros, 2021), la Bm86 de *Rhipicephalus microplus*

(Rodríguez Mallón, y otros, 2020) (Rodríguez Mallon, y otros, 2022) (Rodríguez-Mallon, The Bm86 Discovery: A Revolution in the Development of Anti-Tick Vaccines, 2023) y KLH-P0 de *Megathura crenulata* (Rodríguez-Mallon, y otros, 2012) (Rodríguez-Mallon, y otros, 2015).

La biotecnología es una industria muy regulada. Las entidades regulatorias para el desarrollo de producto biotecnológico bien caracterizado exigen una caracterización exhaustiva desde el punto de vista químico-físico y biológico. Las normas ICHQ6B, consensuadas entre las entidades regulatorias más importantes, pretenden establecer estos parámetros (European Medicines Agency, 1999).

Para la caracterización de las vacunas conjugadas se pueden emplear las mismas herramientas que se utilizan en el estudio de interacciones proteína-proteína mediante la combinación de las reacciones de entrecruzamiento molecular y la Espectrometría de Masas (XL-MS, del inglés cross-linking analyzed by mass spectrometry) (Leitner, y otros, 2010) (Chu, et al., 2004). Uno de los principales desafíos del empleo de esta estrategia es la complejidad de la mezcla peptídica, en la que se encuentran los péptidos entrecruzados tipo 2, los cuales contienen la información sobre los sitios de conjugación.

El grupo de proteómica del CIGB ha identificado los sitios de conjugación en varias vacunas conjugadas al usar softwares que en la proteómica funcional se usan para estudiar la interacción proteína-proteína mediante XL-MS. Entre las diversas herramientas bioinformáticas existentes (Yilmaz, y otros, 2016), el Kojak (Hoopmann, y otros, 2015) y el pLink2 (Yang, y otros, 2012), son las más utilizadas por esta entidad y, a su vez, unas de las más empleadas en distintos laboratorios de proteómica en el mundo (Iacobucci, y otros, 2019).

Una vez que se obtienen las identificaciones automáticas de péptidos lineales y péptidos entrecruzados tipo 2, es necesaria la validación manual de los resultados por los especialistas, siendo la inspección espectro-espectro clave en el proceso de identificación. No obstante, el procesamiento de los datos y su validación manual genera un cuello de botella, ya que el número de posibles combinaciones de enlace péptido-péptido crece cuadráticamente con respecto al número de proteínas (Yu, Li, & Yu, 2016). Por lo que, se hace necesario la utilización de métodos ortogonales para reafirmar las identificaciones, como es el caso del marcado con 18O, el empleo de entrecruzadores escindibles por MS y cualquier otra herramienta que aligere el trabajo manual.

Debido a esto, el equipo de proteómica del CIGB propuso emplear los Cromatogramas Iónicos Extraídos (XIC) (Smoluch & Piechura, 2019) de la Corriente Iónica Total (TIC), como método ortogonal, para evaluar la reproducibilidad en experimentos XL-MS, ya que encontraron que los péptidos entrecruzados tipo 2 presentan, por lo general, un patrón distinto al de los péptidos lineales, al emplear la química del maleimido-tiol para la síntesis de los conjugados, lo que evitaría la asignación incorrecta de un péptido lineal como tipo 2 y viceversa, lo que redundaría en una identificación más confiable de los sitios de conjugación

Sin embargo, adicionar al proceso de validación manual para obtener los XIC de cada péptido sería impracticable, ya que, hasta el momento, no se cuenta con un sistema de alto flujo que permita generar de forma automatizada los XICs de los péptidos identificados. Por ello, el grupo de proteómica del CIGB propuso desarrollar una herramienta capaz de obtener el XIC para cada péptido identificado mediante softwares de proteómica funcional y que, a su vez, no requiera de un costo computacional elevado y pueda ser empleado para evaluar la síntesis de conjugados proteicos, constituyendo el objetivo de la presente investigación. Esto simplificaría considerablemente el análisis de moléculas complejas, como es el caso de las vacunas conjugadas y aliviar los procesos manuales de validación a realizar por los especialistas.

Métodos o Metodologías Computacionales

Un área realmente importante y necesaria de la Bioinformática, es el desarrollo de herramientas computacionales con el fin de obtener resultados en investigaciones biológicas con mayor rapidez y precisión. Por ende, para llevar a cabo dicho proceso, es necesario el empleo de un enfoque de ingeniería de software y, de esta forma, mejorar el proceso de desarrollo del software bioinformático y la calidad de este.

Analizadas las características identificadas por Leyva y Febles (Leyva Vázquez & Febles Rodríguez, 2008), en los procesos de desarrollo de software en el campo de la Bioinformática, se puede deducir que estos procesos emplean un enfoque Ágil. Este proporciona un modelo basado en la estrecha colaboración del equipo involucrado en el proceso de desarrollo del software bioinformático, siendo las condiciones óptimas para un equipo pequeño multidisciplinario, que trabaje en aumentar la calidad del software que se desarrolla. Uno de los métodos que se

emplea para determinar el enfoque de ingeniería de software a utilizar en un proceso de desarrollo de software es la confección de la Estrella de Boehm y Turner (Boehm & Turner, 2013).

En la Figura 1 se puede observar la Estrella de Boehm y Turner aplicada a la presente investigación, en la que se reafirma el empleo de un enfoque ágil en el proceso de desarrollo de software.

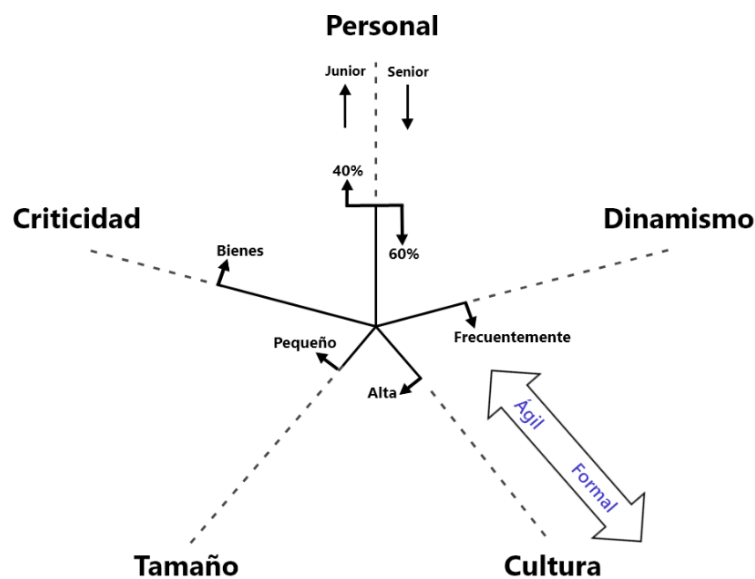


Figura 1. Estrella de Boehm y Turner

Identificado el enfoque, es posible determinar la metodología de Ingeniería de Software a utilizar, en este caso se empleó la metodología ágil, SCRUM, un proceso de gestión que reduce la complejidad en el desarrollo de softwares para satisfacer las necesidades de los clientes (Timkyw, Bournissen, & Tumino, 2020).

Herramientas y Tecnologías

Para el desarrollo del software Xion-C, se empleó NetBeans IDE (NetBeans, 2022) como el Entorno de Desarrollo Integrado, que desde 2007 se ofrece bajo licenciamiento dual de Licencia CDDL (Common Development and Distribution License) y la GPL versión 2. Además, se hizo uso de la aplicación cliente-servidor Jupyter Notebook, ya que se utilizaron dos lenguajes de programación, Java y Python v3.11, ambos lenguajes de primer nivel, bajo Licencia Pública General de GNU y licencia de código abierto Python Software Foundation License (Bisong, 2019), respectivamente.

Como Sistema de Bases de Datos, Xion-C implementa SQLite, uno de los más empleados en todo el mundo, el cual realiza operaciones de manera eficiente y es más rápido que MySQL (Combaudon, 2018) y PostgreSQL (Gilmore & Treat, 2006). Además, SQLite presenta una arquitectura relativamente simple, con una infraestructura conformada por cuatro partes principales: Núcleo, Compilador SQL, Backend y Accesorios (Bhosale, Patil, & Patil, 2015). El enlace con SQLite se establece mediante dos vías, una a través del paquete de Python `sqlite3`, el cual proporciona una interfaz SQL compatible con la especificación DB-API 2.0 descrita por PEP 249 (Python Enhancement Proposals, 2001) y requiere SQLite 3.7.15 o posterior. Por otro lado, mediante la API estandarizada de Conectividad de Bases de Datos Java (JDBC) (JDBC, s.f.), es posible su utilización una vez que se adiciona la biblioteca o paquete SQLite JDBC Driver al proyecto en Java.

Adicionalmente, el software hace uso de diferentes librerías científicas para el correcto funcionamiento de los algoritmos y métodos desarrollados:

- Numpy: paquete fundamental para la computación científica en Python (Numpy: El paquete fundamental para la computación científica con Python, 2022). Utilizado en gran medida, tanto para la creación y manejo de arreglos como para la realización de determinados cálculos.
- Matplotlib: biblioteca para la generación de gráficos a partir de datos contenidos en listas o arreglos en el lenguaje de programación Python y su extensión matemática NumPy (Matplotlib, 2022). Empleado en Xion-C en la visualización de los XICs, TICs y demás gráficos que la herramienta permite generar.
- Pandas: biblioteca de software de código abierto, escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python (Pandas, 2022). Utilizado en la lectura de archivos CSV o Excel provenientes de las salidas de softwares de proteómica funcional, los cuales contienen la información de los péptidos identificados.
- SciPy: proporciona algoritmos para optimización, integración, interpolación, problemas de valor propio, ecuaciones algebraicas, ecuaciones diferenciales, estadísticas y muchas otras clases (SciPy: Algoritmos fundamentales para la computación científica en Python, 2022). Se emplea en el proceso de determinación de picos mediante la interpolación de los datos y la obtención de máximos y mínimos locales, lo que permite la localización de los MS/MS realizados a cada péptido en el cromatograma.

- Scikit-learn: biblioteca para aprendizaje automático de software libre para el lenguaje de programación Python (Scikit-learn, s.f.). Scikit-learn se empleó para la determinación del área de los picos al utilizar la métrica AUC, para determinar el área bajo la curva ROC, mediante la regla trapezoidal.

Flujo de Trabajo

Con el fin de llevar a cabo el cumplimiento de los objetivos propuestos en la investigación, se desarrollaron e implementaron métodos que dieran solución a la problemática planteada. Desde una perspectiva general, el proceso de generación y visualización de XICs implementado, se puede dividir en una secuencia de pasos que lleva a cabo la plataforma, Figura 2.

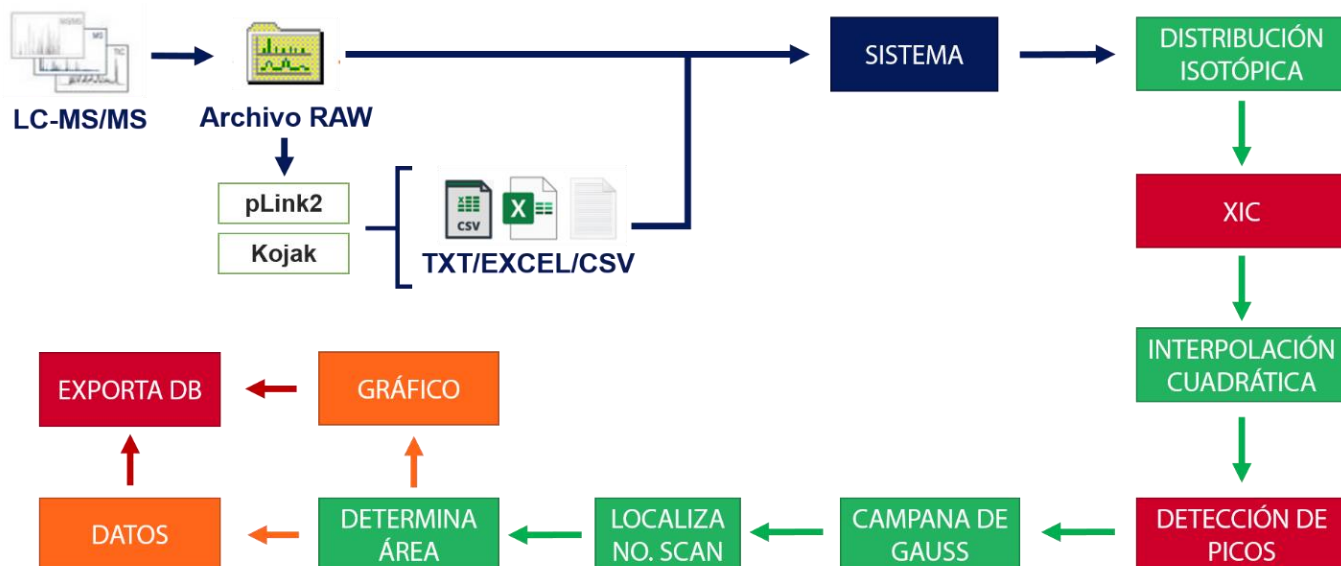


Figura 2. Flujo de trabajo general del proceso de extracción del XIC implementado en el sistema

Algoritmo de extracción del Cromatograma Iónico Extraído

Hasta el momento, las herramientas que se conocen que permiten extraer el XIC de un péptido determinado, generan el cromatograma a partir de un rango de m/z , e imposibilitan al usuario tener la capacidad de elegir varios intervalos de su distribución isotópica y de esta forma eliminar ruido (señales que no pertenecen al péptido que se analiza) en las señales que se obtienen. Esta limitante provoca que al extraer un XIC con una cierta distribución

isotópica y el usuario selecciona un intervalo en el que se contemplen, por ejemplo, los tres primeros picos de esta, se obtendrá el cromatograma para cada valor encontrado en ese intervalo, y se sumará, de forma tal que, al contener un rango amplio, el número de señales ruidosas aumenta. Esto se debe a que las señales que se encuentran entre cada uno de esos picos, y las que van más allá de estos en el intervalo, no pertenecen al péptido que se analiza y, por ende, se está teniendo en cuenta para la extracción del XIC señales “falsas” que provocan ruido en el resultado final.

Por tales motivos, la herramienta desarrollada implementa una nueva estrategia para generar los XICs, de forma tal, que aumenta la confiabilidad y el valor analítico de estos. Este método tiene en cuenta la resolución de los picos de la distribución de iones isotópicos para generar el XIC para cada uno de estos, Figura 3.

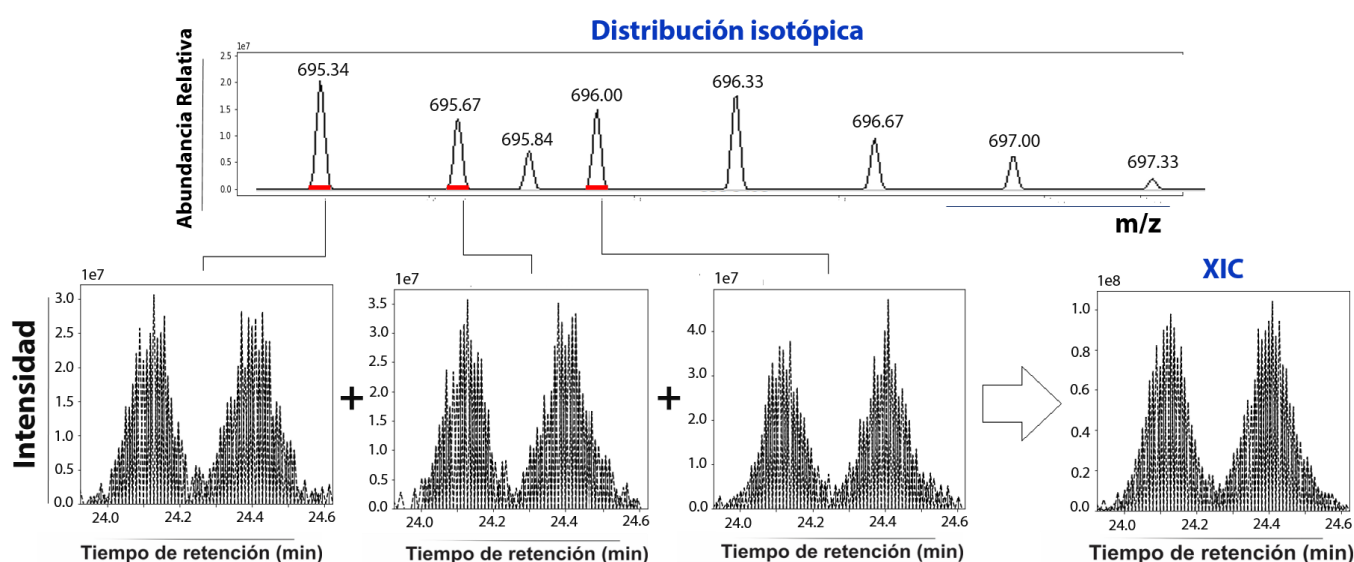


Figura 3. Método de extracción del XIC

De modo general, como se puede observar en la Figura 3, si un usuario desea obtener el XIC de un péptido con un estado de carga específico, el sistema lo primero que realiza es determinar la distribución isotópica del péptido mediante la m/z . Una vez obtenida la distribución isotópica, obtiene la resolución (ancho) del pico monoisotópico, genera el XIC para ese pico tomando como rango su resolución y en caso de que el usuario haya decidido generar el cromatograma teniendo en cuenta el segundo o el tercer pico de la distribución isotópica, entonces se generaría el

XIC para el segundo pico y luego para el tercero. Obtenidos los XICs, se suman y de esta forma se obtiene el cromatograma iónico extraído del péptido que se analiza.

Esto disminuiría la alta probabilidad de obtener señales no deseadas, el tiempo de respuesta y aumentaría la selectividad del péptido, ya que Xion-C genera el XIC teniendo en cuenta solo los picos de la distribución isotópica que el usuario seleccione, desechando intervalos no pertenecientes al péptido, siendo estos de, por lo general, como máximo, 0.06 Da.

Interpolación cuadrática

Con el objetivo de obtener una curva que se aproxime a todas las señales detectadas en el XIC, se lleva a cabo una interpolación mediante splines de los datos, lo cual no es más que la búsqueda de una función que pasa por todos los puntos en un intervalo dado. Por tal motivo, para llevar a cabo la interpolación de las señales obtenidas, se utilizan puntos de muestreo cercanos con el objetivo de que no se afecte la interpolación al contener varios picos cromatográficos. Para ello, primero se obtienen números espaciados uniformemente en un intervalo específico, esto se realiza mediante la función “linspace” de la librería Numpy; luego se emplea la función de la librería SciPy, “interpolate.interp1d”, la cual posibilita crear una aproximación a una función f , en este caso, cuadrática, utilizando los valores del eje X y del eje Y, correspondientes con el TR y la intensidad de las señales, respectivamente.

Una vez obtenida la función f , el sistema procede a evaluar cada uno de los números espaciados uniformemente en dicha función y , de esta forma, lograr una curva que se aproxime a las señales cromatográficas del XIC. Dicha aproximación, constituye el punto de partida para detectar los picos cromatográficos.

Método para la detección de picos y cálculo del área

Realizada la interpolación de los datos, el sistema procede a determinar los máximos y los mínimos locales a lo largo de toda la curva generada empleando la función “argrextrema” del paquete “signal” de la librería Scipy. Posteriormente, se obtienen los tiempos de retención de los espectros MS/MS realizados por el espectrómetro de masas al péptido que se está evaluando en ese instante de tiempo, los cuales se ubican en la curva y se identifican los mínimos locales entre los que se encuentran.

Una vez obtenido cada intervalo de tiempo, en los que se encuentran los MS/MS, se procede a generar la Campana de Gauss por cada intervalo detectado, evaluando la función a lo largo del eje X, mediante la Ecuación [1].

$$f(x) = a * e^{-\frac{(x-b)^2}{2\sigma^2}} \quad \text{Ecuación [1]}$$

Donde “a” es el valor de la intensidad máxima en el intervalo, “b” el centro del intervalo en correspondencia al TR de la intensidad máxima, “x” el valor en el eje X (TR), y “σ” la desviación típica del intervalo.

Una vez, obtenida la Campana de Gauss, se ubican en esta los números de escaneo de los espectros MS/MS al evaluar la función de la Ecuación [1] con el tiempo de retención de cada uno de ellos.

La herramienta desarrollada no se limita a solo generar el cromatograma, determinar los picos y ubicar los espectros MS/MS, sino que, además, calcula el área del pico detectado. Para ello, se emplea la función de la librería científica Scikit-learn, “sklearn.metrics.auc”, la cual calcula el área bajo la curva (AUC) de las señales experimentales en el intervalo de la gaussiana generada, utilizando la Regla del Trapecio usando intervalos múltiples.

Regla del Trapecio usando intervalos múltiples

Sea f una función continua en $[a, b]$. Si se considera que sobre $[a, b]$ hay $n + 1$ puntos igualmente espaciados, $\{x_0, x_1, \dots, x_n\}$, entonces hay n segmentos de longitud $h = \frac{b-a}{n}$. Si $\{x_i = x_{i-1} + h, \text{ con } i = 1 \dots n - 1,$
 $a = x_0 \text{ y } b = x_n,$ entonces:

$$I = \int_a^b f(x)dx = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \dots + \int_{x_{n-1}}^{x_n} f(x)dx \quad \text{Ecuación [2]}$$

Como todo método numérico, la regla del trapecio para intervalos múltiples presenta un error asociado en cada uno de los subintervalos:

$$E = -\frac{h^3}{12}f''(\xi_1) - \frac{h^3}{12}f''(\xi_2) - \dots - \frac{h^3}{12}f''(\xi_n)$$

con $a = x_0 < \xi_1 < x_1 < \xi_2 < x_2 < \dots < x_{n-1} < \xi_n < x_n = b$. Por ende, el error de integración de la regla del trapecio para intervalos múltiples está dado por la Ecuación [3].

$$E \leq \left[\frac{(b-a)h^2}{12} f''(\xi) \right] \quad \text{Ecuación [3]}$$

Además, empleando la métrica AUC, y calculando el área bajo la curva ROC (del inglés, receiver operating characteristic curve), se puede determinar qué tan ajustado está la gaussiana generada en pasos anteriores, con respecto al área de las señales experimentales, así de esta forma obtener una medida de confiabilidad en el resultado.

Resultados y discusión

Implementada la herramienta de software de la presente investigación, es posible estudiar los datos resultantes de un análisis de LC-MS/MS mediante la lectura del archivo RAW. Este archivo provee toda la información recopilada por el espectrómetro de masas, la cual puede ser leída y analizada por Xion-C. Además, permite la entrada al sistema de los archivos de texto, csv y excel provenientes de los softwares de proteómica funcional Kojak y pLink2. En ambos casos es posible determinar las modificaciones postraduccionales (PTMs) detectadas en los péptidos a pesar de la diferencia en los formatos de salida.

Con el objetivo de emplear la herramienta para la extracción de los XICs como método ortogonal para la validación de péptidos entrecruzados tipo 2 y evaluar la reproducibilidad en los conjugados sintetizados mediante la química del maleimido-tiol, se analizaron mediante esta, los conjugados p64K-Cys¹pP0 y Bm86-Cys¹pP0.

Evaluación del método de extracción de XICs

Al analizarse los resultados obtenidos mediante el método desarrollado para la extracción de los cromatogramas, se corroboró la hipótesis de que al emplear esta nueva estrategia se eliminan las señales no pertenecientes al péptido y se aumenta la confiabilidad y la selectividad en el XIC.

En la Figura 4 es posible observar 2 ejemplos en los que el cromatogramas obtuvieron mejores resultados al emplear el método desarrollado, ubicado en la parte superior de la figura.

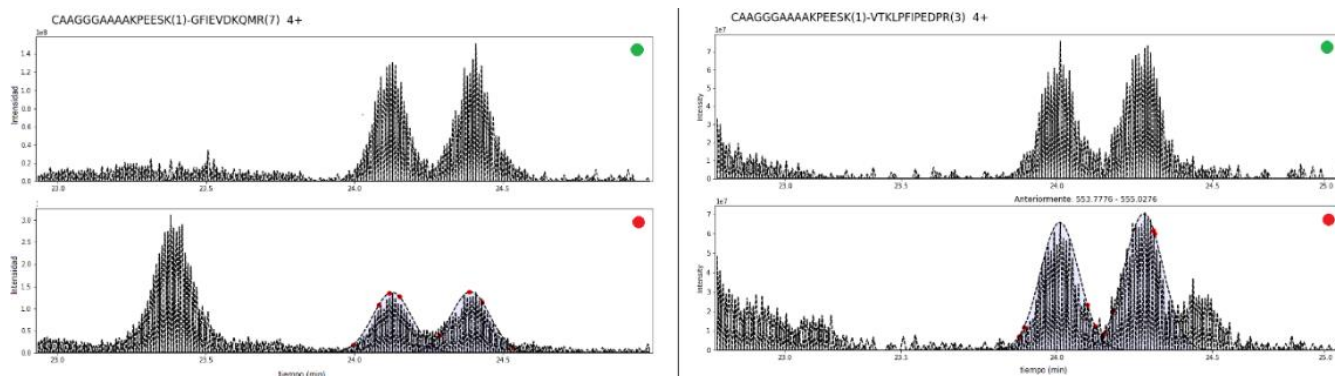


Figura 4. Comparación de los métodos de extracción empleado por softwares existentes (Punto rojo) con el método desarrollado en la presente investigación (Punto verde)

El 95 % de los casos, se comportó como lo previsto, el otro 5 % no presentó diferencias notables. Estos resultados demuestran que, al extraer el XIC utilizando la suma de intervalos más pequeños, pertenecientes a los picos de la distribución de iones isotópicos, se aumenta la selectividad en el proceso de extracción, lo que conlleva al aumento del valor analítico del XIC.

Integración con pLink2

Una de las características esenciales que posee la herramienta desarrollada, es la integración con la base de datos del software de proteómica funcional pLink2. Mediante esta, es posible obtener los datos de todas las modificaciones postraduccionales y de todos los reactivos de entrecruzamiento en las bases de datos del pLink2, lo que reduce el costo en el tiempo, ya que no es necesario insertar manualmente cada una de las modificaciones existentes en la base de datos del software desarrollado.

Por otro lado, esto disminuye considerablemente la aparición de errores durante el análisis de los péptidos identificados, ya que la base de datos se encontrará enriquecida de toda la información necesaria para completar el análisis.

Generación de los XICs de péptidos entrecruzados y lineales del conjugado p64k-Cys¹pP0

Con el propósito de implementar un procedimiento que permita evaluar la complejidad del conjugado p64K-Cys¹pP0 sintetizado y evaluar la reproducibilidad de la síntesis, se obtuvieron los XICs de los péptidos lineales y los entrecruzados tipo 2. Durante la implementación de este proceso, resultó atractivo observar que al superponer los

XICs de los péptidos lineales y los de tipo 2, se obtuvieron cromatogramas con características distintas. Los péptidos tipo 2 presentan mayor cantidad de fracciones (Figura 5), en los que más del 90% de los XICs generados en el estudio eluyeron en dos fracciones contiguas, mientras que los lineales solo lo hicieron en una fracción.

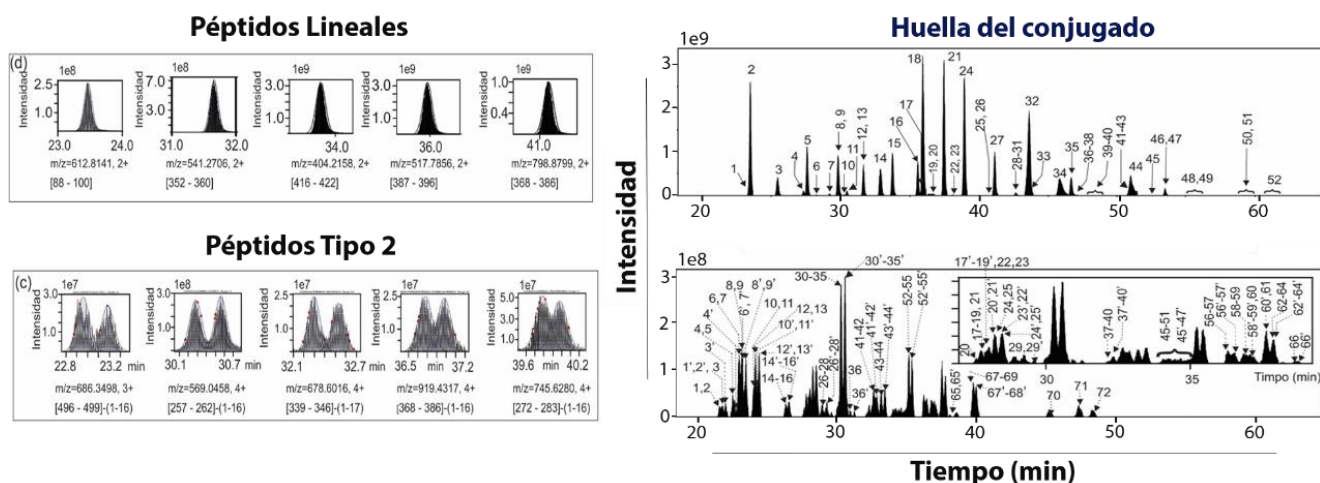


Figura 5. Superposición de todas las XIC del conjugado p64K-Cys1pP0, correspondientes al total de péptidos lineales (a) y péptidos tipo 2 (b). (c) y (d) XICs de diferentes péptidos entrecruzados tipo 2 y lineales, respectivamente

Además, la herramienta permite comparar dos o más conjugados al importar más de un archivo RAW y, de esta forma, generar un gráfico que muestre el comportamiento de la proporción entre las áreas, de aquellos péptidos que eluyeron en dos fracciones. En la Figura 6 es posible observar dicha comparación entre el conjugado p64K-Cys¹pP0 digerido con Tripsina y p64K-Cys¹pP0 digerido con LEP+Tripsina+V8.

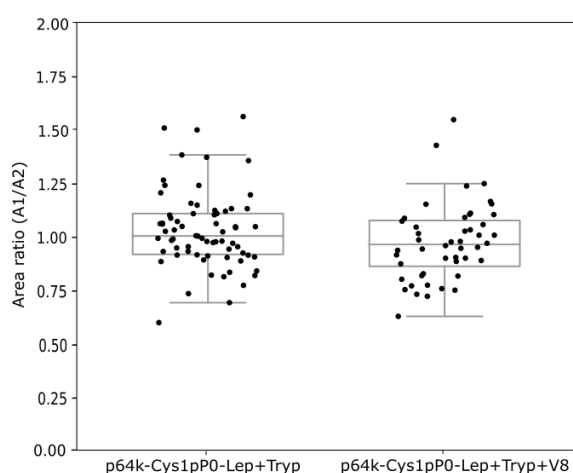


Figura 6. Proporción entre las áreas de las dos fracciones de los péptidos tipo 2 de las digestiones del conjugado p64K-Cys¹pP0

Generación de los XICs de péptidos entrecruzados del conjugado Bm86-Cys1pP0

La resolución lograda por la separación cromatográfica en el análisis LC-MS/MS es crucial en el valor analítico de las XIC. Por ende, tener una noción de qué tipo de digestión realizar, con qué tamaño de gradiente, si se emplean o no contraiones resolutivos o columnas cromatográficas de mayor tamaño, contribuye a resultados notables en menor tiempo.

Mediante la herramienta se analizó el conjugado Bm86-Cys¹pP0 eluido en 60 y 120 min, con el fin de comparar mediante el software los resultados en ambos casos y observar el comportamiento de la diferencia en la separación cromatográfica de aquellos XICs que eluyeron en dos fracciones contiguas.

En la Figura 7, se puede observar el análisis realizado mediante la herramienta desarrollada, en el que la (Figura 7a) representa la diferencia entre los tiempos de retención (Δt) de las fracciones contiguas de los péptidos tipo 2, en tiempos de elución de 60 y 120 min. Además, se muestran 8 XICs pertenecientes a 4 péptidos entrecruzados tipo 2 identificados, 4 en un tiempo de elución de 120 min (Parte superior) y 4 en un tiempo de elución de 60 min (Parte inferior), ubicados por pares de forma vertical.

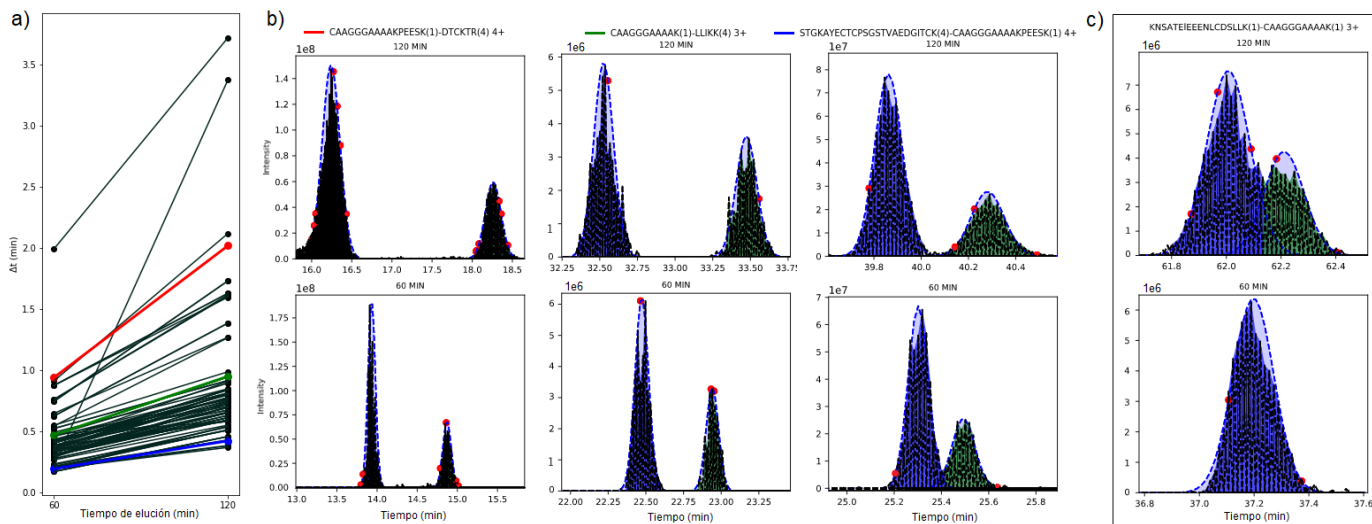


Figura 7. Relación entre la separación de las fracciones contiguas de los péptidos tipo 2 (eje y) correspondientes a los tiempos de elución de 60 y 120 min (eje x) (a). Comparación entre de los XICs resultantes del análisis del conjugado Bm86-Cys1pP0 en los diferentes tiempos de elución (b) y (c)

Es posible observar utilizando el software Xion-C cómo empleando el gradiente de 120 min aumenta la resolución para ambos picos y la separación cromatográfica. Además, de que se logran separar péptidos que en un gradiente de 60 min eluyeron en una sola fracción (Figura 7c). Al realizar esta comparación mediante el software, permite al especialista sacar conclusiones y, en este caso, demostrar la utilidad de emplear un gradiente más largos, no obstante, es necesario analizar otros puntos de vistas como es el caso de la identificación de sitios de conjugación, los cuales en ambos casos fue similar, no presentaron diferencias significativas.

Con las evidencias presentadas es posible apreciar el cumplimiento de los objetivos propuestos en la presente investigación, la cual no se detiene en estos resultados, ya que se continúa trabajando en la implementación de nuevas funcionalidades y en el desarrollo de mejoras para las ya implementadas, así como en la integración con otras herramientas existentes.

Conclusiones

1. El software Xion-C, simplifica considerablemente el análisis de moléculas complejas, como es el caso de las vacunas conjugadas y alivia los procesos manuales de validación a realizar por los especialistas. En más del 85% de los casos, los XICs fueron útiles para respaldar inequívocamente las asignaciones. demostrando que las XIC de los péptidos tipo 2 pueden usarse convenientemente como un elemento de validación ortogonal a la inspección manual y a la detección de iones diagnósticos en los espectros MS/MS, y así asignar de manera más confiable los sitios de conjugación.
2. La herramienta permite superponer los XICs de los péptidos lineales y los de tipo 2 de forma independiente con el fin de obtener una huella dactilar del conjugado. Los péptidos lineales presentan una sola fracción en concordancia con lo esperado. Sin embargo, la mayoría de los péptidos tipo 2 eluyeron en dos fracciones contiguas, separadas con una resolución variable lograda por el sistema cromatográfico de fase reversa.
3. El empleo de la nueva estrategia desarrollada para extraer los XICs aumenta la confiabilidad de las señales y la selectividad del péptido, lo que conlleva a un incremento del valor analítico de los XICs, comparado con los softwares existentes.

Referencias

- Bhosale, S., Patil, M., & Patil, P. (2015, Abril). International Journal of Computer Science and Mobile Computing SQLite: Light Database System. *International Journal of Computer Science and Mobile Computing*, 44, 882-885.
- Bisong, E. (2019). Python. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners* (pp. 71-89). Berkeley, CA. doi:10.1007/978-1-4842-4470-8_9
- Boehm, B., & Turner, R. (2013, Julio). Balancing Agility and Discipline-A Guide for the Perplexed.
- Chu, F., Shan, S.-o., Moustakas, D., Alber, F., Egea, P., Stroud, R., . . . Burlingame, A. (2004, Diciembre). Unraveling the interface of signal recognition particle and its receptor by using chemical cross-linking and tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 16454-16459. doi:10.1073/pnas.0407456101
- Combaudon, S. (2018). *MySQL 5.7 Administración y optimización*. Barcelona.
- Díaz, B., Heynngnezz, L., Beldarraín, A., Iser, Y., Fernández, A., Díaz, J., . . . Sosa, I. (2013, Enero). Adsorption chromatography to purify spacer disaccharide of active pharmaceutical ingredient of QuimiHib®vaccine. *VacciMonitor*, 22, 29-34.
- European Medicines Agency. (1999, Septiembre). *ICH Topic Q6B: Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products*. Retrieved from European Medicines Agency: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-6-b-test-procedures-acceptance-criteria-biotechnological/biological-products-step-5_en.pdf
- Fernández-Santana, V., Cardoso, F., Rodríguez, A., Carmenate, T., Peña, L., Valdés, Y., . . . Verez-Bencomo, V. (2004, Diciembre). Antigenicity and immunogenicity of a synthetic oligosaccharide-protein conjugate vaccine against *Haemophilus influenzae* type b. *Infection and immunity*, 72(12), 7115-23. doi:10.1128/IAI.72.12.7115-7123.2004
- Fernández-Santana, V., Hardy, E., Toledo, M. E., Rodríguez, M. C., Heynngnezz, L., Rodríguez, A., . . . Roy, R. (2004, Julio 23). A synthetic conjugate polysaccharide vaccine against *Haemophilus influenzae* type B. *Science*, 305(5683), 522-5. doi:10.1126/science.1095209
- Gilmore, W. J., & Treat, R. H. (2006). Introducing PostgreSQL. In *Beginning PHP and PostgreSQL 8: From Novice to Professional* (pp. 573-577). doi:10.1007/978-1-4302-0136-6_24
- González, L., Encinosa, P., Machado, W., Pousa, S., Leyva, A., Cano-Argüelles, A. L., . . . Rodríguez-Mallon, A. (2021, Julio). Synthesis, LC-MS/MS analysis, and biological evaluation of two vaccine candidates against ticks based on the antigenic PO peptide from *R. sanguineus* linked to the p64K carrier protein from *Neisseria meningitidis*. *Analytical and Bioanalytical Chemistry*. doi:10.1007/s00216-021-03569-0

- Hermanson, G. T. (2013). Chapter 19 - Vaccines and Immunogen Conjugates. In *Bioconjugate Techniques* (Third ed., pp. 839-865). Boston. doi:doi.org/10.1016/B978-0-12-382239-0.00019-4
- Hoopmann, M., Zelter, A., Johnson, R., Riffle, M., Maccoss, M., Davis, T., & Moritz, R. (2015, Marzo). Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes. *Journal of proteome research*, 14. doi:10.1021/pr501321h
- (n.d.). Retrieved from Scikit-learn: <https://scikit-learn.org/stable/index.html>
- (2022). Retrieved from Matplotlib: <https://matplotlib.org/>
- Iacobucci, C., Piotrowski, C., Aebersold, R., Amaral, B. C., Andrews, P., Bernfur, K., . . . Sinz, A. (2019). First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. *Analytical Chemistry*, 91(11), 6953-6961. doi:10.1021/acs.analchem.9b0065
- JDBC. (n.d.). Retrieved from Oracle: <https://docs.oracle.com/javase/tutorial/jdbc/>
- Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., & Aebersold, R. (2010, Marzo). Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Molecular & cellular proteomics : MCP*, 9, 1634-1649. doi:10.1074/mcp.R000001-MCP201
- Leyva Vázquez, M. Y., & Febles Rodríguez, J. P. (2008, Enero). Métodos Ágiles en el Desarrollo de Software para la Bioinformática. *Serie Científica de la Universidad de las Ciencias Informáticas*, 1.
- NetBeans. (2022). Retrieved from Wikipedia: <https://es.wikipedia.org/wiki/NetBeans>
- Numpy: *El paquete fundamental para la computación científica con Python*. (2022). Retrieved from Numpy: <https://numpy.org/>
- Pandas. (2022). Retrieved from <https://pandas.pydata.org/>
- Python Enhancement Proposals. (2001, Marzo 29). Retrieved from <https://peps.python.org/pep-0249/>
- Rodríguez Mallón, A., Javier González, L., Encinosa Guzmán, P. E., Bechara, G. H., Sanches, G. S., Pousa, S., . . . López Álvarez, J. (2020, Junio 25). Functional and Mass Spectrometric Evaluation of an Anti-Tick Antigen Based on the P0 Peptide Conjugated to Bm86 Protein. *Pathogens*, 9(6). doi:10.3390/pathogens9060513
- Rodríguez Mallon, A., Javier González, L., Encinosa Guzmán, P. E., Estrada García, M. P., Bello Soto, Y., Cabrales, A., . . . Rodríguez Fernández, R. (2022). El conjugado químico pP0-Bm86 como antígeno de una vacuna de amplio espectro contra garrapatas . *Anales de la Academia de Ciencias de Cuba*, 12. Retrieved from http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2304-01062022000200005&nrm=iso
- Rodríguez-Mallon, A. (2023). The Bm86 Discovery: A Revolution in the Development of Anti-Tick Vaccines. *Pathogens*, 12(231). doi:10.3390/pathogens12020231
- Rodríguez-Mallon, A., Encinosa, P. E., Méndez-Pérez, L., Bello, Y., Rodríguez Fernández, R., Garay, H., . . . Estrada, M. P. (2015). High efficacy of a 20 amino acid peptide of the acidic ribosomal protein P0 against the cattle tick,

- Rhipicephalus microplus. *Ticks and Tick-borne Diseases*, 6(4), 530-537.
doi:doi.org/10.1016/j.ttbdis.2015.04.007
- Rodríguez-Mallon, A., Fernández, E., Encinosa, P. E., Bello, Y., Méndez-Pérez, L., Ruiz, L. C., . . . Estrada, M. P. (2012). A novel tick antigen shows high vaccine efficacy against the dog tick, *Rhipicephalus sanguineus*. *Vaccine*, 30(10), 1782-1789. doi:doi.org/10.1016/j.vaccine.2012.01.011
- Saavedra, D., & Crombet, T. (2017). CIMAvax-EGF: A New Therapeutic Vaccine for Advanced Non-Small Cell Lung Cancer Patients. *Frontiers in Immunology*, 8. doi:10.3389/fimmu.2017.00269
- SciPy: *Algoritmos fundamentales para la computación científica en Python*. (2022). Retrieved from SciPy: <https://scipy.org/>
- sd. (1231). s.
- Smoluch, M., & Piechura, K. (2019). Basic Definitions. In M. Smoluch, & K. Piechura, *Mass Spectrometry* (pp. 9-12). doi:10.1002/9781119377368.ch3
- Timkyw, N., Bournissen, J., & Tumino, M. (2020, Octubre). SCRUM como Herramienta Metodológica para el Aprendizaje de la Programación. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*. doi:10.24215/18509959.26.e9
- Toledo-Romaní, M., Verdecia-Sánchez, L., Rodríguez-González, M., Rodríguez-Noda, L., Valenzuela-Silva, C., Paredes-Moreno, B., . . . Fontanies Fernández, M. (2022). Safety and immunogenicity of anti-SARS CoV-2 vaccine SOBERANA 02 in homologous or heterologous scheme: Open label phase I and phase IIa clinical trials. *Vaccine*, 40, 4220-4230. doi:10.1016/j.vaccine.2022.05.082
- Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., Lin, J., Zhang, K., . . . Dong, M.-Q. (2012, Julio). Identification of cross-linked peptides from complex samples. *Nature methods*, 9, 904-906. doi:10.1038/nmeth.2099
- Yilmaz, Ş., Drepper, F., Hulstaert, N., Cernic, M., Gevaert, K., Economou, A., . . . Vandermarliere, E. (2016, Septiembre). Xilmass: A New Approach toward the Identification of Cross-Linked Peptides. *Analytical chemistry*, 88. doi:10.1021/acs.analchem.6b01585
- Yu, F., Li, N., & Yu, W. (2016, Mayo). ECL: An exhaustive search tool for the identification of cross-linked peptides using whole database. *BMC Bioinformatics*, 17. doi:10.1186/s12859-016-1073-y