

**USO Y ADAPTACIÓN DE HERRAMIENTAS BIOINFORMÁTICAS
PARA EL ANÁLISIS METABOLÓMICO DE ORINA BASADO EN
¹H-RMN**

Adrián Hernández Cacho

TRABAJO FINAL DE GRADO BIOTECNOLOGIA

Tutora académica: Dra. María Montserrat Poblet (montserrat.poblet@urv.cat)

En cooperació con: Biosfer Teslab S.L.

Supervisor/es: Sr. Daniel Rodríguez (drodriguez@biosferteslab.com)

16 de Julio de 2020

ÍNDICE

Datos del centro	3
Agradecimientos.....	4
Resumen.....	5
1. INTRODUCCIÓN.....	6
1.1. Metaboloma y metabolitos.....	6
1.1.1. Clasificación de metabolitos.....	6
1.1.2. Metabolómica.....	6
1.2.1. La orina como matriz.....	7
1.2.2. Metabolitos en orina.....	7
1.3. La RMN en metabolómica.....	8
1.3.1. Retos en el análisis de orina por RMN	10
2. HIPÓTESIS Y OBJETIVOS.....	11
3. METODOLOGÍA	11
3.1. Conjunto de muestras.....	11
3.2. Preparación de muestras.....	11
3.3. Análisis RMN	12
4. RESULTADOS.....	12
4.1. Software para cuantificar de manera automática los espectros de orina.....	12
4.1.1. Identificación	12
4.1.2. Referenciado.....	18
4.1.3. Deconvolución	19
4.1.4. Cuantificación	23
4.1.5. Outputs del software	23
4.1.6. Resultado del análisis metabolómico.....	24
4.2. Normalización.....	26
4.3. Análisis estadístico.....	27
4.4. Comparación de los resultados con valores encontrados en la literatura	28
5. DISCUSIÓN	29
5.1. Trabajo futuro	31
6. CONCLUSIONES.....	31
BIBLIOGRAFÍA	32

Yo, Adrián Hernández Cacho, con DNI 39954270G. soy conocedor de la guía de prevención del plagio en la URV *prevenció, detecció i tractament del plagi en la docència: guia per a estudiants* (aprobada en julio 2017) (<http://www.urv.cat/ca/vida-campus/serveis/crai/que-us-oferim/formacio-competencies-nuclears/plagi/>) i afirmo que este TFG no constituye ninguna de las conductas consideradas como plagio para la URV.

Tarragona, 8 de Julio de 2020

A handwritten signature in blue ink, consisting of several overlapping, fluid strokes that form a stylized, somewhat abstract shape.

Datos del centro

Biosfer Teslab S.L. es una spin-off de la universidad Rovira i Virgili (URV) y el instituto de investigaciones sanitarias Pere Virgili. Tiene la sede en Reus y desarrolla su actividad en su oficina de la plaza Prim, 10 y el Center for Omic Sciences (COS). Fue fundada por la Dra. Núria Amigó y el Dr. Roger Mallol, responsables del desarrollo del test Liposcale, producto estrella de la compañía. El test permite caracterizar y cuantificar las diferentes subclases de lipoproteínas presentes en muestras de suero o plasma sanguíneo mediante resonancia magnética nuclear (RMN).

Biosfer Teslab apuesta por el desarrollo y la investigación, por ello, además del Test Liposcale, podemos encontrar en su cartera de servicios la caracterización de metabolitos de bajo peso molecular en suero intacto, test de glicoproteínas en suero intacto, lipidómica de suero, de cultivos celulares y de tejidos (tanto animales como humanos) y caracterización de metabolitos acuosos para cultivos celulares, medios de cultivo y tejidos/biopsias (tanto en animales como humanos). Todo ello gracias a su amplia experiencia en resonancia magnética nuclear y bioinformática. En Biosfer Teslab ofrecen un asesoramiento personalizado desde el diseño experimental hasta el análisis estadístico e interpretación de datos, diferenciados así de los servicios de análisis clásicos. Además, Biosfer cuenta con las certificaciones ISO 9001:2015 e ISO 13485:2016 de productos sanitarios

Agradecimientos

Primero, quiero agradecer a todo el equipo de Biosfer Teslab el increíble apoyo y aceptación que me han mostrado desde el principio. Habéis sido una gran fuente de sabiduría y profesionalidad. He aprendido más de lo que jamás pude imaginar antes de empezar la estancia de prácticas y me habéis hecho ver la investigación biomédica desde un punto de vista hasta ahora desconocido para mí. Núria, gracias por acogerme en tu empresa, es el mejor ambiente de trabajo en el que he estado. Tienes un equipo altamente cohesionado y estoy seguro de que es gracias a ti, eres una gran directora. Miriam y Enrique, me habéis aceptado como uno más desde el principio y habéis compartido vuestra experiencia conmigo, os lo agradezco de corazón. Daniel, has sido el mejor tutor que podía imaginar. Me has transmitido los conocimientos con mucha amabilidad y paciencia. Gracias por hacer que mi entrada en el mundo de la programación haya sido tan enriquecedora. Sara, has estado conmigo al pie del cañón desde el minuto 0 de este trabajo. Gracias a tus consejos y sugerencias he podido sacar todo el potencial que tiene este trabajo, te estoy muy agradecido por todo el esfuerzo que has puesto sobre mí.

A mi tutora académica, la Dra. Montserrat Poblet, por las correcciones y sugerencias realizadas.

A mi familia, mis padres, hermano, tíos y abuelos, tanto a los que están como a los que se fueron. Todos habéis puesto vuestro grano de arena para convertirme en la persona que hoy soy.

A mis amigos, por ayudarme a desconectar en los momentos de estrés. Gracias por todas las risas compartidas, y en especial, gracias a ti, Juan Luis, por estar siempre ahí.

Por último, pero no menos importante, quiero agradecer a mi pareja su apoyo incondicional a lo largo de este trayecto. Amor, sé que he pasado nuestro 6º aniversario pensando en este trabajo, pero los años anteriores ambos pensábamos en nuestros exámenes finales, así que, de alguna forma, estamos acostumbrados. De haberlo sabido me hubiese declarado en un mes menos ajetreado, leñe.

Resumen

La metabolómica es la ciencia que estudia los metabolitos presentes en las células, tejidos o incluso organismos completos. Conocer el estado metabólico de un sistema biológico nos permite saber si este funciona con plena normalidad o sufre alguna anomalía. Para conocer el metaboloma de un sujeto, es necesario poder acceder a este mediante una matriz biológica, como, por ejemplo, la orina. La orina ha sido siempre una importante fuente de información médica ante un posible estado patológico en diferentes órganos o tejidos de un sujeto, ya que, al ser un fluido residual, contiene una gran cantidad de metabolitos procedentes de diferentes vías metabólicas. Además, el proceso de recogida de muestras es sencillo y no invasivo, necesita menos pretratamiento en comparación con otros biofluidos, convirtiéndola en una matriz atractiva para los estudios metabolómicos. Mediante Resonancia Magnética Nuclear (RMN), podemos identificar y cuantificar los metabolitos presentes en la orina, siendo la Resonancia Magnética Nuclear de protón (^1H -RMN) una técnica robusta para dichos estudios debido a su alto rendimiento, reproducibilidad y fácil manipulación. Por otra parte, la cuantificación de metabolitos de bajo peso molecular por RMN es un proceso difícilmente automatizable para muestras de orina debido a su alta variabilidad interindividual.

En este trabajo se pretende automatizar la cuantificación de metabolitos de bajo peso molecular en muestras de orina mediante ^1H -RMN adaptando un software de Biosfer Teslab optimizado para llevar a cabo la misma función en muestras de sangre. Primero, se identificaron los metabolitos presente en las muestras de orina. Para ello, se usó el software TopSpin. Después, se estudió que referenciado es el idóneo para poder cuantificar automáticamente los metabolitos. Se comparan los métodos de referenciado global y referenciado local de señales, se determinó que este último es el idóneo para muestras de orina por ser más preciso y se referenciaron las señales de los espectros. Posteriormente, se deconvolucionaron las señales referenciadas adaptando el software de Biosfer Teslab para cada señal. Se obtuvieron las áreas de las señales y se cuantificó la concentración de los metabolitos.

PALABRAS CLAVE: RMN, METABOLÓMICA, AUTOMATIZACIÓN, BIOINFORMÁTICA

1. INTRODUCCIÓN

1.1. Metaboloma y metabolitos

Los metabolitos son moléculas, generalmente orgánicas, que participan en las reacciones químicas que se producen a nivel celular. Estos sirven como impronta de la actividad bioquímica de los especímenes y pueden ser correlacionados con el fenotipo. En este contexto, el metaboloma se define como el conjunto de metabolitos producidos por las células y ofrece una representación del flujo bioquímico presente en una muestra biológica. Además, aporta una gran cantidad de información útil en el diagnóstico clínico[1].

1.1.1. Clasificación de metabolitos

Los metabolitos se pueden clasificar según diferentes criterios. Principalmente, se suelen dividir en metabolitos primarios y secundarios[2].

Los metabolitos primarios son los que forman parte del metabolismo básico, tienen funciones determinadas y son esenciales para el crecimiento y desarrollo del organismo. Son ejemplos los glúcidos, las proteínas, los ácidos nucleicos y los ácidos grasos[2].

Los metabolitos secundarios, son compuestos orgánicos sintetizados por plantas, hongos y microbios y no forman parte del metabolismo esencial[2]. Aun así, aportan funciones relevantes para el organismo. Un ejemplo de metabolitos secundarios son los antibióticos. Estos no son estrictamente necesarios para el correcto crecimiento y desarrollo del organismo que los produce, pero aportan una ventaja evolutiva que favorece la competitividad del organismo que los produce.

Por otra parte, en el contexto biomédico podemos hablar de metabolitos endógenos y exógenos, ambos relevantes en la identificación de biomarcadores[3]. Los metabolitos endógenos, son aquellos producidos por el propio organismo. Los metabolitos exógenos, en cambio, son aquellos que no son producidos por el propio organismo y pueden ser introducidos mediante la alimentación, como es el caso de la cafeína, presente en el café, té y bebidas energéticas. Otros ejemplos de metabolitos exógenos son los medicamentos y drogas.

1.1.2. Metabolómica

La metabolómica es el estudio de los organismos a través de sus metabolitos[4]. Originalmente se propuso como un método de genómica funcional, pero su utilidad es mucho más extensa. La metabolómica se aplica en microorganismos, plantas y animales. Se usa para comparar mutantes, observar la respuesta de los organismos al estrés del entorno, estudiar los efectos de la manipulación genética en organismos, detección de biomarcadores de enfermedades, etc[4].

En metabolómica, se puede estudiar un conjunto de metabolitos determinados en una muestra (metabolómica dirigida), o se puede usar para determinar todos los metabolitos presentes detectables por la técnica en una muestra (metabolómica no dirigida), como es el caso de este trabajo.

El perfilado de metabolitos es la rama de la metabolómica en la que se cuantifican los metabolitos presentes en una muestra. Es una técnica utilizada ampliamente en medicina para conocer el estado de salud de los pacientes y en investigación para poder detectar con mayor precisión y/o antelación enfermedades en las que haya una manifestación a nivel metabólico de la patología[4].

La metabolómica se puede estudiar a partir de fluidos biológicos, muestras tisulares y cultivos celulares, entre otros. Las técnicas más utilizadas en el perfilado metabólico son la resonancia magnética nuclear (RMN), normalmente de protón, y la espectroscopía de masas (MS), mediante cromatografía de gases o cromatografía líquida en función de los metabolitos que se quieren estudiar.

1.2. Análisis metabolómico de orina

1.2.1. La orina como matriz

Los análisis de orina han sido siempre una fuente de información médica importante. Mucho antes de la aparición de las técnicas analíticas modernas, el color, olor e incluso el gusto de la orina eran usados para diagnosticar diferentes patologías, desde enfermedades renales hasta la diabetes[5]. La orina nos sirve como fuente de información de la funcionalidad de otros órganos a nivel metabólico, convirtiéndola en una matriz rica para las investigaciones biológicas[6].

El proceso de recogida de muestras de orina es sencillo y no invasivo. Además, necesita menos pretratamiento en comparación con otros biofluidos debido a su fácil manipulación y baja concentración de proteínas. Todo ello la convierte en una matriz atractiva para los estudios metabolómicos[7].

1.2.2. Metabolitos en orina

La orina, al ser un fluido residual, contiene una gran cantidad de metabolitos procedentes de diferentes vías metabólicas (metabolitos endógenos), además de compuestos residuales procedentes de la alimentación y la ingesta de fármacos (metabolitos exógenos)[8]. Usando como referencia la Human Metabolome Database (HMDB) [9], se pueden encontrar más de 100 metabolitos en orina. En la figura 1 se mencionan los más comunes, clasificados según su clase u origen.



Figura 1: Metabolitos más comunes en orina ordenados por categorías

1.3. La RMN en metabolómica

La espectroscopia de resonancia magnética nuclear permite estudiar núcleos atómicos con un número impar de protones o neutrones, ya que estos poseen espín y son magnéticamente activos[10]. Al aplicar un campo magnético externo al núcleo se generan dos estados energéticos, α i β , con un ligero exceso de población en el estado de energía inferior α . Una vez establecidos esos dos niveles de energía para el protón, es posible introducir energía en forma de radiación de radiofrecuencia que provocará una transición entre los niveles de energía α i β [10]. La diferencia energética entre los niveles de energía dependerá de la fuerza del campo aplicado B_0 . Una vez el sistema se encuentra en resonancia el protón absorbe la energía y su estado energético aumenta, permitiendo la obtención del espectro[10].

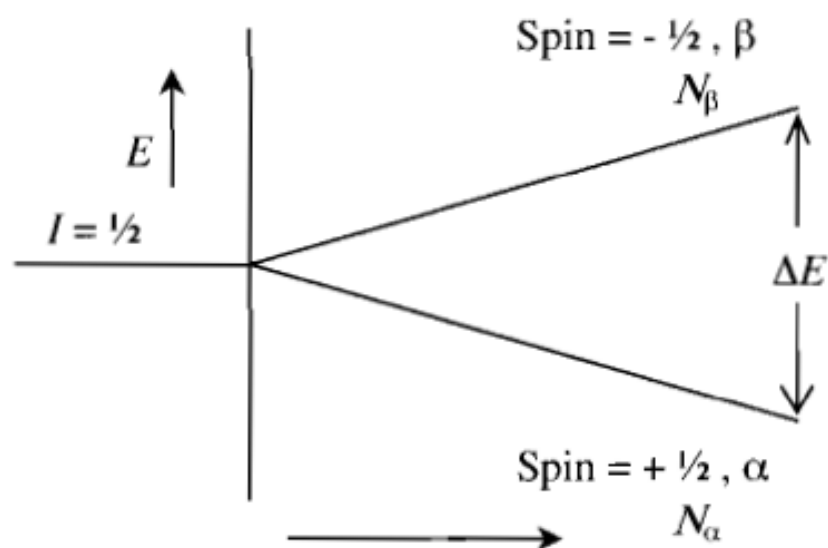


Figura 1: Niveles de energía de un protón en un campo magnético de magnitud B_0 . N es la cantidad de spins en el nivel superior (N_β) e inferior (N_α). A mayor intensidad del campo magnético, mayor es la variación de energía.

Como se menciona anteriormente, la RMN es, juntamente con la MS, una de las dos técnicas por excelencia utilizadas en metabolómica. Como técnica analítica, la RMN es una herramienta muy potente, tanto para la identificación como para la cuantificación de metabolitos. Aunque existan diferentes variantes de la tecnología RMN, la ^1H -RMN se ha convertido en el estándar a la hora de analizar estudios metabolómicos basados en resonancia magnética debido al importante número de ventajas que aporta, tales como su alta reproducibilidad, alto rendimiento, es una técnica no destructiva y no requiere una manipulación o preparación de muestras exhaustiva[11]. Aun así, la principal desventaja de la RMN es la superposición de las señales, ya que el análisis de matrices complejas como la orina genera muchas señales en el espectro y muchas veces se encuentran solapadas, provocando incertezas en la identificación y/o cuantificación de metabolitos[11]. Además, una misma molécula puede resultar en diferentes señales a lo largo del espectro, debido a que en el espectro de resonancia vemos partes de la estructura molecular, debido a que estamos identificando sus protones y no la molécula como tal (Fig. 3). Se obtienen tantas señales como protones químicamente distintos tenga la molécula, y por tanto, cada molécula tendrá un patrón característico de señales en un espectro de resonancia.

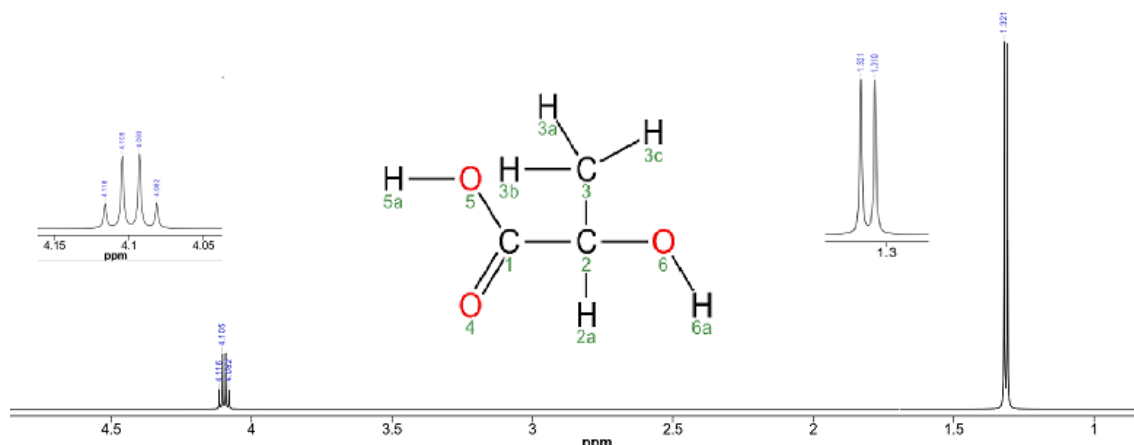


Figura 2: Espectro de ^1H -RMN del ácido láctico y molécula de ácido láctico con los átomos numerados. La señal de doblete en 1.3ppm es generada por los protones 3a, 3b y 3c al acoplarse al protón 2a. El la señal de multiplete en 4.1ppm es generada por el proton 2a al acoplarse a los protones 3a, 3b y 3c.

1.3.1. Retos en el análisis de orina por RMN

Uno de los principales retos en el estudio metabolómico de muestras de orina es su alta variabilidad interindividual. La composición de la orina se ve afectada por factores tales como la dieta, el ejercicio, el estado patológico del sujeto, su edad y género, microbiota y genética entre otros. Esta variabilidad hace difícil encontrar herramientas que se puedan automatizar para extraer información de los espectros de manera rápida y robusta. A este hecho debemos añadir la problemática del solapamiento de señales en los espectros de ^1H -RMN (ver figuras 6, 7 y 8) que dificultan la cuantificación de los metabolitos. Es por ello por lo que se necesitan herramientas matemáticas que puedan extraer la información cuantitativa de los espectros de RMN

Otro de los grandes retos de la metabolómica en orina es la normalización de los resultados para poder comparar unas muestras con otras para evitar así los efectos de dilución de la muestra. La forma más común de normalizar consiste en usar como referencia un metabolito característico de la matriz, normalmente la creatinina, aunque también puede utilizarse la urea. Aun así, se ha observado como la normalización mediante señales que son metabolitos que pueden verse alterados por patologías concretas, como es el caso de la creatinina en sujetos con problemas renales, puede llevarnos a resultados erróneos. Esto es debido a que, si se normaliza relacionando el área de la señal de un determinado metabolito con el área de la señal de la creatinina, por ejemplo, se obtiene la concentración del metabolito determinado en unidades de Concentración de metabolito/Concentración de Creatinina[12]. Por tanto, un sujeto con problemas renales donde la concentración de creatinina se vea alterada, puede obtener resultados de análisis que no se ajusten a la realidad.

Como alternativa a esta limitación, se ha propuesto el método de normalización por cociente probabilístico (PQN) [12], en el cual el espectro se normaliza en base a la dilución más probable. Esta dilución más probable se estima analizando la distribución de los cocientes de las intensidades del espectro respecto las intensidades de un espectro de referencia. Como espectro de referencia, se suele usar el espectro promedio de los controles[12].

2. HIPÓTESIS Y OBJETIVOS

La hipótesis del trabajo es la siguiente: Se puede automatizar la caracterización de metabolitos de bajo peso molecular en muestras de orina por resonancia magnética nuclear.

El presente trabajo tiene como objetivo abordar el principal reto del análisis metabolómico en muestras de orina:

- 1) Realizar un perfilado metabolómico de muestras de orina por RMN de manera automática. Para ello se han definido una serie de etapas a completar:
 - a. Identificación de metabolitos
 - b. Deconvolución de las señales
 - c. Cuantificación absoluta de los metabolitos

Como objetivo secundario:

- 2) Comparar los resultados obtenidos con resultados encontrados en literatura

3. METODOLOGÍA

3.1. Conjunto de muestras

Para el análisis se utilizará un conjunto de 116 muestras de las cuales 57 son control y 59 pertenecen a individuos que presentan una patología. Las muestras pertenecen a un proyecto de investigación en curso y, por lo tanto, no ha sido publicado todavía y es por este motivo por el cual no se pueden dar más detalles ni de la patología ni de la recogida de muestras

3.2. Preparación de muestras

Las muestras de orina se preparan y analizan usando protocolos optimizados y estandarizados[13]. En resumen, las muestras de orina son descongeladas a temperatura ambiente antes de mezclar 540µL de orina con 60µL de tampón (KH₂PO₄, 1.5 M, pH 7.4 en H₂O₂) con 5,8mM TSP y 2mM NaN₃. Se centrifugan las muestras (12000g, 4°C, 5min) para precipitar proteínas en caso de que las haya y se transfieren 550µL de muestra a tubos de RMN

de 5mm. Dichas muestras se incorporan en el robot SampleJet de Bruker que permite la automatización de más de 500 muestras y las mantiene a 4°C.

3.3. Análisis RMN

Los experimentos RMN se llevan a cabo de forma automática a 300K en un espectrómetro Bruker Advance III 600 MHz. El pulso que se utilizó para la adquisición de los espectros fue un NOESY y las muestras fueron adquiridas en 32 scans.

La identificación de metabolitos se realizó usando ChenomX NMR Suite 8.5[14] y literatura[15,16]. La concentración de metabolitos (en mM) se calculó usando la fórmula de Serkova[17]. Además, los metabolitos identificados se verificaron en la HMDB[9].

4. RESULTADOS

4.1. Software para cuantificar de manera automática los espectros de orina

Como se comentó en el apartado 1.3.1, uno de los principales retos de la metabolómica es la cuantificación de metabolitos en muestras de orina. Además, la gran variabilidad que presentan las muestras de orina dificulta la automatización del proceso de cuantificación.

En este trabajo se ha abordado este reto y se ha adaptado un software que Biosfer Teslab usaba para muestras de sangre al análisis de muestras de orina. Para ello se han diseñado una serie de etapas que se explican a continuación:

4.1.1. Identificación

Para poder cuantificar los metabolitos presentes en un conjunto de muestras, primero es necesario identificar que metabolitos están presentes en el conjunto. A continuación, se detallan los pasos que se han seguido para poder identificar los metabolitos presentes en el muestrario.

4.1.1.1. Elección del espectro idóneo mediante TopSpin para generar el “pattern” de metabolitos.

Con el fin de poder identificar la mayor cantidad de metabolitos presentes en un conjunto de muestras, es necesario escoger un espectro que cumpla las siguientes características:

- Se escogerá el espectro con el mayor número de señales
- Se escogerá el espectro con mejor intensidad de señal

En ocasiones, una muestra puede presentar señales únicas no presentes en el resto de muestras. Esto se debe tener en cuenta para identificar la mayor cantidad de metabolitos posibles, ya que, aunque un espectro sea el que tiene la mayor cantidad de señales, no tiene por qué contener todas las señales presentes en el conjunto de los espectros.

TopSpin™ es un programa de Bruker Corporation para analizar y procesar espectros de RMN[18]. Mediante este software, podemos comparar de forma simultánea los distintos espectros de RMN pertenecientes a un grupo de muestras y escoger de forma cualitativa el espectro idóneo para la generar el “pattern” de metabolitos.

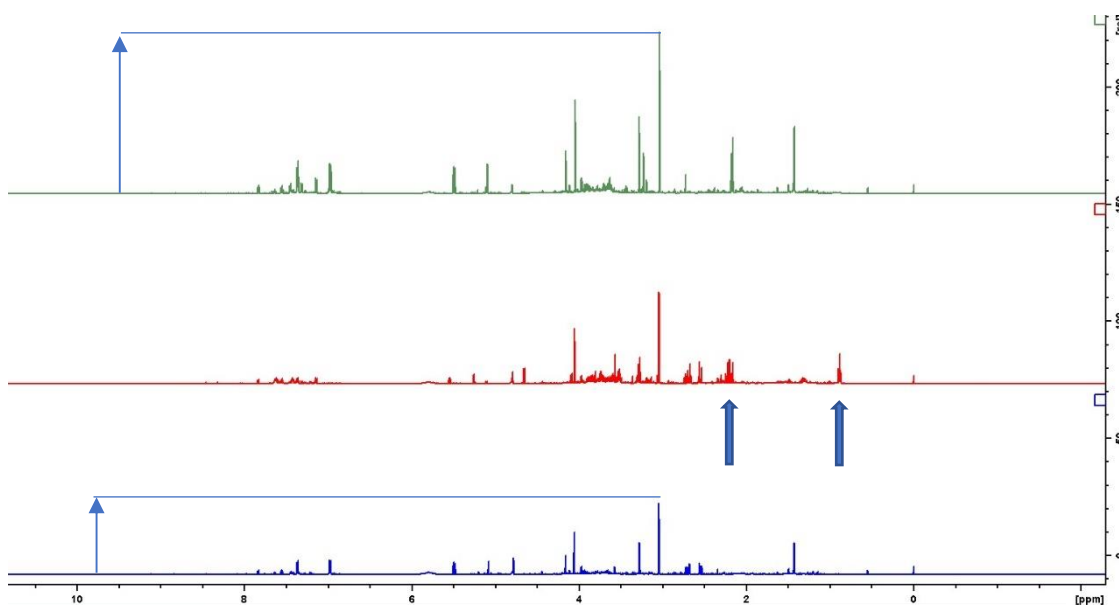


Figura 3. Ejemplo de comparación de espectros de orina mediante TopSpin.

El espectro verde muestra muchas señales con una buena intensidad. El espectro azul muestra aproximadamente las mismas señales que el espectro verde, pero con una menor intensidad. El espectro rojo, no muestra tantas señales, sin embargo, presenta señales que no aparecen en los otros dos. En el caso de este grupo de muestras, el espectro correspondiente a la muestra verde resulta ser el idóneo para realizar el pattern de metabolitos. Además, la identificación se completa con las señales del espectro rojo que no aparecen en el verde.

4.1.1.2. Identificación de metabolitos mediante ChenomX

Una vez definida la muestra que queremos perfilar con el mayor número de metabolitos posibles, cargamos nuestra muestra en el software ChenomX (Fig.5).

ChenomX NMR Suite de ChenomX Inc. es un software que permite identificar metabolitos de espectros de RMN mediante extensas librerías de modelos espectrales[14]. ChenomX te indica qué metabolitos pueden presentar señal en un rango de ppm determinados, además de mostrar la multiplicidad y forma de la señal. Lo ideal es empezar en uno de los extremos del espectro e ir identificando señal a señal qué metabolito la genera, ya que en los extremos existe menos solapamiento de señales que en el centro. Otra opción consiste en buscar los metabolitos más comunes en orina en literatura y mirar si se encuentran en nuestro espectro. Este último método es más rápido, pero menos preciso, ya que, en caso de haber algún metabolito extraño o poco frecuente, lo pasaríamos por alto.

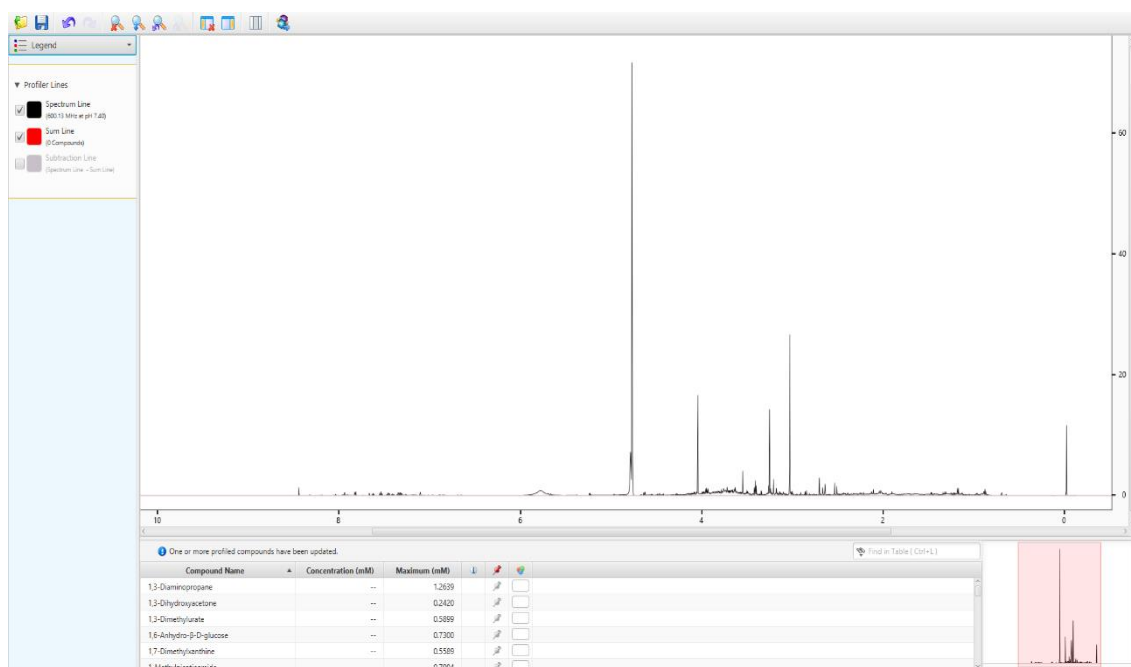


Figura 4: Snapshot de ChenomX NMR Suite 8.5 con la muestra escogida para realizar el pattern de metabolitos

En este trabajo se ha realizado una combinación de los métodos mencionados anteriormente: Primero se observa si los metabolitos más comunes en orina están presentes en el espectro de referencia y posteriormente se intenta identificar las señales a las que no se le ha asignado ningún metabolito de la zona donde hay menos solapamiento (los extremos) a la zona media del espectro.

Al identificar compuestos, es posible que más de un metabolito pueda encajar en una señal. Para determinar con exactitud a qué metabolito corresponde la señal, debemos comprobar que todas las señales del metabolito encajen en el espectro. En otras ocasiones, es posible que una señal esté formada por la combinación de la señal de 2 metabolitos independientes. Veremos ejemplos en las imágenes a continuación.

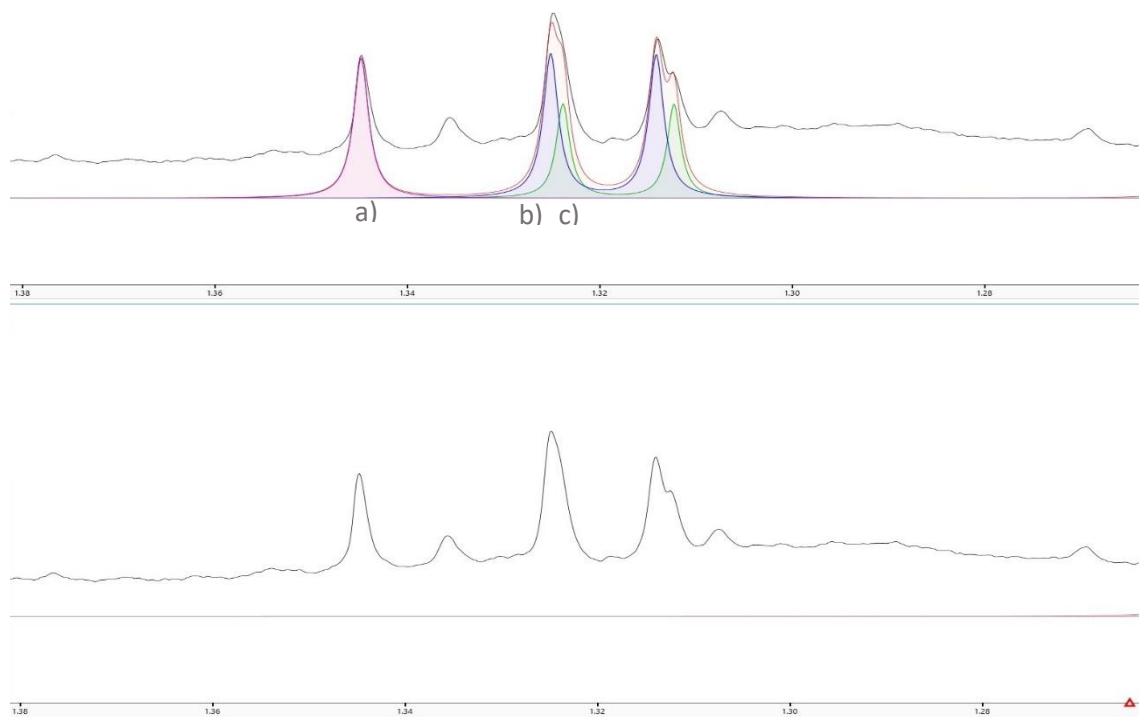


Figura 6: Identificación de compuestos en el rango de 1.28ppm a 1.38ppm. Línea negra: espectro. Área morada: a) 2-hidroxiisobutirato. Área azul: b) Treonina. Área verde: c) Lactato. Área roja: Suma de las áreas

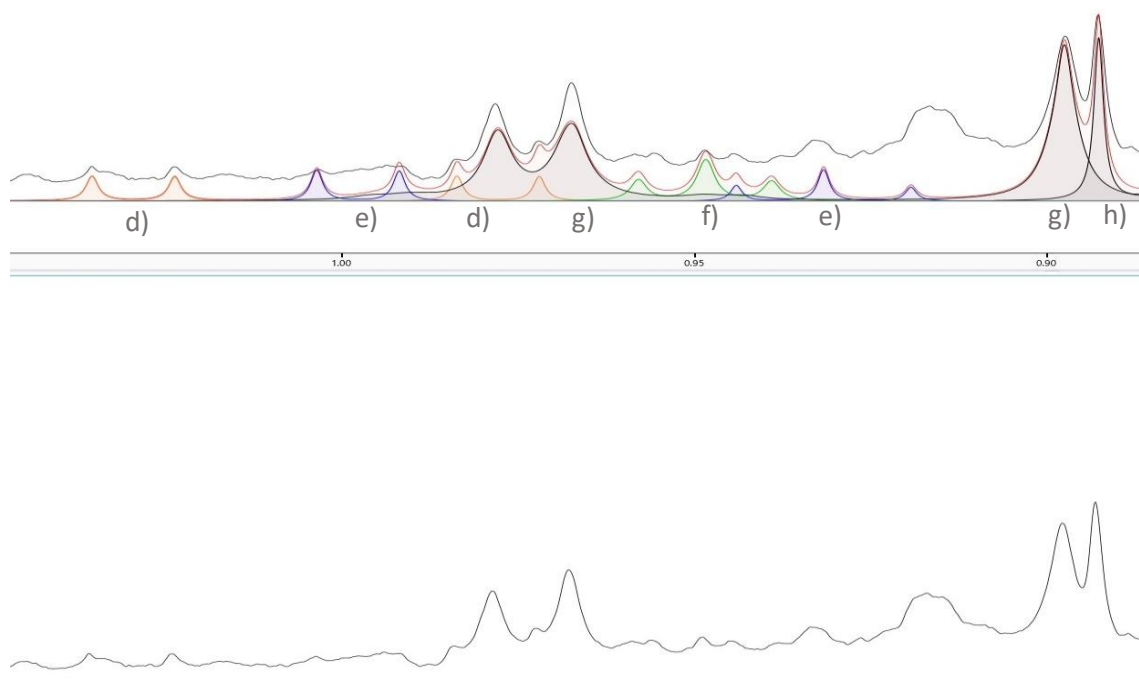


Figura 7: identificación de compuestos en el rango de 0.85ppm a 1.05ppm. Línea negra: espectro. Área naranja: d) Valina. Área azul: e) Isoleucina. Área verde: f) Leucina. Áreas negras: g) ácido isocaproico y h) ácido glicólico

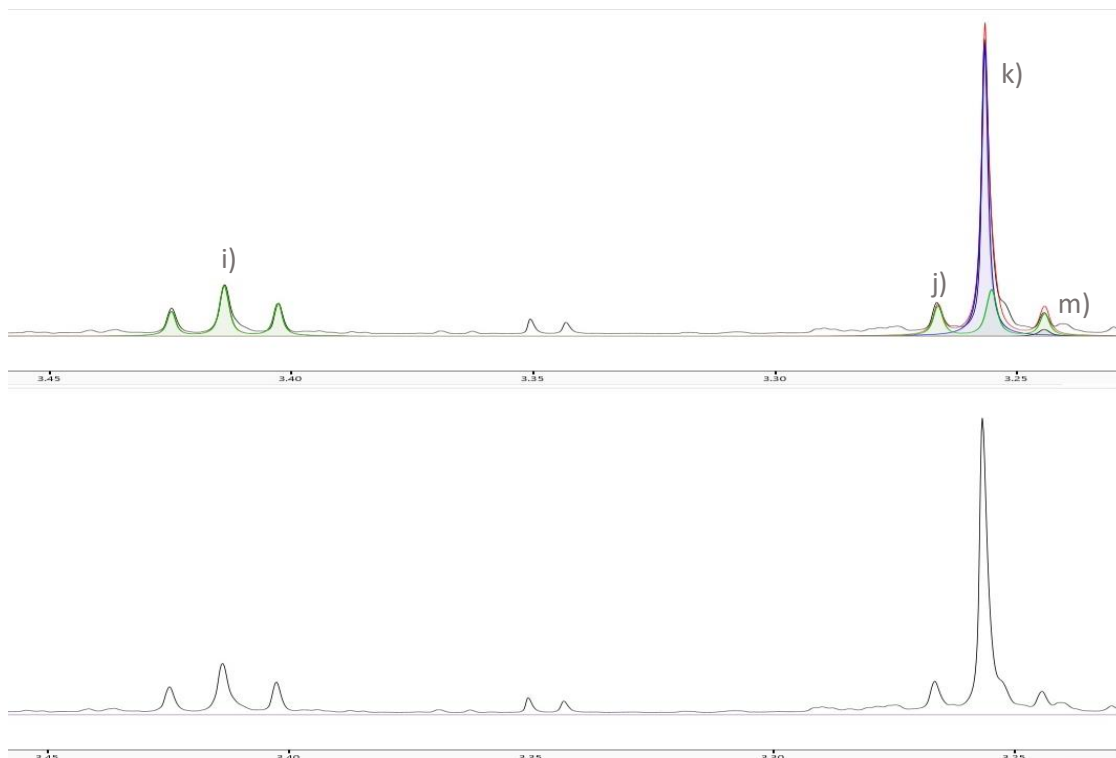


Figura 8: identificación de compuestos en el rango de 3.2ppm a 3.45ppm. Línea negra: espectro.
 Área verde: i) j) Taurina. Área azul: k) TMAO. Área negra: m) betaina.

En las figuras anteriores (6, 7 y 8) podemos observar diferentes retos que encontramos a la hora de identificar metabolitos.

En la figura 6, podemos observar un ejemplo de solapamiento de señales de diferentes metabolitos (lactato (b) y treonina (c)). En este caso, no podemos usar como referencia otras señales de ambos metabolitos en otras regiones del espectro porque estas son muy débiles. Para identificar correctamente cada metabolito en este caso, se usa la anchura de los dobletes. El doblete de lactato es ligeramente más ancho que el doblete de treonina, por tanto, la única forma de que encajen las señales de los metabolitos dadas por ChenomX en nuestra señal problema es asignando la señal más intensa a la treonina y la señal menos intensa y ligeramente desplazada a la derecha al lactato.

En la figura 7 podemos observar la complejidad de la región correspondiente a los aminoácidos ramificados (valina (d), leucina (f) e isoleucina (e)). Las señales son muy débiles, pero se pueden apreciar levemente. Para identificarlos, usamos como referencia el siguiente artículo[16]. La figura 8 se identifican las señales correspondientes a taurina (i) (j), trimetilamina N óxido (TMAO) (k) y betaina (m). En este ejemplo, podemos determinar la intensidad del triplete de taurina solapado (j) al singlete de TMAO (k) gracias al triplete de taurina aislado (i).

- **Metabolitos identificados:**

Para este conjunto de muestras se han podido identificar 50 metabolitos.

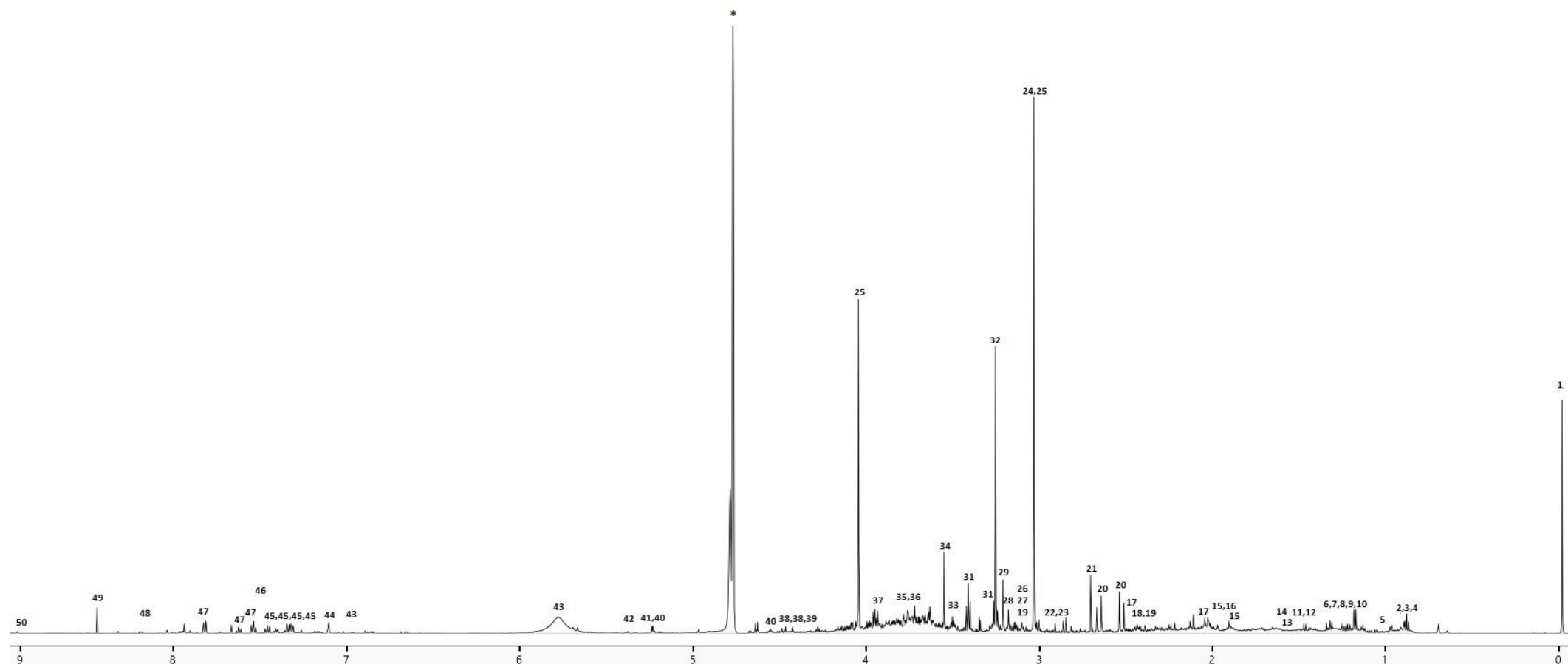


Figura 9: Espectro de RMN de una muestra de orina. Identificaciones: 1.TSP 2.Isocaproato 3.Valorato 4.Glicolato 5.Valina 6.Leucina 7.Isoleucina 8.Acido 3-Hidroxibutírico 9.Acido 3-Aminoisobutirico 10. Ácido 3-Hidroxiiisovalerico 11.Lactato 12.Treonina 13.2-Hidroxiisobutirato 14.Alanina 15.Lisina 16.Acetato 17.Glutamina 18.Succinato 19.Carnitina 20.Citrato 21.Dimetilamina 22.N,N Dimetilglicina 23.Metilhidantoina 24.Creatina 25.Creatinina 26.Etanolamina 27.Colina 28.O-Fosfocolina 29.Sn-Glicerio-Fosfocolina 30.Betaina 31.Betaina 32.Trimetilamina N Oxido (TMAO) 33.Trans-Aconitato 34.Glicina 35.Manitol 36.Etanolamina 37.Glicolato 38.1-Metilnicotinamina 39.Arabinosa 40.Glucoronato 41.Glucosa 42.Alantoina 43.Urea 44.Histidina 45.N-Fenilacetilglicina 46.Indolesulfato 47.Hipurato 48.Hipoxantina 49.Formiato 50.Trigonelina

4.1.2. Referenciado

Debido a factores fisicoquímicos como el pH, se pueden observar variaciones en el desplazamiento químico de las señales de espectros de RMN, sobre todo en fluidos como la orina. Si tenemos un conjunto de muestras, es probable que entre ellas haya ligeras diferencias en el desplazamiento químico para un mismo metabolito, como se puede observar en la figura 10. Puede parecer en la imagen que el desplazamiento

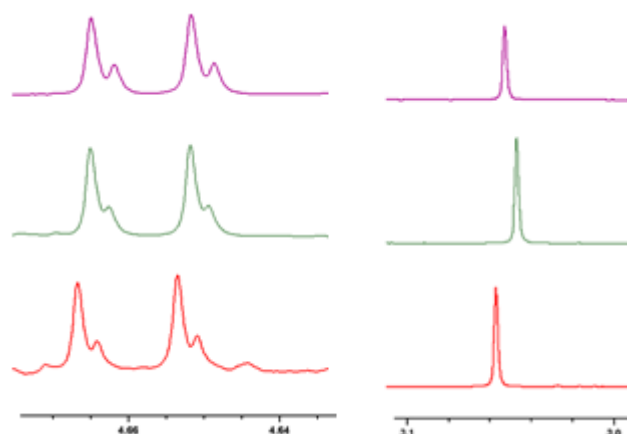


Figura 10: Desplazamientos químicos entre muestras para la misma señal en 4.65ppm y 3.05ppm respectivamente

de la señal es mínimo y que se pueden diferenciar bien las muestras, pero cuando queremos conseguir un método de cuantificación automático y reproducible, el referenciado de señales es esencial.

El referenciado consiste en alinear de la forma más precisa posible las señales de todos los espectros de resonancia magnética nuclear de una muestra. Podemos hacer dos tipos de referenciado: Referenciado global y referenciado local

El referenciado global consiste en alinear únicamente una misma señal única en todos los espectros de un conjunto de muestras. Esta metodología es muy rápida, ya que solo requiere alinear una única señal. Este tipo de referenciado funciona muy bien en matrices como el suero, donde se usa la señal de la glucosa para alinear los espectros, sin embargo, no es aplicable cuando trabajamos con orina, ya que las condiciones fisicoquímicas no afectan a todas las señales de un espectro por igual.

Este es uno de los principales retos de la automatización. Realizar un correcto referenciado de las señales para cuantificar correctamente cada uno de los metabolitos. Por ejemplo, en la figura 10 podemos observar como el doblete de 4.96ppm del espectro lila está ligeramente desplazado a la derecha respecto la misma señal en los otros dos espectros (rojo y verde). Sin embargo, el singlete de 3.05ppm del espectro lila se encuentra a medio camino entre el singlete del espectro rojo y el del espectro verde. Por tanto, alineando todos los espectros a una única señal no se alinean todas las señales de forma correcta.

El referenciado local consiste en el alineamiento de todas las señales de todos los espectros RMN del conjunto de muestras. Es un proceso mucho más costoso que el referenciado global, pero permite alinear todas las muestras de una manera mucho más precisa. Para facilitar el

proceso, solo se alinean las señales de interés, es decir, aquellas que hemos identificado en el apartado 3.1.2.2 Identificación.

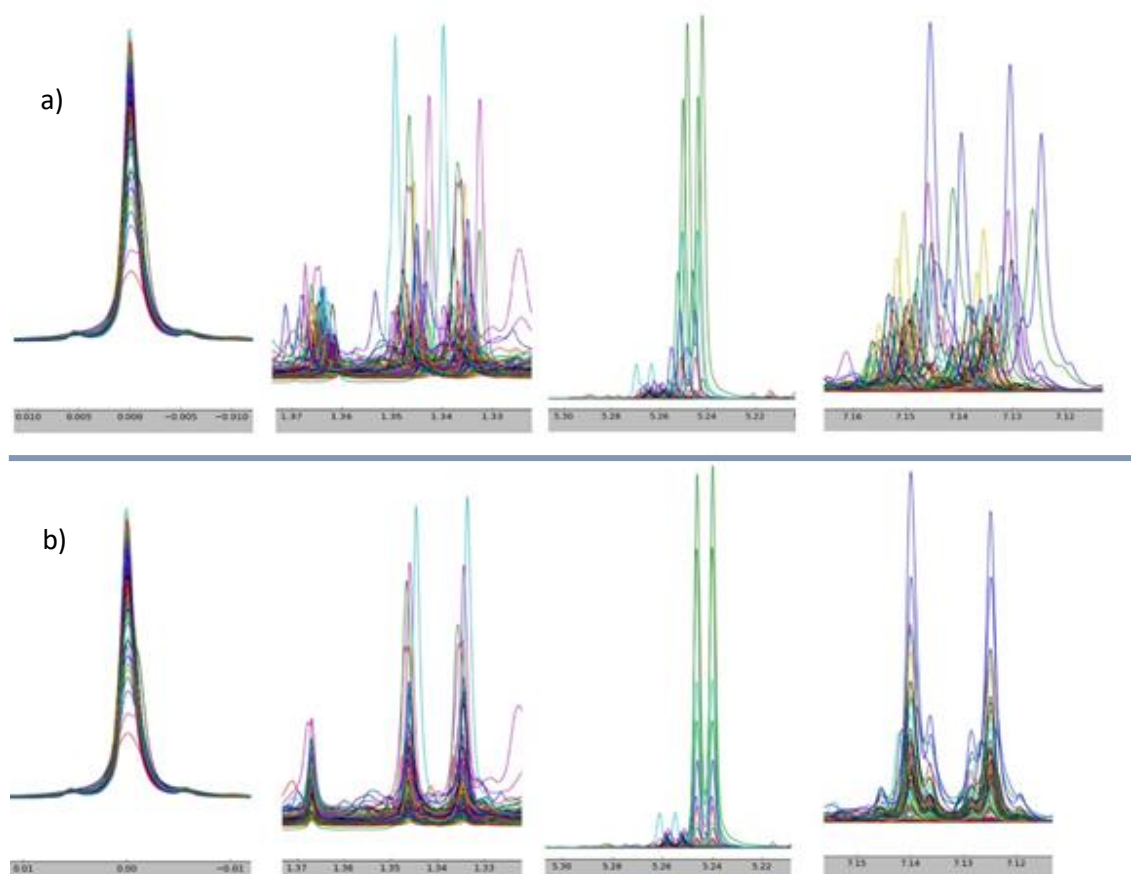


Figura 11: Comparación entre referenciado global por TSP en 0 ppm (a) en las regiones de 0ppm (TSP), 1.35ppm, 5.24ppm y 7.13ppm respectivamente y referenciado local (b)

4.1.3. Deconvolución

La técnica de deconvolución es un proceso matemático basado en algoritmos que permite recrear una ecuación[19]. Podemos deconvolucionar las señales del espectro de RMN para obtener la ecuación que define su forma, calcular su área y, por tanto, la concentración de metabolito asociado a dicha señal, ya que el área de las señales del espectro de RMN es directamente proporcional a la concentración de metabolito que genera la señal.

Teóricamente, la función que describe una señal de espectrometría es una función Lorentziana [20], aunque en la práctica, se pueden observar señales de RMN que se definen mejor según la ecuación Gaussiana [20]. Para solventar esta problemática, se usan funciones pseudo-Voigt basadas en la función de Voigt. Esta función es la convolución (es decir, una mezcla al 50%) de la distribución Gaussiana y Lorentziana, mientras que la función pseudo-Voigt es una combinación lineal de las aproximaciones Gaussianas y Lorentzianas en lugar de una

convolución [20]. Esto permite generar funciones con porcentajes desiguales de tipo de distribución, consiguiendo así una función que permita replicar con mayor fidelidad las señales de RMN.

La función Pseudo-Voigt (PV(x)) se calcula con la formula siguiente[20]:

$$PV(x) = \eta G(x) + (1 - \eta)L(x)$$

Donde G(x) es la función Gaussiana, L(x) es la función Lorentziana y η es la fracción Gaussiana de la función resultante.

En la figura 12 podemos observar diferentes funciones Pseudo-Voigt con distintos valores de η .

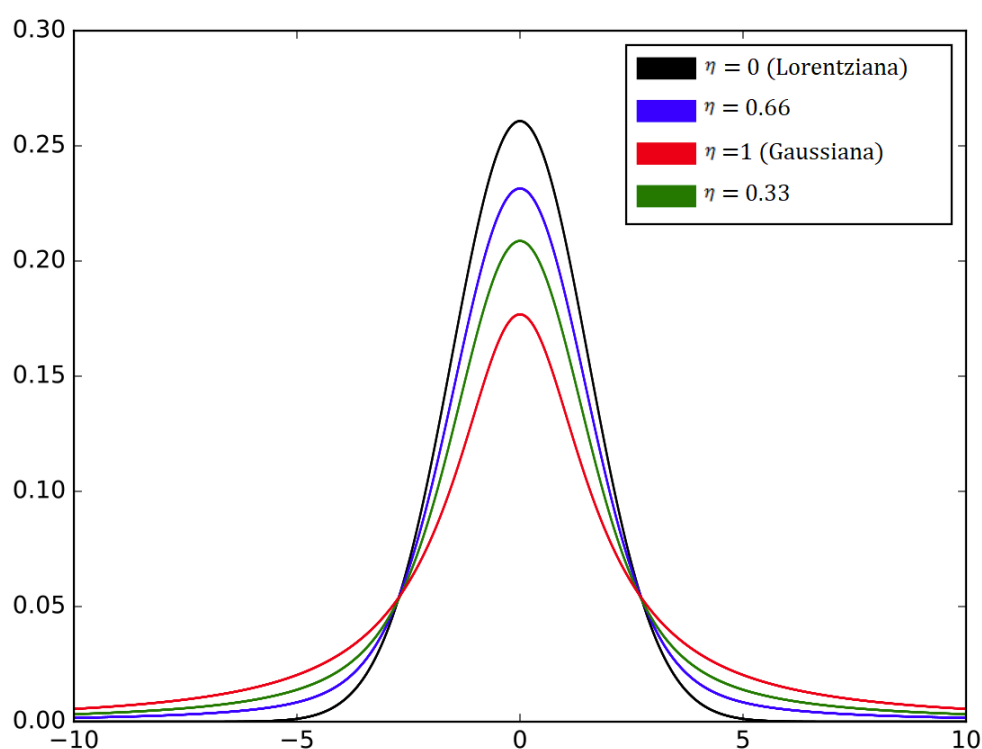


Figura 12: Diferentes representaciones de funciones Pseudo-Voigt con distintos valores de η . Cuando $\eta = 0$, la fracción Gaussiana es nula y por tanto la función es completamente Lorentziana. Cuando $\eta = 1$, la fracción Gaussiana es entera y por tanto la función es completamente Gaussiana. Valores de los ejes X e Y arbitrarios.

La función definida anteriormente se ajusta únicamente a los singletes. Para dobletes, tripletes y multipletes se usan variaciones de la misma función, con tantas repeticiones de la función como multiplicidad tenga la señal y ajustando la altura de las funciones según requiera la señal.

Para deconvolucionar de forma automática múltiples espectros, se usa la herramienta informática de Biosfer Teslab con ligeras modificaciones para cada señal. Primero, se restringe la herramienta informática a la zona de ppm que queremos deconvolucionar, e introducimos parámetros que queremos que adopte la función deconvolucionada. En el apartado anterior

4.1.2. hemos mencionado la importancia de referenciar y alinear las señales. Esto es debido a que la herramienta de Biosfer Teslab es capaz de deconvolucionar espectros de RMN en serie de forma automática, gracias a haber fijado a la misma posición la misma señal de los diferentes espectros que encontramos en un grupo de muestras. A continuación, se muestran ejemplos de deconvoluciones:

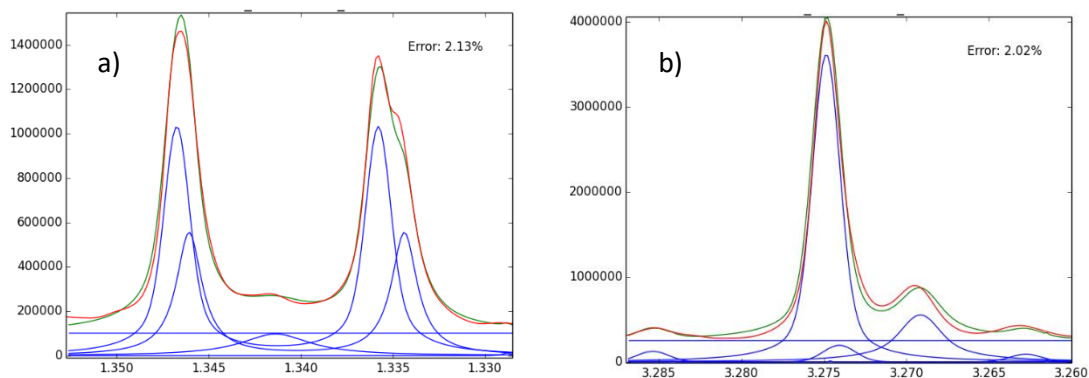


Figura 13: a) deconvolución de las señales de lactato y treonina en la región de 1.355 a 1.33ppm. b) deconvolución de las señales de TMAO, betaina y taurina en la región de 3.285 a 3.26ppm. Eje X: intensidad espectral (unidades arbitrarias). Eje Y: desplazamiento químico (ppm). Línea roja: Espectro original. Línea verde: Función resultado del sumatorio de las deconvoluciones. Líneas azules: deconvoluciones de las señales.

El software, a partir de unos “inputs” necesarios para poder dibujar la función, es capaz de deconvolucionar la señal (línea roja) generando la función pseudo-Voigt que mejor se adapta a esta (líneas azules). Después, el programa suma las funciones pseudo-Voigt de cada señal para generar la deconvolución de la región que hemos delimitado (línea verde).

Este paso es el más costoso de todo el proceso. La deconvolución de una señal simple, como un singlete, puede llevarle al operario unas horas. Sin embargo, la deconvolución de señales complejas en ambientes solapados puede costar días de trabajo, incluso semanas. Además, es necesario tener experiencia para diferenciar una deconvolución correcta de una incorrecta, aunque esta parezca emular a la perfección la señal del espectro RMN. A continuación se muestran varias figuras con ejemplos de deconvoluciones erróneas.

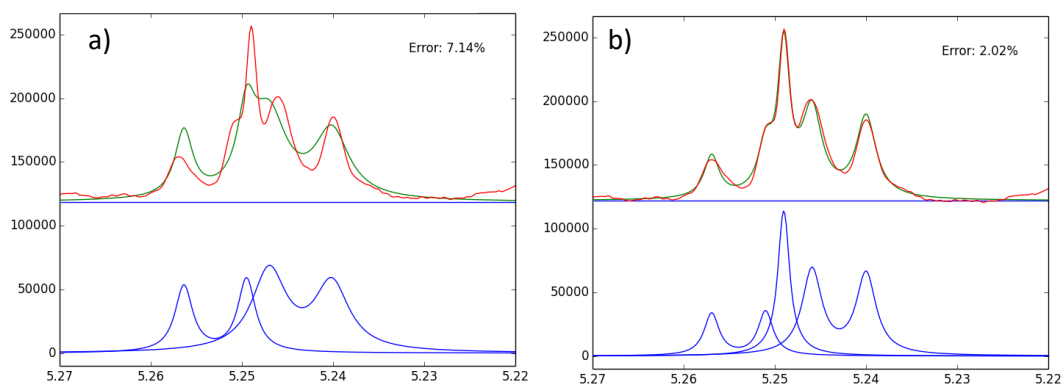


Figura14: Deconvolucion de los dobletes de glucosa y glucuronato en el rango de 5.27 a 5.22ppm. Línea roja = espectro RMN original. Líneas azules = deconvoluciones de las señales. Línea verde = Suma de las deconvoluciones.

En esta zona se encuentran los dobletes de glucosa y glucuronato. Sin embargo, en muchas muestras se puede encontrar un singlete entre los dos dobletes que provoca una deconvolución errónea (Fig14.a). Para solucionar el problema, se añade en el input la función correspondiente a un singlete además de las funciones correspondientes a los dobletes ya mencionados. De esta forma, conseguimos que el programa ajuste correctamente las deconvoluciones (Fig14.b).

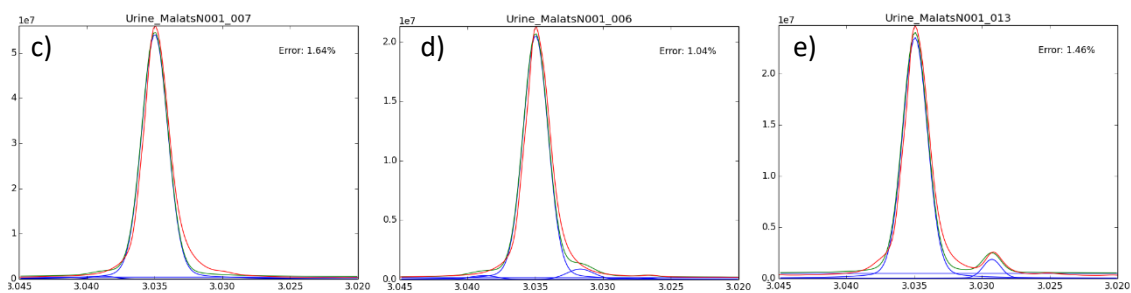


Figura 15: Deconvolución de los singletes de creatinina (3.035ppm) y creatina (3.03ppm) para 3 muestras diferentes. Línea roja = espectro RMN original. Líneas azules = deconvoluciones de las señales. Línea verde = Suma de las deconvoluciones.

En esta región se encuentran los singletes de creatinina y creatina. La señal creatinina (3.035ppm) es claramente más intensa que la señal de creatina (3.03ppm). La figura 15.c muestra una correcta deconvolución, donde encontramos la deconvolución de la señal de creatinina y ninguna deconvolución de la señal de creatina, ya que esta es tan debil que la consideramos por debajo del límite de detección. La figura 15.d muestra una deconvolucion que aunque aparentemente parezca correcta, es errónea porque el software esta intentando generar una función a la derecha de la creatinina para ajustar mejor la suma de las deconvoluciones (línea verde) señal, sin embargo, lo que esta haciendo el software es darnos un dato falso sobre la señal de creatinina ya que le está restando área para generar otra función. Lo lógico sería eliminar del input los datos que generan esa función a la derecha de la creatina,

pero esa función es necesaria ya que corresponde a la deconvolución de la creatina, como se aprecia en la figura 15.e. Hay que recordar que con un único archivo de inputs queremos cuantificar la mayor cantidad de muestras posibles, y eso exige mucha precisión en las funciones que aplicamos para deconvolucionar correctamente una función.

4.1.4. Cuantificación

Además de deconvolucionar las señales, el software de Biosfer Teslab calcula el área de las funciones pseudo-Voigt que genera. El problema es que estas áreas vienen dadas en unidades arbitrarias. El único metabolito que aparece en el espectro, cuya concentración es conocida, es el TSP. Y dado que la concentración es proporcional al área de la señal, podemos usar el área y concentración de TSP para conocer la concentración del resto de metabolitos. Para ello, usamos la formula descrita por Serkova[17]:

$$C_x = \frac{\frac{I_x}{N_x} * C_{tsp}}{I_{tsp} * 9}$$

Donde C_x es la concentración de metabolito, I_x es el área de la señal de dicho metabolito, N_x es el número de núcleos de Hidrogeno que generan la señal, C_{tsp} es la concentración de TSP, I_{tsp} es el área de la señal de TSP y 9 es el número de protones que generan la señal de TSP.

Las áreas de los metabolitos y del TSP (I_x e I_{tsp}) son calculadas por el software de Biosfer Teslab. El número de protones que generan las señales se puede obtener en la HMDB.

4.1.5. Outputs del software

Después de la deconvolución, el software devuelve como output:

- Una carpeta de cada región deconvolucionada que puede ser equivalente a un metabolito o a más de uno. Dentro de esta carpeta encontramos la deconvolución para cada una de las muestras en imágenes como las figuras 14 y 15
- Un Excel con los errores de fitting de las deconvoluciones que nos permitirá ver cuánto se asemeja nuestra deconvolución al espectro de la muestra real. Siendo un error de fitting muy alto una mala cuantificación del metabolito.
- Varios Excel con los valores que el programa ha usado dentro del rango de valores que hemos introducido en los inputs para generar las deconvoluciones. Esto es útil para deconvolucionar mediante ensayo y error, ya que te permite saber que input está provocando una deconvolución errónea

- Un Excel con la tabla de resultados: Esta contiene las áreas de las señales de cada metabolito en unidades arbitrarias

Posteriormente, se deben determinar las muestras cuya deconvolución está por debajo del límite de detección (<LD) y las muestras que no se pueden deconvolucionar con precisión (NA) por la mala calidad del espectro, por dificultades técnicas del software o por la complejidad de la región. Todo esto se debe añadir al archivo Excel para poder normalizar los datos posteriormente.

Las muestras <LD y NA se deben tener en cuenta en los resultados. Los NA se definen como resultados nulos y a los <LD se les asigna un valor de 0,001mmol/L, ya que están presentes en la muestra, pero no podemos cuantificarlos porque la baja intensidad de su señal. Para este estudio se escoge 0,001mmol/L para los <LD para que las muestras aparezcan en el estudio estadístico.

4.1.6. Resultado del análisis metabólico

De los 50 metabolitos identificados, se han podido cuantificar 30. El resultado final del análisis metabólico es el que se muestra en la tabla 1:

Tabla 1: Resultados del análisis estadístico. Resultados expresados con el promedio de concentraciones de cada grupo y la desviación estándar. N es el número de muestras que se han cuantificado y se han tenido en cuenta para calcular el promedio de cada grupo. Resultados NO normalizados.

METABOLITO	ESTADO PATOLÓGICO	N	Concentración (mmol/L)
Alanina	Control	55	5,94 ± 0,66
	Problema	57	7,87 ± 1,28
Citrato	Control	56	42,23 ± 5,17
	Problema	57	52,52 ± 5,62
Urea	Control	57	137,21 ± 6,50
	Problema	59	149,54 ± 7,92
3-Hidroxiisobutirato	Control	54	6,56 ± 2,98
	Problema	54	5,82 ± 2,76
3-Aminoisobutirato	Control	52	4,77 ± 0,56
	Problema	54	5,16 ± 0,54
Acetaminofen	Control	55	9,27 ± 1,59
	Problema	58	14,72 ± 3,84
Carnitina	Control	54	4,34 ± 0,55

	Problema	57	5,69 ± 0,92
<i>Formiato</i>	Control	54	2,80 ± 0,29
	Problema	58	4,36 ± 0,40
<i>Dimetilamina</i>	Control	56	7,32 ± 0,67
	Problema	59	7,15 ± 0,51
<i>Glicina</i>	Control	55	18,08 ± 3,81
	Problema	59	21,10 ± 3,30
<i>Hipurato</i>	Control	54	54,88 ± 8,42
	Problema	56	55,81 ± 7,08
<i>Metilnicotimida</i>	Control	56	3,89 ± 0,63
	Problema	58	4,69 ± 0,57
<i>Sn-glicero-fosfocolina</i>	Control	53	8,06 ± 2,47
	Problema	55	8,95 ± 2,68
<i>Sucinato</i>	Control	52	1,30 ± 0,16
	Problema	58	1,34 ± 0,12
<i>Colina</i>	Control	49	2,08 ± 0,50
	Problema	50	2,72 ± 0,41
<i>O-acetylcarnitine</i>	Control	44	0,53 ± 0,09
	Problema	49	0,63 ± 0,14
<i>Acetato</i>	Control	55	2,79 ± 0,82
	Problema	53	4,44 ± 1,28
<i>Creatina</i>	Control	49	6,48 ± 2,00
	Problema	54	7,50 ± 2,98
<i>Creatinina</i>	Control	57	198,13 ± 17,69
	Problema	58	181,22 ± 14,02
<i>Glucosa</i>	Control	49	109,03 ± 58,63
	Problema	50	111,21 ± 64,13
<i>Glucuronato</i>	Control	48	19,45 ± 4,13
	Problema	49	31,08 ± 7,22

<i>Lactato</i>	Control	21	4,43 ± 1,18
	Problema	24	3,24 ± 0,60
<i>Treonina</i>	Control	19	8,23 ± 2,48
	Problema	24	8,75 ± 2,58
<i>N,N-Dimetilglicina</i>	Control	57	1,23 ± 0,13
	Problema	59	1,50 ± 0,15
<i>N-Metilhidantoina</i>	Control	38	0,29 ± 0,08
	Problema	38	0,29 ± 0,07
<i>3-Hidroxiisovalerate</i>	Control	55	1,19 ± 0,15
	Problema	58	1,26 ± 0,10
<i>Trigonelina</i>	Control	57	4,03 ± 0,52
	Problema	59	4,15 ± 0,52
<i>2-hidroxiisobutirato</i>	Control	52	1,32 ± 0,21
	Problema	55	1,23 ± 0,15
<i>Valina</i>	Control	55	0,90 ± 0,10
	Problema	58	1,09 ± 0,15
<i>Isoleucina</i>	Control	54	0,32 ± 0,05
	Problema	56	0,36 ± 0,06

4.2. Normalización

La normalización de datos es un paso crítico para la comparación de muestras y el análisis estadístico en metabolómica. Por ejemplo, los efectos de dilución de los espectros de RMN pueden hacer imposible la comparación de grupos de muestras si los resultados no están normalizados. Si bien el efecto de dilución no supone un problema en muestras de plasma o tejidos, en orina, es uno de los efectos más dominantes. Como se menciona en el apartado 1.3.1, la normalización por creatinina es el método habitual de normalización en metabolómica de orina, y por ello, en este trabajo se normalizan los resultados por creatinina.

- Normalización por creatinina

Para normalizar por área de creatinina se divide la concentración de cada metabolito por la concentración de creatinina de la misma muestra. Los cálculos se realizan en Excel.

4.3. Analisis estadístico

Para realizar el análisis estadístico se realiza un T-test de muestras independientes y se generan los boxplots correspondientes a los grupos control y problema de cada metabolito mediante el software SPSS V26[21]. Para obtener resultados robustos, se apartan del análisis estadístico los metabolitos con una cantidad de resultados NA superior al 10% del total de muestras (N=116). Los metabolitos que no se tienen en cuenta en el análisis estadístico son: glucosa, glucuronato, lactato, treonina, N-metilhidantoína, creatina y colina. Además, tampoco se tiene en cuenta el acetaminofen por ser un metabolito exógeno.

Tabla 1: P-valor de los datos sin normalizar y normalizados por creatinina entre grupo control y problema después de analizar las medias mediante T-test de variables independientes. No se observan valores significativos. Se usa $p < 0.05$

METABOLITO	NO NORMALIZADOS	CREATININA
Alanina	0,19	0,19
Citrato	0,18	0,11
Urea	0,23	0,91
3-Hidroxiisobutirato	0,86	0,92
3-Aminoisobutirato	0,61	0,24
Carnitina	0,22	0,92
Formiato	0,02(*)	0,08
Dimetilamina	0,84	0,67
Glicina	0,55	0,39
Hipurato	0,93	0,82
Metilnicotimida	0,35	0,65
Sn-glicero-fosfocolina	0,81	0,76
Sucinato	0,85	0,74
O-acetylcarnitina	0,54	0,35
Acetato	0,28	0,06
Creatinina	0,45	-
N,N-Dimetilglicina	0,16	0,06
3-Hidroxiisovalerate	0,32	0,39
Trigonelina	0,18	0,40
2-hidroxiisobutirato	0,46	0,47
Valina	0,24	0,34
Isoleucina	0,57	0,47

Al analizar los datos sin normalizar, se encuentran diferencias significativas entre el grupo control y problema en el formiato (fig.17). Sin embargo, al normalizar por creatinina no se observan diferencias significativas para ningún metabolito.

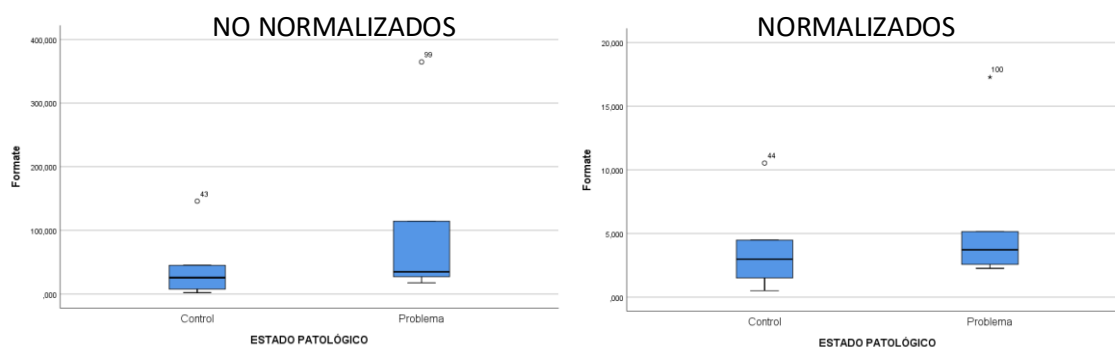


Figura 17: Boxplots de formiato para los grupos controles y problema. A la derecha, resultados no normalizados en los que se observan diferencias significativas. A la izquierda, resultados normalizados por creatinina, no se observan diferencias significativas

4.4. Comparación de los resultados con valores encontrados en la literatura

Para estudiar la robustez de los resultados, se ha buscado en literatura relacionada con metabolómica en orina el umbral de concentraciones en el que se espera encontrar cada uno de los metabolitos identificados. En la tabla 3 se muestran los valores de la concentración de los metabolitos que se han obtenido mediante el software de Biosfer Teslab y los valores encontrados en literatura[8]. Se muestra el valor promedio de las muestras del grupo control y los resultados se expresan normalizados por creatinina ($\mu\text{mol}/\text{mmol}$ de creatinina) con la desviación estándar de cada valor. El umbral esperado se muestra con las mismas unidades.

Tabla 2: Comparación de los resultados obtenidos en el análisis metabolómico normalizados por creatinina y comparación con resultados encontrados en literatura. En negrita, resultados que no concuerdan con la literatura.

METABOLITO	RESULTADO $\mu\text{mol}/\text{mmol}$ de creatinina	UMBRAL ESPERADO [8] $\mu\text{mol}/\text{mmol}$ de creatinina
Alanina	35,78 \pm 4,6	7,1 - 43
Citrato	250,36 \pm 27	49 - 600
Urea	977,9 \pm 75	951,8 - 1740
3-Hidroxiisobutirato	30,85 \pm 12	11,8 – 59,8
3-Aminoisobutirato	25,08 \pm 1,7	1,24 - 32,35
Carnitina	29,40 \pm 4,9	0,7 – 16,4
Formiato	22,27 \pm 3,4	6,9 - 120
Dimetilamina	42,51 \pm 3,6	20,3 – 59,2
Glicina	106,3 \pm 15	44 – 300
Hipurato	301,6 \pm 35	19 – 622
Metilnicotimida	23,68 \pm 2,6	0,2 – 12,0
Sn-glicero-fosfolina	65,07 \pm 30	-
Sucinato	7,97 \pm 0,9	0,3 – 33,3
O-acetylcarnitina	3,88 \pm 0,7	0,4 – 5,9
Acetato	13,63 \pm 3,1	2,5 - 106,0
Creatinina	1	1
N,N-Dimetilglicina	6,92 \pm 0,54	-

<i>3-Hidroxiisovalerate</i>	6,67 ± 0.49	4,1 – 17,2
<i>Trigonelina</i>	21,55 ± 1,9	5,5 – 19,3
<i>2-hidroxiisobutirato</i>	7,54 ± 1,2	1,3 – 4,6
<i>Valina</i>	5,02 ± 0,5	2,7 – 9,8
<i>Isoleucina</i>	1,75 ± 0,2	0,5 – 2,7

Los metabolitos carnitina, metilnicotinamida, trigonelina y 2-hidroxiisobutyrato presentan diferencias respecto el umbral de concentración esperado de la literatura.

5. DISCUSIÓN

En este trabajo he podido automatizar la caracterización de 30 metabolitos de bajo peso molecular en muestras de orina medidas mediante RMN. Este software es una versión piloto de un programa capaz de cuantificar metabolitos en muestras de orina de forma rápida y automática. Este software piloto está optimizado para las 116 de muestras de orina analizadas. Para aplicarlo a cualquier tipo de muestra de orina, será necesario disponer de un grupo de muestras lo suficiente grande como para poder comprobar la reproducibilidad y robustez del método. Además, el software tiene mucho margen de mejora. De los 50 metabolitos identificados, se han podido cuantificar 30. Si bien esto es debido a la complejidad de las señales de los metabolitos no cuantificados, se podría adaptar el software para que determine la presencia o ausencia del metabolito sin necesidad de cuantificarlo, mejorando la calidad del producto y aportando una mayor cantidad de información sobre el experimento.

Dicho software es una adaptación de un software que Biosfer ya tenía para el análisis de metabolitos de bajo peso molecular en suero. A pesar de que al inicio de este trabajo se partía de un software base, hay que destacar que la matriz de orina tiene una serie de peculiaridades que no contiene una matriz de suero; como por ejemplo:

- La orina es una matriz de deshecho y por lo tanto, los metabolitos que encontramos no son homogéneos entre una muestra y otra, lo que dificulta la cuantificación.
- La orina tiene como principal desventaja que no es un fluido de homeostasis como el suero, en el cual el rango de concentraciones está bastante acotado aunque haya enfermedad.
- El pH de la orina sufre variaciones y esto provoca un gran problema en el referenciado de las señales, y es el principal inconveniente de la automatización de la cuantificación de metabolitos.

A estas peculiaridades de la matriz de orina, lo que hacen dificultosa la obtención de un software automático de deconvolución de espectros de RMN, le sumamos una serie de retos personales

a los que me he tenido que enfrentar en cada etapa, y que he ido superando a lo largo del trabajo:

El primero reto vino dado por el lenguaje de programación: Debido a que el software de Biosfer Teslab está escrito en Python, aprender las nociones básicas de dicho lenguaje de programación para posteriormente poder retocar el código según fuese necesario fue la primera lección a aprender.

El segundo reto fue el de identificar los metabolitos: donde el primer objetivo fue ampliar la librería de metabolitos identificados previamente por el equipo de Biosfer Teslab. Este reto fue superado gracias a una búsqueda bibliográfica sobre estudios metabolómicos en orina y el uso de la última versión disponible de ChenomX, que incluye una biblioteca de metabolitos actualizada respecto a la usada previamente por el equipo de Biosfer Teslab.

El tercer y cuarto reto fueron los más duros, el referenciado y la deconvolución de señales:

- En el referenciado: aun habiendo determinado que el referenciado local era la técnica idónea, la varianza del desplazamiento químico de las señales entre las muestras era tan grande que en las zonas con alta densidad de señales resultaba imposible referenciarlas aun usando la técnica de referenciado local. Este reto fue superado gracias a un segundo referenciado local sobre la zona ya referenciada, permitiendo realinear las señales imposibles de alinear a causa del desplazamiento químico. Esto fue una novedad introducida en el software de orinas, ya que en el software de suero no es necesario realizar este paso
- En la deconvolución, cada señal era un reto en sí, ya que describir y delimitar las variables que posteriormente dibujan la función que emula la señal requiere de un alto conocimiento del software, mucha precisión y, sobre todo, paciencia, ya que la deconvolución de una señal compleja con varias señales y diferentes multiplicidades podía tardar horas en generarse para todo el conjunto de muestras. Por tanto, errar a la hora de introducir los inputs en el código podía suponer la pérdida de horas de trabajo.

A pesar de todos los retos que suponía este trabajo se ha obtenido una versión piloto que es capaz de cuantificar 30 metabolitos en las muestras de orina de manera automática. Además, cuando hemos comparado los resultados (en cuantificación absoluta normalizados por creatinina) hemos podido comprobar que los resultados obtenidos se ajustan a los resultados que obtenemos en literatura, a excepción de algún caso, en el que vemos que nuestros resultados estan algo desviados. Cuando Biosfer realice la validación del software con un

conjunto de muestras más grande, será necesario comprobar la concentración de todos los metabolitos, pero en especial de estos para ver si se ajustan al rango de concentración reportado previamente.

5.1. Trabajo futuro

Este software es una versión piloto de un programa capaz de cuantificar metabolitos en muestras de orina de forma rápida y automática. Para pulirlo, será necesario disponer de un grupo de muestras lo suficiente grande como para poder comprobar la reproducibilidad y robustez del método, ya que en este trabajo y para esta versión piloto se han utilizado 116 muestras. Además, el software tiene mucho margen de mejora. De los 50 metabolitos identificados, se han podido cuantificar 30. Si bien esto es debido a la complejidad de las señales de los metabolitos no cuantificados, se podría adaptar el software para que determine la presencia o ausencia del metabolito sin necesidad de cuantificarlo, mejorando la calidad del producto y aportando una mayor cantidad de información sobre el experimento.

6. CONCLUSIONES

Se ha podido automatizar la cuantificación de metabolitos de un conjunto de muestras de orina por resonancia nuclear gracias a la adaptación de un software previo con un propósito similar. Se han obtenido buenos resultados y ha sido posible la cuantificación de 30 metabolitos. Además, los resultados se encuentran dentro de los umbrales de concentración esperados para cada metabolito, según se puede encontrar en literatura. Este software es una versión piloto de un programa con el potencial de facilitar el estudio metabolómico en muestras de orina facilitando procesos como el estudio de biomarcadores o la prevención de patologías.

7. AUTOAVALUACIÓN

Con este trabajo he podido desempeñar un trabajo relacionado con la biotecnología dentro del ámbito de la bioinformática y la biología molecular de sistemas. El trabajo ha superado todas las expectativas y he podido aprender muchos conocimientos interesantes que me serán de utilidad en el futuro para mi desarrollo profesional.

BIBLIOGRAFÍA

- [1] R. A. Silva, T. C. S. Pereira, A. R. Souza, and P. R. Ribeiro, "1H NMR-based metabolite profiling for biomarker identification," *Clin. Chim. Acta*, vol. 502, pp. 269–279, 2020.
- [2] J. Stenesh and J. Stenesh, "Introduction to Metabolism," in *Biochemistry*, Springer US, 1998, pp. 203–219.
- [3] D. S. Wishart, "Emerging applications of metabolomics in drug discovery and precision medicine," *Nat. Rev. Drug Discov.*, vol. 15, no. 7, pp. 473–484, 2016.
- [4] V. Shulaev, "Metabolomics technology and bioinformatics," *Brief. Bioinform.*, vol. 7, no. 2, pp. 128–139, 2006.
- [5] J. A. Armstrong, "Urinalysis in western culture: A brief history," *Kidney Int.*, vol. 71, no. 5, pp. 384–387, 2007.
- [6] J. Jing and Y. Gao, "Urine biomarkers in the early stages of diseases: Current status and perspective," *Discov. Med.*, vol. 25, no. 136, pp. 57–65, Feb. 2018.
- [7] A. Zhang, H. Sun, X. Wu, and X. Wang, "Urine metabolomics," *Clin. Chim. Acta*, vol. 414, pp. 65–69, 2012.
- [8] S. Bouatra *et al.*, "The Human Urine Metabolome," *PLoS One*, vol. 8, no. 9, p. e73076, Sep. 2013.
- [9] "Human Metabolome Database." [Online]. Available: <https://hmdb.ca/>. [Accessed: 25-Jun-2020].
- [10] F. X. . K. D. J. Silverstein, R. M.; Webster, *Proton NMR Spectrometry. In Spectrometry Identification of Organic Compounds, Seventh Edition; Wiley; College of Environmental Science and Forestry: State University of New York, 2005; pp 127-128. .*
- [11] A. H. Emwas *et al.*, "Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review," *Metabolomics*, vol. 11, no. 4, pp. 872–894, 2015.
- [12] F. Dieterle, A. Ross, and H. Senn, "Probabilistic Quotient Normalization as Robust method to account for dilution of complex biological mixtures," *Anal. chem.*, vol. 78, no. 13, pp. 4281–4290, 2006.
- [13] A. C. Dona *et al.*, "Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping," *Anal. Chem.*, vol. 86, no. 19, pp. 9887–9894, Oct. 2014.
- [14] Chenomx Inc, "Products | Chenomx Inc." [Online]. Available: <https://www.chenomx.com/products/>.
- [15] P. Soininen *et al.*, "High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism," *Analyst*, vol. 134, no. 9, pp. 1781–1785, Aug. 2009.
- [16] K. A. Veselkov *et al.*, "Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery," *Anal. Chem.*, vol. 81, no. 1, pp. 56–66, Jan. 2009.
- [17] N. Serkova, T. F. Fuller, J. Klawitter, C. E. Preise, and C. U. Niemann, "1H-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney

transplants," *Kidney Int.*, vol. 67, no. 3, pp. 1142–1151, 2005.

- [18] Bruker Corporation, "Software for NMR Data Analysis and NMR Spectra Data Procession | Bruker." [Online]. Available: <https://www.bruker.com/products/mr/nmr/software/topspin.html>.
- [19] J. Hao, M. Liebeke, W. Astle, M. De Iorio, J. G. Bundy, and T. M. D. Ebbels, "Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN," *Nat. Protoc.*, vol. 9, no. 6, pp. 1416–1427, May 2014.
- [20] S. D. Bruce, J. Higinbotham, I. Marshall, and P. H. Beswick, "An Analytical Derivation of a Popular Approximation of the Voigt Function for Quantification of NMR Spectra," *J. Magn. Reson.*, vol. 142, no. 1, pp. 57–63, 2000.
- [21] "Software SPSS - España | IBM." [Online]. Available: <https://www.ibm.com/es-es/analytics/spss-statistics-software>. [Accessed: 07-Jul-2020].